

Modelli probabilistici

Leo Pasquazzi

Dipartimento di Statistica e Metodi Quantitativi

Università degli Studi di Milano - Bicocca

ORCID iD <https://orcid.org/0000-0002-2467-2667>.

2022-07-27

Indice

1	Definizione di modello probabilistico	3
1.1	Le tre componenti di un modello probabilistico	3
1.2	Metodi di assegnazione delle probabilità	6
1.3	Gli assiomi di Kolmogorov	8
1.4	Algebre e σ -algebre	16
1.5	Le leggi del calcolo delle probabilità	24
1.6	Il teorema di estensione di Hahn-Kolmogorov	37
2	Calcolo combinatorio	44
2.1	Esperimenti in più passi	45
2.2	Disposizioni e permutazioni	47
2.3	Combinazioni	50
2.4	Esercizi sul calcolo combinatorio	52
3	Probabilità condizionata	60
3.1	Definizione di probabilità condizionata	60
3.2	Il teorema della probabilità composta	61
3.3	Eventi indipendenti	64
3.4	Indipendenza globale	68
3.5	La formula della probabilità totale	76
3.6	La formula di Bayes	79
4	Variabili casuali	85
4.1	Definizione di variabile casuale	85
4.2	Funzioni di ripartizione	90
4.3	Distribuzioni discrete e assolutamente continue	104
4.3.1	Funzioni (di massa) di probabilità	104
4.3.2	Funzioni di densità	109
4.4	Quantili	116
4.5	Distribuzioni congiunte	124
4.6	Distribuzioni marginali associate ad una distribuzione congiunta	133
4.7	Distribuzioni congiunte discrete e assolutamente continue	136
4.7.1	Funzioni di massa di probabilità congiunte	137
4.7.2	Funzioni di densità congiunte	143
4.8	Distribuzioni condizionate	152
4.8.1	Distribuzioni marginali e condizionate discrete	153
4.8.2	Distribuzioni marginali e condizionate assolutamente continue*	161
4.8.3	Distribuzioni marginali e condizionate di tipo diverso	165
4.9	Variabili casuali indipendenti	171

4.9.1	Variabili casuali indipendenti: il caso discreto	176
4.9.2	Variabili casuali indipendenti: il caso assolutamente continuo	179
4.10	Successioni infinite di variabili casuali indipendenti	180
5	Valore atteso	181
5.1	Il valore atteso nel caso discreto	182
5.2	Il valore atteso nel caso assolutamente continuo	187
5.3	Proprietà del valore atteso	190
5.4	Applicazioni del valore atteso	194
5.4.1	La definizione di gioco equo e il paradosso di San Pietroburgo	194
5.4.2	La teoria dell'aspettativa morale di Daniel Bernoulli	202
5.4.3	Osservazione sul metodo di assegnazione soggettivo	217
5.4.4	Teorie dell'utilità attesa*	218
6	Varianza e covarianza	219
6.1	Definizioni di varianza e covarianza	219
6.2	Proprietà della varianza e covarianza	222
6.3	Applicazioni	227
6.3.1	Valore atteso e varianza del rendimento di un portafoglio titoli	227
6.3.2	Valore atteso e varianza del rendimento complessivo riferito a più periodi di investimento	234
6.3.3	Teoria del portafoglio di Markowitz	237
7	Momenti e funzione generatrice dei momenti	246
7.1	I momenti	246
7.2	La funzione generatrice dei momenti (fgm)	249
8	Distribuzioni notevoli	253
8.1	Il concetto di distribuzione notevole	253
8.2	La famiglia delle distribuzioni di Bernoulli	254
8.3	La famiglia delle distribuzioni ipergeometriche	261
8.4	La famiglia delle distribuzioni binomiali	267
8.5	La famiglia delle distribuzioni trinomiali	276
8.6	La famiglia delle distribuzioni geometriche (o di Pascal)	284
8.7	La famiglia delle distribuzioni binomiali negative	297
8.8	La famiglia delle distribuzioni di Poisson	309
8.9	La famiglia delle distribuzioni esponenziali	333
8.10	La famiglia delle distribuzioni gamma	340
8.11	La famiglia delle distribuzioni normali	351
8.12	La famiglia delle distribuzioni lognormali	383
8.13	La famiglia delle distribuzioni di Pareto	395

9 Inferenza	400
9.1 Stima puntuale	402
9.1.1 La media campionaria	408
9.1.2 La frequenza relativa campionaria	415
9.1.3 La varianza campionaria e la varianza campionaria corretta	420
9.2 Stima intervallare	425

1 Definizione di modello probabilistico

1.1 Le tre componenti di un modello probabilistico

In questa dispensa parleremo di **esperimenti casuali**, ovvero di processi che si svolgono nel mondo reale i cui esiti finali non possono essere previsti con assoluta certezza. Ovviamente, processi di questo genere possono essere descritti in tanti modi diversi, ma in ambito scientifico si considerano solo descrizioni che possono essere tradotte in un modello matematico con determinate caratteristiche che presenteremo tra breve. Il modello matematico che si ottiene a partire da una tale descrizione viene chiamato **modello probabilistico**.

Definizione 1.1 (Esperimento casuale e modello probabilistico). Un **esperimento casuale** è un *processo* che si svolge nel mondo reale il cui esito finale è incerto. Un **modello probabilistico** è un particolare tipo di modello matematico che descrive un esperimento.

Ora, per descrivere le caratteristiche di un modello probabilistico conviene cominciare dalle sue tre componenti:

- 1) La prima componente di un modello probabilistico è il cosiddetto **spazio campionario**, che in questa dispensa indicheremo sempre con la lettera greca maiuscola Ω (si legge "omega"). Lo spazio campionario di un modello probabilistico è un insieme i cui elementi ω (la lettera greca "omega" minuscola) vengono chiamati **eventi elementari** perché rappresentano tutti i possibili esiti elementari che potrebbero verificarsi al termine dell'esperimento casuale. Lo spazio campionario di un modello probabilistico deve quindi sempre essere *esaustivo* nel senso che ad ogni possibile esito finale dell'esperimento casuale deve corrispondere un evento elementare $\omega \in \Omega$, e gli eventi elementari ω dello spazio campionario devono essere *mutuamente esclusivi* nel senso che al termine dell'esperimento casuale deve verificarsi soltanto *uno* di essi.
- 2) La seconda componente di un modello probabilistico è il cosiddetto **spazio degli eventi** che indicheremo spesso con il simbolo \mathcal{A} , ma per il quale potremmo usare anche altri simboli che introdurremo nel corso di questa dispensa. Lo spazio degli eventi di un modello probabilistico è un *insieme di sottoinsiemi dello spazio campionario* Ω . Com'è facile intuire, i sottoinsiemi di Ω possono essere considerati come **eventi** e questo fatto spiega perché la seconda componente di un modello probabilistico si chiama "spazio degli eventi". Per distinguere gli eventi elementari $\omega \in \Omega$ dagli eventi A che sono invece sottoinsiemi di Ω , questi ultimi vengono a volte chiamati **eventi "composti"**. Chiaramente, il più ampio spazio degli eventi

che si può definire a partire da un dato spazio campionario Ω è il cosiddetto **insieme delle parti** di Ω , ovvero l'insieme di tutti i sottoinsiemi di Ω . In questa dispensa l'insieme delle parti di Ω verrà indicato con $\mathcal{P}(\Omega)$. Come vedremo tra breve, lo spazio degli eventi di un modello probabilistico non deve necessariamente contenere tutti i sottoinsiemi dello spazio campionario Ω , ma deve comunque essere in possesso di una determinata "struttura".

- 3) Infine, la terza e ultima componente di un modello probabilistico è la cosiddetta **funzione di probabilità**. Quest'ultima è una funzione $P(\cdot)$ che associa una probabilità $P(A)$ ad ogni evento A che appartiene allo spazio degli eventi \mathcal{A} . Il *dominio* di una funzione di probabilità è dunque sempre uno spazio di eventi \mathcal{A} che, come abbiamo appena visto, potrebbe anche non contenere tutti i sottoinsiemi dello spazio campionario Ω . In questo caso esisteranno dunque degli eventi $A \subseteq \Omega$ ai quali non risulta associata nessuna probabilità.

Non di rado viene sostenuto che il dominio di una funzione di probabilità $P(\cdot)$ sia uno spazio campionario Ω . *Come abbiamo appena visto, questa affermazione è SBAGLIATA!!!* L'origine di questo errore risiede nel fatto che in molte applicazioni si considera come spazio degli eventi \mathcal{A} l'insieme delle parti dello spazio campionario Ω , e che in questo tipo di contesto si sente spesso parlare della "probabilità associata ad un evento elementare ω " intendendo il valore di $P(\{\omega\})$ e non quello di $P(\omega)$. Infatti, quest'ultimo non è nemmeno definito visto che ω non è un elemento dello spazio degli eventi \mathcal{A} (ovvero del dominio della funzione di probabilità $P(\cdot)$), ma un elemento dello spazio campionario Ω !!!

Nei prossimi esempi considereremo tre esperimenti casuali e definiremo corrispondenti modelli probabilistici.

Esempio 1.1 (Lancio di un dado *regolare* a sei facce). Consideriamo un esperimento casuale che consiste nel lancio di un dado *regolare*. Per descrivere questo esperimento casuale possiamo considerare come

- **spazio campionario** Ω l'insieme

$$\Omega = \{1, 2, 3, 4, 5, 6\},$$

- come **spazio degli eventi** \mathcal{A} l'insieme

$$\begin{aligned} \mathcal{A} = \{ & \emptyset, \{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}, \{1, 2\}, \{1, 3\}, \{1, 4\}, \{1, 5\}, \{1, 6\}, \{2, 3\}, \{2, 4\}, \dots \\ & \dots \{2, 5\}, \{2, 6\}, \{3, 4\}, \{3, 5\}, \{3, 6\}, \{4, 5\}, \{4, 6\}, \{5, 6\}, \{1, 2, 3\}, \{1, 2, 4\}, \dots \\ & \dots \{1, 2, 5\}, \{1, 2, 6\}, \{1, 3, 4\}, \{1, 3, 5\}, \{1, 3, 6\}, \{1, 4, 5\}, \{1, 4, 6\}, \{2, 3, 4\}, \dots \\ & \dots \{2, 3, 5\}, \{2, 3, 6\}, \{3, 4, 5\}, \{3, 4, 6\}, \{4, 5, 6\}, \{1, 2, 3, 4\}, \{1, 2, 3, 5\}, \dots \\ & \dots \{1, 2, 3, 6\}, \{1, 2, 4, 5\}, \{1, 2, 4, 6\}, \{1, 2, 5, 6\}, \{1, 3, 4, 5\}, \{1, 3, 4, 6\}, \dots \\ & \dots \{1, 3, 5, 6\}, \{1, 4, 5, 6\}, \{2, 3, 4, 5\}, \{2, 3, 4, 6\}, \{2, 3, 5, 6\}, \{2, 4, 5, 6\}, \dots \\ & \dots \{3, 4, 5, 6\}, \{1, 2, 3, 4, 5\}, \{1, 2, 3, 5, 6\}, \{1, 2, 4, 5, 6\}, \{1, 3, 4, 5, 6\}, \dots \\ & \dots \{2, 3, 4, 5, 6\}, \{1, 2, 3, 4, 5, 6\} \} \end{aligned}$$

- e come **funzione di probabilità** $P(\cdot)$ la funzione definita come

$$P(A) = \frac{\text{numero di eventi elementari contenuti in } A}{\text{numero di eventi elementari contenuti in } \Omega} = \frac{\#A}{\#\Omega} = \frac{\#A}{6}$$

per $A \in \mathcal{A}$.

Esempio 1.2 (Lancio di un dado *truccato* a sei facce). Consideriamo ora invece un esperimento casuale che consiste nel lancio di un dado truccato e supponiamo di conoscere gli esiti di $n = 1000$ lanci di questo dado che sono stati effettuati in passato. Per descrivere questo esperimento casuale possiamo definire lo spazio campionario Ω e lo spazio di eventi \mathcal{A} come nell'esempio precedente, ma siccome sappiamo che il dado è truccato, sarebbe "sbagliato" definire anche la funzione di probabilità $P(\cdot)$ come nell'esempio precedente. Per ottenere una descrizione "realistica", conviene invece considerare gli esiti degli $n = 1000$ lanci e definire le probabilità degli eventi $A \in \mathcal{A}$ come

$$P(A) = \frac{\text{numero di volte che si è verificato l'evento } A \text{ negli } n = 1000 \text{ lanci}}{n = 1000}.$$

Esempio 1.3 (Scommettitore). Un giornalista sportivo sostiene di essere "*sicuro all'80%*" che la squadra AAA vinca il prossimo campionato di calcio. Ora, siccome il giornalista non è in grado di prevedere con assoluta certezza quale sarà l'esito del prossimo campionato, egli potrebbe considerare il campionato come un esperimento casuale e potrebbe descrivere questo esperimento casuale attraverso un modello probabilistico con

- **spazio campionario** Ω definito come

$$\Omega = \{1, 2, \dots, k\},$$

dove ciascun numero naturale $1, 2, \dots, k$ rappresenta una delle squadre k che partecipano al campionato e dove supponiamo che il numero naturale 1 rappresenti la squadra AAA (la squadra AAA potrebbe anche essere rappresentata da qualsiasi altro numero tra quelli contenuti in Ω);

- **spazio degli eventi** \mathcal{A} definito come

$$\mathcal{A} = \{\emptyset, \{1\}, \{2, 3, \dots, k\}, \Omega\}$$

- **funzione di probabilità** $P(\cdot)$ definita come

$$P(\emptyset) = 0, \quad P(\{1\}) = 0,80, \quad P(\{2, 3, \dots, k\}) = 0,20, \quad P(\Omega) = 1.$$

1.2 Metodi di assegnazione delle probabilità

Confrontando i tre esempi della sezione precedente, ci rendiamo subito conto di una sostanziale differenza:

- nel primo esempio le probabilità degli eventi sono definite come rapporti tra numeri di casi favorevoli e il numero di casi possibili, ...
- ... nel secondo esempio le probabilità degli eventi sono invece definite come frequenze relative, ...
- ... mentre nel terzo e ultimo esempio i valori delle probabilità sembrerebbero rappresentare soltanto un'opinione.

Si noti che queste differenze attengono al modo in cui abbiamo **assegnato** le probabilità agli eventi e al modo in cui queste probabilità devono quindi essere **interpretate**. I tre esempi illustrano i tre **metodi di assegnazione delle probabilità** che di solito vengono usati nelle applicazioni:

- L'Esempio 1.1 sul lancio di un dado regolare illustra un'applicazione del **metodo di assegnazione classico** che dà luogo all'**interpretazione classica della probabilità**. Secondo questa interpretazione la probabilità di un evento è definita come il rapporto tra il numero di casi favorevoli all'evento e il numero di casi possibili purché questi ultimi siano equiprobabili. Questa definizione è basata sul principio di indifferenza e può essere applicata soltanto quando il numero di esiti possibili è finito e quando non c'è motivo di ritenere che uno di questi esiti possa verificarsi "più facilmente" degli altri.

L'interpretazione classica della probabilità venne avanzata da Jacob Bernoulli (1654 - 1705) nel suo libro "*Ars conjectandi*" (pubblicato postumo nell'anno 1713) e da

Abraham de Moivre (1667 - 1754) nel suo libro *"The Doctrine of Chances"* (pubblicato nell'anno 1718). Successivamente venne anche ribadita da Pierre Simon Laplace (1749 - 1827) nel suo *"Essai philosophique sur les probabilités"* (pubblicato nell'anno 1814).

- L'Esempio 1.2 (quello sul dado truccato) illustra invece un'applicazione del **metodo di assegnazione frequentista** che dà luogo all'**interpretazione frequentista** della probabilità. Secondo questa interpretazione la probabilità di un evento rappresenta il limite della frequenza relativa con cui l'evento si è verificato in una successione infinita di repliche indipendenti dell'esperimento casuale. Com'è facile intuire, questo modo di interpretare le probabilità di eventi non si lascia facilmente tradurre in un metodo pratico per assegnare delle probabilità. Infatti, nel mondo reale non si conoscono mai gli esiti di una successione infinita di repliche di un esperimento casuale e pertanto bisogna accontentarsi di frequenze relative che si riferiscono ad un numero finito (possibilmente elevato) di repliche. Ma anche questo modo di procedere dà luogo a dei dubbi: quante repliche dell'esperimento casuale sono necessarie? Sotto quali condizioni un esperimento casuale di cui conosciamo l'esito finale può essere considerato come una *"replica"* dell'esperimento casuale che vogliamo descrivere? Sotto quali condizioni delle repliche possono essere considerate *"indipendenti"*? Che cosa si dovrebbe fare se l'esperimento casuale di interesse non è *"replicabile"*?

Dal punto di vista storico l'interpretazione frequentista prese piede verso la metà del XIX secolo. Nel suo libro *"The logic of chance"*¹ John Venn (1834-1923)² fornisce una descrizione dettagliata di questa interpretazione. Quasi mezzo secolo dopo la pubblicazione della prima edizione del libro di Venn, in un articolo pubblicato nell'anno 1919, Richard von Mises³ propose un'assiomatizzazione della teoria della probabilità basata sull'interpretazione frequentista.

- L'Esempio 1.3, infine, illustra un'applicazione del **metodo di assegnazione soggettivo** che dà luogo all'**interpretazione soggettiva** della probabilità. Questa interpretazione venne proposta in modo indipendente da Frank Plumpton Ramsey⁴ (1903 - 1930) e da Bruno de Finetti⁵ (1906 - 1985). Secondo questa interpretazio-

¹John Venn (1888), *"The logic of Chance, 3rd edition"*, Macmillan & Co, London. Le due precedenti edizioni del libro sono state pubblicate negli anni 1866 e 1876, rispettivamente.

²John Venn è l'inventore degli omonimi diagrammi di Venn, ovvero dei diagrammi che rappresentano insieme attraverso delle curve chiuse che di solito hanno forma ellittica

³Vedi Richard von Mises (1919), *"Grundlagen der Wahrscheinlichkeitsrechnung"*, *Mathematische Zeitschrift*.

⁴ Vedi Ramsey (1926) *"Truth and Probability"*, in Ramsey, 1931, *"the Foundations of Mathematics and other Logical Essays, Ch. VII, p.156-198, edited by R.B. Braithwaite, London: Kegan, Paul, Trench, Trubner & Co., New York: Harcourt, Brace and Company.*

⁵ Vedi De Finetti (1931) *"Sul significato soggettivo della probabilità"* in *"Fundamenta Mathematicae"*, Warszawa 1931, T. XVII, pp. 298-329.

ne, la probabilità di un evento rappresenta semplicemente un *"grado di credenza"* che può variare da soggetto a soggetto. Ramsey e de Finetti sostengono entrambi che il *"grado di credenza"* (ovvero la probabilità) che un soggetto associa ad un determinato evento possa essere determinato attraverso le sue preferenze rispetto a determinate scommesse. Siccome il metodo per ricavare i valori numerici delle probabilità di Ramsey è piuttosto complicato, ci limiteremo ad esporre soltanto quello di de Finetti. Secondo quest'ultimo, la probabilità $P(A)$ che un soggetto associa ad un determinato evento A dovrebbe rappresentare il **prezzo** che il soggetto ritiene **"equo"** per una scommessa dove si vince 1 euro se l'evento A si verifica, e dove non si riceve indietro nulla in caso contrario. Per determinare questo prezzo equo, il soggetto deve immaginare di gestire un banco di scommesse che **acquista** ed **emette** titoli che incorporano il diritto di ricevere 1 euro se l'evento A si verifica, e deve immaginare di essere **costretto** a pagare il prezzo $P(A)$ a chiunque voglia vendergli un titolo della scommessa, e di essere allo stesso tempo anche **costretto** ad accettare il prezzo $P(A)$ per emettere un titolo della scommessa a chiunque lo richieda.

1.3 Gli assiomi di Kolmogorov

Come abbiamo visto nella sezione precedente, le interpretazioni delle probabilità negli Esempi 1.1, 1.2 e 1.3 sono molto diverse. In questa sezione focalizzeremo invece l'attenzione sul fatto che tutte e tre le funzioni di probabilità $P(\cdot)$ che abbiamo definito nei suddetti esempi soddisfano determinate condizioni. Le condizioni in questione sono le seguenti (invitiamo il lettore a verificare che negli Esempi 1.1, 1.2 e 1.3 esse sono effettivamente soddisfatte):

K0) Il dominio \mathcal{A} della funzione di probabilità $P(\cdot)$ (ovvero lo spazio degli eventi) soddisfa le seguenti condizioni:

- a) $\Omega \in \mathcal{A}$;
- b) se $A \subseteq \Omega$ è un evento contenuto in \mathcal{A} , allora \mathcal{A} contiene anche il corrispondente evento complementare $\bar{A} = \{\omega \in \Omega : \omega \notin A\}$;
- c) se A_1 e A_2 sono due eventi contenuti in \mathcal{A} , allora \mathcal{A} contiene anche la loro unione $A_1 \cup A_2$.

K1) Per ogni evento $A \in \mathcal{A}$ la corrispondente probabilità $P(A)$ è un numero reale non negativo.

K2) La probabilità associata allo spazio campionario Ω (**evento certo**) è pari a 1.

K3) (**Additività**). Se A_1 e A_2 sono due **eventi incompatibili** (ovvero $A_1 \cap A_2 = \emptyset$) e entrambi questi eventi appartengono allo spazio degli eventi \mathcal{A} , allora

$$P(A_1 \cup A_2) = P(A_1) + P(A_2).$$

Aggiungendo a queste quattro condizioni anche la condizione ...

K3*) (**Additività numerabile**) Se A_1, A_2, \dots è una successione infinita numerabile⁶ di eventi incompatibili che appartengono tutti allo spazio degli eventi \mathcal{A} , e se anche l'evento $\bigcup_{i=1}^{\infty} A_i$ appartiene allo spazio degli eventi \mathcal{A} , allora

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

... si ottengono quelli che nei libri di testo moderni vengono presentati come **assiomi di Kolmogorov** in onore del matematico russo Andrey Kolmogorov (1903 - 1987) che li ha proposti in una sua celebre monografia.⁷ Gli assiomi di Kolmogorov sono diventati noti e famosi perché hanno finalmente risolto un problema che si trascinava da anni, ovvero quello di unire le principali interpretazioni della probabilità all'interno di una teoria assiomatica.⁸ Kolmogorov non fu l'unico a cimentarsi in questo problema, ma la sua soluzione è quella che è stata accolta con maggiore favore dalla comunità scientifica.⁹ **Sulla base degli assiomi di Kolmogorov si poteva finalmente sviluppare una teoria della probabilità lasciando da parte tutte le diatribe filosofiche sull'interpretazione della probabilità.** I teoremi di questa teoria sono noti come **"leggi del calcolo**

⁶Ricordiamo che, per definizione, un insieme o una successione è **"numerabile"** se gli elementi dell'insieme o della successione possono essere posti in corrispondenza biunivoca con un segmento iniziale dei numeri naturali oppure con l'insieme di tutti i numeri naturali. In modo informale si può dunque dire che un insieme o una successione è numerabile se gli elementi dell'insieme o della successione possono essere "contati". Se gli elementi dell'insieme o della successione possono essere posti in corrispondenza biunivoca con un segmento iniziale dei numeri naturali, si dice che l'insieme o la successione è **"finita"**, mentre se per costruire una corrispondenza biunivoca servissero tutti i numeri naturali, allora si parla di un insieme o di una successione **"infinita numerabile"**. Ovviamente, l'insieme di tutti i numeri naturali, l'insieme dei numeri naturali pari, l'insieme dei numeri naturali dispari e l'insieme di tutti i numeri interi (positivi e negativi) \mathbb{Z} sono tutti esempi di insiemi infiniti numerabili. Un altro noto esempio di insieme infinito numerabile è l'insieme di tutti i numeri razionali \mathbb{Q} . Tuttavia, non tutti gli insiemi infiniti (ovvero insiemi che contengono infiniti elementi) sono numerabili. Un noto esempio di **insieme infinito non numerabile** è l'insieme dei numeri reali \mathbb{R} . Infatti, si può dimostrare che tra \mathbb{R} (o un anche un qualsiasi intervallo aperto di numeri reali) e l'insieme dei numeri naturali \mathbb{N} non esiste nessuna corrispondenza biunivoca (i numeri reali non possono essere "contati"). Chiaramente, sono insiemi infiniti non numerabili anche gli insiemi \mathbb{R}^2 , \mathbb{R}^3 , ecc. i cui elementi sono coppie ordinate, terne ordinate, ecc. di numeri reali.

⁷Vedi A. Kolmogorov (1933), *"Grundbegriffe der Wahrscheinlichkeitsrechnung"*, *Ergebnisse der Mathematik und ihrer Grenzgebiete herausgegeben von der Schriftleitung des "Zentralblatt für Mathematik"*, zweiter Band, Julius Springer Verlag, Berlin.

⁸Questo problema era avvertito da molti matematici verso la fine del XIX secolo e venne posto esplicitamente in una famosa lista di 23 problemi che David Hilbert presentò ad un congresso internazionale di matematici a Parigi nell'anno 1900 (*Hilbert, D. (1902) "Mathematical problems"*, *Bull. Amer. Math. Soc.*, 8, pp. 437-479)

⁹ Il contesto conoscenze che ha condotto agli assiomi di Kolmogorov è ricostruito nell'articolo *Shafer G., Vouk V. (2006) "The Sources of Kolmogorov's Grundbegriffe."* *Statist. Sci.* 21 (1) 70 - 98.

delle probabilità” e queste leggi devono essere soddisfatte da tutte le funzioni di probabilità $P(\cdot)$ che soddisfano i suddetti assiomi. Per esempio, non è difficile dimostrare (omettiamo la dimostrazione) che tutte le funzioni di probabilità che si possono ottenere attraverso il **metodo di assegnazione classico** (quando applicabile), e anche quelle che si possono ottenere attraverso il **metodo di assegnazione frequentista**, soddisfano gli assiomi di Kolmogorov: le leggi del calcolo delle probabilità che sono implicite negli assiomi di Kolmogorov devono quindi essere soddisfatte da tutte le funzioni di probabilità $P(\cdot)$ che si possono ottenere attraverso questi due metodi di assegnazione. D'altra parte, attraverso il **metodo di assegnazione soggettivo** si possono ottenere anche funzioni di probabilità $P(\cdot)$ che non soddisfano gli assiomi di Kolmogorov. Tuttavia, come fanno notare Ramsey⁴ e De Finetti¹¹, se un soggetto assegna le probabilità in modo "razionale", ovvero in modo da non esporsi a scommesse che gli procurerebbero una perdita certa, allora egli dovrà assegnare le probabilità $P(\cdot)$ in modo tale che gli assiomi **K1 - K3** siano soddisfatti, e viceversa. Il teorema che consiste in questa affermazione è noto come **teorema delle scommesse olandesi**.¹⁰

Teorema 1.1 (Teorema delle scommesse olandesi). Sia \mathcal{A} uno spazio di eventi che soddisfa le condizioni poste nell'assioma **K0** e si consideri un soggetto che deve assegnare dei prezzi $P(A)$ a titoli di scommesse sugli eventi $A \in \mathcal{A}$.^a Se il soggetto può essere **costretto** ad acquistare e/o emettere soltanto titoli di scommesse su un numero finito di eventi e il soggetto è "razionale" nel senso che non vuole esporsi a sistemi di scommesse^b che gli procurerebbero una perdita certa, allora deve assegnare i prezzi $P(A)$ in modo tale che siano soddisfatte le condizioni poste negli assiomi **K1 - K3**.

^aRicordiamo che il **titolo di una scommessa** su un dato evento A incorpora il diritto di ricevere dall'emittente 1 euro nel caso in cui l'evento A si verificasse; se l'evento A non si verifica, il titolo non dà luogo a nessun diritto.

^bUn **sistema di scommesse** è un insieme di diritti e obbligazioni che può essere costruito attraverso l'acquisto e/o l'emissione di titoli di scommesse.

Dimostrazione. Secondo l'assioma **K1** si dovrebbe avere $P(A) \geq 0$ per ogni $A \in \mathcal{A}$. Per dimostrare che nessun soggetto "razionale" fisserebbe un prezzo $P(A)$ negativo, basta osservare che se fissasse un prezzo $P(A)$ minore di zero, allora il soggetto sarebbe disposto a pagare $P(A)$ euro per impegnarsi a pagare 1 euro qualora l'evento A si verificasse, e per non ricevere nulla in caso contrario. Fissando un prezzo $P(A)$ negativo egli sarebbe quindi esposto ad una scommessa che gli procura una perdita certa.

Non è nemmeno difficile dimostrare che il prezzo della scommessa sull'evento certo Ω deve essere $P(\Omega) = 1$ così come previsto dall'assioma **K2**. Infatti, nessun soggetto

¹⁰Una "scommessa olandese" (traduzione di "dutch book") è un sistema di scommesse che garantisce un profitto strettamente positivo (oppure, a seconda del punto di vista, una perdita strettamente positiva) per qualunque esito dell'esperimento casuale di riferimento. L'origine di questo termine è controversa.

”razionale” è disposto a pagare più di 1 euro per una scommessa attraverso la quale riceve indietro soltanto 1 euro perché in questo modo realizzerebbe una perdita certa. D'altra parte, nessun soggetto ”razionale” è disposto ricevere meno di 1 euro per una scommessa attraverso la quale dovrà sicuramente pagare 1 euro perché anche in questo caso realizzerebbe una perdita certa.

Consideriamo infine l'assioma [K3](#). Per dimostrare che i prezzi $P(A)$ fissati da un soggetto ”razionale” devono soddisfare anche questo assioma, ipotizzeremo che A_1 e A_2 siano due eventi incompatibili e che i prezzi delle scommesse sugli eventi A_1 , A_2 e $A_1 \cup A_2$ siano fissati in modo tale che

$$P(A_1 \cup A_2) > P(A_1) + P(A_2)$$

oppure in modo tale che

$$P(A_1 \cup A_2) < P(A_1) + P(A_2).$$

Dimostreremo che in entrambi questi casi il soggetto che fissa i prezzi sarebbe esposto ad una perdita certa, e che i prezzi fissati da un soggetto ”razionale” devono quindi necessariamente soddisfare l'uguaglianza

$$P(A_1 \cup A_2) = P(A_1) + P(A_2)$$

così come previsto dall'assioma [K3](#).

Per dimostrare che la disuguaglianza $P(A_1 \cup A_2) > P(A_1) + P(A_2)$ espone colui che fissa i prezzi ad una perdita certa, basta osservare che egli si obbligherebbe ad acquistare il titolo della scommessa sull'evento $A_1 \cup A_2$ ad un prezzo maggiore di quello che incasserebbe per emettere separatamente i titoli delle scommesse sugli eventi A_1 e A_2 . Nel momento in cui acquista il titolo della scommessa sull'evento $A_1 \cup A_2$ e in cui contestualmente emette i titoli delle scommesse sugli eventi A_1 e A_2 in modo separato, conseguirebbe una perdita netta di euro

$$P(A_1 \cup A_2) - P(A_1) - P(A_2) > 0.$$

Al termine dell'esperimento casuale la sua posizione netta sarebbe invece in pareggio: se uno dei due eventi A_1 oppure A_2 si verifica riceve 1 euro perché ha acquistato il titolo della scommessa sull'evento $A_1 \cup A_2$, ma deve anche pagare 1 euro perché ha emesso i titoli delle scommesse sugli eventi A_1 e A_2 . D'altra parte, se nessuno dei due eventi A_1 e A_2 si verifica, non riceve nulla e non paga nulla. Siccome la posizione netta al momento dell'acquisto del titolo della scommessa sull'evento $A_1 \cup A_2$ e della contestuale emissione dei titoli delle scommesse sugli eventi A_1 e A_2 sarebbe negativa, e siccome la posizione netta al termine dell'esperimento casuale sarebbe in ogni caso nulla, possiamo concludere che un soggetto ”razionale” non fisserebbe mai i prezzi in modo tale che $P(A_1 \cup A_2) > P(A_1) + P(A_2)$.

Ragionando in modo speculare si può anche dimostrare che un soggetto "razionale" non fisserebbe mai i prezzi in modo tale che $P(A_1 \cup A_2) < P(A_1) + P(A_2)$.

A questo punto abbiamo dimostrato che il soggetto che fissa i prezzi $P(\cdot)$ delle scommesse è esposto ad una perdita certa se almeno uno degli assiomi **K1 - K3** non è soddisfatto. Rimane da dimostrare che in caso contrario è impossibile costruire un sistema di scommesse che espone il soggetto ad una perdita certa. Per questa parte della dimostrazione ci serviranno alcune leggi del calcolo delle probabilità e per questo motivo la rimandiamo a [dopo](#). \square

Ammettendo la possibilità che un soggetto possa essere costretto ad impegnarsi in scommesse su un numero infinito di eventi, si può anche dimostrare una versione alternativa del [teorema delle scommesse olandesi](#) secondo la quale un soggetto "razionale" assegna le probabilità $P(\cdot)$ in modo tale che sia soddisfatto anche l'assioma **K3***. Tuttavia, de Finetti¹¹ rifiuta questa possibilità schierandosi apertamente contro l'introduzione dell'assioma **K3***: secondo de Finetti, condizioni come quelle poste dall'assioma **K3*** rappresentano delle forzature che potrebbero permettere di determinare in modo "illusorio" le probabilità di alcuni eventi anche se le probabilità di questi eventi sono indeterminate per qualunque soggetto che può scommettere soltanto su un numero finito di eventi (vedi l'Esempio 1.7).

Vista l'obiezione sollevata da de Finetti¹¹, conviene spiegare il motivo per cui Kolmogorov introdusse l'assioma **K3***. Il motivo è contenuto nella premessa della sua monografia⁷ dove Kolmogorov dichiara di voler incorporare la teoria della probabilità nella "**teoria della misura**".¹² L'assioma **K3*** (che Kolmogorov presenta in forma diversa ma comunque equivalente) è funzionale a questo scopo perché permette di interpretare una funzione di probabilità $P(\cdot)$ come una "**misura di insiemi numerabilmente additiva**".¹³ Come fecero notare Kolmogorov e alcuni altri autori,¹⁴ attraverso l'incorporazione nella teoria della misura vengono alla luce molte analogie tra concetti di cui si occupano gli studiosi della probabilità e corrispondenti concetti della teoria della misura. L'incorporazione della teoria della probabilità nella teoria della misura permette quindi di dare un quadro di riferimento preciso a molti problemi di cui si occupano gli studiosi della probabilità. Dopo aver introdotto l'assioma **K3***, Kolmogorov fa alcuni commenti che conviene riportare:

¹¹ Vedi "*Problemi determinati e indeterminati nel calcolo della probabilità*" in "*Rendiconti della R. Accademia Nazionale dei Lincei*", 1930, vol. XII, fasc. 9, pp. 367-373.

¹² Come suggerisce il suo nome, la "*teoria della misura*" è la branca della matematica che studia in modo astratto le proprietà di misure come, per esempio, il peso, la lunghezza, l'area e il volume. La teoria della misura come branca della matematica si sviluppò tra la fine del XIX secolo e l'inizio del XX secolo attraverso alcuni lavori di Émile Borel, Henri Lebesgue, Jon Radon, Constantin Carathéodory e Maurice Fréchet.

¹³ Il concetto di "misura di insiemi" è un concetto della teoria della misura che non definiremo in questa dispensa.

¹⁴ Vedi l'articolo già citato nella nota [9].

- In primo luogo fa notare che se lo spazio degli eventi \mathcal{A} contiene soltanto un numero finito di eventi, l'assioma $K3^*$ è già implicito negli assiomi $K0 - K3$, mentre per spazi di eventi che contengono infiniti eventi esistono anche funzioni di probabilità che $P(\cdot)$ soddisfano soltanto gli assiomi $K0 - K3$ ma non l'assioma $K3^*$.

Per completare questo commento di Kolmogorov conviene aggiungere che, d'altra parte, tutte le funzioni di probabilità $P(\cdot)$ che soddisfano gli assiomi $K0 - K2$ e l'assioma $K3^*$, soddisfano sempre anche l'assioma $K3$, sia quando il loro dominio \mathcal{A} è uno spazio di eventi finito, sia quando contiene infiniti eventi (in altre parole, l'assioma $K3$ è sempre implicito negli altri assiomi di Kolmogorov, sia quando lo spazio di eventi \mathcal{A} è finito, sia quando contiene infiniti eventi).

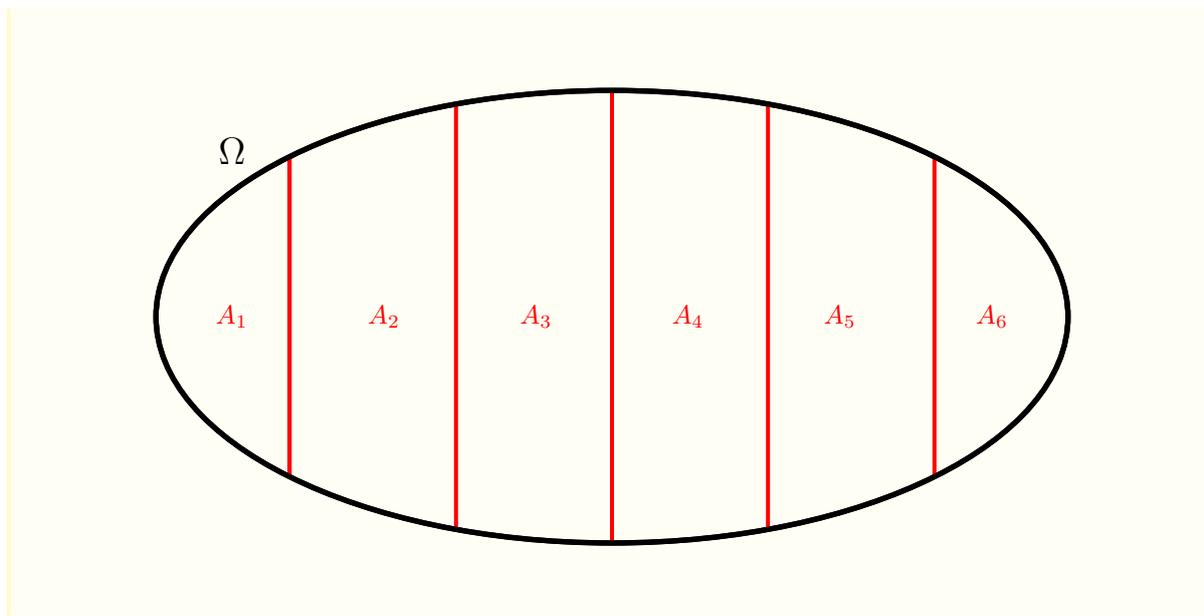
- Dopo l'osservazione al punto precedente, Kolmogorov fa notare che quando si descrivono processi che si svolgono nel mondo reale (ovvero esperimenti casuali) si possono ottenere soltanto spazi di eventi finiti e che pertanto sia molto difficile giustificare l'assioma $K3^*$ facendo riferimento a ciò che accade nel mondo reale (infatti, gli strumenti di misurazione di cui disponiamo nel mondo reale ci permettono di distinguere soltanto un numero finito di esiti elementari ω di un esperimento casuale, e con uno spazio campionario finito si possono ottenere soltanto spazi di eventi che sono anch'essi finiti perché ogni insieme finito ha soltanto un numero finito di sottoinsiemi).
- Infine aggiunge che spazi di eventi infiniti emergono soltanto in descrizioni "idealizzate" di processi che si svolgono nel mondo reale (per esempio quando si assume che una rilevazione numerica possa essere effettuata con precisione arbitraria oppure quando si assume che un esperimento possa essere ripetuto infinite volte) e prosegue dicendo che nel resto della sua monografia⁷ si sarebbe limitato a considerare solo descrizioni idealizzate che soddisfano anche l'assioma $K3^*$, precisando che si trattasse di una scelta *arbitraria* che nelle ricerche fino ad allora effettuate si era sempre rilevata "adeguata allo scopo" (traduzione dell'aggettivo tedesco "zweckmäßig").

Definizione 1.2 (Definizione di modello probabilistico secondo Kolmogorov). Secondo Kolmogorov, un **modello probabilistico** è una descrizione di un esperimento casuale che è composta da

- uno spazio campionario Ω ,
- uno spazio di eventi \mathcal{A} che soddisfa le condizioni **a)**, **b)** e **c)** dell'assioma $K0$
- e una funzione di probabilità $P(\cdot)$ che ha per dominio lo spazio di eventi \mathcal{A} e che soddisfa gli assiomi $K1 - K3$ e anche l'assioma $K3^*$.

Tornando ora al [teorema delle scommesse olandesi](#), conviene osservare che dal punto di vista pratico questo teorema è poco utile per verificare se le probabilità $P(A_1) = p_1$, $P(A_2) = p_2, \dots$ che un soggetto assegna a determinati eventi A_1, A_2, \dots sono effettivamente coerenti con gli assiomi [K0 - K3](#). Per dare un'idea dell'importanza di questa verifica, basti pensare a tutte quelle circostanze dove sentiamo parlare di probabilità $P(A_1) = p_1, P(A_2) = p_2, \dots$ che riflettono l'opinione di un qualche soggetto sull'esito finale di un qualche esperimento casuale: spesso gli eventi A_1, A_2, \dots ai quali si riferiscono le probabilità non formano una collezione di eventi \mathcal{A} che soddisfa le condizioni [a\)](#), [b\)](#) e [c\)](#) dell'assioma [K0](#), e siccome abbiamo tutti un'inclinazione quasi "naturale" ad applicare gli assiomi [K1 - K3](#) (e talvolta anche l'assioma [K3*](#)) a probabilità di qualsiasi tipo, **bisognerebbe sempre verificare se determinate probabilità $P(A_1) = p_1, P(A_2) = p_2, \dots$ possono essere estese ad una funzione di probabilità $P(\cdot)$ che soddisfa gli assiomi [K0 - K3](#) (ed eventualmente anche l'assioma [K3*](#)). Se così non fosse, allora attraverso opportuni ragionamenti che applicano gli assiomi [K1 - K3](#) (ed eventualmente anche l'assioma [K3*](#)) alle probabilità $P(A_i) = p_i$ si potrebbe dimostrare qualsiasi affermazione!!!** (vedi, per esempio, l'Esercizio [1.4](#)). A proposito di questo problema, De Finetti¹¹ fornisce una condizione necessaria e sufficiente affinché determinate probabilità $P(A_1) = p_1, P(A_2) = p_2, \dots$ possano essere estese ad una funzione di probabilità $P(\cdot)$ che soddisfa gli assiomi [K0 - K3](#). Sulla base di questa condizione si può facilmente costruire un *algoritmo* per verificare se le probabilità associate ad un numero finito di eventi possono essere estese. Per non appesantire troppo la trattazione, in questa dispensa non presenteremo la condizione necessaria e sufficiente di de Finetti e non presenteremo nemmeno un *algoritmo* per verificare se determinate probabilità possono essere estese. Di fatto, nelle applicazioni della teoria della probabilità che si trovano in riviste scientifiche l'impiego di un tale algoritmo non è quasi mai necessario perché in tali applicazioni le probabilità vengono quasi sempre assegnate in modo tale che siano soddisfatte le condizioni poste in un qualche **teorema di estensione**. Per fare un esempio di un teorema di estensione tra breve ne enunceremo uno dal contenuto ovvio. Per non appesantire troppo la trattazione non riporteremo la sua dimostrazione anche se sarebbe molto semplice. Prima di enunciare il teorema di estensione dobbiamo definire che cosa si intende per una **partizione di uno spazio campionario Ω** : una partizione di uno spazio campionario Ω è una collezione di eventi incompatibili la cui unione è uguale a Ω (vedi [Figura 1.1](#)).

Figura 1.1. Il grafico mostra gli eventi di una partizione finita di uno spazio campionario Ω .



Formalmente, una collezione di eventi $\mathcal{C} = \{A_1, A_2, \dots\}$ è una partizione di uno spazio campionario Ω se $A_i \cap A_j = \emptyset$ per ogni $i \neq j$ e se $\bigcup_i A_i = \Omega$ ($\bigcup_i A_i$ è l'unione di tutti gli eventi A_i che appartengono alla collezione \mathcal{C}).

Teorema 1.2 (Teorema di estensione per partizioni). Se $\mathcal{C} = \{A_1, A_2, \dots\}$ è una partizione finita o infinita numerabile di uno spazio campionario Ω , e se

$$P(A_1) = p_1, \quad P(A_2) = p_2, \quad \dots$$

sono delle probabilità assegnate agli eventi della partizione \mathcal{C} , allora queste probabilità possono essere estese ad una funzione di probabilità $P(\cdot)$ che soddisfa gli assiomi di Kolmogorov (ovvero gli assiomi **K0** - **K3** con l'aggiunta dell'assioma **K3***) se e solo se

$$P(A_i) = p_i \geq 0 \quad \text{per ogni } A_i \in \mathcal{C}$$

e

$$P(A_1) + P(A_2) + \dots = p_1 + p_2 + \dots = 1.$$

Consideriamo ora un generico spazio campionario Ω e una partizione \mathcal{C} che contiene un numero finito o al più un'infinità numerabile di eventi $A_i \subseteq \Omega$. In generale può esistere più di uno spazio di eventi \mathcal{A} che contiene tutti gli eventi della partizione \mathcal{C} e che soddisfa le condizioni poste nell'assioma **K0**¹⁵ e in questi casi esistono dunque anche più funzioni di probabilità $P(\cdot)$ che soddisfano gli assiomi di Kolmogorov e che estendono determinate

¹⁵Si pensi per esempio al caso dove $\Omega = \{1, 2, 3\}$ e dove $\mathcal{C} = \{\{1, 2\}, \{3\}\}$. In questo caso esistono due spazi di eventi \mathcal{A} che contengono tutti gli eventi della partizione \mathcal{C} e che soddisfano le condizioni poste nell'assioma **K0**: lo spazio di eventi $\mathcal{A} = \{\emptyset, \Omega, \{1, 2\}, \{3\}\}$ e l'insieme delle parti $\mathcal{P}(\Omega)$.

probabilità $P(A_i) = p_i$ assegnate agli eventi della partizione \mathcal{C} (a patto che le probabilità $P(A_i) = p_i$ siano non negative e che sommino a 1 così come previsto dalla condizione necessaria e sufficiente del precedente teorema di estensione). Si noti, tuttavia, che le probabilità $P(A_i) = p_i$ degli eventi della partizione \mathcal{C} **determinano univocamente** le probabilità di tutti gli eventi che sono unioni di eventi A_i che appartengono alla partizione \mathcal{C} . Infatti, se $P(\cdot)$ è una funzione di probabilità (e dunque soddisfa gli assiomi di Kolmogorov) e se il dominio di $P(\cdot)$ contiene una determinata unione di eventi A_i della partizione \mathcal{C} ,¹⁶ allora (come potrebbe sembrare ovvio) la probabilità di tale unione deve essere uguale alla somma delle probabilità $P(A_i) = p_i$ degli eventi A_i che partecipano all'unione:

- nel caso di unioni formate con due eventi $A_i \in \mathcal{C}$, la dimostrazione segue immediatamente dall'assioma **K3**;
- nel caso di unioni formate con infiniti eventi $A_i \in \mathcal{C}$, la dimostrazione segue immediatamente dall'assioma **K3*** (si noti infatti che stiamo assumendo che la partizione \mathcal{C} contenga al più un'infinità numerabile di eventi);
- nel caso di unioni formate con un numero finito $k > 2$ di eventi $A_i \in \mathcal{C}$, la dimostrazione segue immediatamente da una legge del calcolo delle probabilità che enunceremo e dimostreremo tra breve (ovvero della legge **L2**).

1.4 Algebre e σ -algebre

Come abbiamo visto nella Definizione 1.2, lo spazio di eventi di un modello probabilistico è una collezione di eventi \mathcal{A} che soddisfa le tre condizioni **a)**, **b)** e **c)** dell'assioma **K0**. Se Ω è un qualsiasi insieme (non necessariamente lo spazio campionario associato ad un esperimento casuale) e \mathcal{A} è una collezione di sottoinsiemi di Ω che soddisfa le tre condizioni **a)**, **b)** e **c)**, allora la collezione \mathcal{A} viene chiamata **algebra**. Nel caso particolare in cui si considera Ω come spazio campionario associato ad un esperimento casuale, la collezione \mathcal{A} può essere chiamata **algebra di eventi**.

Per comprendere la ”**struttura**” di un'algebra conviene osservare che ogni algebra non è soltanto ”**chiusa**” rispetto all'operazione di **complementazione** (condizione **b)**) e all'operazione di **unione** applicata a coppie di insiemi (condizione **c)**), ma è anche ”**chiusa**” rispetto all'operazione di **intersezione** applicata a coppie insiemi, ovvero che

$$d) \quad A_1, A_2 \in \mathcal{A} \quad \Rightarrow \quad A_1 \cap A_2 \in \mathcal{A}.$$

Per dimostrare questa proprietà delle algebre bisogna fare riferimento alle **leggi di de Morgan** secondo le quali

¹⁶Come vedremo tra breve, questa condizione è sempre soddisfatta se la partizione \mathcal{C} è finita, ma si possono costruire degli esempi con partizioni \mathcal{C} che contengono infiniti eventi A_i dove questa condizione non è soddisfatta.

- $\overline{(A_1 \cap A_2)} = \overline{A_1} \cup \overline{A_2}$ (A_1 e A_2 non si verificano contemporaneamente se e solo se si verifica almeno uno dei due eventi complementari) oppure, più in generale, $\overline{(\bigcap_i A_i)} = \bigcup \overline{A_i}$ (se non si verificano contemporaneamente tutti gli eventi A_i , allora si verifica almeno uno degli eventi complementari $\overline{A_i}$, e viceversa)
- e $\overline{(A_1 \cup A_2)} = \overline{A_1} \cap \overline{A_2}$ (se non si verifica almeno uno dei due eventi A_1 oppure A_2 , allora si verificano entrambi gli eventi complementari $\overline{A_1}$ e $\overline{A_2}$, e viceversa) oppure, più in generale, $\overline{(\bigcup_i A_i)} = \bigcap \overline{A_i}$ (se non si verifica almeno uno degli eventi A_i , allora si verificano tutti gli eventi complementari $\overline{A_i}$, e viceversa).

Dimostrazione (della proprietà **d**) delle algebre).

$$\begin{aligned}
 A_1, A_2 \in \mathcal{A} &\Rightarrow \text{condizione b)} \Rightarrow \overline{A_1}, \overline{A_2} \in \mathcal{A} \Rightarrow \\
 &\Rightarrow \text{condizione c)} \Rightarrow \overline{A_1} \cup \overline{A_2} \in \mathcal{A} \Rightarrow \\
 &\Rightarrow \text{condizione b)} \Rightarrow \overline{(\overline{A_1} \cup \overline{A_2})} \in \mathcal{A} \Rightarrow \\
 \Rightarrow \text{leggi di De Morgan: } &\overline{(\overline{A_1} \cup \overline{A_2})} = A_1 \cap A_2 \Rightarrow A_1 \cap A_2 \in \mathcal{A}.
 \end{aligned}$$

Nell'ultimo passaggio abbiamo usato la **legge di De Morgan** secondo la quale $\overline{(\overline{A_1} \cup \overline{A_2})} = \overline{A_1} \cap \overline{A_2}$ con $\overline{A_i}$ al posto di A_i . Inoltre, abbiamo anche usato il fatto che il complemento del complemento di un insieme è l'insieme di partenza. \square

Chiaramente, il fatto che ogni algebra di eventi sia chiusa rispetto alle operazioni di unione e intersezione applicate a coppie di insiemi, implica che **ogni algebra sia anche chiusa rispetto alle medesime operazioni applicate a qualsiasi sottocollezione finita di insiemi**, ovvero che

- e) se $\mathcal{C} = \{A_1, A_2, \dots, A_k\}$ è una sottocollezione finita di \mathcal{A} , allora l'unione $\bigcup_{i=1}^k A_i$ e l'intersezione $\bigcap_{i=1}^k A_i$ appartengono entrambe a \mathcal{A} .

Dimostrazione (della proprietà **e**) delle algebre). La proprietà **e**) delle algebre può essere facilmente dimostrata applicando ripetutamente le proprietà **c**) e **d**):

$$\begin{aligned}
 A_1 \cup A_2 \in \mathcal{A} &\Rightarrow (A_1 \cup A_2) \cup A_3 = A_1 \cup A_2 \cup A_3 \in \mathcal{A} \Rightarrow \dots \Rightarrow \\
 &\Rightarrow \left(\bigcup_{i=1}^{k-1} A_i \right) \cup A_k = \bigcup_{i=1}^k A_i \in \mathcal{A},
 \end{aligned}$$

e

$$\begin{aligned}
 A_1 \cap A_2 \in \mathcal{A} &\Rightarrow (A_1 \cap A_2) \cap A_3 = A_1 \cap A_2 \cap A_3 \in \mathcal{A} \Rightarrow \dots \Rightarrow \\
 &\Rightarrow \left(\bigcap_{i=1}^{k-1} A_i \right) \cap A_k = \bigcap_{i=1}^k A_i \in \mathcal{A}.
 \end{aligned}$$

Per scrivere una dimostrazione più formale si può fare riferimento al **principio di induzione**. A scopo illustrativo mostreremo solo come si applica il principio di induzione per

dimostrare che

$$A_1, A_2, \dots, A_k \in \mathcal{A} \quad \Rightarrow \quad \bigcup_{i=1}^k A_i \in \mathcal{A} \quad \text{per ogni } k = 1, 2, \dots \quad (1)$$

Secondo il principio di induzione, si può dimostrare che la suddetta implicazione è vera per ogni $k = 1, 2, \dots$ dimostrando che

- i) è vera per $k = 1$ (**base di induzione**)
- ii) e dimostrando che se è vera per un prefissato valore di $k = 1, 2, \dots$ (**ipotesi di induzione**), allora deve essere vera anche per $k + 1$ (**passo di induzione**).

Si noti che per dimostrare la **base di induzione** è sufficiente osservare che per ipotesi A_1 appartiene all'algebra \mathcal{A} . Per compiere il **passo di induzione** ipotizziamo ora che (**ipotesi di induzione**) $\bigcup_{i=1}^k A_i \in \mathcal{A}$ e sfruttiamo la proprietà **c)** delle algebre onde concludere che

$$\begin{aligned} \left(\bigcup_{i=1}^k A_i \right), A_{k+1} \in \mathcal{A} &\Rightarrow \text{[[proprietà c) delle algebre]]} \Rightarrow \\ &\Rightarrow \left(\bigcup_{i=1}^k A_i \right) \cup A_{k+1} = \bigcup_{i=1}^{k+1} A_i \in \mathcal{A}. \end{aligned}$$

Secondo il principio di induzione, l'implicazione (1) è dunque vera per ogni $k = 1, 2, \dots$ □

Come abbiamo appena dimostrato, ogni algebra è chiusa rispetto alle operazioni di unione e intersezione applicate a sottocollezioni finite di insiemi. Tuttavia, come dimostra l'Esempio 1.4 che vedremo tra breve, **un'algebra non è necessariamente chiusa rispetto alle operazioni di unione e intersezione applicate a sottocollezioni infinite di insiemi!!!** Con gli insiemi che appartengono ad un'algebra potrebbe a volte essere possibile formare collezioni infinite la cui unione e/o intersezione non è un insieme che appartiene ancora all'algebra (chiaramente ciò può accadere solo se l'algebra \mathcal{A} contiene infiniti insiemi). Un'algebra che è anche chiusa rispetto alle operazioni di unione e intersezione applicate a sottocollezioni di insiemi che sono **infinite numerabili** viene chiamata **σ -algebra**. Se gli insiemi di una σ -algebra possono essere interpretati come eventi (ovvero se l'insieme universale Ω viene considerato come spazio campionario associato ad un esperimento casuale), allora si parlerà di una **σ -algebra di eventi**. Si noti che per verificare se \mathcal{A} è una **σ -algebra** di sottoinsiemi di un dato insieme Ω , è sufficiente verificare che \mathcal{A} soddisfi le condizioni **a)**, **b)**, **c)** e la condizione

c*) se $\mathcal{C} = \{A_1, A_2, \dots\}$ è una sottocollezione infinita numerabile di \mathcal{A} , allora $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$.

Infatti, usando le versioni generali delle **leggi di De Morgan** si può facilmente dimostrare che le condizioni **b)** e **c*)** implicano anche la condizione

d*) se $\mathcal{C} = \{A_1, A_2, \dots\}$ è una sottocollezione infinita numerabile di \mathcal{A} , allora $\bigcap_{i=1}^{\infty} A_i \in \mathcal{A}$

secondo la quale ogni σ -algebra deve essere anche chiusa rispetto alle operazioni di intersezione applicate a sottocollezioni infinite numerabili.

Dimostrazione (della proprietà d*) delle σ -algre).

$$\begin{aligned} A_1, A_2, \dots \in \mathcal{A} &\Rightarrow \text{condizione b)} \Rightarrow \overline{A_1}, \overline{A_2}, \dots \in \mathcal{A} \Rightarrow \\ &\Rightarrow \text{condizione c*)} \Rightarrow \bigcup_{i=1}^{\infty} \overline{A_i} \in \mathcal{A} \Rightarrow \\ &\Rightarrow \text{condizione b)} \Rightarrow \overline{\left(\bigcup_{i=1}^{\infty} \overline{A_i}\right)} \in \mathcal{A} \Rightarrow \\ \Rightarrow \text{leggi di De Morgan: } &\overline{\left(\bigcup_{i=1}^{\infty} \overline{A_i}\right)} = \bigcap_{i=1}^{\infty} A_i \Rightarrow \bigcap_{i=1}^{\infty} A_i \in \mathcal{A}. \end{aligned}$$

Nell'ultimo passaggio abbiamo usato la legge di De Morgan secondo la quale $\overline{\left(\bigcup_{i=1}^{\infty} \overline{A_i}\right)} = \bigcap_{i=1}^{\infty} A_i$ con $\overline{A_i}$ al posto di A_i . Inoltre, abbiamo di nuovo usato il fatto che il complemento del complemento di un insieme è l'insieme di partenza. \square

Anche se a prima vista potrebbe sembrare che le condizioni c*) e d*) siano implicite nelle condizioni c) e d) (e che ogni algebra sia dunque allo stesso tempo anche una σ -algebra), il prossimo esempio dimostra che questo non è vero.

Esempio 1.4 (Esempio di un'algebra che non è una σ -algebra). Consideriamo un esperimento casuale che consiste in una successione **infinita** di lanci di una moneta. Anche se non potremo mai osservare l'esito finale di questo esperimento casuale, possiamo comunque tentare di descrivere questo esperimento casuale considerando come **eventi elementari** ω tutte le sequenze infinite di numeri "0" e "1", come per esempio la sequenza

$$\omega = (1, 0, 0, 0, 1, 1, 0, 1, 1, 1, 0, 0, 1, 0, 1, \dots).$$

Il numero "0" o "1" che compare nell' i -esima posizione di una sequenza di questo tipo ci dice se l'esito dell' i -esimo lancio è "testa" (in questo caso conveniamo che l' i -esima posizione sia occupata dal numero "1") oppure "croce" (in questo caso conveniamo che l' i -esima posizione sia occupata dal numero "0"). Come **spazio campionario** di un esperimento casuale che consiste in una successione infinita di lanci di una moneta possiamo quindi considerare l'insieme Ω di tutte le sequenze ω del suddetto tipo.

Ora, per definire uno spazio di eventi \mathcal{A} che soddisfi i condizioni a), b) e c) che lo qualificano come un'algebra di eventi, e che non soddisfi allo stesso tempo anche la

condizione c^*) che lo qualificherebbe come una σ -algebra di eventi, consideriamo **lo spazio di eventi \mathcal{A} che contiene tutti gli eventi che descrivono soltanto gli esiti di un numero finito di lanci**, ovvero tutti gli eventi $A \subseteq \Omega$ per i quali dopo un prefissato numero finito n_A di lanci possiamo sempre stabilire (qualunque sia l'esito degli n_A lanci) se si sono verificati o meno (si noti che il valore di n_A non è lo stesso per tutti gli eventi $A \in \mathcal{A}$).

Alcuni esempi di eventi che appartengono allo spazio di eventi \mathcal{A} sono:

- l'evento secondo il quale nel primo lancio si ottiene testa: questo evento è rappresentato dal sottoinsieme di Ω che contiene tutte le sequenze ω che hanno il numero "1" in prima posizione;
- l'evento secondo il quale nell' i -esimo lancio si ottiene testa: questo evento è rappresentato dal sottoinsieme di Ω che contiene tutte le sequenze ω che hanno il numero "1" in i -esima posizione;
- l'evento secondo il quale nei primi $n = 10$ lanci si ottiene $x = 5$ volte il lato con la testa: questo evento è rappresentato dal sottoinsieme di Ω che contiene tutte le sequenze ω nelle quali compare 5 volte il numero "1" e $10 - 5 = 5$ volte il numero "0" nelle prime dieci posizioni;
- l'evento secondo il quale si ottiene per la prima volta testa al quinto lancio: questo evento è rappresentato dal sottoinsieme di Ω che contiene tutte le sequenze ω nelle quali le prime quattro posizioni sono tutte occupate dal numero "0", e nelle quali la quinta posizione è occupata dal numero "1".
- ecc.

D'altra parte, l'evento secondo il quale si ottiene sempre soltanto l'esito testa in tutti i lanci non appartiene allo spazio di eventi \mathcal{A} perché anche se osservassimo sempre soltanto l'esito testa in una sequenza finita di lanci, non possiamo essere sicuri che nei lanci futuri osserveremo sempre soltanto l'esito testa. Altri esempi di eventi che non appartengono allo spazio di eventi \mathcal{A} sono:

- l'evento secondo il quale si ottiene testa in tutti i lanci pari e croce in tutti i lanci dispari: questo evento è rappresentato dal sottoinsieme di Ω che contiene soltanto la sequenza nella quale tutte le posizioni pari sono occupate dal numero "1" e tutte le posizioni dispari sono occupate dal numero "0";
- l'evento secondo il quale da un certo lancio in poi si ottiene sempre soltanto l'esito testa: questo evento è rappresentato dal sottoinsieme di Ω che contiene tutte le sequenze che contengono soltanto un numero finito di zeri.

- ecc.

Ora, per dimostrare che lo spazio di eventi \mathcal{A} soddisfa le condizioni a), b) e c) che lo qualificano come un'algebra di eventi basta osservare che

- la condizione a) è soddisfatta, ovvero che $\Omega \in \mathcal{A}$, perché già prima di effettuare il primo lancio sappiamo che l'evento Ω si verificherà, ovvero sappiamo che al termine dell'esperimento casuale si ottiene una sequenza $\omega \in \Omega$;
- la condizione b) è soddisfatta, ovvero che

$$A \in \mathcal{A} \quad \Rightarrow \quad \bar{A} \in \mathcal{A},$$

perché se dopo un prefissato numero finito n_A di lanci si può stabilire se un determinato evento A si è verificato o meno, allora dopo lo stesso numero di lanci si può anche stabilire se si è verificato l'evento complementare \bar{A} ;

- la condizione c) è soddisfatta, ovvero che

$$A_1, A_2 \in \mathcal{A} \quad \Rightarrow \quad A_1 \cup A_2 \in \mathcal{A},$$

perché se dopo un prefissato numero finito n_1 di lanci possiamo stabilire se l'evento A_1 si è verificato o meno, e dopo un prefissato numero finito n_2 di lanci possiamo stabilire se l'evento A_2 si è verificato o meno, allora dopo $n = \max\{n_1, n_2\}$ lanci possiamo anche stabilire se almeno uno di questi due eventi si è verificato oppure se nessuno dei due si è verificato, ovvero possiamo stabilire se si è verificato l'evento $A_1 \cup A_2$.

Questi ragionamenti dimostrano quindi che lo spazio di eventi \mathcal{A} soddisfa le condizioni a), b) e c) che lo qualificano come un'algebra di eventi. D'altra parte, non è difficile dimostrare che lo spazio di eventi \mathcal{A} non soddisfa anche la condizione c*) che lo qualificerebbe come una σ -algebra di eventi. Infatti, se indichiamo con A_i l'evento secondo il quale l'esito dell' i -esimo lancio è testa, allora $A_i \in \mathcal{A}$ per ogni $i = 1, 2, \dots$ (vedi sopra), ma l'evento $B = \bigcup_{i=1}^{\infty} A_i$ non appartiene all'algebra di eventi \mathcal{A} perché sulla base di un prefissato numero finito di lanci non possiamo sempre stabilire se l'evento B si è verificato o meno: infatti, se in una sequenza finita di lanci osserviamo sempre soltanto l'esito croce, non possiamo escludere che in almeno uno dei lanci futuri non osserveremo l'esito testa (si noti che l'evento $B = \bigcup_{i=1}^{\infty} B_i$ è l'evento secondo il quale in almeno uno degli infiniti lanci si ottiene l'esito testa).

Nei prossimi due esempi introdurremo alcune σ -algebre che per la teoria della probabilità sono di fondamentale importanza.

Esempio 1.5 (La σ -algebra di Borel su \mathbb{R}). Consideriamo l'insieme dei numeri reali \mathbb{R} . Come è noto, $\Omega = \mathbb{R}$ è un insieme infinito *non numerabile* e con un insieme così ricco di elementi si possono definire tantissime collezioni di sottoinsiemi che hanno la struttura di un'algebra e/o di una σ -algebra.

Come vedremo in questa dispensa, in ambito probabilistico (e non solo) sono di fondamentale importanza le seguenti collezioni di sottoinsiemi di \mathbb{R} :

- La collezione $\mathcal{B} = \{(-\infty, x] : x \in \mathbb{R}\}$ che contiene tutti gli intervalli chiusi e illimitati inferiormente, ovvero tutti gli intervalli del tipo $(-\infty, x]$ con $x \in \mathbb{R}$,
- ... e la cosiddetta **σ -algebra di Borel** che indicheremo con $\sigma(\mathcal{B})$ e che può essere *definita* come la più piccola σ -algebra che include la collezione \mathcal{B} che abbiamo definito al punto precedente.

Si può dimostrare che la σ -algebra di Borel è strettamente inclusa nell'insieme delle parti $\mathcal{P}(\mathbb{R})$ (ovvero che esistono sottoinsiemi di \mathbb{R} che non appartengono a $\sigma(\mathcal{B})$). Ciò nonostante, la σ -algebra di Borel è una collezione di sottoinsiemi di \mathbb{R} molto ampia che contiene tutti i sottoinsiemi di \mathbb{R} che potrebbero avere una qualche rilevanza pratica nelle applicazioni. Infatti, non è difficile dimostrare che $\sigma(\mathcal{B})$ contiene tutti gli intervalli di numeri reali (per intervalli di numeri reali intendiamo intervalli aperti/chiusi/semiaperti che possono essere limitati o anche illimitati) e anche tutte le unioni di collezioni numerabili (finite e infinite) di intervalli. In molti libri di testo la σ -algebra di Borel (su \mathbb{R}) viene definita come la più piccola σ -algebra che contiene tutti i sottoinsiemi aperti di \mathbb{R} . Si può dimostrare che questa definizione è equivalente alla definizione che abbiamo dato poco sopra.

Esempio 1.6 (La σ -algebra di Borel su \mathbb{R}^k). Consideriamo ora invece \mathbb{R}^k , ovvero l'insieme di tutti i vettori (x_1, x_2, \dots, x_k) composti da un numero finito k di numeri reali. Ovviamente, anche \mathbb{R}^k è un insieme infinito non numerabile e anche con \mathbb{R}^k si possono costruire tantissime collezioni di sottoinsiemi che hanno la struttura di un'algebra e/o di una σ -algebra. Come vedremo più avanti, in ambito probabilistico (e non solo) sono di fondamentale importanza le seguenti collezioni:

- La collezione $\mathcal{B}_k = \{\prod_{i=1}^k (-\infty, x_i] : (x_1, x_2, \dots, x_k) \in \mathbb{R}^k\}$ che contiene tutti i prodotti cartesiani di intervalli chiusi e illimitati inferiormente, ovvero tutti i sottoinsiemi $A = \prod_{i=1}^k (-\infty, x_i]$ di \mathbb{R}^k che si possono ottenere fissando un punto $(x_1, x_2, \dots, x_k) \in \mathbb{R}^k$ e prendendo come elementi di A tutti i punti $(x'_1, x'_2, \dots, x'_k) \in \mathbb{R}^k$ tali che $x'_1 \leq x_1, x'_2 \leq x_2, \dots$ e $x'_k \leq x_k$ (se $k = 2$, l'insieme $A = \prod_{i=1}^2 (-\infty, x_i] = (-\infty, x_1] \times (-\infty, x_2]$ è l'insieme di tutti i punti (x'_1, x'_2) che in un piano cartesiano si trovano a "sud-ovest" rispetto al punto

(x_1, x_2) .

- La **σ -algebra di Borel su \mathbb{R}^k** che può essere definita come la più piccola σ -algebra che include la collezione \mathcal{B}_k che abbiamo definito al punto precedente. D'ora in poi indicheremo questa σ -algebra con $\sigma(\mathcal{B}_k)$ tranne nel caso in cui $k = 1$, dove scriveremo semplicemente $\sigma(\mathcal{B})$ così come abbiamo fatto nel precedente Esempio 1.5.

Com'è facile intuire, anche nel caso $k > 1$ la σ -algebra $\sigma(\mathcal{B}_k)$ è una collezione molto ampia che contiene tutti i sottoinsiemi di \mathbb{R}^k che potrebbero avere una qualche rilevanza pratica nelle applicazioni, ma anche per $k > 1$ la σ -algebra di Borel $\sigma(\mathcal{B}_k)$ è strettamente inclusa nell'insieme delle parti $\mathcal{P}(\mathbb{R}^k)$. Infine, aggiungiamo che anche per $k > 1$ la σ -algebra di Borel $\sigma(\mathcal{B}_k)$ potrebbe essere definita come la più piccola σ -algebra che contiene tutti i sottoinsiemi aperti di \mathbb{R}^k (in altre parole: la più piccola σ -algebra che contiene tutti i sottoinsiemi aperti di \mathbb{R}^k coincide con la più piccola σ -algebra che include la collezione \mathcal{B}_k che abbiamo definito poco sopra).

Negli ultimi due esempi abbiamo applicato un metodo che viene spesso utilizzato per definire delle σ -algre: in entrambi gli esempi siamo partiti da un insieme Ω e da una ristretta collezione \mathcal{A} di sottoinsiemi di Ω , e sulla base di \mathcal{A} abbiamo definito una σ -algebra $\sigma(\mathcal{A})$ stabilendo che $\sigma(\mathcal{A})$ fosse la più piccola σ -algebra che contiene tutti i sottoinsiemi di Ω che appartengono alla collezione di partenza \mathcal{A} . Tuttavia, a pensarci bene potremmo aver commesso una leggerezza, perché **non ci siamo preoccupati di verificare se la σ -algebra $\sigma(\mathcal{A})$ esiste, e in caso affermativo, se essa è unica!!!!** La questione dell'*esistenza* e dell'*unicità* di $\sigma(\mathcal{A})$ è di fondamentale importanza perché ...

- ... se $\sigma(\mathcal{A})$ non dovesse esistere (ovvero se per ogni σ -algebra che include \mathcal{A} ne esiste sempre un'altra che è strettamente inclusa nella prima), allora assumendo l'esistenza di una più piccola σ -algebra che include la collezione di insiemi \mathcal{A} avremmo *introdotto una contraddizione* e, com'è noto, *facendo leva su una contraddizione si può dimostrare qualsiasi affermazione!!!!* ...
- ... se $\sigma(\mathcal{A})$ non fosse unica (ovvero se esistono due o più σ -algre diverse che includono \mathcal{A} tali che nessuna sottocollezione di queste σ -algre è ancora una σ -algebra che include \mathcal{A}), allora il simbolo $\sigma(\mathcal{A})$ si riferirebbe a tante σ -algre diverse.

Come si desume dal prossimo teorema, la nostra preoccupazione sull'esistenza e l'unicità delle σ -algre di Borel è infondata.

Teorema 1.3 (Esistenza e unicità della σ -algebra generata da una collezione di sottoinsiemi). Se \mathcal{A} è una qualunque collezione di sottoinsiemi di un dato insieme

Ω , allora **esiste** sempre un'**unica** più piccola σ -algebra che include la collezione \mathcal{A} . Questa σ -algebra viene chiamata " **σ -algebra generata da \mathcal{A}** ". La σ -algebra generata da una data collezione \mathcal{A} viene indicata con il simbolo $\sigma(\mathcal{A})$.

Chiaramente, se la collezione \mathcal{A} ha già la struttura di una σ -algebra, allora $\sigma(\mathcal{A}) = \mathcal{A}$, ma come si può facilmente immaginare (vedi il prossimo esercizio), in generale la σ -algebra generata da una collezione di sottoinsiemi di un dato insieme Ω potrebbe contenere molti più sottoinsiemi di Ω della collezione di partenza.

Esercizio 1.1. Sia $\Omega = \{1, 2, 3, 4\}$ e \mathcal{A} la collezione di sottoinsiemi di Ω che contiene gli insiemi $\{1\}$, $\{2\}$, $\{3, 4\}$. Si elenchino tutti gli insiemi che appartengono a $\sigma(\mathcal{A})$.

La soluzione viene lasciata per esercizio.

1.5 Le leggi del calcolo delle probabilità

In questa sezione presenteremo le principali **leggi del calcolo delle probabilità**, ovvero alcuni fondamentali **teoremi** che possono essere dimostrati a partire dagli assiomi di Kolmogorov. Negli enunciati di queste leggi assumeremo sempre che $P(\cdot)$ sia una funzione di probabilità che soddisfa gli assiomi **K0 - K3** e l'assioma **K3***. Come previsto nell'assioma **K0**, assumeremo quindi che il dominio \mathcal{A} della funzione di probabilità $P(\cdot)$ sia un'algebra di eventi e che dunque sia chiuso rispetto a operazioni di unione e intersezione applicate a qualsiasi sottocollezione finita di eventi (proprietà **e**) delle algebre). Prima di procedere conviene aggiungere un'altra osservazione: in realtà, tutte le leggi che non coinvolgono infiniti eventi possono essere dimostrate senza l'assioma **K3***. Il seguente elenco contiene dunque soltanto due leggi del calcolo delle probabilità la cui dimostrazione necessita anche dell'assioma **K3***: la legge **L10** e la legge **L11** (per queste due leggi non forniremo una dimostrazione).

L1) La probabilità dell'**evento impossibile** è sempre nulla: $P(\emptyset) = 0$.

Dimostrazione. Siccome \emptyset e Ω sono due eventi incompatibili ($\emptyset \cap \Omega = \emptyset$), possiamo concludere che

$$P(\Omega) = P(\emptyset \cup \Omega) = \text{[[assioma K3]]} = P(\emptyset) + P(\Omega) \Rightarrow P(\emptyset) = 0.$$

□

L2) (Additività finita) Se A_1, A_2, \dots, A_k sono $k < \infty$ eventi incompatibili che appartengono tutti all'algebra \mathcal{A} , allora si ha

$$P\left(\bigcup_{i=1}^k A_i\right) = \sum_{i=1}^k P(A_i).$$

Dimostrazione. La legge L2 può essere facilmente dimostrata usando ripetutamente l'assioma K3. Per scrivere una dimostrazione rigorosa basta invocare il principio di induzione. □

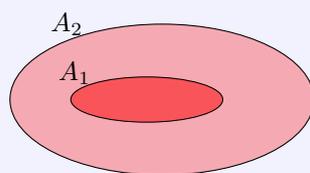
L3) Per ogni $A \in \mathcal{A}$ si ha $P(\bar{A}) = 1 - P(A)$.

Dimostrazione. Siccome $A \in \mathcal{A} \Rightarrow \bar{A} \in \mathcal{A}$ (condizione b) delle algebre), $\Omega = A \cup \bar{A} \in \mathcal{A}$ (condizione a) delle algebre) e $A \cap \bar{A} = \emptyset$, possiamo concludere che

$$\begin{aligned} P(\Omega) = P(A \cup \bar{A}) &= [[\text{assioma K3}]] = P(A) + P(\bar{A}) \Rightarrow P(\bar{A}) = P(\Omega) - P(A) \Rightarrow \\ &\Rightarrow [[\text{assioma K2}]] \Rightarrow P(\bar{A}) = 1 - P(A). \end{aligned}$$

□

L4) Se $A_1, A_2 \in \mathcal{A}$ e se $A_1 \subseteq A_2$, allora $0 \leq P(A_1) \leq P(A_2)$. In particolare, per ogni $A \in \mathcal{A}$ si deve avere $0 \leq P(A) \leq 1$.



Dimostrazione. Siccome (si veda il diagramma di Venn)

$$A_1 \subseteq A_2 \Rightarrow A_2 = A_1 \cup (A_2 \cap \bar{A}_1),$$

e siccome gli eventi A_1 e $A_2 \cap \bar{A}_1$ sono incompatibili e appartengono entrambi a \mathcal{A} (ricordiamo che tutte le algebre sono chiuse rispetto a operazioni di complementazione e anche rispetto a operazioni di unione e intersezione applicate a sottocollezioni finite), possiamo concludere che

$$P(A_2) = [[\text{assioma K3}]] = P(A_1) + P(\bar{A}_1 \cap A_2).$$

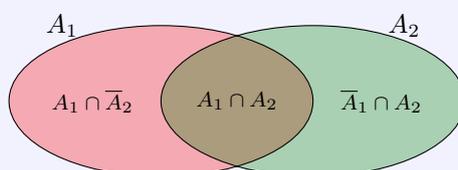
Tenendo presente che una probabilità non può mai essere negativa (assioma **K1**), vediamo dunque che $0 \leq P(A_1) \leq P(A_2)$. Questo dimostra la prima parte della legge **L4**. Ponendo ora $A_1 = A$ e $A_2 = \Omega$, otteniamo anche le disuguaglianze

$$0 \leq P(A) \leq P(\Omega) = \text{[[assioma K2]]} = 1.$$

□

L5) Se A_1, A_2 sono due eventi che appartengono entrambi all'algebra \mathcal{A} (A_1 e A_2 possono essere compatibili o anche incompatibili), allora la probabilità della loro unione può essere calcolata come

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2).$$



Dimostrazione. Siccome (si veda il diagramma di Venn)

$$A_1 = (A_1 \cap \bar{A}_2) \cup (A_1 \cap A_2)$$

e siccome i due eventi $(A_1 \cap \bar{A}_2)$ e $(A_1 \cap A_2)$ sono incompatibili e appartengono entrambi all'algebra \mathcal{A} (si ricordi che le algebre sono chiuse rispetto a unioni e intersezioni applicate a sottocollezioni finite), possiamo concludere che

$$P(A_1) = P((A_1 \cap \bar{A}_2) \cup (A_1 \cap A_2)) = \text{[[assioma K3]]} = P(A_1 \cap \bar{A}_2) + P(A_1 \cap A_2). \quad (2)$$

Scambiando i ruoli degli eventi A_1 e A_2 otteniamo anche l'uguaglianza

$$P(A_2) = P(\bar{A}_1 \cap A_2) + P(A_1 \cap A_2). \quad (3)$$

Usando ora il fatto che (si verifichi sul diagramma di Venn)

$$A_1 \cup A_2 = (A_1 \cap \bar{A}_2) \cup (A_1 \cap A_2) \cup (\bar{A}_1 \cap A_2),$$

e il fatto che $(A_1 \cap \bar{A}_2)$, $(A_1 \cap A_2)$ e $(\bar{A}_1 \cap A_2)$ sono eventi mutuamente incompatibili che appartengono all'algebra \mathcal{A} (si ricordi che tutte le algebre sono chiuse rispetto all'operazione di complementazione e anche rispetto all'operazione di intersezione applicata a sottocollezioni finite di eventi), vediamo che secondo l'assioma **K3** deve essere soddisfatta anche l'uguaglianza

$$P(A_1 \cup A_2) = P((A_1 \cap \bar{A}_2) \cup (A_1 \cap A_2) \cup (\bar{A}_1 \cap A_2)) \\ \text{[[assioma K3]]} = P(A_1 \cap \bar{A}_2) + P(A_1 \cap A_2) + P(\bar{A}_1 \cap A_2).$$

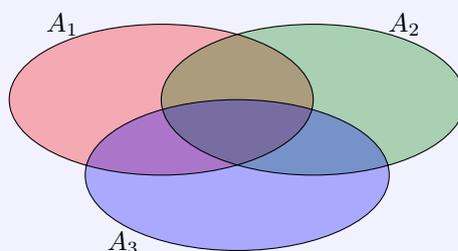
Combinando questo risultato con le precedenti equazioni (2) e (3) otteniamo dunque l'uguaglianza

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$$

che volevamo dimostrare. \square

L6) (Formula di inclusione-esclusione) Se $A_1, A_2, A_3 \in \mathcal{A}$ (gli eventi A_1, A_2 e A_3 possono essere compatibili o anche incompatibili), allora la probabilità della loro unione può essere calcolata come

$$\begin{aligned} P(A_1 \cup A_2 \cup A_3) = & P(A_1) + P(A_2) + P(A_3) + \\ & - P(A_1 \cap A_2) - P(A_1 \cap A_3) - P(A_2 \cap A_3) + \\ & + P(A_1 \cap A_2 \cap A_3). \end{aligned}$$



Più in generale, la probabilità dell'unione di un qualsiasi numero finito k di eventi (che appartengono tutti all'algebra \mathcal{A} e che possono essere compatibili o anche incompatibili) può essere calcolata come

- 1) $(-1)^{1-1}$ =più la somma delle probabilità dei singoli eventi
- 2) $(-1)^{2-1}$ =meno la somma delle probabilità di tutte le "intersezioni a due"
- 3) $(-1)^{3-1}$ =più la somma delle probabilità di tutte le "intersezioni a tre"
- 4) $(-1)^{4-1}$ =meno ... ,
- ⋮) ...
- n) più/meno (a seconda del segno di $(-1)^{k-1}$) la probabilità dell'intersezione di tutti gli k eventi.

Dimostrazione. La legge L6 può essere dimostrata attraverso il principio di induzione. Omettiamo i dettagli della dimostrazione. \square

L7) (Disuguaglianza di Boole per due eventi) Se $A_1, A_2 \in \mathcal{A}$, allora

$$P(A_1 \cup A_2) \leq P(A_1) + P(A_2).$$

Dimostrazione. La dimostrazione segue immediatamente dalla legge L5 e dal fatto che $P(A_1 \cap A_2) \geq 0$ per l'assioma K1. \square

L8) (Disuguaglianza di Boole per k eventi) Se $A_1, A_2, \dots, A_k \in \mathcal{A}$, allora

$$P\left(\bigcup_{i=1}^k A_i\right) \leq \sum_{i=1}^k P(A_i).$$

Questa disuguaglianza rimane valida anche per una successione infinita di eventi a patto che $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$ (come abbiamo visto nella sezione precedente, questa condizione è sempre soddisfatta se \mathcal{A} è una σ -algebra).

Dimostrazione. Anche la legge L8 può essere facilmente dimostrata attraverso il principio di induzione. Omettiamo i dettagli della dimostrazione. \square

L9) (Disuguaglianza di Bonferroni) Se $A_1, A_2, \dots, A_k \in \mathcal{A}$, allora

$$P\left(\bigcap_{i=1}^k A_i\right) \geq 1 - \sum_{i=1}^k P(\bar{A}_i).$$

Questa disuguaglianza rimane valida anche per una successione infinita di eventi a patto che $\bigcap_{i=1}^{\infty} A_i \in \mathcal{A}$ (come abbiamo visto nella sezione precedente, questa condizione è sempre soddisfatta se \mathcal{A} è una σ -algebra).

Dimostrazione. Applicando la disuguaglianza di Boole otteniamo

$$P\left(\bigcup_{i=1}^k \bar{A}_i\right) \leq \sum_{i=1}^k P(\bar{A}_i), \quad (4)$$

e siccome (leggi di De Morgan)

$$\bigcup_{i=1}^k \bar{A}_i = \overline{\left(\bigcap_{i=1}^k A_i\right)},$$

possiamo concludere che (legge L3)

$$P\left(\bigcup_{i=1}^k \bar{A}_i\right) = 1 - P\left(\bigcap_{i=1}^k A_i\right).$$

Sostituendo questo risultato nella (4) otteniamo

$$1 - P\left(\bigcap_{i=1}^k A_i\right) \leq \sum_{i=1}^k P(\bar{A}_i)$$

e quest'ultima disuguaglianza può essere riscritta come

$$P\left(\bigcap_{i=1}^k A_i\right) \geq 1 - \sum_{i=1}^k P(\bar{A}_i).$$

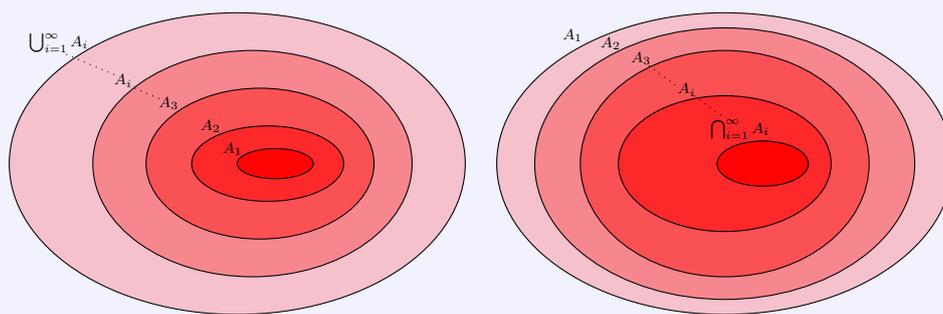
□

L10) Se $A_1 \subseteq A_2 \subseteq \dots$ è una successione infinita numerabile e "non decrescente" di eventi che appartengono tutti all'algebra \mathcal{A} , e se anche l'unione $\bigcup_{i=1}^{\infty} A_i$ appartiene all'algebra \mathcal{A} , allora

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{n \rightarrow \infty} P(A_n).$$

L11) Se $A_1 \supseteq A_2 \supseteq \dots$ è una successione infinita numerabile e "non crescente" di eventi che appartengono tutti all'algebra \mathcal{A} , e se anche l'intersezione $\bigcap_{i=1}^{\infty} A_i$ appartiene all'algebra \mathcal{A} , allora

$$P\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_{n \rightarrow \infty} P(A_n).$$



Si noti che le condizioni $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$ e $\bigcap_{i=1}^{\infty} A_i \in \mathcal{A}$ sono entrambe soddisfatte se \mathcal{A} è una σ -algebra.

Per non appesantire troppo la trattazione omettiamo le dimostrazioni delle leggi L10 e L11.

Avendo visto le principali leggi del calcolo delle probabilità, possiamo finalmente dimostrare anche la seconda parte del [teorema delle scommesse olandesi](#).

Dimostrazione (della seconda parte del [teorema delle scommesse olandesi](#)). Per completare la dimostrazione del [teorema delle scommesse olandesi](#) dobbiamo ancora dimostrare che **un soggetto non è esposto a scommesse che gli procurerebbero una perdita certa se fissa i prezzi $P(A)$ in modo tale che siano soddisfatte le condizioni poste negli assiomi K1 - K3** (si noti che in questo caso i prezzi delle scommesse devono anche soddisfare le leggi L1 - L9 perché queste leggi necessitano soltanto degli assiomi K1 - K3 per le loro dimostrazioni). Dimosteremo questa affermazione sotto l'ipotesi che un soggetto possa essere costretto ad impegnarsi soltanto in scommesse su un numero finito di eventi (ipotesi che nel mondo reale è senz'altro soddisfatta). Assumiamo dunque che un soggetto sia impegnato in scommesse su un determinato numero finito k di eventi. Supponiamo inizialmente che $k = 1$ e indichiamo con A_1 l'evento sul quale il soggetto ha scommesso e con S_1 il numero di titoli di scommesse sull'evento A_1 che il soggetto ha acquistato (se $S_1 > 0$) o emesso (se $S_1 < 0$). In quanto segue non escludiamo la possibilità che S_1 non sia un numero intero (il soggetto può quindi acquistare e/o emettere anche soltanto una quota parte di un titolo). Siccome stiamo ipotizzando che $P(A_1)$ rappresenti il prezzo di un singolo titolo, possiamo dire quanto segue:

- Se $S_1 > 0$, ovvero se il soggetto acquista S_1 titoli al prezzo $P(A_1)$ cadauno, allora il soggetto ha un esborso immediato di $S_1 \times P(A_1)$ euro e riceverà S_1 euro nel caso in cui l'evento A_1 si verificasse.
- D'altra parte, se $S_1 < 0$, ovvero se il soggetto emette $|S_1| = -S_1$ titoli di scommesse sull'evento A_1 , allora il soggetto ha un incasso immediato di $|S_1| \times P(A_1) = -S_1 \times P(A_1)$ euro e dovrà pagare $|S_1| = -S_1$ euro nel caso in cui l'evento A_1 si verificasse.

Non è difficile rendersi conto che in entrambi i casi appena considerati il guadagno netto del soggetto è dato da

$$G = \begin{cases} S_1 - S_1 P(A_1) & \text{se si verifica l'evento } A_1 \\ -S_1 P(A_1) & \text{se si verifica l'evento } \bar{A}_1. \end{cases}$$

Supponiamo ora per assurdo che il guadagno netto G sia negativo in ogni evenienza, ovvero che i valori di $S_1 - S_1 P(A_1)$ e $-S_1 P(A_1)$ siano entrambi negativi. Per dimostrare che questo non è possibile se i prezzi $P(\cdot)$ soddisfano gli assiomi K1 - K3, osserviamo che se il guadagno netto G fosse negativo in ogni evenienza, allora dovrebbe essere negativo anche il **guadagno netto atteso** che è definito come

$$\bar{G} = [S_1 - S_1 P(A_1)] \times P(A_1) + [-S_1 \times P(A_1)] \times P(\bar{A}_1).$$

Ma siccome

$$\begin{aligned} \bar{G} &= [S_1 - S_1 P(A_1)] \times P(A_1) + [-S_1 P(A_1)] \times P(\bar{A}_1) \\ &= S_1 \times P(A_1) - S_1 P(A_1) \times [P(A_1) + P(\bar{A}_1)] \\ \text{[[assioma K3]]} &= S_1 \times P(A_1) - S_1 P(A_1) \times P(A_1 \cup \bar{A}_1) \\ \text{[[} A_1 \cup \bar{A}_1 = \Omega \text{]]} &= S_1 \times P(A_1) - S_1 P(A_1) \times P(\Omega) \\ \text{[[assioma K2]]} &= S_1 \times P(A_1) - S_1 P(A_1) \times 1 = 0, \end{aligned}$$

possiamo concludere che il guadagno netto G non può essere negativo in ogni evenienza, ovvero che il soggetto non è esposto a scommesse che gli procurerebbero una perdita certa se i prezzi $P(\cdot)$ soddisfano gli assiomi K1 - K3 e se acquista e/o emette titoli di scommesse che si riferiscono ad un unico evento.

Generalizzando il precedente ragionamento si può anche dimostrare che il soggetto non è esposto a scommesse che gli procurerebbero una perdita certa se i prezzi $P(\cdot)$ soddisfano gli assiomi K1 - K3 e se acquista e/o emette titoli di scommesse che si riferiscono ad un generico numero finito k di eventi. Per rendere più chiaro il modo in cui il ragionamento deve essere generalizzato, conviene considerare prima il caso in cui $k = 2$ e soltanto dopo il caso generico con $k \geq 2$.

Consideriamo dunque un soggetto che acquista e/o emette titoli di scommesse che si riferiscono a $k = 2$ eventi che indichiamo con A_1 e A_2 . Per $i = 1, 2$ indichiamo con S_i il numero di titoli di scommesse sull'evento A_i che il soggetto acquista (se $S_i > 0$) o emette (se $S_i < 0$). Anche in questo caso ammettiamo la possibilità che i valori di S_1 e/o S_2 non siano dei numeri interi. Usando la notazione appena introdotta possiamo esprimere il guadagno netto di questo soggetto come

$$G = \begin{cases} S_1 + S_2 - S_1P(A_1) - S_2P(A_2) & \text{se si verifica l'evento } A_1 \cap A_2 \\ S_1 - S_1P(A_1) - S_2P(A_2) & \text{se si verifica l'evento } A_1 \cap \bar{A}_2 \\ S_2 - S_1P(A_1) - S_2P(A_2) & \text{se si verifica l'evento } \bar{A}_1 \cap A_2 \\ -S_1P(A_1) - S_2P(A_2) & \text{se si verifica l'evento } \bar{A}_1 \cap \bar{A}_2. \end{cases}$$

Si noti che i quattro casi che occorre distinguere per esprimere il guadagno netto G sono mutuamente incompatibili e che uno e soltanto uno di essi deve verificarsi. In altre parole, gli eventi $(A_1 \cap A_2)$, $(A_1 \cap \bar{A}_2)$, $(\bar{A}_1 \cap A_2)$ e $(\bar{A}_1 \cap \bar{A}_2)$ formano una partizione finita dello spazio campionario Ω e per la legge [L2](#) deve quindi essere soddisfatta l'uguaglianza

$$P(A_1 \cap A_2) + P(A_1 \cap \bar{A}_2) + P(\bar{A}_1 \cap A_2) + P(\bar{A}_1 \cap \bar{A}_2) = P(\Omega) = [\text{assioma K2}] = 1. \quad (5)$$

Assumiamo ora per assurdo che il guadagno netto G sia negativo in ogni evenienza. Anche in questo caso possiamo dimostrare che questa ipotesi non è compatibile con gli assiomi [K1](#) - [K3](#) dimostrando che il guadagno netto atteso

$$\begin{aligned} \bar{G} &= [S_1 + S_2 - S_1P(A_1) - S_2P(A_2)] \times P(A_1 \cap A_2) + \\ &+ [S_1 - S_1P(A_1) - S_2P(A_2)] \times P(A_1 \cap \bar{A}_2) + \\ &+ [S_2 - S_1P(A_1) - S_2P(A_2)] \times P(\bar{A}_1 \cap A_2) + \\ &+ [-S_1P(A_1) - S_2P(A_2)] \times P(\bar{A}_1 \cap \bar{A}_2). \end{aligned}$$

è nullo: infatti, se il guadagno netto fosse negativo in ogni evenienza, allora tutte le espressioni tra parentesi quadre sarebbero negative, e siccome i pesi $P(A_1 \cap A_2)$, $P(A_1 \cap \bar{A}_2)$, $P(\bar{A}_1 \cap A_2)$ e $P(\bar{A}_1 \cap \bar{A}_2)$ non possono essere negativi (assioma [K1](#)) e in forza della (5) almeno uno di essi deve essere positivo, il valore di \bar{G} dovrebbe essere negativo. Tuttavia, semplici passaggi algebrici mostrano che

$$\begin{aligned} \bar{G} &= S_1 \times [P(A_1 \cap A_2) + P(A_1 \cap \bar{A}_2)] + S_2 \times [P(A_1 \cap A_2) + P(\bar{A}_1 \cap A_2)] + \\ &+ [-S_1P(A_1) - S_2P(A_2)] \times [P(A_1 \cap A_2) + P(A_1 \cap \bar{A}_2) + P(\bar{A}_1 \cap A_2) + P(\bar{A}_1 \cap \bar{A}_2)] \\ &[[A_1 = (A_1 \cap A_2) \cup (A_1 \cap \bar{A}_2), A_2 = (A_1 \cap A_2) \cup (\bar{A}_1 \cap A_2), \text{assioma K3, equ. (5)}]] \\ &= S_1 \times P(A_1) + S_2 \times P(A_2) + [-S_1P(A_1) - S_2P(A_2)] \times P(\Omega) \\ &[[\text{assioma K2}]] \\ &= S_1 \times P(A_1) + S_2 \times P(A_2) + [-S_1P(A_1) - S_2P(A_2)] \times 1 = 0 \end{aligned}$$

e quindi possiamo concludere che un soggetto non è esposto a scommesse che gli procurerebbero una perdita certa se i prezzi $P(\cdot)$ soddisfano gli assiomi [K1](#) - [K3](#) e se acquista e/o emette titoli di scommesse che si riferiscono a $k = 2$ eventi.

A questo punto possiamo finalmente generalizzare. Consideriamo dunque un soggetto che acquista e/o emette titoli di scommesse che si riferiscono ad un generico numero finito k di eventi. Come in precedenza indichiamo con A_1, A_2, \dots, A_k i k eventi, e con S_i il numero di titoli di scommesse sull'evento A_i che il soggetto acquista (se $S_i > 0$) o emette (se $S_i < 0$). Notiamo che attraverso i k eventi A_i si può costruire una partizione finita dello spazio campionario Ω definendo gli eventi della partizione come

$$E_1 \cap E_2 \cap \dots \cap E_k$$

con $E_i = A_i$ oppure $E_i = \bar{A}_i$. Si noti che la partizione è composta al più da 2^k intersezioni del suddetto tipo perché ci sono 2 modi per definire ciascuno dei k eventi E_i (quindi se $k = 1$ la partizione è composta solo dai $2^1 = 2$ eventi A_1 e \bar{A}_1 ; se $k = 2$ la partizione è composta dai $2^2 = 4$ eventi $A_1 \cap A_2$, $A_1 \cap \bar{A}_2$, $\bar{A}_1 \cap A_2$ e $\bar{A}_1 \cap \bar{A}_2$; se $k = 3$ la partizione è composta dai $2^3 = 8$ eventi ...). Siccome alcune delle suddette intersezioni potrebbero essere vuote, il numero di eventi della partizione potrebbe essere strettamente minore di 2^k . Ora, siccome il numero degli eventi della partizione è finito, possiamo applicare la legge **L2** onde concludere che la somma dei prezzi delle scommesse associate agli eventi della partizione deve essere uguale al prezzo della scommessa sull'evento certo Ω e quindi uguale a 1 (assioma **K2**):

$$\sum_{E_1, E_2, \dots, E_k} P(E_1 \cap E_2 \cap \dots \cap E_k) = P(\Omega) = 1. \quad (6)$$

La sommatoria che compare nella precedente formula si estende tutti i modi possibili per definire gli eventi E_i (ricordiamo che ciascun evento E_i può essere definito solo come $E_i = A_i$ oppure come $E_i = \bar{A}_i$).

Supponiamo ora che al termine dell'esperimento casuale si verifichino gli eventi $A_{i_1}, A_{i_2}, \dots, A_{i_r}$ (dove $i_1 < i_2 < \dots < i_r$ sono r pedici compresi tra 1 e k) e che tutti gli altri eventi A_i non si verifichino. Si noti che questa ipotesi equivale ad assumere che si verifichi l'evento $E_1 \cap E_2 \cap \dots \cap E_k$ con

$$E_{i_1} = A_{i_1}, \quad E_{i_2} = A_{i_2}, \quad \dots, \quad E_{i_r} = A_{i_r}$$

e con tutti gli altri eventi E_i definiti come $E_i = \bar{A}_i$ (questa ipotesi non esclude il caso in cui $E_i = \bar{A}_i$ per ogni $i = 1, 2, \dots, k$, ovvero in caso in cui non si verifica nessuno dei k eventi A_i). In questo caso il guadagno netto G sarebbe dato da

$$G = (S_{i_1} + S_{i_2} + \dots + S_{i_r}) - \sum_{i=1}^k S_i P(A_i)$$

perché $\sum_{i=1}^k S_i P(A_i)$ è l'importo che il soggetto ha pagato (se positivo) oppure incassato (se negativo) nel momento in cui ha acquistato e/o emesso i titoli delle scommesse, e perché $S_{i_1} + S_{i_2} + \dots + S_{i_r}$ è la somma che il soggetto incassa o paga per le vincite associate alle scommesse sugli eventi $A_{i_1}, A_{i_2}, \dots, A_{i_r}$ che si sono verificati (visto che gli altri eventi A_i non si sono verificati, per i corrispondenti titoli il soggetto non riceve e non paga nulla). Da questo ragionamento desumiamo che se si verifica un generico evento $E_1 \cap E_2 \cap \dots \cap E_k$ della partizione di Ω , allora il corrispondente guadagno netto è dato da

$$G = \sum_{i: E_i = A_i} S_i - \sum_{i=1}^k S_i P(A_i),$$

dove la sommatoria $\sum_{i: E_i = A_i}$ è estesa a tutti gli indici i tali che $E_i = A_i$.

Chiaramente, se il guadagno netto G fosse negativo *in ogni evenienza* (ovvero se per qualsiasi modo di definire gli eventi E_i si avesse $G < 0$), allora dovrebbe essere negativo anche il guadagno netto atteso

$$\bar{G} = \sum_{E_1, E_2, \dots, E_k} \left[\sum_{i: E_i = A_i} S_i - \sum_{i=1}^k S_i P(A_i) \right] \times P(E_1 \cap E_2 \cap \dots \cap E_k),$$

perché in forza dell'assioma **K1** nessun peso $P(E_1 \cap E_2 \cap \dots \cap E_k)$ non può essere negativo e perché la condizione (6) impone che almeno uno di essi sia positivo (ricordiamo che la sommatoria $\sum_{E_1, E_2, \dots, E_k}$ è estesa a tutti i modi possibili per definire gli eventi E_i , ovvero a tutti gli eventi $E_1 \cap E_2 \cap \dots \cap E_k$ della partizione di Ω che abbiamo introdotto poco sopra). *Per dimostrare che è impossibile che il guadagno netto sia negativo in ogni evenienza, è dunque sufficiente dimostrare che $\bar{G} = 0$* così come faremo nelle prossime righe. A tal fine osserviamo in primo luogo che il guadagno netto atteso può essere scomposto

come

$$\begin{aligned}\bar{G} &= \sum_{E_1, E_2, \dots, E_k} \left[\sum_{i: E_i = A_i} S_i - \sum_{i=1}^k S_i P(A_i) \right] \times P(E_1 \cap E_2 \cap \dots \cap E_k) \\ &= \sum_{E_1, E_2, \dots, E_k} \sum_{i: E_i = A_i} S_i \times P(E_1 \cap E_2 \cap \dots \cap E_k) + \\ &\quad + \sum_{E_1, E_2, \dots, E_k} \left[- \sum_{i=1}^k S_i P(A_i) \right] \times P(E_1 \cap E_2 \cap \dots \cap E_k)\end{aligned}$$

Usando l'uguaglianza nella (6) possiamo facilmente verificare che il termine nell'ultima riga può essere semplificato (si noti che l'espressione tra parentesi quadre è una costante che non dipende dal modo in cui sono definiti gli eventi E_i):

$$\begin{aligned}\sum_{E_1, E_2, \dots, E_k} \left[- \sum_{i=1}^k S_i P(A_i) \right] \times P(E_1 \cap E_2 \cap \dots \cap E_k) &= \\ &= \left[- \sum_{i=1}^k S_i P(A_i) \right] \times \sum_{E_1, E_2, \dots, E_k} P(E_1 \cap E_2 \cap \dots \cap E_k) \\ \text{[[equ. (6)]]} &= \left[- \sum_{i=1}^k S_i P(A_i) \right] \times P(\Omega) \\ \text{[[assioma K2]]} &= \left[- \sum_{i=1}^k S_i P(A_i) \right] \times 1 = - \sum_{i=1}^k S_i P(A_i).\end{aligned}$$

Per completare la dimostrazione dobbiamo dunque soltanto più dimostrare che

$$\sum_{E_1, E_2, \dots, E_k} \sum_{i: E_i = A_i} S_i \times P(E_1 \cap E_2 \cap \dots \cap E_k) = \sum_{i=1}^k S_i P(A_i).$$

Per dimostrare questa uguaglianza osserviamo in primo luogo che nella doppia sommatoria al primo membro il fattore S_i compare soltanto nei termini dove $E_i = A_i$ e in nessun altro termine. Quindi, immaginando di scrivere per esteso tutti i termini della doppia sommatoria, possiamo raccogliere ciascun S_i e moltiplicarlo per la somma dei corrispondenti pesi $P(E_1 \cap E_2 \cap \dots \cap E_k)$ (ovvero per la somma di tutti i pesi per i quali $E_i = A_i$). Così facendo vediamo che la doppia sommatoria può essere riscritta come

$$\begin{aligned}& S_1 \times \sum_{E_1, E_2, \dots, E_k: E_1 = A_1} P(E_1 \cap E_2 \cap \dots \cap E_k) + \\ & + S_2 \times \sum_{E_1, E_2, \dots, E_k: E_2 = A_2} P(E_1 \cap E_2 \cap \dots \cap E_k) + \\ & + \dots + \\ & + S_k \times \sum_{E_1, E_2, \dots, E_k: E_k = A_k} P(E_1 \cap E_2 \cap \dots \cap E_k) = \\ & = \sum_{i=1}^k S_i \times \sum_{E_1, E_2, \dots, E_k: E_i = A_i} P(E_1 \cap E_2 \cap \dots \cap E_k).\end{aligned}$$

A questo punto dobbiamo solo più dimostrare che

$$\sum_{E_1, E_2, \dots, E_k: E_i = A_i} P(E_1 \cap E_2 \cap \dots \cap E_k) = P(A_i) \quad \text{per ogni } i = 1, 2, \dots, k.$$

A tal fine basta applicare la legge L2. Infatti, la legge L2 può essere applicata e conduce al risultato desiderato perché ...

- ...l'evento A_i è l'unione di tutti gli eventi $E_1 \cap E_2 \cap \dots \cap E_k$ nei quali $E_i = A_i$:

$$A_i = \bigcup_{E_1, E_2, \dots, E_k: E_i = A_i} (E_1 \cap E_2 \cap \dots \cap E_k) \quad (7)$$

(per rendersi conto che questa uguaglianza è vera basta notare che se si verifica una delle intersezioni $E_1 \cap E_2 \cap \dots \cap E_k$ con $E_i = A_i$, allora dev'essersi ovviamente verificato anche l'evento A_i ; viceversa, se si verifica l'evento A_i , allora deve essersi verificata una delle suddette intersezioni: per stabilire quale, bisogna sapere quali degli altri eventi A_j si sono verificati e quali no) ...

- ...e perché l'unione nella (7) coinvolge soltanto un numero finito di eventi incompatibili visto gli eventi $E_1 \cap E_2 \cap \dots \cap E_k$ formano una partizione finita dello spazio campionario Ω .

□

Nei prossimi esercizi illustreremo alcune applicazioni delle leggi del calcolo delle probabilità. Per risolvere questi esercizi conviene visualizzare gli eventi attraverso degli opportuni diagrammi di Venn.

Esercizio 1.2. Sia $P(\cdot)$ una funzione di probabilità (ovvero una funzione che soddisfa gli assiomi di Kolmogorov) e si assuma che

$$P(A) = 0,7, \quad P(B) = 0,5, \quad P(A \cap \bar{B}) = 0,3.$$

Si determini il valore di $P(A \cup B)$.

Soluzione:

Secondo la legge L5,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \quad (8)$$

Per determinare il valore di $P(A \cup B)$ dobbiamo quindi determinare il valore di $P(A \cap B)$. Tenendo presente che (assioma K3)

$$P(A) = P(A \cap B) + P(A \cap \bar{B}),$$

vediamo che

$$\begin{aligned} P(A \cap B) &= P(A) - P(A \cap \bar{B}) \\ [[\text{dati noti della consegna}]] &= 0,7 - 0,3 = 0,4, \end{aligned} \quad (9)$$

e sostituendo questo risultato nella (8) otteniamo la probabilità richiesta:

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ [[\text{dati noti della consegna}]] &= 0,7 + 0,5 - 0,4 = 0,8. \end{aligned}$$

Esercizio 1.3. Sia A_1 l'evento secondo il quale domani piove, A_2 l'evento secondo il quale dopodomani piove e A_3 l'evento secondo il quale tra tre giorni piove. Secondo un servizio meteo, si ha

$$P(A_1) = P(A_2) = P(A_3) = 0,3;$$

$$P(A_1 \cap A_2) = 0,2;$$

$$P(A_1 \cap A_2 \cap A_3) = 0,1.$$

- Qual è la probabilità che piova in almeno uno dei prossimi due giorni?
- Qual è la probabilità che non piova in nessuno dei prossimi due giorni?
- Qual è la probabilità che non piova in almeno uno dei prossimi tre giorni?
- Qual è la probabilità che piova in entrambi i prossimi due giorni e che il giorno successivo non piova?
- Qual è la probabilità che tra tre giorni piova ma che non piova in tutti e tre i giorni considerati?

Soluzione:

- Qual è la probabilità che piova in almeno uno dei prossimi due giorni?

L'evento secondo il quale in almeno uno dei prossimi due giorni piove può essere rappresentato come $A_1 \cup A_2$. Usando la legge L5 vediamo che

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$$

$$[[\text{dati noti della consegna}]] = 0,3 + 0,3 - 0,2 = 0,4.$$

- Qual è la probabilità che non piova in nessuno dei prossimi due giorni?

L'evento secondo il quale in nessuno dei prossimi due giorni piove può essere rappresentato come $\bar{A}_1 \cap \bar{A}_2$. Applicando le leggi di De Morgan vediamo che questo evento può essere anche espresso come $\bar{A}_1 \cap \bar{A}_2 = \overline{(A_1 \cup A_2)}$. La probabilità richiesta è dunque data da

$$P(\bar{A}_1 \cap \bar{A}_2) = P(\overline{(A_1 \cup A_2)})$$

$$[[\text{legge L3}]] = 1 - P(A_1 \cup A_2)$$

$$[[\text{risposta quesito a)}]] = 1 - 0,4 = 0,6.$$

- c) Qual è la probabilità che non piova in almeno uno dei prossimi tre giorni?

L'evento secondo il quale in almeno uno dei prossimi tre giorni non piove può essere rappresentato come $\bar{A}_1 \cup \bar{A}_2 \cup \bar{A}_3$. Secondo le [leggi di De Morgan](#) questo evento può essere anche espresso come

$$\bar{A}_1 \cup \bar{A}_2 \cup \bar{A}_3 = \overline{(A_1 \cap A_2 \cap A_3)}.$$

La probabilità richiesta è dunque data da

$$\begin{aligned} P(\bar{A}_1 \cup \bar{A}_2 \cup \bar{A}_3) &= P\left(\overline{(A_1 \cap A_2 \cap A_3)}\right) \\ &[[\text{legge L3}]] = 1 - P(A_1 \cap A_2 \cap A_3) \\ &[[\text{dati noti della consegna}]] = 1 - 0,1 = 0,9. \end{aligned}$$

- d) Qual è la probabilità che piova in entrambi i prossimi due giorni e che il giorno successivo non piova?

L'evento secondo il quale in entrambi i prossimi due giorni piove e il giorno successivo non piove può essere espresso come $A_1 \cap A_2 \cap \bar{A}_3$. Rappresentando gli eventi A_1 , A_2 e A_3 attraverso un diagramma di Venn non è difficile rendersi conto che

$$A_1 \cap A_2 = (A_1 \cap A_2 \cap A_3) \cup (A_1 \cap A_2 \cap \bar{A}_3)$$

e che i due eventi tra parentesi al secondo membro sono incompatibili. In virtù dell'assioma [K3](#) deve quindi valere l'uguaglianza

$$P(A_1 \cap A_2) = P(A_1 \cap A_2 \cap A_3) + P(A_1 \cap A_2 \cap \bar{A}_3).$$

Da questa uguaglianza deduciamo che la probabilità richiesta è data da

$$\begin{aligned} P(A_1 \cap A_2 \cap \bar{A}_3) &= P(A_1 \cap A_2) - P(A_1 \cap A_2 \cap A_3) \\ &[[\text{dati noti della consegna}]] = 0,2 - 0,1 = 0,1. \end{aligned}$$

- e) Qual è la probabilità che tra tre giorni piova ma che non piova in tutti e tre i giorni considerati?

La risposta a quest'ultimo quesito la lasciamo per esercizio.

Il prossimo esempio illustra una situazione dove le probabilità di determinati eventi non sono coerenti con gli assiomi di Kolmogorov e dove (forse) questo fatto non è immediatamente ovvio.

Esercizio 1.4 (Probabilità non coerenti con gli assiomi di Kolmogorov). Si A_1 l'evento secondo il quale al termine dell'anno in corso il prezzo del titolo X è maggiore del prezzo odierno, e sia A_2 il medesimo evento con riferimento al titolo Y . Secondo un investitore,

$$P(A_1) = 0,9, \quad P(A_2) = 0,8, \quad P(A_1 \cap A_2) = 0,1.$$

Si calcoli il valore di $P(A_1 \cup A_2)$.

Soluzione:

Secondo la legge L5 la probabilità richiesta dovrebbe essere data da

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2) = 0,9 + 0,8 - 0,1 = 1,6.$$

Siccome questo risultato contraddice la legge L4 (secondo la quale la probabilità di un evento deve essere compresa nell'intervallo $[0, 1]$), possiamo concludere che le probabilità assegnate dall'investitore non sono coerenti con gli assiomi di Kolmogorov, ovvero che non possono essere estese ad una funzione di probabilità $P(\cdot)$ che soddisfa gli assiomi di Kolmogorov (visto che abbiamo ricavato la contraddizione usando soltanto le leggi L4 e L5, e visto che queste due leggi possono essere dimostrate usando soltanto gli assiomi K0 - K3, possiamo anche concludere che le probabilità assegnate dall'investitore non possono essere nemmeno estese ad una funzione di probabilità che soddisfi soltanto gli assiomi K0 - K3).

1.6 Il teorema di estensione di Hahn-Kolmogorov

Come abbiamo visto nella Definizione 1.2, un modello probabilistico è una descrizione di un esperimento casuale che è composta da

- uno spazio campionario Ω ,
- uno spazio di eventi \mathcal{A} che soddisfa le condizioni a), b) e c) dell'assioma K0, ovvero da uno spazio di eventi \mathcal{A} che possiede la struttura di un'algebra,
- e da una funzione di probabilità $P(\cdot)$ che ha per dominio lo spazio di eventi \mathcal{A} e che soddisfa gli assiomi K1 - K3 e anche l'assioma K3*.

Secondo il **Teorema di estensione di Hahn-Kolmogorov**, la funzione di probabilità $P(\cdot)$ di un modello probabilistico può essere sempre **estesa in un unico modo** ad una funzione di probabilità $P^*(\cdot)$ che soddisfa anch'essa tutti gli assiomi di Kolmogorov e che ha per dominio la σ -algebra $\sigma(\mathcal{A})$ (ricordiamo che $\sigma(\mathcal{A})$ è la più piccola σ -algebra che

contiene la collezione di eventi \mathcal{A} - vedi il Teorema 1.3). Per maggiore chiarezza, conviene enunciare il [teorema di estensione di Hahn-Kolmogorov](#) in modo più formale:

Teorema 1.4 (Teorema di estensione di Hahn-Kolmogorov). Se il dominio \mathcal{A} di una funzione di probabilità $P(\cdot)$ ha la struttura di un'algebra (così come previsto dall'assioma [K0](#)) e la funzione di probabilità $P(\cdot)$ soddisfa gli assiomi di Kolmogorov (ovvero gli assiomi [K0 - K3](#) e anche l'assioma [K3*](#)), allora **esiste un'unica** funzione di probabilità $P^*(\cdot)$ che

- ha per dominio $\sigma(\mathcal{A})$
- che è un'estensione della funzione di probabilità $P(\cdot)$ nel senso che

$$P^*(A) = P(A) \quad \text{per ogni } A \in \mathcal{A},$$

- e che soddisfa anch'essa tutti gli assiomi di Kolmogorov.

Chiaramente, la conclusione del [teorema di estensione di Hahn-Kolmogorov](#) è banale se lo spazio di eventi \mathcal{A} possiede già in partenza la struttura di una σ -algebra: infatti, in questo caso si avrebbe $\sigma(\mathcal{A}) = \mathcal{A}$ e $P^*(A) = P(A)$ per ogni $A \in \mathcal{A} = \sigma(\mathcal{A})$. In generale, tuttavia, la conclusione del [teorema di estensione di Hahn-Kolmogorov](#) è tutt'altro che banale. Infatti, come abbiamo visto nella Sezione 1.4, in generale $\sigma(\mathcal{A})$ potrebbe essere uno spazio di eventi molto più ampio di \mathcal{A} , e secondo il [teorema di estensione di Hahn-Kolmogorov](#) i valori $P(A)$ delle probabilità degli eventi $A \in \mathcal{A}$ **determinano univocamente** le probabilità $P^*(A)$ di tutti gli eventi $A \in \sigma(\mathcal{A})$ che non si trovano nel dominio \mathcal{A} della funzione di probabilità $P(\cdot)$ di partenza. Com'è facile immaginare, la dimostrazione del [teorema di estensione di Hahn-Kolmogorov](#) dipende crucialmente dall'assioma [K3*](#), ovvero da un assioma che de Finetti¹¹ rifiuta proprio perché rende possibile la dimostrazione di teoremi come questo (la dimostrazione di questo teorema è piuttosto lunga e può essere trovata in testi specializzati¹⁷). Il prossimo esempio illustra un'applicazione del [teorema di estensione di Hahn-Kolmogorov](#) e chiarisce perché De Finetti rifiuta l'assioma [K3*](#).

Esempio 1.7. Consideriamo ancora l'esperimento casuale che consiste in una successione infinita di lanci di una moneta. Come abbiamo visto nell'Esempio 1.4, l'esito finale di questo esperimento casuale (esito finale che non potremo mai osservare) può essere descritto attraverso una sequenza infinita di numeri "0" e "1" dove l' i -esima posizione della sequenza è occupata dal numero "1" se l'esito dell' i -esimo lancio è croce, e dove l' i -esima posizione della sequenza è occupata dal numero

¹⁷Vedi, per esempio, *Patrick Billingsley (1995), "Theory and Measure", third edition, Wiley.*

"0" se l'esito dell' i -esimo lancio è testa. Come spazio campionario Ω consideriamo dunque l'insieme di tutte queste sequenze.

Consideriamo ora l'algebra di eventi \mathcal{A} che abbiamo definito nell'Esempio 1.4. Siccome questa algebra contiene soltanto eventi che descrivono l'esito di un numero finito di lanci, e siccome in un numero finito di lanci si può ottenere soltanto un numero finito di esiti diversi (si noti infatti che in n lanci si possono ottenere soltanto $2 \times 2 \times \dots \times 2 = 2^n$ esiti diversi perché ciascuno degli n lanci può dare luogo soltanto a 2 esiti diversi), possiamo definire la probabilità di un evento $A \in \mathcal{A}$ come rapporto tra il numero di casi favorevoli e il numero di casi possibili:

$$P(A) = \frac{\left(\begin{array}{c} \text{numero di esiti dei primi } n \text{ lanci che sono} \\ \text{favorevoli all'evento } A \end{array} \right)}{\text{numero di esiti possibili dei primi } n \text{ lanci}} \quad (10)$$

$$= \frac{1}{2^n} \times \left(\begin{array}{c} \text{numero di esiti dei primi } n \text{ lanci che sono} \\ \text{favorevoli all'evento } A \end{array} \right).$$

Si può dimostrare (omettiamo la dimostrazione) che la funzione di probabilità $P(\cdot)$ che abbiamo appena definito soddisfa gli assiomi **K0 - K3** e anche l'assioma **K3*** e in virtù del [teorema di estensione di Hahn-Kolmogorov](#) possiamo dunque concludere che esista un'unica funzione di probabilità $P^*(\cdot)$ che

- ha per dominio $\sigma(\mathcal{A})$
- che è un'estensione della funzione di probabilità $P(\cdot)$ nel senso che

$$P^*(A) = P(A) \quad \text{per ogni } A \in \mathcal{A}, \quad (11)$$

- e che soddisfa anch'essa tutti gli assiomi di Kolmogorov.

Ora, nell'Esempio 1.4 abbiamo visto che \mathcal{A} è un'algebra di eventi ma non una σ -algebra e quindi possiamo concludere che $\sigma(\mathcal{A})$ contiene degli eventi che non sono contenuti nell'algebra \mathcal{A} . Un esempio di un evento che appartiene alla σ -algebra $\sigma(\mathcal{A})$ ma non all'algebra \mathcal{A} è l'evento secondo il quale in tutti gli infiniti lanci si ottiene sempre soltanto l'esito testa. Si noti che questo evento può essere espresso come

$$B = \bigcap_{i=1}^{\infty} A_i$$

dove A_i è l'evento secondo il quale l'esito dell' i -esimo lancio è testa. Secondo il [teorema di estensione di Hahn-Kolmogorov](#), la probabilità $P^*(B)$ dell'evento B è dunque **univocamente** determinata dalle probabilità $P(\cdot)$ degli eventi che appartengono all'algebra \mathcal{A} . Il valore di $P^*(B)$ può dunque essere ricavato a partire dalle probabilità $P(\cdot)$ di eventi $A \in \mathcal{A}$ attraverso gli assiomi di Kolmogorov e/o le

leggi del calcolo delle probabilità. Nel caso dell'evento B questo fatto può essere facilmente verificato. Infatti, per la legge L4 si deve avere

$$P^*(B) = P^*\left(\bigcap_{i=1}^{\infty} A_i\right) \leq P^*\left(\bigcap_{i=1}^n A_i\right) \quad \text{per ogni } n = 1, 2, \dots,$$

e siccome l'evento $\bigcap_{i=1}^n A_i$ appartiene all'algebra di eventi \mathcal{A} (si noti infatti che questo evento descrive soltanto gli esiti dei primi n lanci), possiamo concludere che

$$P^*\left(\bigcap_{i=1}^n A_i\right) = P\left(\bigcap_{i=1}^n A_i\right) \quad \text{per ogni } n = 1, 2, \dots$$

e quindi che

$$P^*(B) \leq P\left(\bigcap_{i=1}^n A_i\right) \quad \text{per ogni } n = 1, 2, \dots \quad (12)$$

Tenendo presente che

$$P\left(\bigcap_{i=1}^n A_i\right) = \frac{\left(\begin{array}{c} \text{numero di esiti dei primi } n \text{ lanci che sono} \\ \text{favorevoli all'evento } \bigcup_{i=1}^n A_i \end{array}\right)}{\text{numero di esiti possibili dei primi } n \text{ lanci}} = \frac{1}{2^n} = \frac{1}{2^n}$$

(infatti, in n lanci c'è solo un modo per ottenere sempre soltanto l'esito testa), vediamo dunque che

$$\begin{aligned} P^*(B) &\leq \frac{1}{2^n} \quad \text{per ogni } n = 1, 2, \dots \quad \Rightarrow \\ \Rightarrow P^*(B) &\leq \lim_{n \rightarrow \infty} \frac{1}{2^n} = 0 \quad \Rightarrow \text{[[assioma K1]]} \Rightarrow P^*(B) = 0. \end{aligned}$$

L'**unica** funzione di probabilità $P^*(\cdot)$ che estende la funzione di probabilità $P(\cdot)$ alla σ -algebra $\sigma(\mathcal{A})$ assegna quindi probabilità nulla all'evento $B = \bigcap_{i=1}^{\infty} A_i$ (ovvero all'evento secondo il quale in una serie di infiniti lanci si ottiene sempre soltanto l'esito testa). Questo risultato può essere interpretato in modo diverso dicendo che **le probabilità (10) assegnate agli eventi dell'algebra \mathcal{A} implicano che la probabilità dell'evento $B = \bigcap_{i=1}^{\infty} A_i$ sia pari a 0.**

Vediamo ora invece come De Finetti avrebbe determinato la probabilità dell'evento $B = \bigcap_{i=1}^{\infty} A_i$. A tal fine ricordiamo che De Finetti interpreta le probabilità di eventi come prezzi di titoli di scommesse che incorporano il diritto di ricevere 1 euro se l'evento di riferimento si verifica e che non danno luogo a nessun diritto in caso contrario. Secondo De Finetti,¹¹ **le probabilità assegnate a degli eventi di uno spazio di eventi \mathcal{A} sono sempre "ammissibili" se attraverso di esse non è possibile costruire sistemi di scommesse che danno luogo ad una perdita certa.** A proposito di questa definizione di "probabilità ammissibili" è

importante tenere presente che **secondo De Finetti un sistema di scommesse è un insieme di diritti e/o obblighi che può essere costruito attraverso l'acquisto e/o l'emissione di titoli che si riferiscono soltanto ad un numero finito di eventi**. Dopo aver definito che cosa intende per **"probabilità ammissibili"**, nel suo articolo [11] De Finetti considera il caso dove \mathcal{A} è una generica classe di eventi (non necessariamente un'algebra) e dove $P(\cdot)$ è una funzione che associa probabilità ammissibili agli eventi $A \in \mathcal{A}$. Poi considera un evento B che non appartiene alla classe di eventi \mathcal{A} e si chiede se sulla base delle probabilità $P(A)$ assegnate agli eventi $A \in \mathcal{A}$ si può determinare in modo univoco un valore ammissibile $P^*(B)$ per la probabilità dell'evento B (ovvero per il prezzo del titolo della scommessa sull'evento B) oppure se si può quantomeno determinare un insieme di valori ammissibili per $P^*(B)$. In altre parole, De Finetti vuole verificare se partendo da determinate probabilità $P(A)$ assegnate agli eventi $A \in \mathcal{A}$ di cui a priori è noto che sono "ammissibili" è sempre possibile determinare la probabilità $P^*(B)$ di un evento B che non appartiene allo spazio di eventi \mathcal{A} in modo tale che le probabilità $P(A)$ degli eventi $A \in \mathcal{A}$ e la probabilità $P^*(B)$ dell'evento B siano **nel loro insieme** ammissibili (ovvero in modo tale che con i prezzi $P(A)$ dei titoli delle scommesse sugli eventi $A \in \mathcal{A}$ e il prezzo $P^*(B)$ del titolo della scommessa sull'evento B non sia possibile costruire un sistema di scommesse che dà luogo ad una perdita certa). De Finetti risolve questo problema dimostrando che esiste sempre un prezzo ammissibile $P^*(B)$, ma che potrebbero esserne anche infiniti. Infatti, De Finetti dimostra che l'insieme dei prezzi ammissibili $P^*(B)$ è un intervallo chiuso: le probabilità $P(A)$ degli eventi $A \in \mathcal{A}$ determinano dunque **in modo univoco** la probabilità $P^*(B)$ di un evento B che non appartiene allo spazio di eventi \mathcal{A} se e solo se i due estremi di questo intervallo coincidono. Nel caso particolare in cui lo spazio di eventi \mathcal{A} possiede la struttura di un'algebra gli estremi dell'intervallo che contiene tutti i valori ammissibili $P^*(B)$ sono dati da

$$\text{estremo inferiore} = \sup_{A \in \mathcal{A}: A \subseteq B} P(A) \quad (13)$$

e da

$$\text{estremo superiore} = \inf_{A \in \mathcal{A}: A \supseteq B} P(A) \quad (14)$$

e il prezzo di $P^*(B)$ è univocamente determinato se e solo se questi due estremi coincidono e in questo caso è dato dal loro valore comune.

Consideriamo ora il caso dove \mathcal{A} è l'algebra che contiene tutti gli eventi che descrivono l'esito di un numero finito di lanci di una moneta e dove $B = \bigcap_{i=1}^{\infty} A_i$ è l'evento secondo il quale in una successione infinita di lanci si ottiene sempre

soltanto l'esito testa. Siccome

$$B = \bigcap_{i=1}^{\infty} A_i \subseteq \bigcap_{i=1}^n A_i \quad \text{per ogni } n = 1, 2, \dots,$$

possiamo concludere che anche secondo l'approccio di De Finetti la probabilità $P^*(B)$ deve soddisfare tutte le disuguaglianze (12), ovvero tutte le disuguaglianze

$$P^*(B) \leq P\left(\bigcap_{i=1}^n A_i\right) \quad \text{per ogni } n = 1, 2, \dots$$

Se le probabilità $P(A)$ degli eventi $A \in \mathcal{A}$ sono definite come nella (10) (si ricordi che queste probabilità soddisfano tutti gli assiomi di Kolmogorov e per il [teorema delle scommesse olandesi](#) sono quindi "ammissibili" nel senso della definizione di De Finetti che abbiamo riportato poco sopra), possiamo quindi concludere che

$$P^*(B) \leq \lim_{n \rightarrow \infty} \frac{1}{2^n} = 0$$

come in precedenza. Siccome le probabilità $P(A)$ definite nella (10) sono tutte non negative, possiamo dunque concludere che secondo De Finetti esiste un unico valore ammissibile per $P^*(B)$ e che questo valore è dato da

$$P^*(B) = P^*\left(\bigcap_{i=1}^{\infty} A_i\right) = 0,$$

ovvero dal valore comune dei due estremi definiti nelle formule (13) e (14). Quindi possiamo concludere che anche secondo l'approccio di De Finetti le probabilità (10) assegnate agli eventi dell'algebra \mathcal{A} implicano che la probabilità dell'evento $B = \bigcap_{i=1}^{\infty} A_i$ sia nulla.

Per fare un esempio di un evento B' la cui probabilità $P^*(B')$ non è univocamente determinata dalle probabilità $P(A)$ degli eventi A che appartengono all'algebra di eventi \mathcal{A} , conviene considerare l'evento

$$B' = \bigcup_{i=1}^{\infty} \left(\bigcap_{j=i}^{\infty} A_j \right).$$

Si noti che questo evento B' può essere descritto come l'evento secondo il quale da un certo lancio in poi si ottiene sempre il lato testa. Ovviamente, dopo qualsiasi numero finito di lanci, **qualunque sia il loro esito**, non possiamo mai dire se questo evento si è verificato o meno e quindi possiamo concludere che l'evento B' non appartiene all'algebra di eventi \mathcal{A} e inoltre possiamo anche concludere che

$$A \in \mathcal{A} \quad \text{e} \quad A \neq \emptyset \quad \Rightarrow \quad A \cap B' \neq \emptyset \quad \text{e} \quad A \cap \overline{B'} \neq \emptyset.$$

Da quest'ultima osservazione possiamo immediatamente dedurre che l'algebra \mathcal{A} contiene soltanto un unico evento che è un sottoinsieme di B' , ovvero l'evento \emptyset e che l'algebra \mathcal{A} contiene anche soltanto un unico evento che include l'evento B' , ovvero l'evento certo Ω . **Secondo l'approccio di De Finetti la probabilità $P^*(B')$ è quindi completamente indeterminata nel senso che può essere qualsiasi numero compreso tra 0 e 1** (si noti che la probabilità di questo evento B' è completamente indeterminata anche se al posto delle probabilità definite nella (10) considerassimo un qualsiasi altro modo "ammissibile" di definire le probabilità degli eventi $A \in \mathcal{A}$).

D'altra parte, siccome l'evento $B' = \bigcup_{i=1}^{\infty} \left(\bigcap_{j=i}^{\infty} A_j \right)$ appartiene alla σ -algebra $\sigma(\mathcal{A})$, secondo il teorema di estensione **teorema di estensione di Hahn-Kolomogorov** la probabilità dell'evento B' è **univocamente determinata dalle probabilità degli eventi $A \in \mathcal{A}$** e, come dimostreremo tra breve, se le probabilità degli eventi $A \in \mathcal{A}$ sono definite come nella formula (10), il valore di $P^*(B')$ è nullo.

Per dimostrare che l'unica funzione di probabilità $P^*(\cdot)$ che estende la funzione di probabilità $P(\cdot)$ definita nella formula (10) alla σ -algebra di eventi $\sigma(\mathcal{A})$ assegna probabilità nulla all'evento B' , basta osservare che

$$P^*(B') = P^* \left(\bigcup_{i=1}^{\infty} \left(\bigcap_{j=i}^{\infty} A_j \right) \right) = \text{[[legge L10]]} = \lim_{n \rightarrow \infty} P^* \left(\bigcup_{i=1}^n \left(\bigcap_{j=i}^{\infty} A_j \right) \right)$$

e che il limite sul lato destro è nullo perché

•

$$0 \leq \text{[[assioma K1]]} \leq P^* \left(\bigcup_{i=1}^n \left(\bigcap_{j=i}^{\infty} A_j \right) \right) \leq \text{[[legge L8]]} \leq \sum_{i=1}^n P^* \left(\bigcap_{j=i}^{\infty} A_j \right)$$

• e perché per ogni $k = 1, 2, \dots$ si ha

$$\begin{aligned} 0 &\leq \text{[[assioma K1]]} \leq P^* \left(\bigcap_{j=i}^{\infty} A_j \right) \leq \\ &\leq \text{[[legge L4]]} \leq P^* \left(\bigcap_{j=i}^{i+k} A_j \right) = \text{[[} \bigcap_{j=i}^{i+k} A_j \in \mathcal{A} \text{ e formula (11)]]} = \\ &= P \left(\bigcap_{j=i}^{i+k} A_j \right) = \text{[[formula (10)]]} = \frac{2^{i-1} \times 1^{k+1}}{2^{i+k}} = \frac{1}{2^{k+1}} \end{aligned}$$

e $1/2^{k+1} \rightarrow 0$ se $k \rightarrow \infty$.

Si noti che per ottenere $P^*(B') = 0$ abbiamo invocato la legge [L10](#) che non può essere dimostrata usando soltanto gli assiomi [K0 - K3](#) senza invocare un ulteriore assioma come, per esempio, l'assioma [K3*](#). Nella precedente dimostrazione siamo quindi partiti dalle probabilità degli eventi $A \in \mathcal{A}$ e attraverso l'assioma [K3*](#) siamo riusciti a determinare un **unico** valore numerico per la probabilità $P^*(B')$ anche se secondo De Finetti la probabilità dell'evento B' non può essere determinata a partire dalle probabilità degli eventi $A \in \mathcal{A}$. In questo senso, secondo De Finetti, l'assioma [K3*](#) rappresenta dunque una forzatura che talvolta permette di determinare le probabilità di eventi che secondo la sua impostazione sono completamente indeterminate.

Siccome tutte le funzioni di probabilità che hanno per dominio un'algebra (e che soddisfano gli assiomi di Kolmogorov) possono essere estese soltanto in un unico modo alla più piccola σ -algebra che include l'algebra di partenza (questo è ciò che afferma la conclusione del [teorema di estensione di Hahn-Kolmogorov](#)), possiamo concludere che non c'è nulla da perdere se per la descrizione di esperimenti casuali si considerano solo le funzioni di probabilità che hanno per dominio una σ -algebra e si escludono tutte quelle il cui dominio è soltanto un'algebra (semmai si assegnano delle probabilità a degli eventi aggiuntivi le cui probabilità, secondo De Finetti, sono indeterminate). Per questo motivo, soprattutto nei testi che trattano la teoria della probabilità come branca della teoria della misura, viene spesso ipotizzato già in partenza che il dominio di una funzione di probabilità $P(\cdot)$ sia una σ -algebra piuttosto che un'algebra come previsto dall'assioma [K0](#). D'ora in poi, anche noi seguiremo questa consuetudine: **d'ora in poi assumeremo dunque sempre che il dominio \mathcal{A} di una funzione di probabilità $P(\cdot)$ sia una σ -algebra di eventi**. Siccome un'algebra che contiene soltanto un numero finito di insiemi è sempre anche una σ -algebra (perché la condizione [c*](#)) è vacuamente soddisfatta), questa ipotesi esclude soltanto funzioni di probabilità che hanno per dominio delle algebre infinite che non soddisfano la condizione [c*](#).

2 Calcolo combinatorio

Nell'[enciclopedia Treccani](#), sotto la voce "calcolo combinatorio" si legge ...

càlculo combinatòrio (o analisi combinatoria) Parte dell'aritmetica che ha come scopo principale quello di contare i raggruppamenti di varia specie che si possono formare con oggetti o simboli. I suoi procedimenti e i suoi risultati (coefficienti binomiali, determinanti, gruppi di sostituzioni) trovano applicazione nell'algebra e sono di utilità in tutti i campi della matematica. Il c.c. offre inoltre i mezzi per risolvere alcune questioni fondamentali del calcolo delle probabilità. Fra i primi matematici che se ne interessarono ricordiamo B.

Pascal, G. W. Leibniz, G. Bernoulli (1655-1705), A. De Moivre (1667-1754).
I raggruppamenti di oggetti che l'analisi c. considera più frequentemente sono le disposizioni, le permutazioni, le combinazioni.

Come si desume dal contenuto di questa voce, i metodi del calcolo combinatorio trovano applicazione anche nel calcolo delle probabilità, tant'è vero che gli studiosi citati nella suddetta voce vengono di solito anche citati (insieme ad altri studiosi) come fondatori della teoria della probabilità (forse con l'eccezione di Leibniz al quale viene comunque riconosciuto un ruolo importante di stimolo e incoraggiamento nella fase iniziale della teoria della probabilità).

In questa sezione presenteremo dunque alcune nozioni del calcolo combinatorio che sono spesso utili per determinare probabilità di eventi che sono definite attraverso il metodo di assegnazione classico, ovvero come rapporti tra *numeri di casi favorevoli* e un *numero di casi possibili*.

2.1 Esperimenti in più passi

Per approcciarci alle applicazioni del calcolo combinatorio in campo probabilistico conviene partire da un ipotetico **esperimento casuale suddiviso in un numero finito k di passi** con

- n_1 esiti possibili al primo passo,
- n_2 esiti possibili al secondo passo,...
- ...
- n_k esiti possibili all'ultimo (k -esimo) passo.

Non è difficile rendersi conto che il numero complessivo di eventi elementari che possono verificarsi al termine di un esperimento casuale che si svolge in k passi è dato da

$$n_1 \times n_2 \times \cdots \times n_k. \quad (15)$$

Infatti, gli esiti di un tale esperimento possono essere rappresentati mediante sequenze di lunghezza k i cui elementi descrivono gli esiti dei singoli passi dell'esperimento. Siccome il primo elemento della sequenza può essere scelto in n_1 modi diversi, e può essere combinato con ciascuno degli n_2 modi possibili per scegliere il secondo elemento della sequenza e così via, possiamo concludere che il numero totale di sequenze possibili è dato dal prodotto nella (15).

Per visualizzare lo svolgimento di un esperimento suddiviso in passi è spesso utile costruire un **diagramma ad albero** come quello che vedremo nella soluzione del prossimo esercizio.

Esercizio 2.1 (Esperimento in più passi). Si consideri un esperimento casuale che consiste nel

- lancio di una moneta, ...
- ... seguito dall'estrazione di una pallina da un'urna che contiene una pallina gialla, una verde ed una rossa, ...
- ... che a sua volta è seguita dal lancio di un tetraedro con facce numerate da 1 a 4.

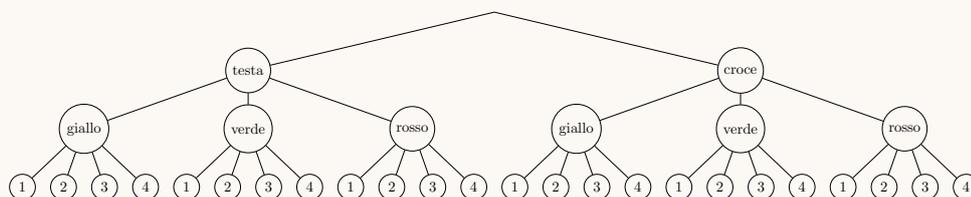
Con riferimento all'esperimento casuale in questione:

- Si costruisca un diagramma ad albero che visualizza lo svolgimento dell'esperimento e si elenchino gli eventi elementari che descrivono tutti i possibili esiti finali. Quanti sono?
- Si definisca un modello probabilistico (Ω, \mathcal{A}, P) che descrive l'esperimento casuale descritto nella consegna. Secondo quale metodo si dovrebbero assegnare le probabilità?

Soluzione:

- Si costruisca un diagramma ad albero che visualizza lo svolgimento dell'esperimento e si elenchino gli eventi elementari che descrivono tutti i possibili esiti finali. Quanti sono?

Il diagramma ad albero può essere rappresentato come



... e gli eventi elementari ω che possono verificarsi al termine di questo esperimento casuale possono essere rappresentati attraverso terne (sequenze di lunghezza tre) del tipo

$$\omega = (\text{testa}, \text{giallo}, 1),$$

dove il primo elemento della sequenza indica il lato *testa* oppure il lato *croce* della moneta, il secondo elemento indica uno dei tre colori *giallo*, *verde* oppure *rosso* delle palline presenti nell'urna, e il terzo ed ultimo elemento della sequenza è uno dei primi quattro numeri naturali positivi riportati sulle quattro facce del tetraedro. Per tener conto di tutti i possibili esiti finali di questo esperimento casuale bisogna quindi fare riferimento ad uno spazio campionario Ω che contiene $2 \times 3 \times 4 = 24$ eventi elementari ω . A ciascun evento elementare $\omega \in \Omega$ corrisponde un unico percorso dall'alto verso il basso lungo i nodi dell'albero.

- b) Si definisca un modello probabilistico (Ω, \mathcal{A}, P) che descrive l'esperimento casuale descritto nella consegna. Secondo quale metodo si dovrebbero assegnare le probabilità?

Siccome lo spazio campionario Ω definito nella risposta al quesito precedente è un insieme finito, e siccome la descrizione dell'esperimento casuale non fornisce indizi sulla possibilità che qualche evento elementare $\omega \in \Omega$ si possa verificare "più facilmente" di altri, applicheremo il metodo di assegnazione classico. Ricordiamo che attraverso questo metodo possiamo assegnare una probabilità ad ogni $A \subseteq \Omega$ ponendo

$$P(A) = \frac{\#A}{\#\Omega} = \frac{\#A}{24}.$$

Applicando il metodo classico otteniamo dunque una funzione di probabilità che ha come dominio \mathcal{A} l'insieme delle parti $\mathcal{P}(\Omega)$ e che assegna probabilità

$$P(\{\omega\}) = \frac{1}{24}$$

ad ogni sottoinsieme di Ω che contiene un unico evento elementare.

2.2 Disposizioni e permutazioni

Un caso particolare di esperimento suddiviso in passi consiste nell'estrazione, una alla volta, di k oggetti da un'urna che contiene n oggetti.

- a) Se le estrazioni avvengono **senza reimmissione** (si noti che in questo caso l'esito di ciascuna estrazione influisce sugli esiti possibili delle estrazioni successive), allora ci saranno
- n esiti possibili alla prima estrazione,
 - $n - 1$ esiti possibili alla seconda estrazione,...
 - ...

- $n - k + 1$ esiti possibili alla k -esima (e ultima) estrazione.

Ragionando come nel caso di un generico esperimento suddiviso in passi, vediamo dunque che lo spazio campionario Ω associato a un esperimento casuale che consiste in k estrazioni senza reimmissione deve contenere

$$n \times (n - 1) \times (n - 2) \times \cdots \times (n - k + 1)$$

eventi elementari. Gli eventi elementari in questione vengono chiamati **disposizioni senza ripetizione di n oggetti presi k alla volta**. Non è difficile rendersi conto che due disposizioni senza ripetizione possono distinguersi

- per la presenza di oggetti diversi,
- e/o per l'ordine in cui si presentano gli oggetti.

Nel caso particolare in cui $k = n$, le disposizioni senza ripetizione vengono chiamate **permutazioni** (senza ripetizione). Il numero di permutazioni senza ripetizione di n oggetti viene solitamente indicato con $n!$ (si legge n fattoriale). Ovviamente,

$$n! = n \times (n - 1) \times (n - 2) \times \cdots \times 2 \times 1.$$

Siccome ogni permutazione contiene tutti gli oggetti presenti nell'urna, due permutazioni possono distinguersi solamente per l'ordine in cui si presentano gli stessi oggetti.

- b) Se d'altra parte le estrazioni avvengono **con reimmissione** (si noti che in questo caso gli esiti possibili in tutte le estrazioni sono sempre gli stessi), allora il numero di esiti possibili per ciascuna delle k estrazioni è sempre pari ad n , e pertanto possiamo concludere che lo spazio campionario Ω contiene

$$n \times n \times \cdots \times n = n^k$$

eventi elementari che vengono chiamati **disposizioni con ripetizione di n elementi presi k alla volta**. Si noti che nel caso di estrazioni con reimmissione il numero k di oggetti estratti potrebbe anche eccedere il numero n di oggetti presenti nell'urna. Ovviamente, due disposizioni con ripetizione possono distinguersi

- per la presenza di oggetti diversi,
- per il numero di volte che contengono alcuni oggetti
- e/o per l'ordine in cui si presentano gli oggetti.

Il riquadro sottostante riassume le principali proprietà delle disposizioni.

Definizione 2.1 (Disposizioni con e senza ripetizione). Sia U un insieme di n oggetti (come per esempio un'urna che contiene palline numerate), e sia Ω è l'insieme di tutte le sequenze ordinate di k oggetti che si possono formare con gli n oggetti appartenenti all'insieme U :

- Se escludiamo la possibilità che un oggetto possa ripetersi più di una volta all'interno di una stessa sequenza, allora il numero di sequenze ordinate possibili è uguale a

$$\#\Omega = D(n, k) = n \times (n - 1) \times (n - 2) \times \cdots \times (n - k + 1).$$

In questo caso le sequenze vengono chiamate **disposizioni senza ripetizione di n oggetti presi k alla volta**. Nel caso particolare in cui $k = n$ le sequenze vengono chiamate **permutazioni** e il loro numero complessivo viene indicato con $n!$ (si legge n fattoriale).

- Se invece ammettiamo anche la possibilità che un oggetto possa ripetersi più di una volta all'interno di una stessa sequenza, allora il numero di sequenze ordinate possibili è uguale a

$$\#\Omega = D_r(n, k) = n \times n \times \cdots (k \text{ volte}) \times \cdots \times n = n^k.$$

In questo caso le sequenze vengono chiamate **disposizioni con ripetizione di n oggetti presi k alla volta**.

Esercizio 2.2 (Disposizioni con e senza ripetizione). Da un'urna che contiene $n = 9$ palline numerate vengono estratte $k = 3$ palline una dopo l'altra senza riporre le palline estratte nell'urna.

- Si descriva lo spazio campionario Ω associato a questo esperimento casuale. Quanti eventi elementari contiene Ω ?
- Si assegnino delle probabilità agli eventi di $\mathcal{P}(\Omega)$.
- Si risponda alle precedenti domande ipotizzando che le estrazioni avvengano con riposizione.

Soluzione:

- Si descriva lo spazio campionario Ω associato a questo esperimento casuale. Quanti eventi elementari contiene Ω ?

Lo spazio campionario Ω è costituito dall'insieme di tutte le disposizioni senza ripetizione che si possono formare con $n = 9$ palline prendendone $k = 3$ alla

volta. Il numero totale di disposizioni di questo tipo ammonta a

$$\#\Omega = 9 \times 8 \times 7 = 504.$$

- b) Si assegnino delle probabilità agli eventi di $\mathcal{P}(\Omega)$.

Siccome lo spazio campionario Ω è finito, e siccome le informazioni contenute nella consegna non forniscono nessun indizio circa la possibilità che qualche disposizione $\omega \in \Omega$ possa verificarsi "più facilmente" di altre, assegneremo le probabilità in base al metodo classico. Secondo questo metodo di assegnazione, la probabilità di ogni evento che contiene soltanto una disposizione ω è data da

$$P(\{\omega\}) = \frac{1}{504} = 0,00198,$$

e la probabilità di un generico evento $A \subseteq \Omega$ è quindi uguale al rapporto tra il numero $\#A$ di casi favorevoli ed il numero $\#\Omega = 504$ di casi possibili.

- c) Si risponda alle precedenti domande ipotizzando che le estrazioni avvengano con riposizione.

Nel caso di estrazioni con riposizione lo spazio campionario Ω sarebbe formato da

$$\#\Omega = 9 \times 9 \times 9 = 9^3 = 729$$

disposizioni con ripetizione. Secondo il metodo di assegnazione classico, la probabilità associata a ciascun evento che contiene soltanto una delle suddette disposizioni sarebbe dunque data da

$$P(\{\omega\}) = \frac{1}{729} = 0,00137,$$

e la probabilità associata a qualsiasi altro evento $A \subseteq \Omega$ sarebbe invece uguale al rapporto tra il numero $\#A$ di casi favorevoli ed il numero $\#\Omega = 729$ di casi possibili.

2.3 Combinazioni

Consideriamo ora invece un esperimento casuale consiste nell'estrazione in blocco di k oggetti da un'urna che contiene n oggetti.

In questo caso lo spazio campionario Ω è composto da eventi elementari ω che sono degli insiemi non ordinati di k oggetti diversi scelti tra gli n oggetti presenti nell'urna. Gli eventi elementari in questione vengono chiamati **combinazioni di n oggetti presi k alla volta** e il numero di combinazioni possibili viene solitamente indicato con il simbolo $\binom{n}{k}$. Ovviamente, due combinazioni diverse possono distinguersi solo per la presenza di

almeno un oggetto diverso. Non è difficile dimostrare che il numero di combinazioni diverse che si possono formare con n oggetti prendendone k alla volta è dato da

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \quad [[\text{NB: per convenzione, } 0! = 1]].$$

Dimostrazione. Ciascuna combinazione contiene esattamente k oggetti diversi e k oggetti diversi possono essere ordinati in $k!$ modi diversi (si vedano le permutazioni). Ordinando tutte le $\binom{n}{k}$ combinazioni diverse in tutti i $k!$ modi possibili si ottengono tutte le disposizioni di n oggetti presi k alla volta. Quindi possiamo scrivere

$$\binom{n}{k} \times k! = n(n-1) \cdots (n-k+1).$$

Tenendo presente che il secondo membro può essere espresso come

$$\frac{n!}{(n-k)!}$$

vediamo che

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

come volevamo dimostrare. □

Definizione 2.2 (Combinazioni). Sia U un insieme di n oggetti (come per esempio un'urna che contiene n palline numerate) e sia Ω l'insieme di tutti i sottoinsiemi di U che contengono k oggetti diversi. Questi sottoinsiemi di U vengono chiamati **combinazioni di n oggetti presi k alla volta** e il numero di tali combinazioni è dato da

$$\#\Omega = \binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

Esercizio 2.3 (Combinazioni). Da un'urna che contiene $n = 90$ palline numerate si estraggono in blocco $k = 6$ palline.

- a) Quanti eventi elementari contiene lo spazio campionario associato a questo esperimento casuale?
- b) Si definisca un modello probabilistico (Ω, \mathcal{A}, P) per questo tipo di esperimento casuale. Secondo quale metodo si dovrebbero assegnare le probabilità?

Soluzione:

- a) Quanti eventi elementari contiene lo spazio campionario associato a questo esperimento casuale?

Lo spazio campionario Ω contiene

$$\binom{90}{6} = \frac{90!}{6!(90-6)!} = \frac{90 \times 89 \times 88 \times 87 \times 86 \times 85}{6 \times 5 \times 4 \times 3 \times 2 \times 1} = 622.614.630$$

combinazioni di palline diverse.

- b) Si definisca un modello probabilistico (Ω, \mathcal{A}, P) per questo tipo di esperimento casuale. Secondo quale metodo si dovrebbero assegnare le probabilità?

Siccome lo spazio campionario Ω è finito e le informazioni contenute nella consegna non forniscono alcun indizio sulla possibilità che alcune combinazioni $\omega \in \Omega$ si possano verificare "più facilmente" di altre, la probabilità dovrebbero essere assegnate secondo il metodo classico. La probabilità associata a ciascuna combinazione di palline (o meglio a ciascun evento che contiene soltanto una combinazione di palline) dovrebbe quindi essere definita come

$$P(\{\omega\}) = \frac{1}{622.614.630} = 1,606 \times 10^{-9},$$

e le probabilità associate agli altri eventi $A \subseteq \Omega$ dovrebbero dunque essere date dal rapporto tra il numero di casi favorevoli $\#A$ ed il numero di casi possibili $\#\Omega = 622.614.630$.

2.4 Esercizi sul calcolo combinatorio

Esercizio 2.4 (Probabilità e calcolo combinatorio). Si consideri un mazzo di 32 carte ciascuna delle quali è contrassegnata con uno degli otto valori 7, 8, 9, 10, J, Q, K oppure A, e con uno dei quattro semi cuori, quadri, fiori oppure picche.

- A) Dal mazzo vengono estratte cinque carte. Le estrazioni avvengono una dopo l'altra e senza reinserire le carte estratte nel mazzo.
- A₁) Qual è la probabilità che le cinque carte estratte siano tutte di cuori?
- A₂) Qual è la probabilità che le prime due carte estratte siano entrambe di cuori?
- A₃) Qual è la probabilità che l'ultima carta estratta sia di cuori?
- B) Si risponda alle domande del punto A) ipotizzando che dopo ciascuna estrazione la carta estratta venga reinserita nel mazzo.
- C) Dal mazzo vengono estratte in blocco cinque carte.

- C_1) Qual è la probabilità di ottenere quattro carte del valore A ?
- C_2) Qual è la probabilità di ottenere tre carte del valore A e due carte del valore K ?
- C_3) Qual è la probabilità di ottenere due carte del valore A , due carte del valore K e una carta del valore Q ?

Soluzione:

- A) Dal mazzo vengono estratte cinque carte. Le estrazioni avvengono una dopo l'altra e senza reinserire le carte estratte nel mazzo.

Estraendo una alla volta cinque carte senza reinserimento, si ottiene una delle

$$32 \times 31 \times 30 \times 29 \times 28 = 24.165.120$$

disposizioni (senza ripetizione) possibili. Siccome non c'è motivo di ritenere che alcune di esse si possano verificare "più facilmente" di altre, assegneremo le probabilità in base al metodo classico: per ogni $A \subseteq \Omega$ definiamo quindi il corrispondente valore di $P(A)$ come rapporto tra il numero $\#A$ di casi favorevoli e il numero $\#\Omega = 24.165.120$ di casi possibili. Nelle risposte ai quesiti A_1 , A_2 e A_3 determineremo i valori che si ottengono per questo rapporto per gli eventi ivi considerati.

- A_1) Qual è la probabilità che le cinque carte estratte siano tutte di cuori?

Siccome nel mazzo ci sono $32/4 = 8$ carte di cuori, la probabilità di ottenere cinque carte di cuori (evento A_1) è data da

$$\begin{aligned} P(A_1) &= \frac{8 \times 7 \times 6 \times 5 \times 4}{32 \times 31 \times 30 \times 29 \times 28} \\ &= \frac{6720}{24.165.120} = 0,000278. \end{aligned}$$

Infatti, la prima carta di cuori può essere scelta in 8 modi diversi e, dopo aver scelto la prima carta di cuori, rimangono 7 modi per scegliere la seconda carta di cuori, ecc..

- A_2) Qual è la probabilità che le prime due carte estratte siano entrambe di cuori?

La probabilità che le prime due carte estratte siano entrambe di cuori è data da

$$\begin{aligned} P(A_2) &= \frac{8 \times 7 \times 30 \times 29 \times 28}{32 \times 31 \times 30 \times 29 \times 28} \\ &= \frac{1.364.160}{24.165.120} = 0,05645 \end{aligned}$$

perché ci sono 8 modi per scegliere la prima carta di cuori tra le 8 carte di cuori nel mazzo, 7 modi per scegliere la seconda carta di cuori tra le rimanenti 7 carte di cuori, 30 modi per scegliere la terza carta (che eventualmente potrebbe anche essere di cuori) tra le 30 carte rimaste nel mazzo, 29 modi per

A_3) Qual è la probabilità che l'ultima carta estratta sia di cuori?

La probabilità che l'ultima (ovvero la quinta) carta estratta sia di cuori è data da

$$\begin{aligned} P(A_3) &= \frac{8 \times 31 \times 30 \times 29 \times 28}{32 \times 31 \times 30 \times 29 \times 28} \\ &= \frac{6.041.280}{24.165.120} = 0,25 \end{aligned}$$

perché ci sono 8 modi per scegliere una carta di cuori da posizionare nella quinta posizione di una disposizione che è favorevole all'evento A_3 , 31 modi per scegliere tra le carte rimaste nel mazzo una carta (che eventualmente potrebbe anche essere di cuori) da posizionare nella quarta posizione, 30 modi per scegliere tra le carte rimaste nel mazzo una carta (che eventualmente potrebbe anche essere di cuori) da posizionare della terza posizione, ecc..

Si osservi che la probabilità che l'ultima carta estratta sia di cuori è uguale alla probabilità che la prima carta estratta sia di cuori. Più in generale, la probabilità di ottenere una carta di cuori in una qualsiasi posizione prefissata coincide con la probabilità di ottenere una carta di cuori nella prima estrazione.

B) Si risponda alle domande del punto A) ipotizzando che dopo ciascuna estrazione la carta estratta venga reinserita nel mazzo.

Nel caso di estrazioni con reinserimento lo spazio campionario contiene

$$\begin{aligned} 32 \times 32 \times 32 \times 32 \times 32 &= \\ &= 32^5 = 33.554.432 \end{aligned}$$

disposizioni con ripetizione. Nelle risposte ai quesiti B_1 , B_2 e B_3 assumeremo che nessuna di queste disposizioni possa verificarsi "più facilmente" delle altre e quindi determineremo le probabilità che corrispondono al metodo di assegnazione classico.

B_1) Qual è la probabilità che le cinque carte estratte siano tutte di cuori?

Siccome nel mazzo ci sono $32/4 = 8$ carte di cuori, la probabilità di ottenere cinque carte di cuori (evento B_1) è data da

$$P(B_1) = \frac{8 \times 8 \times 8 \times 8 \times 8}{32 \times 32 \times 32 \times 32 \times 32} = \frac{32.768}{33.554.432} = 0,000976.$$

Infatti, ci sono 8 modi per scegliere la prima carta tra quelle di cuori che sono presenti nel mazzo, 8 modi per scegliere la seconda carta tra quelle di cuori che sono presenti nel mazzo, ecc..

B_2) Qual è la probabilità che le prime due carte estratte siano entrambe di cuori?

La probabilità che le prime due carte estratte siano entrambe di cuori è data da

$$\begin{aligned} P(B_2) &= \frac{8 \times 8 \times 32 \times 32 \times 32}{32 \times 32 \times 32 \times 32 \times 32} \\ &= \frac{2.097.152}{33.554.432} = 0,0625 \end{aligned}$$

perché ci sono 8 modi per scegliere la prima carta di cuori, 8 modi per scegliere la seconda carta di cuori, 32 modi per scegliere la terza carta (che eventualmente potrebbe anche essere di cuori), 32 modi per ...

B_3) Qual è la probabilità che l'ultima carta estratta sia di cuori?

La probabilità che l'ultima carta estratta sia di cuori è data da

$$\begin{aligned} P(B_3) &= \frac{32 \times 32 \times 32 \times 32 \times 8}{32 \times 32 \times 32 \times 32 \times 32} \\ &= \frac{8.388.608}{33.554.432} = 0,25 \end{aligned}$$

perché ci sono 8 modi per scegliere una carta di cuori da posizionare nella quinta posizione di una disposizione che è favorevole all'evento B_3 , 32 modi per scegliere una qualunque carte (che eventualmente potrebbe anche essere di cuori) da posizionare nella quarta posizione, 32 modi per scegliere una qualunque carta (che eventualmente potrebbe anche essere di cuori) da posizionare nella terza posizione, ecc..

C) Dal mazzo vengono estratte in blocco cinque carte.

Nel caso di un'estrazione in blocco di cinque carte, lo spazio campionario contiene

$$\binom{32}{5} = \frac{32!}{5!(32-5)!} = 201.376$$

combinazioni. Nelle risposte ai quesiti C_1 , C_2 e C_3 assumeremo che nessuna di queste combinazioni possa verificarsi "più facilmente" delle altre e quindi determineremo le probabilità che corrispondono al metodo di assegnazione classico.

C_1) Qual è la probabilità di ottenere quattro carte del valore A ?

La probabilità di ottenere quattro carte del valore A è data da

$$P(C_1) = \frac{\binom{4}{4} \times \binom{28}{1}}{\binom{32}{5}} = \frac{1 \times 28}{201.376} = 0,000139,$$

perché ci sono $\binom{4}{4} = 1$ modi per scegliere 4 carte del valore A tra le 4 carte del valore A nel mazzo e $\binom{28}{1} = 28$ modi per scegliere una carta tra le 28 carte restanti.

C_2) Qual è la probabilità di ottenere tre carte del valore A e due carte del valore K ?

La probabilità di ottenere tre carte del valore A e due carte del valore K è data da

$$\begin{aligned} P(C_2) &= \frac{\binom{4}{3} \times \binom{4}{2}}{\binom{32}{5}} \\ &= \frac{4 \times 6}{201.376} = 0,000119. \end{aligned}$$

Infatti, ci sono $\binom{4}{3} = 4$ modi per scegliere 3 carte del valore A tra le 4 carte del valore A disponibili nel mazzo, e ciascuno di questi modi può essere combinato con uno dei $\binom{4}{2}$ modi possibili per scegliere 2 carte del valore K tra le 4 carte del valore K disponibili nel mazzo.

C_3) Qual è la probabilità di ottenere due carte del valore A , due carte del valore K e una carta del valore Q ?

La probabilità di ottenere due carte del valore A , due carte del valore K ed una carta del valore Q è data da

$$\begin{aligned} P(C_3) &= \frac{\binom{4}{2} \times \binom{4}{2} \times \binom{4}{1}}{\binom{32}{5}} \\ &= \frac{6 \times 6 \times 4}{201.376} \\ &= \frac{144}{201.376} = 0,000715, \end{aligned}$$

perché ci sono $\binom{4}{2}$ modi per scegliere due carte del valore A tra le quattro che sono presenti nel mazzo, $\binom{4}{2}$ modi per scegliere due carte del valore K tra le quattro che sono presenti nel mazzo e infine $\binom{4}{1}$ modi per scegliere una carta del valore Q tra le quattro che sono presenti nel mazzo.

Esercizio 2.5 (Probabilità e calcolo combinatorio). Ad un concorso canoro per bambini la classifica finale viene redatta in modo casuale. Al concorso partecipano 6 bambini.

- a) Qual è la probabilità che il bambino Marco vinca il concorso?
- b) Qual è la probabilità che il bambino Marco finisca in una delle prime tre posizioni?
- c) Qual è la probabilità che Marco e Chiara finiscano entrambi nelle prime tre posizioni?

Soluzione:

Siccome per ipotesi tutte le $6! = 720$ classifiche hanno la stessa probabilità di verificarsi, calcoleremo le probabilità richieste come rapporti tra il numero di casi favorevoli ed il numero di casi possibili (che è sempre uguale a $6! = 720$).

- a) Qual è la probabilità che il bambino Marco vinca il concorso?

La probabilità che Marco vinca il concorso (evento M_1) è data da

$$P(M_1) = \frac{1 \times 5 \times 4 \times 3 \times 2 \times 1}{6!} = \frac{1}{6}$$

perché dopo aver posizionato Marco in prima posizione, rimangono 5 modi per scegliere il bambino in seconda posizione; ...

... e ciascun modo di riempire le prime due posizioni può essere combinato con 4 modi diversi per riempire la terza posizione; ...

... e ciascun modo di riempire le prime tre posizioni può essere combinato con 3 modi diversi per riempire la quarta posizione; ecc..

- b) Qual è la probabilità che il bambino Marco finisca in una delle prime tre posizioni?

Per determinare la probabilità che Marco finisca in una delle prime tre posizioni (evento $M_1 \cup M_2 \cup M_3$) conviene innanzitutto osservare che gli eventi M_1 , M_2 e M_3 sono incompatibili e che la probabilità dell'unione di questi tre eventi deve dunque essere data da

$$P(M_1 \cup M_2 \cup M_3) = P(M_1) + P(M_2) + P(M_3).$$

Siccome $P(M_1) = P(M_2) = P(M_3)$ (invitiamo il lettore a verificare), e siccome

$$P(M_1) = \frac{1}{6} \quad (\text{vedi la risposta al quesito a)),$$

possiamo concludere che

$$P(M_1 \cup M_2 \cup M_3) = 3 \times \frac{1}{6} = \frac{1}{2}.$$

- c) Qual è la probabilità che Marco e Chiara finiscano entrambi nelle prime tre posizioni?

Il numero di casi favorevoli all'evento secondo il quale Marco e Chiara finiscono entrambi nelle prime tre posizioni è dato da

$$\binom{3}{2} \times 2 \times 4 \times 3 \times 2 \times 1 = 144$$

perché ci sono $\binom{3}{2}$ modi per scegliere, tra le prime tre posizioni, le due posizioni da occupare con Marco e Chiara; ...

... per ciascun modo di scegliere le due posizioni da occupare con Marco e Chiara ci sono 2 modi diversi per posizionare Marco e Chiara nelle due posizioni scelte (prima Marco e poi Chiara oppure viceversa); ...

... e una volta posizionati Marco e Chiara rimangono $4! = 4 \times 3 \times 2 \times 1$ modi per riempire le quattro posizioni che sono ancora vuote.

Dividendo il numero di casi favorevoli per il numero di casi possibili (che, come abbiamo già visto, è uguale $6! = 720$) otteniamo la probabilità richiesta:

$$P(A) = \frac{\binom{3}{2} \times 2 \times 4 \times 3 \times 2 \times 1}{6!} = \frac{144}{720} = 0,2$$

Esercizio 2.6 (Probabilità e calcolo combinatorio). Quanti sottoinsiemi diversi si possono ottenere con un insieme Ω che contiene un numero finito $\#\Omega$ di elementi?

Soluzione:

Per rispondere alla domanda conviene partire da un caso particolare, per esempio dal caso particolare dove $\#\Omega = 4$ e dove l'insieme Ω può dunque essere rappresentato come

$$\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}.$$

Consideriamo un qualunque sottoinsieme di questo insieme Ω , per esempio l'insieme

$$A = \{\omega_1, \omega_3, \omega_4\}.$$

Notiamo che a questo sottoinsieme di Ω corrisponde la sequenza $(1, 0, 1, 1)$ perché questa sequenza ci dice che l'elemento ω_1 è contenuto nell'insieme A (si noti l'uno nella prima posizione della sequenza), che l'elemento ω_2 non è contenuto nell'insieme A (si noti lo zero nella seconda posizione della sequenza), e che gli elementi ω_3 e

ω_4 sono entrambi contenuti nell'insieme A (si notino i due uni nella terza e quarta posizione della sequenza). Allo stesso modo, all'insieme

$$B = \{\omega_1, \omega_4\}$$

corrisponde la sequenza $(1, 0, 0, 1)$ che è diversa dalla sequenza che corrisponde ad A . Da questo ragionamento deduciamo che ad ogni sottoinsieme di Ω corrisponde un'unica sequenza e che le sequenze corrispondenti a due sottoinsiemi diversi non possono essere identiche. Ovviamente vale anche il viceversa: a ciascuna sequenza di lunghezza $\#\Omega = 4$ che è formata solo da zeri e uni corrisponde un unico sottoinsieme di Ω e a due sequenze diverse corrispondono sempre due sottoinsiemi di Ω diversi. Ora siccome ciascuna delle $\#\Omega = 4$ posizioni di una delle suddette sequenze può essere occupata soltanto in due modi diversi (ovvero con uno 0 oppure con un 1), il numero complessivo di tali sequenze (e quindi anche il numero complessivo di sottoinsiemi di Ω) deve essere dato da

$$2 \times 2 \times 2 \times 2 = 2^4 = 16.$$

Ovviamente questo conteggio include anche l'insieme vuoto \emptyset (che corrisponde alla sequenza $(0, 0, 0, 0)$) e l'insieme Ω (che corrisponde alla sequenza $(1, 1, 1, 1)$).

Generalizzando il precedente ragionamento non è difficile rendersi conto che il numero complessivo di sottoinsiemi diversi che si possono formare con un generico insieme Ω che contiene un numero finito $\#\Omega$ di elementi deve essere dato da

$$2 \times 2 \times \dots (\#\Omega \text{ volte}) \dots \times 2 = 2^{\#\Omega}.$$

Ora, siccome il numero complessivo di sottoinsiemi diversi di Ω deve essere anche uguale a

$$\sum_{x=0}^{\#\Omega} \binom{\#\Omega}{x} = \binom{\#\Omega}{0} + \binom{\#\Omega}{1} + \dots + \binom{\#\Omega}{x} + \dots + \binom{\#\Omega}{\#\Omega}$$

(infatti il numero complessivo di sottoinsiemi diversi di Ω deve coincidere con il numero complessivo di sottoinsiemi che contengono $x = 0$ elementi + il numero complessivo di sottoinsiemi che contengono $x = 1$ elemento e così via), possiamo concludere che

$$2^{\#\Omega} = \binom{\#\Omega}{0} + \binom{\#\Omega}{1} + \dots + \binom{\#\Omega}{x} + \dots + \binom{\#\Omega}{\#\Omega} = \sum_{x=0}^{\#\Omega} \binom{\#\Omega}{x}.$$

Si noti che questo risultato è un caso particolare del cosiddetto *teorema binomiale* secondo il quale

$$(a + b)^n = \sum_{x=0}^n \binom{n}{x} a^x b^{n-x}$$

(basta porre $a = b = 1$ e $n = \#\Omega$).

3 Probabilità condizionata

3.1 Definizione di probabilità condizionata

Consideriamo un esperimento casuale che consiste nel lancio di due dadi e indichiamo con (Ω, \mathcal{A}, P) un modello probabilistico che descrive questo esperimento casuale. Definiamo dunque ...

- ... lo spazio campionario Ω come

$$\Omega = \left\{ \begin{array}{cccccc} (1, 1), & (1, 2), & \dots, & \dots, & (1, 6) \\ (2, 1), & (2, 2), & \dots, & \dots, & (2, 6) \\ \vdots, & \dots, & \ddots, & \dots, & \vdots \\ (6, 1), & (6, 2), & \dots, & \dots, & (6, 6) \end{array} \right\},$$

- ... la σ -algebra di eventi come $\mathcal{A} = \mathcal{P}(\Omega)$,
- ... e la funzione di probabilità $P(\cdot)$ come quella che corrisponde al metodo di assegnazione classico:

$$P(A) = \frac{\#A}{\#\Omega} = \frac{\#A}{6 \times 6} = \frac{\#A}{36} \quad \text{per ogni } A \subseteq \Omega.$$

Consideriamo ora l'evento A secondo il quale la somma dei punteggi è pari a dodici. Ovviamente, $A = \{(6, 6)\}$, e per valutare quanto sia "probabile" ottenere una somma di dodici calcoliamo

$$P(A) = \frac{\#A}{\#\Omega} = \frac{\#A}{6 \times 6} = \frac{1}{36}.$$

Supponiamo ora che i due dadi vengano lanciati uno dopo l'altro e che l'esito del primo lancio sia il punteggio sei (evento B). In questo caso, sarebbe ancora "corretto" valutare la "probabilità" di ottenere dodici (evento A) attraverso $P(A) = 1/36$??? Ovviamente la risposta a questa domanda è negativa. Infatti, per ottenere una valutazione "corretta" della "probabilità" di ottenere dodici (evento A) bisognerebbe fare riferimento alla cosiddetta **probabilità condizionata** che nel caso in questione è data da

$$P(A|B) = \frac{\#(A \cap B)}{\#B} = \frac{\#\{(6, 6)\}}{\#\{(6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6)\}} = \frac{1}{6}.$$

A questo punto potremmo chiederci se la probabilità condizionata calcolata secondo la precedente formula restituisce il risultato "corretto" anche nel caso in cui l'esito del primo lancio fosse il punteggio cinque (evento C). Non è difficile verificare che la risposta è affermativa. Infatti, se l'esito del primo lancio è il punteggio cinque, la precedente formula restituisce la probabilità condizionata

$$P(A|C) = \frac{\#(A \cap C)}{\#C} = \frac{\#\emptyset}{\#\{(5, 1), (5, 2), (5, 3), (5, 4), (5, 5), (5, 6)\}} = \frac{0}{6} = 0$$

(si noti che $A \cap C = \emptyset$ perché $A = \{(6, 6)\}$ e C è l'evento che contiene tutte le coppie ordinate con il punteggio 5 in prima posizione). Sulla scia di queste considerazioni possiamo definire il concetto di **probabilità condizionata** come segue:

Definizione 3.1 (Probabilità condizionata). Sia $P(\cdot)$ una funzione di probabilità che soddisfa gli assiomi di Kolmogorov, e siano A e B due eventi che appartengono al dominio \mathcal{A} della funzione di probabilità $P(\cdot)$. Se $P(B) > 0$, allora la **probabilità condizionata** dell'evento A dato l'evento B è definita come

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

In parole: la probabilità condizionata $P(A|B)$ è il rapporto tra

- la **probabilità congiunta** degli eventi A e B (ovvero $P(A \cap B)$)
- e la **probabilità marginale** dell'evento B (ovvero $P(B)$).

3.2 Il teorema della probabilità composta

Partendo dalla definizione di probabilità condizionata

$$P(A_2|A_1) = \frac{P(A_1 \cap A_2)}{P(A_1)}$$

ricaviamo immediatamente la formula

$$P(A_1 \cap A_2) = P(A_1) \times P(A_2|A_1). \quad (16)$$

A stretto rigore, questa formula ha senso solo se $P(A_1) > 0$ (altrimenti la probabilità condizionata $P(A_2|A_1)$ non sarebbe definita), ma può essere ritenuta valida anche nel caso in cui $P(A_1) = 0$ a patto che il prodotto $P(A_1) \times P(A_2|A_1) = 0 \times P(A_2|A_1)$ venga interpretato come nullo. Infatti, dalla legge L4 discende che $P(A_1) = 0 \Rightarrow P(A_1 \cap A_2) = 0$.

Consideriamo ora invece la probabilità condizionata

$$P(A_3|A_1 \cap A_2) = \frac{P(A_1 \cap A_2 \cap A_3)}{P(A_2 \cap A_1)},$$

dove assumiamo che $P(A_1 \cap A_2) > 0$ (altrimenti la probabilità condizionata al primo membro non sarebbe definita). Moltiplicando ambo i membri di questa uguaglianza per $P(A_1 \cap A_2)$, otteniamo

$$P(A_1 \cap A_2 \cap A_3) = P(A_1 \cap A_2) \times P(A_3|A_1 \cap A_2)$$

e ricordando l'uguaglianza (16) (si noti che nel caso in questione la probabilità condizionata $P(A_2|A_1)$ che appare al secondo membro della (16) è definita perché stiamo assumendo

che $P(A_1 \cap A_2) > 0$ e perché, secondo la legge L4, $A_1 \cap A_2 \subseteq A_1 \Rightarrow 0 < P(A_1 \cap A_2) \leq P(A_1)$, vediamo che

$$P(A_1 \cap A_2 \cap A_3) = P(A_1) \times P(A_2|A_1) \times P(A_3|A_1 \cap A_2). \quad (17)$$

Non è difficile verificare che anche questa formula può essere considerata valida anche quando una o entrambe le probabilità condizionate $P(A_2|A_1)$ e $P(A_3|A_1 \cap A_2)$ non sono definite, a patto che in tal caso il prodotto al secondo membro venga considerato nullo.

Iterando il ragionamento che ci ha condotto alla formula (17) otteniamo la dimostrazione del cosiddetto **teorema della probabilità composta**:

Teorema 3.1 (Teorema della probabilità composta). La probabilità congiunta

$$P(A_1 \cap A_2 \cap \cdots \cap A_k \cap A_{k+1})$$

può essere **fattorizzata** come

$$P(A_1 \cap A_2 \cap \cdots \cap A_k \cap A_{k+1}) = P(A_1) \times P(A_2|A_1) \times P(A_3|A_1 \cap A_2) \times \cdots \times P(A_{k+1}|A_1 \cap A_2 \cap \cdots \cap A_k).$$

A stretto rigore questa formula sarebbe valida solo se tutte le probabilità condizionate al secondo membro sono definite, ma può essere ritenuta valida anche in caso contrario a patto che il prodotto al secondo membro venga considerato nullo quando almeno una delle probabilità condizionate non è definita.

Esercizio 3.1. Un'urna contiene 10 palline bianche e 5 palline rosse. Dall'urna vengono estratte, una dopo l'altra, 4 palline. Dopo ogni estrazione vengono reinserite nell'urna due palline del colore della pallina estratta.

Supponendo che in ciascuna delle 4 estrazioni tutte le palline presenti nell'urna abbiano la medesima probabilità (condizionata) di essere estratte, a quanto ammonta il valore della probabilità congiunta degli eventi

$$B_1 = \{\text{nella prima estrazione si ottiene una pallina bianca}\},$$

$$B_2 = \{\text{nella seconda estrazione si ottiene una pallina bianca}\},$$

$$R_3 = \{\text{nella terza estrazione si ottiene una pallina rossa}\},$$

$$B_4 = \{\text{nella quarta estrazione si ottiene una pallina bianca}\} ?$$

Risposta:

Ovviamente, l'esperimento casuale in questione può essere interpretato come un esperimento che si svolge in $k = 4$ passi.

- (Prima estrazione) Siccome prima di procedere alla prima estrazione l'urna contiene 10 palline bianche e 5 palline rosse, e siccome stiamo ipotizzando che ad ogni estrazione ciascuna pallina presente nell'urna abbia la medesima probabilità (condizionata) di essere estratta, la probabilità di ottenere una pallina bianca nella prima estrazione deve essere data da

$$P(B_1) = \frac{10}{10 + 5} = \frac{10}{15}.$$

- (Seconda estrazione) Secondo la descrizione dell'esperimento casuale, se nella prima estrazione si ottiene una pallina bianca (evento B_1), allora prima di procedere alla seconda estrazione (secondo passo) nell'urna devono essere reinserite due palline bianche.

Siccome stiamo ipotizzando che ad ogni estrazione ciascuna pallina presente nell'urna abbia la medesima probabilità (condizionata) di essere estratta, la probabilità condizionata di ottenere una pallina bianca nella seconda estrazione (evento B_2), **dato che nella prima estrazione è stata ottenuta una pallina bianca (evento B_1)**, deve essere data da

$$P(B_2|B_1) = \frac{10 - 1 + 2}{(10 - 1 + 2) + 5} = \frac{11}{16}.$$

- (Terza estrazione) Secondo la descrizione dell'esperimento casuale, se in entrambe le prime due estrazioni si ottiene una pallina bianca (evento $B_1 \cap B_2$), allora prima di procedere alla terza estrazione nell'urna devono esserci $10 - 1 + 2 - 1 + 2 = 12$ palline bianche e 5 palline rosse.

Siccome stiamo ipotizzando che ad ogni estrazione ciascuna pallina presente nell'urna abbia la medesima probabilità (condizionata) di essere estratta, la probabilità condizionata di ottenere una pallina rossa nella terza estrazione (evento R_3), **dato che nelle prime due estrazioni sono state ottenute due palline bianche (evento $B_1 \cap B_2$)**, deve essere data da

$$P(R_3|B_1 \cap B_2) = \frac{5}{(10 - 1 + 2 - 1 + 2) + 5} = \frac{5}{17}.$$

- (Quarta estrazione) Secondo la descrizione dell'esperimento casuale, se nelle prime due estrazioni si ottengono due palline bianche e nella terza estrazione si ottiene una pallina rossa (evento $B_1 \cap B_2 \cap R_3$), allora prima di procedere alla quarta e ultima estrazione nell'urna dovrebbero esserci $10 - 1 + 2 - 1 + 2 = 12$ palline bianche e $5 - 1 + 2 = 6$ palline rosse.

Siccome stiamo ipotizzando che ad ogni estrazione ciascuna pallina presente nell'urna abbia la medesima probabilità (condizionata) di essere estratta, la

probabilità condizionata di ottenere una pallina bianca alla quarta estrazione (evento B_4), **dato che sono state ottenute due palline bianche nelle prime due estrazioni e dato che è stata ottenuta una pallina rossa nella terza estrazione (evento $B_1 \cap B_2 \cap R_3$)**, deve essere data da

$$P(B_4|B_1 \cap B_2 \cap R_3) = \frac{10 - 1 + 2 - 1 + 2}{(10 - 1 + 2 - 1 + 2) + (5 - 1 + 2)} = \frac{12}{18}.$$

Avendo determinato i valori di

$$P(B_1) = \frac{10}{15}, \quad P(B_2|B_1) = \frac{11}{16},$$

$$P(R_3|B_1 \cap B_2) = \frac{5}{17}, \quad P(B_4|B_1 \cap B_2 \cap R_3) = \frac{12}{18},$$

possiamo finalmente calcolare il valore della probabilità congiunta $P(B_1 \cap B_2 \cap R_3 \cap B_4)$. Applicando il [teorema della probabilità composta](#) vediamo che il valore di $P(B_1 \cap B_2 \cap R_3 \cap B_4)$ è dato da

$$\begin{aligned} P(B_1 \cap B_2 \cap R_3 \cap B_4) &= \\ &= P(B_1) \times P(B_2|B_1) \times P(R_3|B_1 \cap B_2) \times P(B_4|B_1 \cap B_2 \cap R_3) \\ &= \frac{12}{18} \times \frac{5}{17} \times \frac{11}{16} \times \frac{10}{15} = 0,0899. \end{aligned}$$

3.3 Eventi indipendenti

Nel nostro modo di esprimerci quotidiano consideriamo due eventi A e B come "indipendenti" se il verificarsi di uno dei due eventi non modifica la nostra opinione su quanto sia "probabile" che si verifichi anche l'altro. Questo concetto informale di "indipendenza" può essere tradotto in una definizione matematica dicendo che due eventi A e B che appartengono entrambi al dominio \mathcal{A} di una funzione di probabilità $P(\cdot)$ sono indipendenti se

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = P(A) \quad \text{nel caso in cui } P(B) > 0$$

e

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = P(B) \quad \text{nel caso in cui } P(A) > 0.$$

Tuttavia, questo modo di definire eventi indipendenti è un po' complicato e per questo motivo viene di solito preferita la seguente definizione che, non è difficile verificarlo, è equivalente a quella che abbiamo appena dato:

Definizione 3.2 (Eventi indipendenti). Due eventi A e B si dicono **indipendenti** se la loro probabilità congiunta è uguale al prodotto delle loro probabilità marginali,

ovvero se

$$P(A \cap B) = P(A) \times P(B).$$

Quando si analizzano eventi indipendenti torna spesso utile il seguente teorema:

Teorema 3.2 (Implicazioni dell'indipendenza tra eventi). Se A e B sono una coppia di eventi indipendenti, allora è indipendente anche

- la coppia di eventi A e \bar{B} ,
- la coppia di eventi \bar{A} e B
- e la coppia di eventi \bar{A} e \bar{B} .

Dimostrazione. Consideriamo la tabella a doppia entrata

	B	\bar{B}	Totale
A	$P(A \cap B)$	$P(A \cap \bar{B})$	$P(A)$
\bar{A}	$P(\bar{A} \cap B)$	$P(\bar{A} \cap \bar{B})$	$P(\bar{A})$
Totale	$P(B)$	$P(\bar{B})$	1

Notiamo che per l'assioma **K3** le probabilità nelle righe e colonne marginali devono coincidere con le corrispondenti somme per riga e per colonna. Inoltre, per la legge **L3** si deve anche avere

$$P(\bar{A}) = 1 - P(A) \quad \text{e} \quad P(\bar{B}) = 1 - P(B).$$

Supponiamo ora che gli eventi A e B siano indipendenti, ovvero assumiamo che

$$P(A \cap B) = P(A) \times P(B).$$

Allora, per i vincoli sulle probabilità marginali, devono valere le seguenti uguaglianze:

$$\begin{aligned} P(A \cap \bar{B}) &= P(A) - P(A \cap B) \\ &= P(A) - P(A) \times P(B) \\ &= P(A) \times (1 - P(B)) \\ &= P(A) \times P(\bar{B}) \end{aligned}$$

(questo dimostra che gli eventi A e \bar{B} sono indipendenti),

$$\begin{aligned} P(\bar{A} \cap B) &= P(B) - P(A \cap B) \\ &= P(B) - P(A) \times P(B) \\ &= (1 - P(A)) \times P(B) \\ &= P(\bar{A}) \times P(B) \end{aligned} \tag{18}$$

(questo dimostra che gli eventi \bar{A} e B sono anch'essi indipendenti) e

$$\begin{aligned} P(\bar{A} \cap \bar{B}) &= P(\bar{A}) - P(\bar{A} \cap B) \\ [[\text{per la (18)}]] &= P(\bar{A}) - P(\bar{A}) \times P(B) \\ &= P(\bar{A}) \times (1 - P(B)) \\ &= P(\bar{A}) \times P(\bar{B}) \end{aligned}$$

(questo dimostra che anche gli eventi \bar{A} e \bar{B} sono indipendenti). □

Esercizio 3.2 (Indipendenza e diversificazione). Un risparmiatore ha investito in due titoli azionari. La probabilità che ad un anno dall'investimento il primo titolo presenti un rendimento positivo (evento A) è pari a 0,60. D'altra parte, la probabilità che ad un anno dall'investimento il secondo titolo presenti un rendimento positivo (evento B) è soltanto pari a 0,55.

- a) Supponendo che $P(A|B) = 0,90$, qual è la probabilità che ad un anno dall'investimento almeno uno dei due titoli presenti un rendimento positivo?
- b) Supponendo che gli eventi A e B siano indipendenti (ovvero che $P(A|B) = P(A) = 0,60$), qual è la probabilità che ad un anno dall'investimento almeno uno dei due titoli presenti un rendimento positivo?
- c) Supponendo che $P(A|B) = 0,40$, qual è la probabilità che ad un anno dall'investimento almeno uno dei due titoli presenti un rendimento positivo?

Soluzione:

In tutti e tre i quesiti è richiesta la probabilità che ad un anno dall'investimento almeno uno dei due titoli presenti un rendimento positivo. Per la legge L5, la probabilità richiesta deve essere data da

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

in tutti e tre i casi, ma, come vedremo tra breve, i valori di $P(A \cap B)$ (e dunque anche quelli di $P(A \cup B)$) corrispondenti alle ipotesi fatte nei tre quesiti sono molto diversi tra di loro.

- a) Supponendo che $P(A|B) = 0,90$, qual è la probabilità che ad un anno dall'investimento almeno uno dei due titoli presenti un rendimento positivo?

Se $P(A|B) = 0,90$, si ottiene

$$P(A \cap B) = P(A|B) \times P(B) = 0,90 \times 0,55 = 0,495,$$

e di conseguenza si ha $P(A \cup B) = 0,60 + 0,55 - 0,495 = 0,655$.

- b) Supponendo ora invece che gli eventi A e B siano indipendenti (ovvero che $P(A|B) = P(A) = 0,60$), qual è la probabilità che ad un anno dall'investimento almeno uno dei due titoli presenti un rendimento positivo?

Se A e B sono indipendenti, si ottiene

$$P(A \cap B) = P(A) \times P(B) = 0,60 \times 0,55 = 0,33,$$

e la probabilità dell'unione è quindi data da

$$P(A \cup B) = 0,60 + 0,55 - 0,33 = 0,82.$$

- c) Supponendo che $P(A|B) = 0,40$, qual è la probabilità che ad un anno dall'investimento almeno uno dei due titoli presenti un rendimento positivo?

Se $P(A|B) = 0,40$, si ottiene

$$P(A \cap B) = P(A|B) \times P(B) = 0,40 \times 0,55 = 0,22,$$

e la probabilità dell'unione è quindi data da

$$P(A \cup B) = 0,60 + 0,55 - 0,22 = 0,93.$$

Vale la pena notare che i valori ipotizzati per $P(A|B)$ nei quesiti a), b) e c) sono molto diversi tra di loro:

$$\text{a) } P(A|B) = 0,9; \quad \text{b) } P(A|B) = P(A) = 0,6; \quad \text{c) } P(A|B) = 0,4.$$

Ne consegue il valore di $P(A \cap B) = P(A|B) \times P(B)$ che si ottiene sotto l'ipotesi del quesito a) è più elevato di quello che si ottiene sotto l'ipotesi del quesito b) che a sua volta è più elevato di quello che si ottiene sotto l'ipotesi del quesito c). Esaminando la tabella a doppia entrata

	B	\bar{B}	Totale
A	$P(A \cap B)$	$P(A \cap \bar{B})$	$P(A) = 0,6$
\bar{A}	$P(\bar{A} \cap B)$	$P(\bar{A} \cap \bar{B})$	$P(\bar{A}) = 0,4$
	$P(B) = 0,55$	$P(\bar{B}) = 0,45$	1

(19)

ci rendiamo conto che a valori più elevati di $P(A \cap B)$ corrispondono valori più piccoli di $P(A \cap \bar{B})$ e $P(\bar{A} \cap B)$. Quindi possiamo dire che a valori più elevati di $P(A \cap B)$ corrisponde un livello di diversificazione più basso.

3.4 Indipendenza globale

Nel nostro modo di esprimerci quotidiano consideriamo tre eventi A_1 , A_2 e A_3 come "indipendenti" se il verificarsi di uno qualunque di essi non modifica la nostra opinione su quanto sia "probabile" che si verifichi uno degli altri due oppure anche entrambi, e se il verificarsi di due qualunque di essi non modifica la nostra opinione su quanto sia "probabile" che si verifichi anche il terzo. Questo concetto informale di "indipendenza" riferito a tre eventi può essere tradotto in una definizione matematica dicendo che tre eventi A_1 , A_2 e A_3 che appartengono tutti al dominio di una data funzione di probabilità $P(\cdot)$ sono indipendenti se sono soddisfatte tutte le seguenti condizioni:

$$\begin{aligned}
 P(A_1) > 0 &\Rightarrow \begin{cases} P(A_2|A_1) = \frac{P(A_1 \cap A_2)}{P(A_1)} = P(A_2); \\ P(A_3|A_1) = \frac{P(A_1 \cap A_3)}{P(A_1)} = P(A_3); \\ P(A_2 \cap A_3|A_1) = \frac{P(A_1 \cap A_2 \cap A_3)}{P(A_1)} = P(A_2 \cap A_3); \end{cases} \\
 P(A_2) > 0 &\Rightarrow \begin{cases} P(A_1|A_2) = \frac{P(A_1 \cap A_2)}{P(A_2)} = P(A_1); \\ P(A_3|A_2) = \frac{P(A_2 \cap A_3)}{P(A_2)} = P(A_3); \\ P(A_1 \cap A_3|A_2) = \frac{P(A_1 \cap A_2 \cap A_3)}{P(A_2)} = P(A_1 \cap A_3); \end{cases} \\
 P(A_3) > 0 &\Rightarrow \begin{cases} P(A_1|A_3) = \frac{P(A_1 \cap A_3)}{P(A_3)} = P(A_1); \\ P(A_2|A_3) = \frac{P(A_2 \cap A_3)}{P(A_3)} = P(A_2); \\ P(A_1 \cap A_2|A_3) = \frac{P(A_1 \cap A_2 \cap A_3)}{P(A_3)} = P(A_1 \cap A_2); \end{cases} \\
 P(A_1 \cap A_2) > 0 &\Rightarrow P(A_3|A_1 \cap A_2) = \frac{P(A_1 \cap A_2 \cap A_3)}{P(A_1 \cap A_2)} = P(A_3); \\
 P(A_1 \cap A_3) > 0 &\Rightarrow P(A_2|A_1 \cap A_3) = \frac{P(A_1 \cap A_2 \cap A_3)}{P(A_1 \cap A_3)} = P(A_2); \\
 P(A_2 \cap A_3) > 0 &\Rightarrow P(A_1|A_2 \cap A_3) = \frac{P(A_1 \cap A_2 \cap A_3)}{P(A_2 \cap A_3)} = P(A_1).
 \end{aligned}$$

Chiaramente, questo modo di definire terne di eventi indipendenti è molto complicato e per questo motivo viene di solito preferita la seguente definizione che, non è difficile verificarlo, è equivalente:

Definizione 3.3 (Indipendenza globale di tre eventi). Tre eventi A_1 , A_2 , A_3 si dicono **globalmente indipendenti** se la probabilità congiunta di qualunque coppia o terna che si può formare con i tre eventi è uguale al prodotto delle corrispondenti probabilità marginali, ovvero se sono soddisfatte tutte e quattro le seguenti

uguaglianze:

$$\begin{aligned} P(A_1 \cap A_2) &= P(A_1) \times P(A_2), \\ P(A_1 \cap A_3) &= P(A_1) \times P(A_3), \\ P(A_2 \cap A_3) &= P(A_2) \times P(A_3), \\ P(A_1 \cap A_2 \cap A_3) &= P(A_1) \times P(A_2) \times P(A_3) \end{aligned}$$

Con riferimento alla precedente definizione potrebbe sorgere il dubbio che l'ultima uguaglianza sia ridondante, ovvero che sia già implicita nelle precedenti tre uguaglianze. In altre parole, potrebbe sorgere il dubbio che tre eventi siano sempre globalmente indipendenti se ciascuna coppia che può essere formata con i tre eventi è una coppia di eventi indipendenti nel senso della Definizione 3.2. Nel prossimo esempio vedremo che questo non è vero.

Esempio 3.1 (Indipendenza globale - primo esempio). In questo esempio vedremo come nell'ambito di un semplicissimo esperimento casuale si possono definire tre eventi che sono indipendenti a coppie ma che non sono globalmente indipendenti. A tal fine consideriamo l'esperimento casuale che consiste in tre lanci di una moneta e definiamo

- A_1 come l'evento secondo il quale gli esiti dei primi due lanci sono uguali tra di loro,
- A_2 l'evento secondo il quale gli esiti del primo e del terzo lancio sono uguali tra di loro,
- e A_3 come l'evento secondo il quale gli esiti degli ultimi due lanci sono uguali tra di loro.

Chiaramente, gli eventi elementari che possono verificarsi al termine di questo esperimento casuale possono essere rappresentati attraverso delle terne ordinate i cui elementi sono delle "T" (per testa) oppure delle "C" (per croce). Siccome ogni elemento di una terna può essere scelto in 2 modi diversi, possiamo concludere che lo spazio campionario Ω contiene

$$\#\Omega = 2 \times 2 \times 2 = 2^3 = 8$$

eventi elementari. Ora, siccome non c'è motivo di ritenere che una sequenza di esiti si verifichi "più facilmente" delle altre, le probabilità di tutti gli eventi associati a questo esperimento casuale dovrebbero essere definite secondo quanto previsto dal metodo di assegnazione classico:

- Siccome il numero di casi favorevoli all'evento A_1 è dato da

$$\#A_1 = 2 \times 1 \times 2 = 4$$

(infatti, il primo elemento di una terna contenuta nell'evento A_1 può essere scelto in 2 modi diversi; il secondo elemento della terna può invece essere scelto solo in un unico modo in quanto deve coincidere con il primo elemento della terna; il terzo e ultimo elemento della terna può di nuovo essere scelto in due modi diversi), secondo il metodo di assegnazione classico la probabilità dell'evento A_1 deve essere data da

$$P(A_1) = \frac{\#A_1}{\#\Omega} = \frac{4}{8} = \frac{1}{2}.$$

- Ragionando come al punto precedente vediamo che

$$\#A_2 = 2 \times 2 \times 1 = 4 \quad \text{e che} \quad \#A_3 = 2 \times 2 \times 1 = 4.$$

Secondo il metodo di assegnazione classico i valori di $P(A_2) = \#A_2/\#\Omega$ e $P(A_3) = \#A_3/\#\Omega$ sono dunque identici a quello di $P(A_1) = \#A_1/\#\Omega = 1/2$:

$$P(A_1) = P(A_2) = P(A_3) = \frac{4}{8} = \frac{1}{2}. \quad (20)$$

- A questo punto consideriamo l'intersezione $A_1 \cap A_2$. Siccome l'evento $A_1 \cap A_2$ si verifica se e solo se gli esiti di tutti e tre i lanci sono uguali, il numero di casi favorevoli all'evento $A_1 \cap A_2$ deve essere dato da

$$\#(A_1 \cap A_2) = 2 \times 1 \times 1 = 2.$$

Ne consegue che, secondo il metodo di assegnazione classico,

$$P(A_1 \cap A_2) = \frac{\#(A_1 \cap A_2)}{\#\Omega} = \frac{2}{8} = \frac{1}{4},$$

e siccome questo risultato coincide con il valore di

$$P(A_1) \times P(A_2) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4},$$

possiamo concludere che A_1 e A_2 sono una coppia di eventi indipendenti.

- Non è difficile rendersi conto che anche gli eventi $A_1 \cap A_3$ e $A_2 \cap A_3$ si verificano se e solo se gli esiti di tutti e tre i lanci sono uguali e che quindi

$$A_1 \cap A_2 = A_1 \cap A_3 = A_2 \cap A_3 \quad \Rightarrow \quad [[\text{metodo classico}]]$$

$$\Rightarrow \quad P(A_1 \cap A_2) = P(A_1 \cap A_3) = P(A_2 \cap A_3) = \frac{1}{4}.$$

Ricordando la (20) vediamo dunque che anche la coppia di eventi A_1 e A_3 e la coppia di eventi A_2 e A_3 sono entrambe coppie di eventi indipendenti.

Finora abbiamo dunque dimostrato che gli eventi A_1 , A_2 e A_3 sono indipendenti a coppie (ovvero che tutte le coppie di eventi che si possono formare con A_1 , A_2 e A_3 sono coppie di eventi indipendenti). Per verificare che gli eventi A_1 , A_2 e A_3 non sono globalmente indipendenti dobbiamo dunque verificare che

$$P(A_1 \cap A_2 \cap A_3) \neq P(A_1) \times P(A_2) \times P(A_3).$$

Ricordando che $P(A_1) = P(A_2) = P(A_3) = \frac{1}{2}$, vediamo che il prodotto delle probabilità marginali al secondo membro è dato da

$$P(A_1) \times P(A_2) \times P(A_3) = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8}.$$

Rimane da determinare il valore di $P(A_1 \cap A_2 \cap A_3)$. A tal fine osserviamo che anche l'evento $A_1 \cap A_2 \cap A_3$ si verifica se e solo se gli esiti di tutti e tre i lanci sono uguali. Quindi possiamo concludere che (sempre secondo il metodo di assegnazione classico)

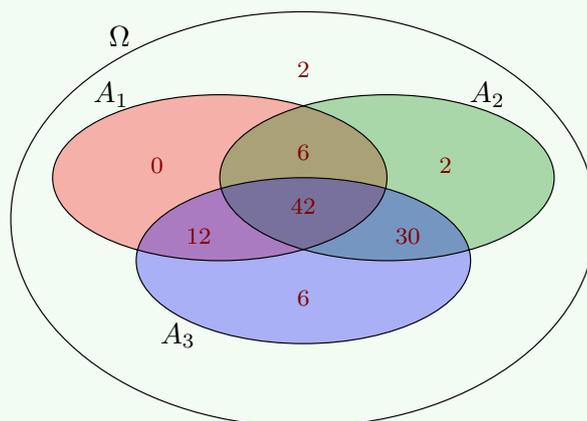
$$P(A_1 \cap A_2 \cap A_3) = P(A_1 \cap A_2) = P(A_1 \cap A_3) = P(A_2 \cap A_3) = \frac{1}{4}.$$

Siccome questo risultato è diverso dal prodotto delle probabilità marginali dei tre eventi, questo risultato ci dice che gli eventi A_1 , A_2 e A_3 non sono globalmente indipendenti.

Il prossimo esempio chiarisce che il caso di tre eventi che sono indipendenti a coppie ma che non sono globalmente indipendenti non è un caso così eccezionale come si potrebbe pensare.

Esempio 3.2 (Indipendenza globale - secondo esempio). Consideriamo un'urna che contiene 100 palline delle quali 60 sono contrassegnate con un punto nero, 80 hanno la superficie ruvida e 90 contengono una sorpresa. Le palline contrassegnate con un punto nero e che hanno anche la superficie ruvida sono 48, le palline contrassegnate con un punto nero e che contengono anche una sorpresa sono 54 e le palline con superficie ruvida che contengono anche una sorpresa sono 72. Le palline che presentano tutte e tre le suddette caratteristiche sono invece "soltanto" 42.

Usando le informazioni sulla composizione dell'urna otteniamo il grafico sottostante:



Ω = urna; A_1 = palline con punto nero;
 A_2 = palline ruvide; A_3 = palline con sorpresa

Si noti che

$$\begin{aligned} \#(A_1 \cap A_2 \cap A_3) &= 42, & \#(A_1 \cap A_2) &= 42 + 6 = 48, \\ \#(A_1 \cap A_3) &= 42 + 12 = 54, & \#(A_2 \cap A_3) &= 42 + 30 = 72 \\ \#A_1 &= 0 + 6 + 42 + 12 = 60, & \#A_2 &= 2 + 6 + 42 + 30 = 80, \\ \#A_3 &= 6 + 12 + 42 + 30 = 90 \end{aligned}$$

come nella precedente descrizione.

Supponiamo ora che dall'urna venga estratta una pallina in modo casuale e consideriamo i seguenti eventi:

- l'evento A_1 secondo il quale la pallina estratta è contrassegnata con un punto nero;
- l'evento A_2 secondo il quale la pallina estratta ha la superficie ruvida;
- l'evento A_3 secondo il quale la pallina estratta contiene una sorpresa.

Come vedremo tra breve, se tutte le palline presenti nell'urna hanno la medesima probabilità di essere estratte, gli eventi A_1 , A_2 e A_3 sono indipendenti a coppie ma non globalmente indipendenti. Infatti, se tutte le palline presenti nell'urna hanno la medesima probabilità di essere estratte, otteniamo

$$\begin{aligned} P(A_1) &= 0,60, & P(A_2) &= 0,80, & P(A_3) &= 0,90, \\ P(A_1 \cap A_2) &= 0,48, & P(A_1 \cap A_3) &= 0,54, \\ P(A_2 \cap A_3) &= 0,72 & \text{e} & P(A_1 \cap A_2 \cap A_3) &= 0,42. \end{aligned}$$

Siccome

$$P(A_1) \times P(A_2) = 0,60 \times 0,80 = 0,48 = P(A_1 \cap A_2),$$

$$P(A_1) \times P(A_3) = 0,60 \times 0,90 = 0,54 = P(A_1 \cap A_3)$$

$$P(A_2) \times P(A_3) = 0,80 \times 0,90 = 0,72 = P(A_2 \cap A_3),$$

possiamo dunque concludere che **gli eventi A_1 , A_2 e A_3 sono indipendenti a coppie**. Tuttavia,

$$\begin{aligned} P(A_1) \times P(A_2) \times P(A_3) &= 0,60 \times 0,80 \times 0,90 \\ &= \mathbf{0,432} \\ &\neq 0,42 = P(A_1 \cap A_2 \cap A_3) \end{aligned}$$

e per questo motivo **gli eventi A_1 , A_2 e A_3 non sono globalmente indipendenti**. Si noti che affinché gli eventi A_1 , A_2 e A_3 risultino globalmente indipendenti, l'urna dovrebbe contenere **43,2** palline che presentano tutte e tre le caratteristiche (con punto nero, ruvide e con sorpresa)!!!! Chiaramente, una siffatta urna non esiste.

Nel prossimo esercizio vedremo invece un esempio del contesto nell'ambito del quale sono tipicamente definiti tre eventi globalmente indipendenti. Il contesto in questione è quello di un esperimento casuale che si svolge in tre passi dove gli esiti dei singoli passi non si influenzano a vicenda.

Esercizio 3.3 (Indipendenza globale - terzo esempio). Un esperimento casuale che si svolge in tre passi consiste nel

- i) nel lancio di una moneta al primo passo,
- ii) lancio di un tetraedro con 4 facce numerate al secondo passo
- iii) e nel lancio di un dado con 6 facce numerate al terzo e ultimo passo.

Si considerino

- l'evento A_1 secondo il quale l'esito del lancio della moneta è testa;
- l'evento A_2 secondo il quale l'esito del lancio del tetraedro è la faccia che riporta il numero 4;
- e l'evento A_3 secondo il quale l'esito del lancio del dado è la faccia che riporta il numero 6.

Si verifichi che gli eventi A_1 , A_2 e A_3 sono globalmente indipendenti.

Risposta:

Gli eventi elementari che possono verificarsi al termine dell'esperimento casuale in questione possono essere rappresentati attraverso delle terne $\omega = (\omega_1, \omega_2, \omega_3)$ dove

- ω_1 descrive l'esito del lancio della moneta ("T" per testa oppure "C" per croce),
- ω_2 descrive l'esito del lancio del tetraedro (gli esiti del lancio del tetraedro possono essere identificati con i quattro numeri naturali 1, 2, 3, 4),
- e dove ω_3 descrive l'esito del lancio del dado (gli esiti del lancio del dado possono essere identificati con i sei numeri naturali 1, 2, 3, 4, 5, 6).

Lo spazio campionario associato all'esperimento casuale in questione contiene dunque

$$\#\Omega = 2 \times 4 \times 6 = 48$$

eventi elementari. Siccome non c'è motivo di ritenere che alcuni di questi eventi elementari possano verificarsi "più facilmente" di altri, assumeremo che per qualsiasi $A \subseteq \Omega$ la corrispondente probabilità $P(A)$ sia definita secondo il metodo di assegnazione classico. Le probabilità degli eventi A_1 , A_2 e A_3 sono dunque date da

$$P(A_1) = \frac{1 \times 4 \times 6}{48} = \frac{1}{2}, \quad P(A_2) = \frac{2 \times 1 \times 6}{48} = \frac{1}{4},$$

$$P(A_3) = \frac{2 \times 4 \times 1}{48} = \frac{1}{6}$$

e le probabilità congiunte delle $\binom{3}{2} = \frac{3!}{2!(3-1)!} = 3$ coppie di eventi che si possono formare con gli eventi A_1 , A_2 e A_3 sono date da

$$P(A_1 \cap A_2) = \frac{1 \times 1 \times 6}{48} = \frac{1}{8}, \quad P(A_1 \cap A_3) = \frac{1 \times 4 \times 1}{48} = \frac{1}{12}$$

$$P(A_2 \cap A_3) = \frac{2 \times 1 \times 1}{48} = \frac{1}{24}.$$

Siccome ciascuna di queste tre probabilità congiunte coincide con il prodotto delle corrispondenti probabilità marginali (la verifica è immediata), possiamo concludere che tutte le coppie che si possono formare con gli eventi A_1 , A_2 e A_3 sono coppie di eventi indipendenti.

Per verificare se gli eventi A_1 , A_2 e A_3 sono anche globalmente indipendenti rimane quindi soltanto da verificare se

$$P(A_1 \cap A_2 \cap A_3) = P(A_1) \times P(A_2) \times P(A_3).$$

Questa verifica è immediata: infatti, non è difficile rendersi conto che con il metodo classico si ottiene

$$P(A_1 \cap A_2 \cap A_3) = \frac{1 \times 1 \times 1}{48} = \frac{1}{48}$$

e siccome

$$P(A_1) \times P(A_2) \times P(A_3) = \frac{1}{2} \times \frac{1}{4} \times \frac{1}{6} = \frac{1}{48}.$$

possiamo concludere che gli eventi A_1 , A_2 e A_3 sono globalmente indipendenti.

Prima di concludere questo esercizio osserviamo che in modo del tutto analogo a quello appena visto si può verificare che sono anche globalmente indipendenti tutte le terne di eventi B_1 , B_2 e B_3 dove

- B_1 è un qualunque evento che riguarda solo l'esito del lancio della moneta (ovvero soltanto il primo passo dell'esperimento),
- B_2 è un qualunque evento che riguarda solo l'esito del lancio del tetraedro (ovvero il secondo passo dell'esperimento)
- e B_3 è un qualunque evento che riguarda solo l'esito del lancio del dado (ovvero il terzo e ultimo passo dell'esperimento).

Ovviamente, la definizione di "indipendenza globale" può essere estesa anche al caso di più di tre eventi:

Definizione 3.4 (Indipendenza globale - definizione generale). Se $\mathcal{C} = \{A_1, A_2, \dots, A_k\}$ è una collezione finita di $k \geq 2$ eventi, allora si dice che \mathcal{C} è **una collezione di eventi globalmente indipendenti** se per ogni sottocollezione $\mathcal{D} \subseteq \mathcal{C}$, la probabilità congiunta degli eventi appartenenti a \mathcal{D} coincide con il prodotto delle probabilità marginali degli stessi eventi, ovvero se sono soddisfatte tutte le seguenti condizioni:

- 1) Per ogni coppia di eventi si ha (con k eventi si possono formare $\binom{k}{2}$ coppie di eventi)

$$P(A_{i_1} \cap A_{i_2}) = P(A_{i_1}) \times P(A_{i_2});$$

- 2) per ogni combinazione di tre eventi si ha (con k eventi si possono formare $\binom{k}{3}$ combinazioni di eventi presi tre alla volta)

$$P(A_{i_1} \cap A_{i_2} \cap A_{i_3}) = P(A_{i_1}) \times P(A_{i_2}) \times P(A_{i_3})$$

- 3) ...

- k-1) per l'unica combinazione che contiene tutti i k eventi si ha (con k eventi si può formare un'unica combinazione di eventi presi k alla volta: infatti, $\binom{k}{k} = 1$ per qualsiasi $k = 1, 2, \dots$)

$$P(A_1 \cap A_2 \cap \dots \cap A_k) = P(A_1) \times P(A_2) \times \dots \times P(A_k).$$

Se \mathcal{C} è una collezione infinita di eventi, allora si dice che \mathcal{C} è una collezione di eventi globalmente indipendenti se ogni sottocollezione finita $\mathcal{D} \subseteq \mathcal{C}$ è una collezione globalmente indipendente.

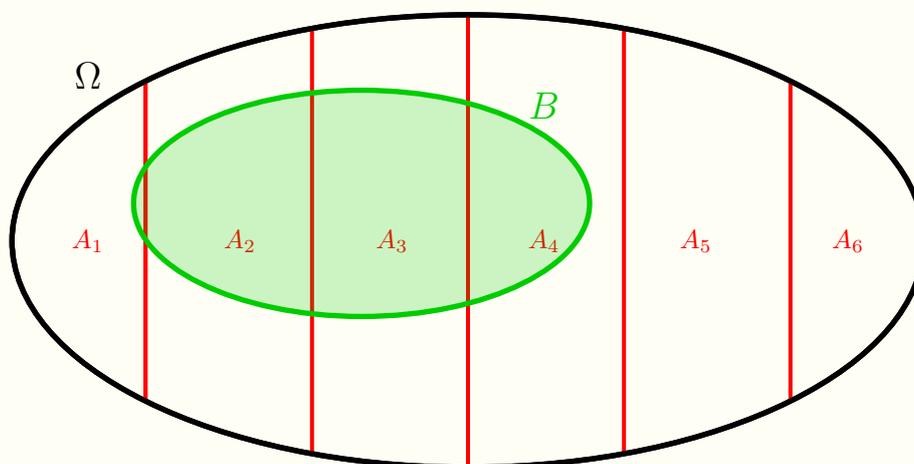
3.5 La formula della probabilità totale

Consideriamo una **partizione** finita $\mathcal{C} = \{A_1, A_2, \dots, A_k\}$ di uno spazio campionario Ω e un ulteriore evento B . Siccome gli eventi di una partizione sono mutuamente incompatibili e la loro unione è uguale a Ω , possiamo concludere che

$$B = (B \cap A_1) \cup (B \cap A_2) \cup \dots \cup (B \cap A_k)$$

e che gli eventi $B \cap A_i$ al secondo membro della precedente uguaglianza sono incompatibili (vedi Figura 3.1).

Figura 3.1. Il grafico mostra come attraverso una partizione dello spazio campionario si può ottenere una partizione di un altro evento B .



Consideriamo ora una funzione di probabilità $P(\cdot)$ (che soddisfa gli assiomi di Kolmogorov) e il cui dominio contiene tutti gli eventi A_i della partizione \mathcal{C} e anche l'evento B . Applicando la legge L2 (oppure l'assioma K3) alla scomposizione dell'evento B , otteniamo

$$P(B) = \sum_{i=1}^k P(B \cap A_i),$$

e ipotizzando che

$$P(A_i) > 0 \quad \text{per ogni } i = 1, 2, \dots, k,$$

e che tutte le k probabilità condizionate $P(B|A_i)$ siano dunque definite, possiamo esprimere i termini della suddetta sommatoria come

$$P(B \cap A_i) = P(B|A_i) \times P(A_i),$$

e in questo modo otteniamo la cosiddetta **formula della probabilità totale**:

$$P(B) = \sum_{i=1}^k P(B|A_i) \times P(A_i).$$

Teorema 3.3 (Formula della probabilità totale). Sia $\mathcal{C} = \{A_1, A_2, \dots, A_k\}$ una **partizione** finita di uno spazio campionario Ω e si assuma che tutti gli eventi della partizione \mathcal{C} appartengano al dominio di una data funzione di probabilità $P(\cdot)$ (che soddisfa gli assiomi di Kolmogorov). Se

$$P(A_i) > 0 \quad \text{per ogni } i = 1, 2, \dots, k,$$

allora

$$P(B) = \sum_{i=1}^k P(B|A_i) \times P(A_i)$$

per ogni evento B che appartiene al dominio di $P(\cdot)$.

Questa formula è nota come **formula della probabilità totale**.

Si noti che il ragionamento che ci ha condotti alla formula della probabilità totale può essere applicato anche ad una partizione $\mathcal{C} = \{A_1, A_2, \dots\}$ che è costituita da un'**infinità numerabile** di eventi A_i (basta invocare l'assioma **K3*** al posto dell'assioma **K3**). La formula della probabilità totale vale dunque anche nel caso di una partizione infinita numerabile.

Esercizio 3.4 (Formula della probabilità totale). Secondo i dati forniti da un'azienda farmaceutica, un test per individuare un'infezione virale ...

- ...rileva correttamente la presenza del virus nel 95% dei casi,
- ...rileva erroneamente la presenza del virus (falsi positivi) nell'1% dei casi.

[[In ambito sanitario la percentuale di diagnosi corrette su individui infetti viene chiamata "**sensibilità**" del test, mentre la percentuale di diagnosi corrette su individui sani viene chiamata "**specificità**" del test]]

Si supponga che il 20% della popolazione sia infetta dal virus. Qual è la probabilità che un individuo scelto casualmente risulti positivo al test?

Soluzione:

Indichiamo con A l'evento secondo il quale l'individuo scelto è infetto dal virus e con B l'evento secondo il quale il test rileva la presenza del virus. Dalle informazioni nella consegna si desume che

$$P(B|A) = 0,95, \quad P(B|\bar{A}) = 0,01, \quad P(A) = 0,20.$$

Siccome A e \bar{A} formano una partizione dello spazio campionario Ω (Ω può essere interpretato come l'insieme di tutti gli individui che appartengono alla popolazione e A può essere interpretato come il sottoinsieme degli individui che sono infetti),

possiamo applicare la formula della probabilità totale onde ottenere

$$\begin{aligned} P(B) &= P(B|A) \times P(A) + P(B|\bar{A}) \times P(\bar{A}) \\ &= P(B|A) \times P(A) + P(B|\bar{A}) \times [1 - P(A)] \\ &= 0,95 \times 0,20 + 0,01 \times (1 - 0,20) = 0,198. \end{aligned}$$

Esercizio 3.5 (Formula della probabilità totale). Un gioco consiste nel lancio di un dado e a seconda del punteggio ottenuto viene poi estratta una pallina da una delle tre urne qui di seguito descritte:

- 1) se nel lancio del dado si ottiene il punteggio 1 (evento A_1), la pallina viene estratta da un'urna che contiene 2 palline bianche e 3 palline nere;
- 2) se nel lancio del dado si ottiene il punteggio 2 oppure il punteggio 3 (evento A_2), la pallina viene estratta da un'urna che contiene 2 palline bianche e 4 palline nere;
- 3) in tutti gli altri casi (evento A_3), la pallina viene estratta da un'urna che contiene 2 palline bianche e 6 palline nere.

Qual è la probabilità di ottenere una pallina bianca (evento B)?

Soluzione:

Dalla descrizione del gioco desumiamo che

$$P(A_1) = \frac{1}{6}, \quad P(A_2) = \frac{2}{6} = \frac{1}{3}, \quad P(A_3) = \frac{3}{6} = \frac{1}{2}$$

e che

$$P(B|A_1) = \frac{2}{2+3} = \frac{2}{5}, \quad P(B|A_2) = \frac{2}{2+4} = \frac{1}{3}, \quad P(B|A_3) = \frac{2}{2+6} = \frac{1}{4}.$$

Sostituendo questi valori nella formula della probabilità totale otteniamo la probabilità richiesta:

$$\begin{aligned} P(B) &= P(B|A_1) \times P(A_1) + P(B|A_2) \times P(A_2) + P(B|A_3) \times P(A_3) \\ &= \left(\frac{2}{5} \times \frac{1}{6}\right) + \left(\frac{1}{3} \times \frac{1}{3}\right) + \left(\frac{1}{4} \times \frac{1}{2}\right) \\ &= \frac{1}{15} + \frac{1}{9} + \frac{1}{8} = \frac{327}{1080} = 0,303. \end{aligned}$$

3.6 La formula di Bayes

Come abbiamo appena visto, se $\mathcal{C} = \{A_1, A_2, \dots, A_k\}$ è una partizione di uno spazio campionario $\Omega = A_1 \cup A_2 \cup \dots \cup A_k$, e se le probabilità degli eventi A_i sono tutte positive, allora la probabilità di un qualunque evento B può essere calcolata attraverso la [formula della probabilità totale](#)

$$P(B) = \sum_{i=1}^k P(B|A_i) \times P(A_i).$$

In questo paragrafo ricaveremo la cosiddetta **formula di Bayes** che ci dice come a partire dalle quantità coinvolte nella [formula della probabilità totale](#) si possono calcolare le cosiddette **probabilità a posteriori**

$$P(A_1|B), \quad P(A_2|B), \quad \dots, \quad P(A_k|B).$$

Il motivo per cui queste probabilità vengono chiamate "probabilità a posteriori" diverrà chiaro tra breve quando vedremo alcuni esempi.

Ora, prima di ricavare la formula di Bayes conviene osservare che le probabilità a posteriori

$$P(A_i|B) = \frac{P(A_i \cap B)}{P(B)}, \quad i = 1, 2, \dots, k,$$

sono definite solo se $P(B) > 0$, e che le probabilità condizionate

$$P(B|A_i) = \frac{P(A_i \cap B)}{P(A_i)}, \quad i = 1, 2, \dots, k,$$

sono definite solo se le probabilità marginali $P(A_i)$ sono tutte positive. Siccome la formula di Bayes coinvolge entrambi questi tipi di probabilità condizionate, possiamo concludere che la formula di Bayes riguarda solo situazioni dove le probabilità

$$P(B), \quad P(A_1), \quad P(A_2), \quad \dots, \quad P(A_k)$$

sono tutte strettamente positive.

A questo punto siamo pronti per ricavare la formula di Bayes che, come abbiamo già accennato, è una formula per calcolare le probabilità a posteriori $P(A_i|B)$ a partire dalle probabilità condizionate $P(B|A_i)$ e dalle probabilità marginali $P(A_i)$. Per ottenere la formula di Bayes basta osservare che il **numeratore** della probabilità a posteriori

$$P(A_i|B) = \frac{P(A_i \cap B)}{P(B)}$$

può essere espresso come

$$P(A_i \cap B) = P(B|A_i) \times P(A_i),$$

e che il **denominatore** $P(B)$ può essere espresso attraverso la **formula della probabilità totale**:

$$P(B) = \sum_{j=1}^k P(B|A_j) \times P(A_j).$$

La **formula di Bayes** è quindi data da

$$P(A_i|B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(B|A_i) \times P(A_i)}{\sum_{j=1}^k P(B|A_j) \times P(A_j)}.$$

Teorema 3.4 (Formula di Bayes). Sia $\mathcal{C} = \{A_1, A_2, \dots, A_k\}$ una **partizione** (finita o infinita numerabile) di uno spazio campionario Ω e si assuma che tutti gli eventi della partizione \mathcal{C} appartengano al dominio di una data funzione di probabilità $P(\cdot)$ (che soddisfa gli assiomi di Kolmogorov).

Se B è un evento con $P(B) > 0$ e se $P(A_i) > 0$ per ogni $A_i \in \mathcal{C}$, allora

$$P(A_i|B) = \frac{P(B|A_i) \times P(A_i)}{\sum_{j=1}^k P(B|A_j) \times P(A_j)} \quad \text{per ogni } i = 1, 2, \dots, k.$$

Questa formula è nota come **formula di Bayes**.

Esercizio 3.6 (Formula di Bayes - continuazione dell'Esercizio 3.4). Consideriamo ancora la consegna dell'Esercizio 3.4 che, per comodità, ripetiamo:

Secondo i dati forniti da un'azienda farmaceutica, un test per individuare un'infezione virale ...

- ... rileva correttamente la presenza del virus nel 95% dei casi,
- ... rileva erroneamente la presenza del virus (falsi positivi) nell'1% dei casi.

Si supponga che il 20% della popolazione sia infetta dal virus.

In questo esercizio risponderemo alle seguenti domande:

- a) Ipotizzando che un individuo scelto casualmente risulti positivo al test, qual è la probabilità condizionata che l'individuo in questione sia effettivamente infetto dal virus?
- b) Come cambierebbe la risposta al quesito a) se soltanto l'1% della popolazione fosse infetto?

Risposte:

Per rispondere ai due quesiti conviene ricordare che nell'Esercizio 3.4 abbiamo utilizzato la formula della probabilità totale per calcolare la probabilità che un

individuo scelto casualmente risulti positivo al test (si ricordi che abbiamo definito A come l'evento secondo il quale l'individuo scelto è infetto dal virus e che abbiamo definito B come l'evento secondo il quale l'individuo scelto è positivo al test):

$$\begin{aligned} P(B) &= P(B|A) \times P(A) + P(B|\bar{A}) \times P(\bar{A}) \\ &= P(B|A) \times P(A) + P(B|\bar{A}) \times [1 - P(A)] \\ &= 0,95 \times 0,20 + 0,01 \times (1 - 0,20) = 0,198. \end{aligned}$$

- a) **Ipotizzando che un individuo scelto casualmente risulti positivo al test, qual è la probabilità condizionata che l'individuo in questione sia effettivamente infetto dal virus?**

Usando la formula di Bayes vediamo che la probabilità a posteriori dell'evento secondo il quale l'individuo scelto è infetto dal virus è data da

$$\begin{aligned} P(A|B) &= \frac{P(B|A) \times P(A)}{P(B|A) \times P(A) + P(B|\bar{A}) \times P(\bar{A})} \\ &= \frac{0,95 \times 0,20}{0,95 \times 0,20 + 0,01 \times (1 - 0,20)} = 0,960. \end{aligned}$$

Il valore di questa probabilità a posteriori ci dice che **una volta constatata la positività al test (evento B), la probabilità condizionata di avere a che fare con un individuo infetto (evento A) è pari al 96%**.

- b) **Come cambierebbe la risposta al quesito a) se soltanto l'1% della popolazione fosse infetto?**

Se soltanto l'1% della popolazione fosse infetto, allora la probabilità totale dell'evento B (un individuo scelto casualmente risulta positivo al test) sarebbe data da ([formula della probabilità totale](#))

$$\begin{aligned} P(B) &= P(B|A) \times P(A) + P(B|\bar{A}) \times P(\bar{A}) \\ &= P(B|A) \times P(A) + P(B|\bar{A}) \times [1 - P(A)] \\ &= 0,95 \times 0,01 + 0,01 \times (1 - 0,01) = 0,095 + 0,099 = 0,194, \end{aligned}$$

e la probabilità condizionata $P(A|B)$ (ovvero la probabilità a posteriori dell'evento secondo il quale un individuo scelto casualmente è infetto dato che risulta positivo al test) sarebbe invece data da ([formula di Bayes](#))

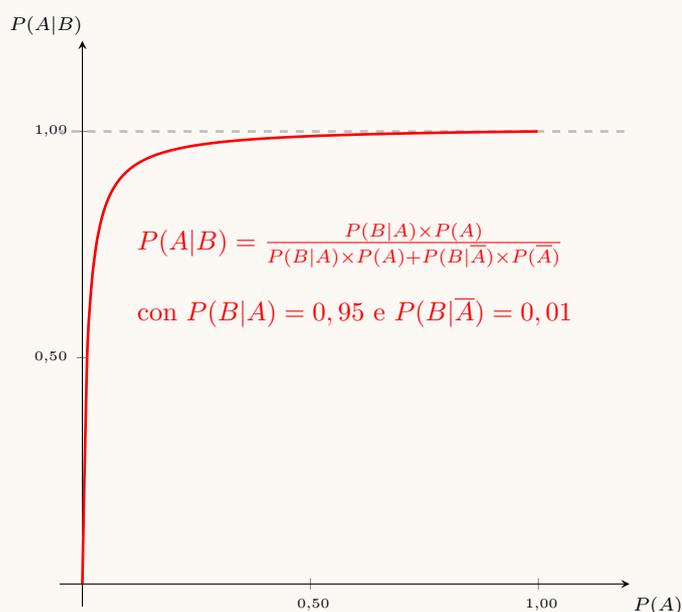
$$\begin{aligned} P(A|B) &= \frac{P(B|A) \times P(A)}{P(B|A) \times P(A) + P(B|\bar{A}) \times P(\bar{A})} \\ &= \frac{0,95 \times 0,01}{0,95 \times 0,01 + 0,01 \times (1 - 0,01)} = \frac{0,095}{0,095 + 0,099} = 0,499. \end{aligned}$$

Si noti che questo valore è molto più basso di quello che abbiamo ottenuto nella risposta al quesito precedente dove abbiamo considerato l'ipotesi che

una fetta molto più ampia della popolazione sia infetta dal virus. Il grafico sottostante aiuta a capire questo effetto. Esso mostra come a parità di **sensibilità** e di **specificità** del test (ovvero a parità di $P(B|A) = 0,95$ e di $P(\bar{B}|\bar{A}) = 1 - P(B|\bar{A}) = 1 - 0,01 = 0,99$) la probabilità a posteriori

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B|A) \times P(A) + P(B|\bar{A}) \times P(\bar{A})}$$

varia al variare della quota di popolazione infetta (ovvero al variare della probabilità $P(A)$):



Si noti che la probabilità a posteriori $P(A|B)$ è una funzione strettamente crescente della quota di popolazione infetta $P(A)$ e che al variare di $P(A)$ nell'intervallo $(0, 1)$ la probabilità a posteriori $P(A|B)$ può assumere qualsiasi valore nell'intervallo $(0, 1)$. Si noti, inoltre, che la probabilità a posteriori $P(A|B)$ è molto più sensibile a variazioni di $P(A)$ quando il valore iniziale di $P(A)$ è prossimo a zero rispetto a quando il valore iniziale di $P(A)$ è prossimo a 1. Quest'ultima osservazione è una conseguenza del fatto che il valore di $P(B|A) = 0,95$ è molto più elevato del valore di $P(B|\bar{A}) = 0,01$.

Esercizio 3.7 (Formula di Bayes - continuazione dell'Esercizio 3.5). Consideriamo ancora la consegna dell'Esercizio 3.5 che, per comodità, ripetiamo:

Un gioco consiste nel lancio di un dado e a seconda del punteggio ottenuto viene poi estratta una pallina da una delle tre urne qui di seguito descritte:

- 1) se nel lancio del dado si ottiene il punteggio 1 (evento A_1), la pallina viene estratta da un'urna che contiene 2 palline bianche e 3 palline nere;
- 2) se nel lancio del dado si ottiene il punteggio 2 oppure il punteggio 3 (evento A_2), la pallina viene estratta da un'urna che contiene 2 palline bianche e 4 palline nere;
- 3) in tutti gli altri casi (evento A_3), la pallina viene estratta da un'urna che contiene 2 palline bianche e 6 palline nere.

In questo esercizio risponderemo alla seguente domanda:

Si supponga che al termine dell'esperimento casuale il punteggio del dado non sia noto e che il colore della pallina estratta dall'urna sia bianco (evento B). Come si può valutare se è più verosimile che la pallina provenga dalla prima urna (evento A_1), dalla seconda urna (evento A_2) oppure dalla terza urna (evento A_3)?

Risposta:

Per rispondere al quesito dobbiamo confrontare i valori delle probabilità a posteriori

$$P(B|A_1), \quad P(B|A_2) \quad \text{e} \quad P(B|A_3).$$

Per calcolare queste probabilità condizionate conviene ricordare che nell'Esercizio 3.5 abbiamo applicato la [formula della probabilità](#) totale onde calcolare la probabilità dell'evento B :

$$\begin{aligned} P(B) &= P(B|A_1) \times P(A_1) + P(B|A_2) \times P(A_2) + P(B|A_3) \times P(A_3) \\ &= \left(\frac{2}{5} \times \frac{1}{6}\right) + \left(\frac{1}{3} \times \frac{1}{3}\right) + \left(\frac{1}{4} \times \frac{1}{2}\right) \\ &= \frac{1}{15} + \frac{1}{9} + \frac{1}{8} = \frac{327}{1080} = 0,303. \end{aligned}$$

Questa formula contiene tutte le quantità che ci servono per calcolare le probabilità a posteriori attraverso la [formula di Bayes](#):

$$\begin{aligned} P(A_1|B) &= \frac{P(B|A_1) \times P(A_1)}{P(B|A_1) \times P(A_1) + P(B|A_2) \times P(A_2) + P(B|A_3) \times P(A_3)} \\ &= \frac{\frac{1}{15}}{\frac{1}{15} + \frac{1}{9} + \frac{1}{8}} = 0,220; \end{aligned}$$

$$P(A_2|B) = \frac{\frac{1}{9}}{\frac{1}{15} + \frac{1}{9} + \frac{1}{8}} = 0,367;$$

$$P(A_3|B) = \frac{\frac{1}{8}}{\frac{1}{15} + \frac{1}{9} + \frac{1}{8}} = 0,413.$$

Confrontando i valori delle tre probabilità a posteriori

$$P(A_1|B) = 0,220, \quad P(A_2|B) = 0,367, \quad P(A_3|B) = 0,413,$$

vediamo che l'ipotesi più verosimile è quella secondo la quale la pallina estratta proviene dalla terza urna (evento A_3).

I due esercizi chiariscono perché nelle applicazioni che coinvolgono la [formula della probabilità totale](#) e la [formula di Bayes](#) le probabilità marginali

$$P(A_1), \quad P(A_2), \quad \dots, \quad P(A_k)$$

vengono chiamate **probabilità a priori**, mentre le probabilità condizionate

$$P(A_1|B), \quad P(A_2|B), \quad \dots, \quad P(A_k|B)$$

vengono chiamate **probabilità a posteriori**:

- le **probabilità a priori**

$$P(A_1), \quad P(A_2), \quad \dots, \quad P(A_k)$$

si riferiscono infatti ad eventi che non possiamo osservare e vengono assegnate "a priori" sulla base di ciò che sappiamo o crediamo di sapere sullo svolgimento di un esperimento casuale (la percentuale di individui infetti nella popolazione, la probabilità che l'estrazione verrà eseguita da un'urna piuttosto che da un'altra, ...),

- ... mentre le **probabilità a posteriori**

$$P(A_1|B), \quad P(A_2|B), \quad \dots, \quad P(A_k|B)$$

possono essere interpretate come valori rivisti delle probabilità a priori $P(A_1)$, $P(A_2)$, ..., $P(A_k)$, valori rivisti alla luce del fatto che alla fine dell'esperimento abbiamo osservato l'evento B (un individuo risulta positivo ad un test virale, la pallina estratta dall'urna è di colore bianco, ...).

Nell'ultimo esercizio di questa sezione vedremo come attraverso le probabilità a priori e quelle a posteriori si può valutare l'incidenza di un fattore di rischio.

Esercizio 3.8. Secondo uno studio, il 10% degli abitanti di una regione vive in zone con bassa qualità dell'aria e il 30% degli abitanti che sono affetti da disturbi polmonari vive in una queste zone. Si valuti se per un abitante che vive in una zona con bassa qualità dell'aria il rischio di essere affetto da disturbi polmonari è

maggiore rispetto ad un abitante che vive in altre zone della regione.

Risposta:

Per rispondere consideriamo un esperimento che consiste nella scelta casuale di un abitante della regione e indichiamo

- con A l'evento secondo il quale l'abitante scelto vive in una zona con bassa qualità dell'aria
- e con B l'evento secondo il quale l'abitante scelto è affetto da disturbi polmonari.

In considerazione dei dati nella consegna possiamo porre

$$P(A) = 0,10 \quad \text{e} \quad P(A|B) = 0,30.$$

Per valutare se il rischio di essere affetti da disturbi polmonari è maggiore per gli abitanti che vivono in zone con bassa qualità dell'aria, dobbiamo confrontare i valori di $P(B|A)$ e $P(B|\bar{A})$. Siccome

$$P(B|A) > P(B|\bar{A}) \quad \Leftrightarrow \quad \frac{P(B|A)}{P(B|\bar{A})} > 1,$$

possiamo rispondere al quesito calcolando il valore del rapporto $\frac{P(B|A)}{P(B|\bar{A})}$. Dalla [formula di Bayes](#) deduciamo che

$$\begin{aligned} P(A|B) &= \frac{P(B|A) \times P(A)}{P(B|A) \times P(A) + P(B|\bar{A}) \times P(\bar{A})} \Rightarrow \\ \Rightarrow P(A|B) &= \frac{P(A)}{P(A) + \frac{P(B|\bar{A})}{P(B|A)} \times P(\bar{A})} \Rightarrow \dots \Rightarrow \\ \Rightarrow \frac{P(B|A)}{P(B|\bar{A})} &= \frac{P(A|B)}{1 - P(A|B)} \times \frac{1 - P(A)}{P(A)} = \frac{0,30}{1 - 0,30} \times \frac{1 - 0,10}{0,10} = 3,857. \end{aligned}$$

Questo risultato ci dice che per un abitante di una zona con bassa qualità dell'aria il "rischio" di essere affetto da disturbi polmonari è di **3,857 volte più elevato** che per un abitante di un'altra zona della regione.

4 Variabili casuali

4.1 Definizione di variabile casuale

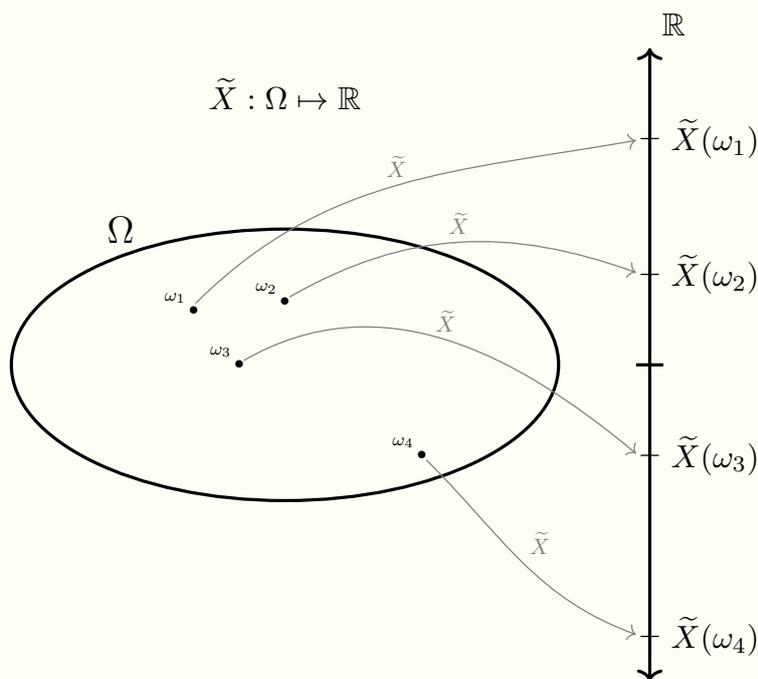
Consideriamo una funzione

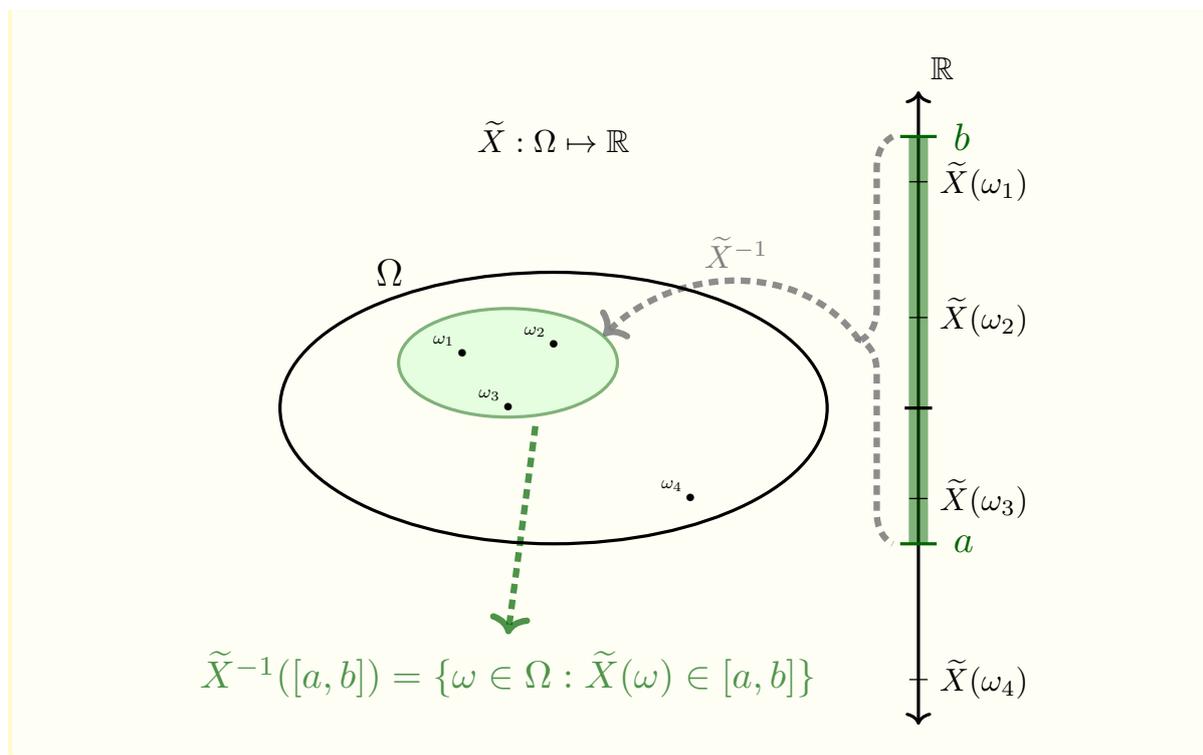
$$\tilde{X} : \Omega \mapsto \mathbb{R}$$

che ad ogni evento elementare ω di uno spazio campionario Ω associa un numero reale $x \in \mathbb{R}$. Funzioni di questo tipo forniscono una descrizione numerica dell'esito finale di un esperimento casuale (ovvero dell'evento elementare $\omega \in \Omega$ che si verifica al termine di un esperimento casuale) e vengono chiamate **variabili casuali**. Per distinguere variabili casuali da altri tipi di funzione, indicheremo le prime con delle lettere maiuscole soprassegnate da una "tilde" (come, per esempio \tilde{X} , \tilde{Y} , \tilde{Z} , ecc.).

Definizione 4.1. Una **variabile casuale** è una funzione $\tilde{X} : \Omega \mapsto \mathbb{R}$ che ad ogni evento elementare ω di uno spazio campionario Ω associa un numero reale $x \in \mathbb{R}$.

Figura 4.1. Il primo grafico mostra i valori che una variabile casuale $\tilde{X} : \Omega \mapsto \mathbb{R}$ associa ad alcuni eventi elementari $\omega \in \Omega$. Il secondo grafico mostra un evento che è definito come controimmagine dell'intervallo reale $[a, b]$ tramite la variabile casuale $\tilde{X} : \Omega \mapsto \mathbb{R}$.





Una variabile casuale potrebbe descrivere

- la vincita associata ad una scommessa circa l'esito finale di un esperimento casuale
- il numero di volte che si ottiene 6 in un certo numero di lanci di un dado
- il prezzo di un'azione in un determinato istante temporale
- ...

Variabili casuali vengono spesso utilizzate per definire eventi come per esempio l'evento

$$\tilde{X}^{-1}(B) = \{\omega \in \Omega : \tilde{X}(\omega) \in B\}, \quad \text{con } B \subseteq \mathbb{R},$$

che contiene tutti gli eventi elementari $\omega \in \Omega$ ai quali la variabile casuale $\tilde{X} : \Omega \mapsto \mathbb{R}$ associa un numero reale che appartiene ad un dato sottoinsieme B di \mathbb{R} . Il secondo grafico in Figura 4.1 mostra, per esempio, l'evento

$$\tilde{X}^{-1}([a, b]) = \{\omega \in \Omega : \tilde{X}(\omega) \in [a, b]\}.$$

Per semplificare la notazione, d'ora in poi indicheremo eventi come

$$\tilde{X}^{-1}(B) = \{\omega \in \Omega : \tilde{X}(\omega) \in B\}$$

anche con $\{\tilde{X} \in B\}$, e se B è un intervallo, come per esempio nel caso in cui $B = [a, b]$, scriveremo anche $\{a \leq \tilde{X} \leq b\}$.

Consideriamo ora un modello probabilistico (Ω, \mathcal{A}, P) . Quando attraverso una variabile casuale $\tilde{X} : \Omega \mapsto \mathbb{R}$ definiamo degli eventi del tipo $\{\tilde{X} \in B\}$ potrebbe sorgere un problema: non è sempre detto che l'evento

$$\{\tilde{X} \in B\} = \{\omega \in \Omega : \tilde{X}(\omega) \in B\}$$

sia contenuto nella σ -algebra di eventi \mathcal{A} che è il dominio della funzione di probabilità $P(\cdot)$. Attraverso variabili casuali si potrebbero quindi definire degli eventi ai quali la funzione di probabilità $P(\cdot)$ del modello probabilistico (Ω, \mathcal{A}, P) di riferimento non associa nessuna probabilità. Il prossimo esempio illustra questo problema.

Esempio 4.1 (Problema con eventi legati a variabili casuali). Consideriamo un esperimento casuale che consiste nel lancio di un dado truccato con 6 facce numerate da 1 a 6 e supponiamo di conoscere soltanto la probabilità di ottenere un punteggio pari e che la probabilità di questo evento sia uguale a $0,6$.

Per descrivere questo esperimento casuale possiamo dunque considerare un modello probabilistico (Ω, \mathcal{A}, P) dove

- lo spazio campionario Ω è definito come

$$\Omega = \{1, 2, 3, 4, 5, 6\},$$

- lo spazio degli eventi \mathcal{A} è definito come

$$\mathcal{A} = \{\Omega, \emptyset, \{1, 3, 5\}, \{2, 4, 6\}\},$$

- e dove la funzione di probabilità $P(\cdot)$ è definita come

$$P(\Omega) = 1, \quad P(\emptyset) = 0, \quad P(\{1, 3, 5\}) = 0,4, \quad P(\{2, 4, 6\}) = 0,6.$$

Per verificare che (Ω, \mathcal{A}, P) è effettivamente un modello probabilistico invitiamo il lettore a verificare che ...

- ... lo spazio di eventi \mathcal{A} soddisfa le condizioni [a\)](#), [b\)](#) e [c\)](#) che lo qualificano come un'algebra (e siccome lo spazio di eventi \mathcal{A} è finito possiamo dunque concludere che esso sia anche una σ -algebra), ...
- ... e che la funzione di probabilità $P(\cdot)$ soddisfa gli assiomi di Kolmogorov (si tenga presente che lo spazio di eventi \mathcal{A} è finito e che quindi l'assioma [K3*](#) è implicito negli assiomi [K0](#) - [K3](#) - vedi i commenti sull'assioma [K3*](#) di Kolmogorov nella Sezione [1.3](#)).

Per fare un'esempio di un evento che è definito attraverso una variabile casuale e al quale non risulta assegnata nessuna probabilità, basta considerare la variabile casuale $\tilde{X} : \Omega \mapsto \mathbb{R}$ che è definita come

$$\tilde{X}(\omega) = \omega \quad \text{per} \quad \omega \in \Omega,$$

ovvero la variabile casuale che restituisce il punteggio che si ottiene lanciando il dado. Attraverso questa variabile casuale possiamo infatti definire l'evento

$$\{\tilde{X} = 6\} = \{\omega \in \Omega : \tilde{X}(\omega) = 6\} = \{6\} \subseteq \Omega$$

che non appartiene allo spazio di eventi \mathcal{A} e al quale non risulta quindi assegnata nessuna probabilità.

Consideriamo ora invece la variabile casuale $\tilde{Y} : \Omega \mapsto \mathbb{R}$ che è definita come

$$\tilde{Y}(\omega) = \begin{cases} 1 & \text{se } \omega = 1, 3, 5; \\ 0 & \text{se } \omega = 2, 4, 6. \end{cases}$$

Per questa variabile casuale possiamo facilmente verificare che tutti gli eventi del tipo

$$\{\tilde{Y} \in B\} \quad \text{con } B \subseteq \mathbb{R}$$

sono contenuti nello spazio di eventi \mathcal{A} e che a tutti questi eventi risulta pertanto assegnata una probabilità. Infatti,

- se $1 \in B$ e $0 \notin B$, allora l'evento $\{\tilde{Y} \in B\}$ è l'evento $\{1, 3, 5\}$ e questo evento appartiene allo spazio di eventi \mathcal{A} ;
- allo stesso modo: $1 \notin B$ e $0 \in B \Rightarrow \{\tilde{Y} \in B\} = \{2, 4, 6\} \in \mathcal{A}$;
- $0 \in B$ e $1 \in B \Rightarrow \{\tilde{Y} \in B\} = \Omega \in \mathcal{A}$;
- $0 \notin B$ e $1 \notin B \Rightarrow \{\tilde{Y} \in B\} = \emptyset \in \mathcal{A}$.

Consideriamo ora la collezione di tutti gli eventi $\{\tilde{X} \in B\}$ che possono essere definiti attraverso una data variabile casuale \tilde{X} . Siccome l'insieme dei numeri reali \mathbb{R} ha tantissimi sottoinsiemi B e la maggior parte di essi sono talmente complicati che è impossibile visualizzarli sulla retta reale, la maggior parte degli eventi $\{\tilde{X} \in B\}$ non sono d'interesse nelle applicazioni e quindi non ha nessuna rilevanza se ad essi risulta assegnata una probabilità o meno. Di fatto, gli eventi del tipo $\{\tilde{X} \in B\}$ che nelle applicazioni potrebbero avere una qualche rilevanza corrispondono **sempre** a sottoinsiemi B di \mathbb{R} che appartengono alla σ -algebra di Borel. Infatti, nell'Esempio 1.5 abbiamo visto che la σ -algebra di Borel è un'ampia collezione di sottoinsiemi di \mathbb{R} che contiene tutti i sottoinsiemi di \mathbb{R} ai quali siamo abituati a pensare (intervalli di ogni tipo, unioni di intervalli, ecc.). Lo spazio di eventi che contiene tutti gli eventi del tipo $\{\tilde{X} \in B\}$ con $B \in \sigma(\mathcal{B})$ (ricordiamo che $\sigma(\mathcal{B})$ è il simbolo che abbiamo introdotto per indicare la σ -algebra di Borel) è noto come **σ -algebra generata dalla variabile casuale \tilde{X}** e viene indicato con $\sigma(\tilde{X})$. Come suggerisce il suo nome, questo spazio di eventi soddisfa le quattro condizioni a), b), c) e c*) che lo qualificano come una σ -algebra.

Lemma 4.1 (σ -algebra generata da una variabile casuale). Se Ω è uno spazio campionario e $\tilde{X} : \Omega \mapsto \mathbb{R}$ è una variabile casuale definita su Ω , allora l'insieme di tutti gli eventi del tipo

$$\{\tilde{X} \in B\} \quad \text{con } B \in \sigma(\mathcal{B})$$

è una σ -algebra e viene chiamato **σ -algebra generata dalla variabile casuale \tilde{X}** . La σ -algebra generata da una data variabile casuale \tilde{X} viene indicata con il simbolo $\sigma(\tilde{X})$ e, attraverso il linguaggio formale della matematica, può essere definita come

$$\sigma(\tilde{X}) = \left\{ \{\tilde{X} \in B\} : B \in \sigma(\mathcal{B}) \right\}.$$

Se \mathcal{A} è una σ -algebra di eventi e se \mathcal{A} contiene tutti gli eventi che appartengono a $\sigma(\tilde{X})$ (ovvero se $\sigma(\tilde{X}) \subseteq \mathcal{A}$), allora si dice che la variabile casuale \tilde{X} è **\mathcal{A} -misurabile**. Una variabile casuale \tilde{X} è \mathcal{A} -misurabile se e solo se

$$\{\tilde{X} \leq x\} \in \mathcal{A} \quad \text{per ogni } x \in \mathbb{R}.$$

4.2 Funzioni di ripartizione

Definizione 4.2 (Funzione di ripartizione). Se (Ω, \mathcal{A}, P) un modello probabilistico e $\tilde{X} : \Omega \mapsto \mathbb{R}$ è una variabile casuale **\mathcal{A} -misurabile**, allora si definisce come **funzione di ripartizione** di \tilde{X} la funzione

$$F(x) = P(\{\tilde{X} \leq x\}) \quad \text{per } x \in \mathbb{R}.$$

[[Si noti che il dominio di una funzione di ripartizione è sempre l'insieme dei numeri reali \mathbb{R} (infatti, si può dimostrare che la probabilità $P(\{\tilde{X} \leq x\})$ è definita per ogni $x \in \mathbb{R}$ se e solo se la variabile casuale \tilde{X} è **misurabile** rispetto alla σ -algebra del modello probabilistico di riferimento (vedi la parte finale del Lemma 4.1).]]

La funzione di ripartizione di una variabile casuale è di fondamentale importanza perché, come vedremo nel prossimo teorema (di cui omettiamo la dimostrazione), la funzione di ripartizione **determina univocamente** le probabilità di tutti gli eventi che appartengono alla **σ -algebra generata dalla variabile casuale di riferimento**. Per questo motivo la funzione di ripartizione viene spesso anche chiamata **distribuzione di probabilità** o semplicemente **distribuzione** della variabile casuale.

Teorema 4.1. Se le funzioni di ripartizione di due variabili casuali \tilde{X} e \tilde{Y} sono identiche,^a allora, per ogni $B \in \sigma(\mathcal{B})$, la probabilità dell'evento $\{\tilde{X} \in B\}$ deve essere identica alla probabilità dell'evento $\{\tilde{Y} \in B\}$.

“Le variabili casuali \tilde{X} e \tilde{Y} potrebbero anche essere definite nell’ambito di due modelli probabilistici diversi.

Il fatto che la funzione di ripartizione $F(x) = P(\{\tilde{X} \leq x\})$ determini univocamente le probabilità di tutti gli eventi che appartengono a $\sigma(\tilde{X}) = \{\{\tilde{X} \in B\} : B \in \sigma(\mathcal{B})\}$ significa che i valori delle probabilità di tutti gli eventi che appartengono a $\sigma(\tilde{X})$ sono **impliciti** nei valori delle probabilità degli eventi del tipo $\{\tilde{X} \leq x\}$ con $x \in \mathbb{R}$. Questo fatto ha due importanti implicazioni:

- 1) Significa che, per un qualunque $B \in \sigma(\mathcal{B})$, il valore di $P(\{\tilde{X} \in B\})$ può, almeno in teoria, essere “ricavato” dalla funzione di ripartizione di \tilde{X} applicando gli assiomi di Kolmogorov e le conseguenti leggi del calcolo delle probabilità. Come abbiamo già detto, per questo motivo la funzione di ripartizione di una variabile casuale viene spesso anche chiamata “distribuzione di probabilità” o “distribuzione” della variabile casuale.
- 2) Significa che per **assegnare** delle probabilità a tutti gli eventi che appartengono a $\sigma(\tilde{X})$ in modo tale che le condizioni poste dagli assiomi di Kolmogorov siano soddisfatte, è sufficiente assegnare probabilità “coerenti” a tutti gli eventi del tipo $\{\tilde{X} \leq x\}$ con $x \in \mathbb{R}$. Per probabilità “coerenti” intendiamo delle probabilità che possono essere estese ad una funzione di probabilità $P(\cdot)$ che soddisfa gli assiomi di Kolmogorov. Verso la fine di questa sezione vedremo un importante teorema che caratterizza tutti i possibili modi per assegnare delle probabilità “coerenti” a tutti gli eventi del tipo $\{\tilde{X} \leq x\}$ con $x \in \mathbb{R}$ (vedi il Teorema 4.4).

Il prossimo teorema descrive le quattro proprietà fondamentali che tutte le funzioni di ripartizione devono soddisfare.

Teorema 4.2 (Proprietà delle funzioni di ripartizione). La funzione di ripartizione di una variabile casuale è sempre una funzione $F(\cdot)$ che soddisfa tutte e quattro le seguenti proprietà:

F0) Il dominio di $F(\cdot)$ è l’insieme dei numeri reali \mathbb{R} .

F1) $F(\cdot)$ è una funzione monotona non decrescente:

$$a < b \quad \Rightarrow \quad F(a) \leq F(b).$$

F2) In ogni punto $x \in \mathbb{R}$ la funzione $F(\cdot)$ è continua da destra:

$$F(x) = \lim_{h \rightarrow 0^+} F(x + h) \quad \text{per ogni } x \in \mathbb{R}.$$

F3) $F(x)$ tende a zero per x che tende a $-\infty$, e $F(x)$ tende a 1 per x che tende a $+\infty$:

$$\lim_{x \rightarrow -\infty} F(x) = 0, \quad \text{e} \quad \lim_{x \rightarrow +\infty} F(x) = 1.$$

Dimostrazione. La proprietà **F0** segue direttamente dal modo in cui abbiamo definito le funzioni di ripartizione (vedi la Definizione 4.2).

Consideriamo ora invece le proprietà **F1 - F3**. Le dimostrazioni di queste proprietà sono basate sugli assiomi di Kolmogorov e le conseguenti leggi del calcolo delle probabilità. Per non appesantire troppo la trattazione dimostreremo soltanto la proprietà **F1**. A tal fine è sufficiente dimostrare che

$$P(\{a < \tilde{X} \leq b\}) = F(b) - F(a). \quad (21)$$

Infatti, in virtù dell'assioma **K1** il valore della probabilità al primo membro non può essere negativo e pertanto l'uguaglianza (21) implica che ogni funzione di ripartizione deve essere non decrescente così come previsto dalla proprietà **F1**.

Ora, per dimostrare l'uguaglianza (21), osserviamo innanzitutto che

$$\{\tilde{X} \leq b\} = \{a < \tilde{X} \leq b\} \cup \{\tilde{X} \leq a\},$$

e che i due eventi al secondo membro sono incompatibili. In virtù dell'assioma **K3** possiamo dunque scrivere

$$P(\{\tilde{X} \leq b\}) = P(\{a < \tilde{X} \leq b\}) + P(\{\tilde{X} \leq a\}),$$

e isolando $P(\{a < \tilde{X} \leq b\})$ al primo membro di questa equazione otteniamo

$$P(\{a < \tilde{X} \leq b\}) = P(\{\tilde{X} \leq b\}) - P(\{\tilde{X} \leq a\}).$$

Tenendo presente che per definizione $P(\{\tilde{X} \leq b\}) = F(b)$ e $P(\{\tilde{X} \leq a\}) = F(a)$, otteniamo l'uguaglianza (21) che volevamo dimostrare.¹⁸ \square

Esercizio 4.1 (Funzione di ripartizione). Si definisca un modello probabilistico (Ω, \mathcal{A}, P) per un esperimento casuale che consiste nel lancio di un dado regolare e si ricavi la funzione di ripartizione della variabile casuale \tilde{X} che descrive il punteggio del lancio.

Risposta:

Come abbiamo già visto nell'Esempio 1.1, per descrivere un esperimento casuale

¹⁸Chiaramente avremmo potuto dimostrare la proprietà **F1** in modo più diretto facendo riferimento alla legge **L4**, ma in questo modo non avremmo ottenuto l'uguaglianza (21) che tra breve ci servirà.

che consiste nel lancio di un dado regolare possiamo considerare

- come spazio campionario l'insieme

$$\Omega = \{1, 2, 3, 4, 5, 6\},$$

- come σ -algebra di eventi l'insieme delle parti $\mathcal{P}(\Omega)$
- e come funzione di probabilità $P(\cdot)$ quella che corrisponde al metodo di assegnazione classico:

$$P(A) = \frac{\#A}{\#\Omega} = \frac{\#A}{6} \quad \text{per ogni } A \subseteq \Omega.$$

Con riferimento a questo modello probabilistico $(\Omega, \mathcal{A} = \mathcal{P}(\Omega), P)$, la variabile casuale $\tilde{X} : \Omega \mapsto \mathbb{R}$ che descrive il punteggio del lancio può essere definita come

$$\tilde{X}(\omega) = \omega \quad \text{per } \omega \in \Omega,$$

e la sua funzione di ripartizione è quindi data da

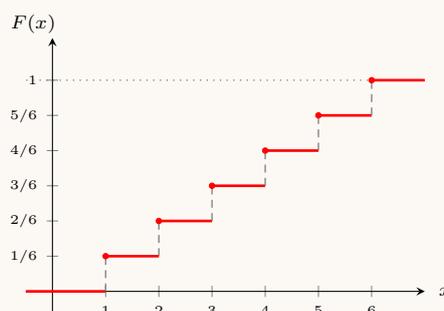
$$F(x) = P(\{\tilde{X} \leq x\}) = \frac{\#\{\tilde{X} \leq x\}}{\#\Omega} = \frac{\#\{\tilde{X} \leq x\}}{6} \quad \text{per } x \in \mathbb{R}.$$

Tenendo presente che

$$\#\{\tilde{X} \leq x\} = \lfloor x \rfloor \quad \text{per } x \in \mathbb{R},$$

(il simbolo $\lfloor x \rfloor$ indica il più grande numero intero che non supera x), vediamo che la funzione di ripartizione $F(x)$ può essere espressa in modo più esplicito come

$$F(x) = \begin{cases} 0 & \text{per } x < 0 \\ 1/6 & \text{per } 1 \leq x < 2 \\ 2/6 & \text{per } 2 \leq x < 3 \\ 3/6 & \text{per } 3 \leq x < 4 \\ 4/6 & \text{per } 4 \leq x < 5 \\ 5/6 & \text{per } 5 \leq x < 6 \\ 1 & \text{per } x \geq 6 \end{cases}$$



Non è difficile verificare che la funzione $F(x)$ soddisfa tutte e quattro le proprietà

- **F0** (il dominio di $F(x)$ è l'insieme dei numeri reali \mathbb{R}),
- **F1** ($F(x)$ è non decrescente),
- **F2** ($F(x)$ è continuità da destra) e
- **F3** ($\lim_{x \rightarrow -\infty} F(x) = 0$ e $\lim_{x \rightarrow \infty} F(x) = 1$)

che tutte le funzioni di ripartizione devono soddisfare (vedi il Teorema 4.2).

Esercizio 4.2 (Funzione di ripartizione). Un'urna contiene cinque palline delle quali tre sono di colore bianco e due di colore nero. Dall'urna vengono estratte, una dopo l'altra tutte e cinque le palline (le palline estratte non vengono quindi riposte nell'urna). Sia \tilde{X} la variabile casuale che restituisce il numero dell'estrazione nella quale si ottiene per la prima volta una pallina nera. Si ricavi la funzione di ripartizione di \tilde{X} .

La risposta viene lasciata per esercizio.

Dimostrando la proprietà **F1** abbiamo visto che

$$P(\{a < \tilde{X} \leq b\}) = F(b) - F(a).$$

Nel prossimo esercizio vedremo come a partire da una funzione di ripartizione si ricavano le probabilità di molti altri eventi che appartengono alla σ -algebra generata dalla variabile casuale di riferimento.

Esercizio 4.3 (Probabilità di eventi legati ad una variabile casuale). Sia \tilde{X} una variabile casuale con funzione di ripartizione data da

$$F(x) = \begin{cases} 0 & \text{per } x < 1 \\ 1 - 0.5 \times e^{-(x-1)} & \text{per } x \geq 1 \end{cases}$$

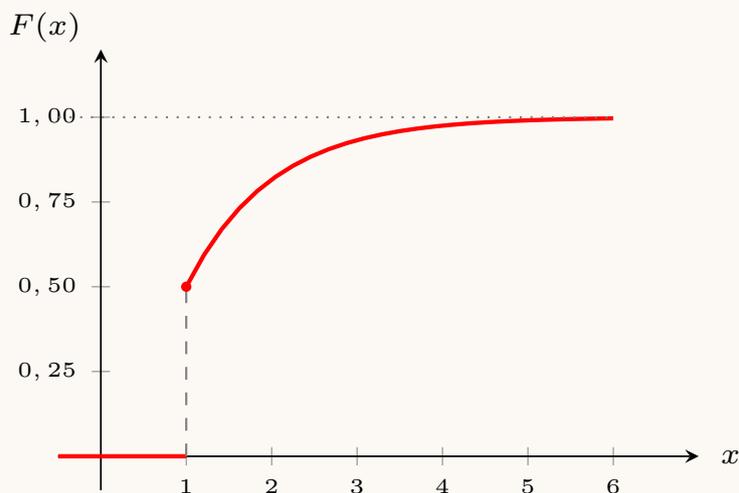
- Si ricavi il valore di $P(\{\tilde{X} \leq 2\})$.
- Si ricavi il valore di $P(\{\tilde{X} > 2\})$.
- Si ricavi il valore di $P(\{2 < \tilde{X} \leq 3\})$.
- Si ricavi il valore di $P(\{\tilde{X} = 3\})$.
- Si ricavi il valore di $P(\{2 < \tilde{X} < 3\})$ e lo si confronti con il valore di $P(\{2 < X \leq 3\})$ ricavato al punto c).
- Si ricavi il valore di $P(\{2 \leq \tilde{X} \leq 3\})$.
- Si ricavi il valore di $P(\{\tilde{X} = 1\})$.
- Si ricavi il valore di $P(\{1 < \tilde{X} \leq 3\})$.
- Si ricavi il valore di $P(\{1 \leq \tilde{X} \leq 3\})$.

Risposte:

Prima di rispondere ai quesiti conviene visualizzare l'andamento della funzione di ripartizione

$$F(x) = \begin{cases} 0 & \text{per } x < 1 \\ 1 - 0.5 \times e^{-(x-1)} & \text{per } x \geq 1 \end{cases}$$

tracciando il suo grafico:



Non è difficile verificare che $F(x)$ soddisfa le quattro proprietà **F0 - F4** che tutte le funzioni di ripartizione devono soddisfare (vedi il Teorema 4.2). Infatti:

- Il dominio di $F(x)$ è l'insieme dei numeri reali così come previsto dalla proprietà **F0**.
- $F(x)$ è una funzione non decrescente come previsto dalla proprietà **F1**.
- $F(x)$ è continua da destra dato che $F(x)$ è continua in ogni punto $x \in \mathbb{R}$ tranne che nel punto $x = 1$ dove comunque è continua da destra. Questo conferma che $F(x)$ soddisfa anche la proprietà **F2**.
- Siccome $F(x) = 0$ per ogni $x < 0$, possiamo concludere che

$$\lim_{x \rightarrow -\infty} F(x) = 0.$$

Inoltre, non è difficile verificare che $\lim_{x \rightarrow +\infty} F(x) = 1$. I due limiti confermano che $F(x)$ soddisfa anche la proprietà **F3**.

Passiamo ora alle risposte dei quesiti:

- a) Si ricavi il valore di $P(\{\tilde{X} \leq 2\})$.

Dalla definizione di funzione di ripartizione discende immediatamente che

$$P(\{\tilde{X} \leq 2\}) = F(2) = 1 - 0,5 \times e^{-(2-1)} = 0,816.$$

b) Si ricavi il valore di $P(\{\tilde{X} > 2\})$.

Tenendo presente che

$$\{\tilde{X} > 2\} = \overline{\{\tilde{X} \leq 2\}}$$

e usando la legge L3, vediamo che

$$\begin{aligned} P(\{\tilde{X} > 2\}) &= 1 - P(\{\tilde{X} \leq 2\}) \\ &= 1 - F(2) \\ &= 1 - 0,816 = 0,184. \end{aligned}$$

c) Si ricavi il valore di $P(\{2 < \tilde{X} \leq 3\})$.

Sostituendo $a = 2$ e $b = 3$ nella formula (21), otteniamo

$$\begin{aligned} P(\{2 < \tilde{X} \leq 3\}) &= F(3) - F(2) \\ &= [1 - 0,5 \times e^{-(3-1)}] - [1 - 0,5 \times e^{-(2-1)}] \\ &= 0,932 - 0,816 = 0,116. \end{aligned}$$

d) Si ricavi il valore di $P(\{\tilde{X} = 3\})$.

In primo luogo osserviamo che

$$\{\tilde{X} = 3\} = \bigcap_{n=1}^{\infty} \left\{ 3 - \frac{1}{n} < \tilde{X} \leq 3 \right\},$$

e che gli eventi coinvolti nell'intersezione al secondo membro sono annidati uno dentro all'altro formando una successione di eventi "non crescente" così come previsto dall'ipotesi nella legge L11. Dalla legge L11 discende dunque che

$$\begin{aligned} P(\{\tilde{X} = 3\}) &= P\left(\bigcap_{n=1}^{\infty} \left\{ 3 - \frac{1}{n} < \tilde{X} \leq 3 \right\}\right) \\ &= \lim_{n \rightarrow \infty} P\left(\left\{ 3 - \frac{1}{n} < \tilde{X} \leq 3 \right\}\right). \end{aligned}$$

Per determinare il limite nell'ultima riga notiamo innanzitutto che, come previsto dalla formula (21),

$$P\left(\left\{ 3 - \frac{1}{n} < \tilde{X} \leq 3 \right\}\right) = F(3) - F\left(3 - \frac{1}{n}\right)$$

per ogni $n = 1, 2, \dots$, e che (si ricordi che la funzione di ripartizione $F(x)$ è continua nel punto $x = 3$)

$$\lim_{n \rightarrow \infty} F\left(3 - \frac{1}{n}\right) = F(3).$$

Combinando questi risultati otteniamo

$$\begin{aligned}
 P(\{\tilde{X} = 3\}) &= \lim_{n \rightarrow \infty} P\left(\left\{3 - \frac{1}{n} < \tilde{X} \leq 3\right\}\right) \\
 &= \lim_{n \rightarrow \infty} \left[F(3) - F\left(3 - \frac{1}{n}\right) \right] \\
 &= F(3) - \lim_{n \rightarrow \infty} F\left(3 - \frac{1}{n}\right) \\
 &= F(3) - F(3) = 0
 \end{aligned}$$

Prima di procedere al prossimo quesito osserviamo che nella dimostrazione dell'uguaglianza $P(\{\tilde{X} = 3\}) = 0$ abbiamo usato soltanto il fatto che la funzione di ripartizione $F(x)$ è continua nel punto $x = 3$. Procedendo allo stesso modo si può anche dimostrare che $P(\{\tilde{X} = x\}) = 0$ per qualsiasi altro punto $x \in \mathbb{R}$ dove la funzione di ripartizione $F(x)$ è continua. Siccome $F(x)$ è continua in ogni punto $x \in \mathbb{R}$ tranne che nel punto $x = 1$, possiamo dunque concludere che

$$P(\{\tilde{X} = x\}) = 0 \quad \text{per ogni } x \neq 1.$$

Questo risultato potrebbe sembrare in contraddizione con la risposta al quesito b), dove abbiamo visto che $P(\{\tilde{X} > 2\}) = 0,184$. Infatti, siccome $P(\{\tilde{X} = x\}) = 0$ per ogni $x > 2$ e

$$\{\tilde{X} > 2\} = \bigcup_{x > 2} \{\tilde{X} = x\}, \quad (22)$$

si potrebbe pensare che in virtù dell'assioma **K3*** la probabilità $P(\{\tilde{X} > 2\})$ debba essere nulla. Per rendersi conto che questo ragionamento è sbagliato, basta osservare che l'assioma **K3*** non può essere applicato all'unione nella (22). Infatti, l'unione al secondo membro della (22) coinvolge un'infinità **non numerabile** di eventi mentre l'assioma **K3*** riguarda solamente unioni di collezioni infinite **numerabili** di eventi (come si calcolerebbe la somma di tutte le probabilità $P(\{\tilde{X} = x\})$ con $x > 2$? ... si tratterebbe di una somma di un'infinità non numerabile di termini ...).

- e) Si ricavi il valore di $P(\{2 < \tilde{X} < 3\})$ e lo si confronti con il valore di $P(\{2 < \tilde{X} \leq 3\})$ ricavato al punto c).

Per rispondere al quesito osserviamo che

$$\{2 < \tilde{X} \leq 3\} = \{\tilde{X} = 3\} \cup \{2 < \tilde{X} < 3\},$$

e che i due eventi che partecipano all'unione al secondo membro sono incompatibili. In virtù dell'assioma **K3** possiamo dunque scrivere

$$P(\{2 < \tilde{X} \leq 3\}) = P(\{\tilde{X} = 3\}) + P(\{2 < \tilde{X} < 3\}),$$

e da questa uguaglianza deduciamo che

$$\begin{aligned} P(\{2 < \tilde{X} < 3\}) &= P(\{2 < \tilde{X} \leq 3\}) - P(\{\tilde{X} = 3\}) \\ \text{[[quesito d)]]} &= P(\{2 < \tilde{X} \leq 3\}) - 0 \\ \text{[[quesito c)]]} &= 0,116. \end{aligned}$$

Il valore di $P(\{2 < \tilde{X} < 3\}) = 0,116$ poteva essere ottenuto anche per altra via, ovvero facendo riferimento alla legge L10. Infatti, non è difficile rendersi conto che

$$\{2 < \tilde{X} < 3\} = \bigcup_{n=1}^{\infty} \left\{ 2 < \tilde{X} \leq 3 - \frac{1}{n} \right\},$$

e che gli eventi coinvolti nell'unione formano una successione infinita "non decrescente" così come previsto dall'ipotesi nella legge L10. Applicando la legge L10 vediamo dunque che

$$\begin{aligned} P(\{2 < \tilde{X} < 3\}) &= P\left(\bigcup_{n=1}^{\infty} \left\{ 2 < \tilde{X} \leq 3 - \frac{1}{n} \right\}\right) \\ &= \lim_{n \rightarrow \infty} P\left(\left\{ 2 < \tilde{X} \leq 3 - \frac{1}{n} \right\}\right) \\ \text{[[formula (21)]]} &= \lim_{n \rightarrow \infty} \left[F\left(3 - \frac{1}{n}\right) - F(2) \right] \\ &= \left[\lim_{n \rightarrow \infty} F\left(3 - \frac{1}{n}\right) \right] - F(2) \\ &\text{[[} F(x) \text{ è continua nel punto } x = 3 \text{]]} \\ &= F(3) - F(2) \\ \text{[[formula (21)]]} &= P(\{2 < \tilde{X} < 3\}) \\ \text{[[quesito c)]]} &= 0,116 \end{aligned}$$

come già visto in precedenza.

f) Si ricavi il valore di $P(\{2 \leq \tilde{X} \leq 3\})$.

Siccome la funzione di ripartizione $F(x)$ è continua nel punto $x = 2$, possiamo ragionare come nella risposta al quesito d) onde ottenere

$$P(\{\tilde{X} = 2\}) = 0.$$

A questo punto basta osservare che

$$\{2 \leq \tilde{X} \leq 3\} = \{\tilde{X} = 2\} \cup \{2 < \tilde{X} \leq 3\}$$

e che i due eventi al secondo membro sono incompatibili onde concludere che (assioma K3)

$$\begin{aligned} P(\{2 \leq \tilde{X} \leq 3\}) &= P(\{\tilde{X} = 2\}) + P(\{2 < \tilde{X} \leq 3\}) \\ &= 0 + 0,116 = 0,116. \end{aligned}$$

Questo risultato coincide con il valore di $P(\{2 < \tilde{X} \leq 3\})$ e quello di $P(\{2 < \tilde{X} < 3\})$ che abbiamo ottenuto nelle risposte ai quesiti c) ed e).

g) Si ricavi il valore di $P(\{\tilde{X} = 1\})$.

Anche in questo caso possiamo ragionare come nella risposta al quesito d). Tuttavia, $x = 1$ è un punto di discontinuità della funzione di ripartizione $F(x)$ e per questo motivo, come vedremo tra breve, la probabilità $P(\{X = 1\})$ è positiva.

Vediamo dunque come si ricava il valore numerico di $P(\{X = 1\})$. Innanzitutto osserviamo che

$$\{\tilde{X} = 1\} = \bigcap_{n=1}^{\infty} \left\{ 1 - \frac{1}{n} < \tilde{X} \leq 1 \right\}$$

e che gli eventi coinvolti nell'intersezione formano una successione infinita "non crescente" così come previsto dall'ipotesi nella legge L11. Applicando la legge L11 otteniamo dunque

$$\begin{aligned} P(\{\tilde{X} = 1\}) &= P\left(\bigcap_{n=1}^{\infty} \left\{ 1 - \frac{1}{n} < \tilde{X} \leq 1 \right\}\right) \\ [[\text{legge L11}]] &= \lim_{n \rightarrow \infty} P\left(\left\{ 1 - \frac{1}{n} < \tilde{X} \leq 1 \right\}\right) \\ [[\text{formula (21)}]] &= \lim_{n \rightarrow \infty} \left[F(1) - F\left(1 - \frac{1}{n}\right) \right] \\ &= F(1) - \lim_{n \rightarrow \infty} F\left(1 - \frac{1}{n}\right). \end{aligned}$$

Questa volta, tuttavia,

$$\lim_{n \rightarrow \infty} F\left(1 - \frac{1}{n}\right) = 0 < F(1),$$

e pertanto concludiamo che

$$\begin{aligned} P(\{\tilde{X} = 1\}) &= F(1) - 0 = F(1) \\ &= 1 - 0,5 \times e^{-(1-1)} \\ &= 1 - 0,5 = 0,5. \end{aligned}$$

Notiamo che il valore di $P(\{\tilde{X} = 1\})$ coincide esattamente con l'altezza del salto della funzione di ripartizione $F(x)$ nel punto $x = 1$.

Anche il risultato ottenuto al punto d) può essere interpretato in questo senso: siccome nel punto $x = 3$ la funzione di ripartizione $F(x)$ è continua, e il salto della funzione di ripartizione nel punto $x = 3$ è dunque nullo, si ha $P(\{\tilde{X} = 3\}) = 0$.

h) Si ricavi il valore di $P(\{1 < \tilde{X} \leq 3\})$.

Usando la formula (21) otteniamo

$$\begin{aligned} P(\{1 < \tilde{X} \leq 3\}) &= F(3) - F(1) \\ &= [1 - 0,5 \times e^{-(3-1)}] - [1 - 0,5 \times e^{-(1-1)}] \\ &= 0,932 - 0,5 = 0,432. \end{aligned}$$

i) Si ricavi il valore di $P(\{1 \leq \tilde{X} \leq 3\})$.

Siccome

$$\{1 \leq \tilde{X} \leq 3\} = \{\tilde{X} = 1\} \cup \{1 < \tilde{X} \leq 3\},$$

e siccome i due eventi al secondo membro sono incompatibili, possiamo concludere che (assioma K3)

$$\begin{aligned} P(\{1 \leq \tilde{X} \leq 3\}) &= P(\{\tilde{X} = 1\}) + P(\{1 < \tilde{X} \leq 3\}) \\ [[\text{quesiti g) e h) }]] &= 0,5 + 0,432 = 0,932. \end{aligned}$$

Nel precedente esercizio abbiamo visto come a partire dalla funzione di ripartizione di una variabile casuale \tilde{X} si ricavano le probabilità di alcuni eventi che appartengono a $\sigma(\tilde{X})$. Generalizzando i ragionamenti contenuti nelle risposte ai quesiti dell'esercizio, si possono dimostrare alcune regole generali che esporremo nel prossimo teorema. D'ora in poi applicheremo queste regole anche senza richiamarle in modo esplicito.

Teorema 4.3 (Calcolo delle probabilità con funzioni di ripartizione). Se $F(x) = P(\{\tilde{X} \leq x\})$ è la funzione di ripartizione di una variabile casuale \tilde{X} , allora

i) $P(\{\tilde{X} > x\}) = 1 - F(x)$ per ogni $x \in \mathbb{R}$;

ii) per ogni intervallo semiaperto del tipo $(a, b]$ si ha

$$P(\{a < \tilde{X} \leq b\}) = F(b) - F(a);$$

iii) per ogni $x \in \mathbb{R}$ si ha

$$P(\{\tilde{X} = x\}) = F(x) - \lim_{h \rightarrow 0^+} F(x-h)$$

(il limite al secondo membro deve essere interpretato come il limite di $F(t)$ quando t si avvicina al punto x da sinistra). Ne consegue che:

– $P(\{\tilde{X} = x\}) = 0$ se x è un punto di continuità della funzione di ripartizione $F(\cdot)$;

- $P(\{\tilde{X} = x\})$ è uguale all'altezza del salto della funzione di ripartizione nel punto x se x è un punto di discontinuità.
- $P(\{a < \tilde{X} \leq b\}) = P(\{a \leq \tilde{X} \leq b\}) = P(\{a \leq \tilde{X} < b\}) = P(\{a < \tilde{X} < b\}) = F(b) - F(a)$ se $x = a$ e $x = b$ sono entrambi punti di continuità della funzione di ripartizione $F(x) = P(\{\tilde{X} \leq x\})$.

A questo punto conviene parlare della fase di assegnazione delle probabilità. Consideriamo dunque un esperimento casuale al quale è associato uno spazio campionario Ω , e supponiamo di voler **assegnare** delle probabilità **coerenti** a tutti gli eventi che appartengono alla σ -algebra generata da una data variabile casuale $\tilde{X} : \Omega \mapsto \mathbb{R}$, ovvero di voler definire una funzione di probabilità $P(\cdot)$ che abbia come dominio $\sigma(\tilde{X})$ e che soddisfi gli assiomi di Kolmogorov. Per evitare inutili complicazioni assumiamo che ...

- ... l'esito finale dell'esperimento casuale possa essere descritto da un unico numero reale ω e che lo spazio campionario Ω sia quindi l'insieme dei numeri reali \mathbb{R} ...
- ... e che la variabile casuale $\tilde{X} : \Omega \mapsto \mathbb{R}$ sia definita come $\tilde{X}(\omega) = \omega$ (\tilde{X} è quindi la variabile casuale che all'esito finale $\omega \in \Omega = \mathbb{R}$ associa esattamente lo stesso numero reale ω).

Si noti che nel caso che stiamo considerando si ha

$$\tilde{X}(\omega) \in B \Leftrightarrow \omega \in B$$

e questo significa che

$$\{\tilde{X} \in B\} = \{\omega \in \Omega : \tilde{X}(\omega) \in B\} = B \quad \text{per ogni } B \subseteq \mathbb{R}.$$

Nel caso che stiamo esaminando gli eventi che appartengono $\sigma(\tilde{X})$ sono dunque esattamente i sottoinsiemi B di \mathbb{R} che appartengono alla σ -algebra di Borel su \mathbb{R} e il problema che stiamo affrontando è quindi quello di definire una funzione di probabilità $P(\cdot)$ che abbia per dominio la σ -algebra di Borel $\sigma(\mathcal{B})$ e che soddisfi gli assiomi di Kolmogorov. Ora, in virtù del Teorema 4.2 possiamo dire che qualsiasi modo di **assegnare** le probabilità deve dare luogo ad una funzione di ripartizione

$$F(x) = P(\{\tilde{X} \leq x\})$$

che soddisfa tutte e quattro le proprietà **F0 - F3**. **Viceversa, si può anche dimostrare (omettiamo la dimostrazione) che la funzione di ripartizione di \tilde{X} può essere una qualunque funzione $F(\cdot)$ che soddisfa tutte e quattro le proprietà **F0 - F3!!!****

Si noti che questo risultato è di grande interesse perché ci permette di caratterizzare l'insieme di tutte le possibili funzioni di ripartizione di \tilde{X} come l'insieme di tutte le

funzioni $F(\cdot)$ che soddisfano le proprietà **F0 - F3**. E siccome la funzione di ripartizione di \tilde{X} **determina univocamente** le probabilità di tutti gli eventi che appartengono a $\sigma(\tilde{X})$ che nel caso che stiamo esaminando coincide con la σ -algebra di Borel $\sigma(\mathcal{B})$ (Teorema 4.1), questo risultato ci permette anche di **caratterizzare l'insieme di tutte le funzioni di probabilità $P(\cdot)$ che possono essere definite sulla σ -algebra di Borel $\sigma(\mathcal{B})$!!!** (ovvero l'insieme di tutte le possibili funzioni di probabilità $P(\cdot)$ che soddisfano gli assiomi di Kolmogorov e che hanno per dominio la σ -algebra di Borel $\sigma(\mathcal{B})$). Vista l'importanza di questi risultati, conviene dedicargli un apposito teorema:

Teorema 4.4 (Teorema di esistenza per funzioni di ripartizione). Se $F(\cdot)$ è una qualunque funzione che soddisfa le quattro proprietà **F0 - F3**, allora

- a) esiste un'unica funzione di probabilità $P(\cdot)$ che ha per dominio la σ -algebra di Borel $\sigma(\mathcal{B})$ (e che soddisfa gli assiomi di Kolmogorov) tale che

$$P((-\infty, x]) = F(x) \quad \text{per ogni } x \in \mathbb{R}$$

e di conseguenza

- b) esistono un modello probabilistico (Ω, \mathcal{A}, P) e una corrispondente variabile casuale $\tilde{X} : \Omega \mapsto \mathbb{R}$ tali che

$$P(\{\tilde{X} \leq x\}) = F(x) \quad \text{per ogni } x \in \mathbb{R}$$

(si ricordi che la funzione di probabilità di un modello probabilistico deve sempre soddisfare gli assiomi di Kolmogorov).

In virtù di questi fatti, tutte le funzioni $F(\cdot)$ che soddisfano le quattro proprietà **F0 - F3** possono essere chiamate "funzione di ripartizione" o "distribuzione" anche se non si riferiscono ad una specifica variabile casuale.

Commento sul Teorema 4.4: Si noti che il precedente teorema può essere interpretato come un **teorema di estensione**. Infatti, la prima parte della conclusione ci dice che le probabilità $P((-\infty, x]) = F(x)$ possono essere **estese in un unico modo** ad una funzione di probabilità $P(\cdot)$ che soddisfa gli assiomi di Kolmogorov e che ha per dominio la σ -algebra di Borel $\sigma(\mathcal{B})$.

A proposito del precedente teorema e delle considerazioni che lo precedono, conviene chiarire che se $\Omega \neq \mathbb{R}$ e/o $\tilde{X} : \Omega \mapsto \mathbb{R}$ non è la funzione identità (ovvero la funzione $\tilde{X}(\omega) = \omega$), allora potrebbe anche non essere possibile scegliere come funzione di ripartizione di \tilde{X} una qualunque funzione $F(x)$ che soddisfa le quattro proprietà **F0 - F3**, ma soltanto

alcune di tali funzioni che soddisfano anche determinate **condizioni aggiuntive** volte ad escludere la possibilità di **conflitti di assegnazione**. Per spiegare il problema basta considerare un esempio molto banale, come quello dove $\Omega = \{\omega_1, \omega_2\}$ e dove la variabile casuale $\tilde{X} : \Omega \mapsto \mathbb{R}$ è la funzione definita come

$$\tilde{X}(\omega) = \begin{cases} 1 & \text{se } \omega = \omega_1 \\ 2 & \text{se } \omega = \omega_2. \end{cases}$$

In questo caso si ha

$$\{\omega_1\} = \{\omega \in \Omega : \tilde{X}(\omega) \leq 1\} = \{\omega \in \Omega : \tilde{X}(\omega) \leq 1, 5\}$$

e la funzione di ripartizione di \tilde{X} non può dunque essere una qualunque funzione $F(x)$ che soddisfa le quattro proprietà **F0 - F3** ma deve necessariamente soddisfare anche la condizione $F(1) = F(1, 5)$ perché altrimenti si avrebbe

$$\begin{aligned} P(\{\omega_1\}) &= P(\{\omega \in \Omega : \tilde{X}(\omega) \leq 1\}) = F(1) \neq \\ &\neq F(1, 5) = P(\{\omega \in \Omega : \tilde{X}(\omega) \leq 1, 5\}) = P(\{\omega_1\}), \end{aligned}$$

e all'evento $\{\omega_1\}$ risulterebbero quindi associate due probabilità diverse (conflitto di assegnazione)!!!

Generalizzando questo esempio non è difficile rendersi conto che se esistono due valori x_1 e x_2 tali che

$$\{\omega \in \Omega : \tilde{X}(\omega) \leq x_1\} = \{\omega \in \Omega : \tilde{X}(\omega) \leq x_2\}, \quad (23)$$

allora la funzione di ripartizione di \tilde{X} non può mai essere una qualunque funzione $F(x)$ che soddisfa le quattro proprietà **F0 - F3**, ma dovrà necessariamente soddisfare anche il vincolo $F(x_1) = F(x_2)$ onde evitare l'insorgere di un conflitto di assegnazione.

D'altra parte, si può dimostrare (omettiamo la dimostrazione) che se la corrispondenza

$$x \in \mathbb{R} \quad \mapsto \quad \{\omega \in \Omega : \tilde{X}(\omega) \leq x\} \quad (24)$$

è **iniettiva** (ovvero a due valori distinti di x corrispondono sempre due insiemi $\{\omega \in \Omega : \tilde{X}(\omega) \leq x\}$ diversi), allora la funzione di ripartizione di \tilde{X} può effettivamente essere una qualunque funzione $F(x)$ che soddisfa le quattro proprietà **F0 - F3** senza che ciò introduca delle contraddizioni con gli assiomi di Kolmogorov, dei conflitti di assegnazione o qualunque altro tipo di contraddizione.

Non è difficile dimostrare che la corrispondenza nella (24) è iniettiva **se e solo se** per ogni $x \in \mathbb{R}$ esiste almeno un $\omega \in \Omega$ tale che $\tilde{X}(\omega) = x$, ovvero se e solo se la funzione $\tilde{X} : \Omega \mapsto \mathbb{R}$ è **suriettiva** (la dimostrazione della parte "se" è quasi immediata, mentre per la parte "solo se" conviene fare un ragionamento per assurdo).

4.3 Distribuzioni discrete e assolutamente continue

Come abbiamo appena visto, la funzione di ripartizione di una variabile casuale determina univocamente le probabilità di tutti gli eventi che appartengono alla σ -algebra generata dalla variabile casuale (vedi il Teorema 4.1) e per assegnare delle probabilità che soddisfanno gli assiomi di Kolmogorov a tutti gli eventi che appartengono alla σ -algebra generata da una variabile casuale è sufficiente specificare la sua funzione di ripartizione (vedi il Teorema 4.4 e i successivi commenti). Tuttavia, l'assegnazione di probabilità attraverso la scelta di una funzione di ripartizione è poco intuitiva e per questo motivo nelle applicazioni la funzione di ripartizione viene spesso specificata in modo indiretto scegliendo ...

- ... una opportuna **funzione (di massa) di probabilità** ...
- ... oppure una opportuna **funzione di densità**.

4.3.1 Funzioni (di massa) di probabilità

Definizione 4.3 (Funzione di massa di probabilità). Una **funzione di massa di probabilità** (o semplicemente **funzione di probabilità^a**) è una qualunque funzione $p(x)$ che soddisfa tutte e tre le seguenti proprietà:

p0) Il dominio di $p(x)$ è l'insieme dei numeri reali.

p1) $0 \leq p(x) \leq 1$ per ogni $x \in \mathbb{R}$ e l'insieme $S = \{x \in \mathbb{R} : p(x) > 0\}$ è un sottoinsieme finito o al più infinito numerabile di \mathbb{R} .

p2) $\sum_{x \in S} p(x) = 1$.

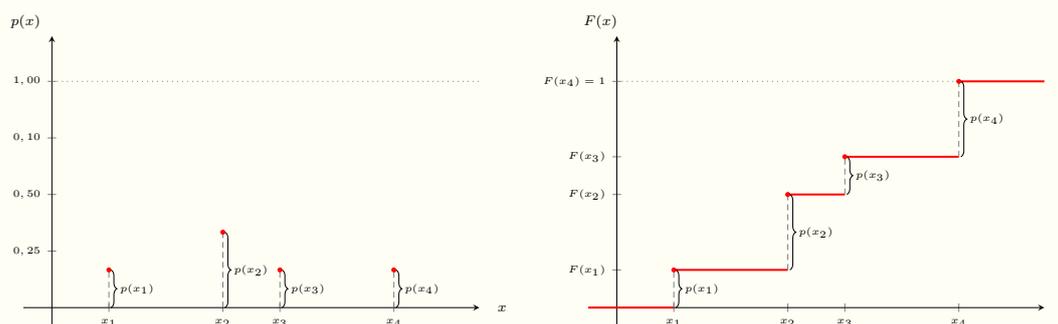
L'insieme S , ovvero l'insieme di tutti i punti x dove una funzione di probabilità $p(x)$ assume valori strettamente positivi, viene chiamato **supporto**. Siccome la proprietà p1 prevede che il supporto S di una funzione di massa di probabilità sia un insieme finito o al più infinito numerabile, i punti del supporto possono essere posti in corrispondenza biunivoca con un segmento iniziale dei numeri naturali positivi o con tutti i numeri naturali positivi. I punti del supporto S possono dunque essere indicati con $x_1, x_2, \dots, x_i, \dots$

^aSi noti che il termine **"funzione di probabilità"** ha un doppio significato perché potrebbe riferirsi sia ad una funzione di probabilità che è definita su una σ -algebra di eventi che ad una funzione di massa di probabilità il cui dominio è invece sempre l'insieme dei numeri reali \mathbb{R} . Di solito il contesto non lascia dubbi sul tipo di "funzione di probabilità" di cui si sta parlando.

Come si desume dai due grafici nella Figura 4.2, ad ogni **funzione di massa di probabilità** $p(x)$ corrisponde un'unica **funzione di ripartizione** $F(x)$ "a gradini" che ha dei "salti" di altezza $p(x_i)$ in corrispondenza dei punti x_1, x_2, \dots che

appartengono al suo supporto S , e viceversa. Le funzioni di ripartizione di questo tipo vengono chiamate **”discrete”** e anche le variabili casuali che hanno una funzione di ripartizione a gradini vengono chiamate **”discrete”**. Ovviamente, per **supporto** di una funzione di ripartizione discreta oppure di una variabile casuale discreta si intende sempre il supporto della corrispondente funzione di massa di probabilità.

Figura 4.2. Il grafico mostra come ad ogni **funzione di massa di probabilità** (grafico sul lato sinistro) corrisponde un’unica **funzione di ripartizione** a gradini (grafico sul lato destro) e viceversa.



Supponiamo ora che \tilde{X} sia una variabile casuale discreta con **funzione di massa di probabilità** $p(x)$. In questo caso si deve avere:

-

$$P(\{\tilde{X} = x\}) = p(x) \quad \text{per ogni } x \in \mathbb{R}$$

(per ottenere questa uguaglianza basta ricordare che per ogni $x \in \mathbb{R}$ il corrispondente valore di $p(x)$ è l’altezza del salto della funzione di ripartizione di \tilde{X} nel punto x e applicare il punto iii) della conclusione del Teorema 4.3), ...

- ... e se S è il supporto di \tilde{X} (si ricordi che la proprietà p1 impone che il supporto sia un sottoinsieme finito o al più infinito numerabile di \mathbb{R}), si deve avere

$$\begin{aligned} P(\{\tilde{X} \in S\}) &= \text{[[legge L2 oppure assioma K3*]]} = \\ &= \sum_{x \in S} P(\{\tilde{X} = x\}) = \sum_{x \in S} p(x) = \text{[[proprietà p2]]} = 1 \quad \Rightarrow \\ &\Rightarrow P(\{\tilde{X} \notin S\}) = 0. \end{aligned}$$

Per completare il quadro sulle variabili casuali discrete conviene aggiungere un’ultima osservazione (che non dimostreremo anche se la dimostrazione non sarebbe difficile): se \tilde{X} è una qualunque variabile casuale (di cui a priori non sappiamo se è discreta o meno), e se esiste un sottoinsieme S di \mathbb{R} che è finito o infinito numerabile e tale che

$$P(\{\tilde{X} \in S\}) = 1,$$

allora la funzione di ripartizione di \tilde{X} deve necessariamente essere discreta e \tilde{X} deve quindi essere una variabile casuale discreta.

Esercizio 4.4. Sia θ un numero reale e si consideri la funzione $p_\theta(x)$ definita come

$$p_\theta(x) = \begin{cases} \frac{x}{\theta} & \text{per } x = 1, 2, \dots, 10 \\ 0 & \text{altrimenti.} \end{cases}$$

Per quali valori di θ la corrispondente funzione $p_\theta(x)$ è una funzione di massa di probabilità?

Risposta:

Per rispondere al quesito dobbiamo determinare l'insieme dei valori di θ per i quali la corrispondente funzione $p_\theta(x)$ soddisfa tutte e tre le proprietà **p0**, **p1** e **p2**.

Per quanto concerne la proprietà **p0**, è immediato constatare che essa è soddisfatta per ogni $\theta \in \mathbb{R}$, mentre la proprietà **p1** è soddisfatta se e solo se $\theta > 10$.

Per quanto concerne invece la proprietà **p2**, notiamo che

$$\sum_{x=1}^{10} p_\theta(x) = \sum_{x=1}^{10} \frac{x}{\theta} = \frac{1}{\theta} \sum_{x=1}^{10} x = \frac{1}{\theta} \times \frac{10 \times (10 + 1)}{2} = \frac{55}{\theta}.$$

Quindi, la proprietà **p2** è soddisfatta se e solo se

$$\frac{55}{\theta} = 1 \quad \Leftrightarrow \quad \theta = 55.$$

Dalle precedenti considerazioni deduciamo che esiste un unico valore di θ per il quale la corrispondente funzione $p_\theta(x)$ è una funzione di massa di probabilità e che tale valore è $\theta = 55$.

Esercizio 4.5. Secondo un meteorologo, per i prossimi tre giorni a partire da domani la probabilità di pioggia raddoppia ogni giorno ed entro i prossimi tre giorni pioverà sicuramente.

- Si determini la funzione di massa di probabilità della variabile casuale \tilde{X} che ci dice tra quanti giorni inizierà a piovere.
- Si determini la funzione di ripartizione di \tilde{X} .
- Si calcoli il valore di $P(\{\tilde{X} > 1\})$.

Soluzione:

- a) Si determini la funzione di massa di probabilità della variabile casuale \tilde{X} che ci dice tra quanti giorni inizierà a piovere.

Dalle informazioni contenute nella consegna deduciamo che la variabile casuale \tilde{X} può assumere solamente i valori $x = 1$, $x = 2$ e $x = 3$, e pertanto possiamo concludere che deve trattarsi di una variabile casuale discreta. Indichiamo dunque con

$$p(x) = P(\{\tilde{X} = x\}), \quad x \in \mathbb{R},$$

la funzione di massa di probabilità della variabile casuale \tilde{X} . Dalle informazioni contenute nella consegna deduciamo che $p(x)$ deve soddisfare le seguenti condizioni:

$$p(2) = 2p(1) \quad \text{e} \quad p(3) = 2p(2) = 4p(1).$$

Siccome il valore di $\sum_{x \in S} p(x)$ deve essere uguale a 1 (proprietà p2 delle funzioni di massa di probabilità), possiamo concludere che

$$p(1) + p(2) + p(3) = p(1) + 2p(1) + 4p(1) = 7p(1) = 1$$

e che quindi $p(1) = 1/7$, $p(2) = 2/7$ e $p(3) = 4/7$. La funzione di probabilità di \tilde{X} è quindi definita come

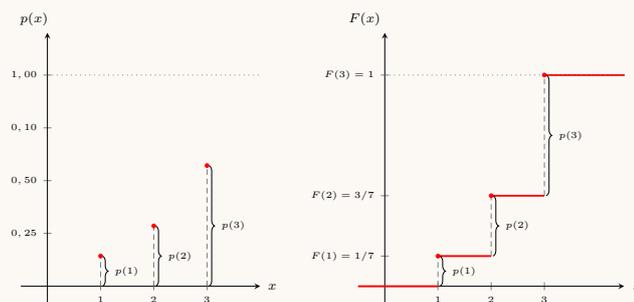
$$p(x) = P(\{\tilde{X} = x\}) = \begin{cases} 1/7 & \text{per } x = 1, \\ 2/7 & \text{per } x = 2, \\ 4/7 & \text{per } x = 3, \\ 0 & \text{altrimenti.} \end{cases}$$

- b) Si determini la funzione di ripartizione di \tilde{X} .

Cumulando i valori assunti dalla funzione di massa di probabilità $p(x)$, vediamo che la funzione di ripartizione di \tilde{X} deve essere data da

$$F(x) = P(\{\tilde{X} \leq x\}) = \sum_{t \in S: t \leq x} p(t) = \begin{cases} 0 & \text{per } x < 1 \\ p(1) = 1/7 & \text{per } 1 \leq x < 2 \\ p(1) + p(2) = 3/7 & \text{per } 2 \leq x < 3 \\ p(1) + p(2) + p(3) = 1 & \text{per } x \geq 3 \end{cases}$$

I due grafici sottostanti mostrano la funzione di probabilità $p(x)$ e la corrispondente funzione di ripartizione $F(x)$.



c) Si calcoli il valore di $P(\{\tilde{X} > 1\})$.

Facendo riferimento alla funzione di ripartizione possiamo calcolare la probabilità richiesta come

$$P(\{\tilde{X} > 1\}) = 1 - F(1) = 1 - p(1) = 1 - \frac{1}{7} = \frac{6}{7}.$$

Si noti che in alternativa, senza ricorrere alla funzione di ripartizione, possiamo calcolare la probabilità richiesta anche a partire dalla funzione di massa di probabilità $p(x)$:

$$P(\{\tilde{X} > 1\}) = P(\{\tilde{X} = 2\}) + P(\{\tilde{X} = 3\}) = p(2) + p(3) = \frac{2}{7} + \frac{4}{7} = \frac{6}{7}.$$

Esercizio 4.6. Si tracci il grafico della funzione di ripartizione discreta

$$F(x) = \begin{cases} 0 & \text{per } x < 0 \\ 0,3 & \text{per } 0 \leq x < 2 \\ 0,7 & \text{per } 2 \leq x < 5 \\ 1 & \text{per } x \geq 5 \end{cases}$$

e si ricavi la corrispondente funzione di probabilità.

La soluzione viene lasciata come esercizio.

In questo paragrafo abbiamo visto soltanto esempi di distribuzioni discrete con **supporto** finito. Più avanti, quando studieremo le caratteristiche di alcune distribuzioni notevoli, vedremo anche esempi di distribuzioni discrete con supporto infinito (numerabile).

4.3.2 Funzioni di densità

Definizione 4.4 (Funzione di densità). Una **funzione di densità** è una qualunque funzione $f(x)$ che soddisfa tutte e tre le seguenti proprietà:

- f0) il dominio di $f(x)$ è l'insieme dei numeri reali.
- f1) $f(x) \geq 0$ per ogni $x \in \mathbb{R}$.
- f2) $f(x)$ è una funzione integrabile^a e $\int_{-\infty}^{\infty} f(x)dx = 1$.

L'insieme S di tutti i punti $x \in \mathbb{R}$ dove una funzione di densità $f(\cdot)$ è positiva viene chiamato **supporto** della funzione di densità: $S = \{x \in \mathbb{R} : f(x) > 0\}$.

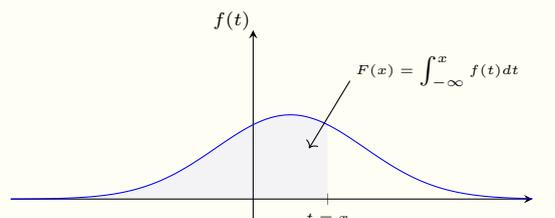
^aTutti gli integrali che compariranno in questa dispensa possono essere interpretati come degli integrali di Riemann (eventualmente impropri) e in questo caso il termine "funzione integrabile" deve essere interpretato come "funzione Riemann-integrabile". Altrimenti, volendo interpretare gli integrali come integrali di Lebesgue, il termine "funzione integrabile" può essere interpretato come "funzione misurabile". Siccome le funzioni che si incontrano nelle applicazioni sono sempre Riemann-integrabili (se limitate a sottoinsiemi compatti di \mathbb{R}), e pertanto sono anche misurabili, **in questa dispensa non ci preoccuperemo mai di verificare se una data funzione $f : \mathbb{R} \mapsto \mathbb{R}$ è effettivamente "integrabile" e daremo sempre per scontato che ciò sia vero.**

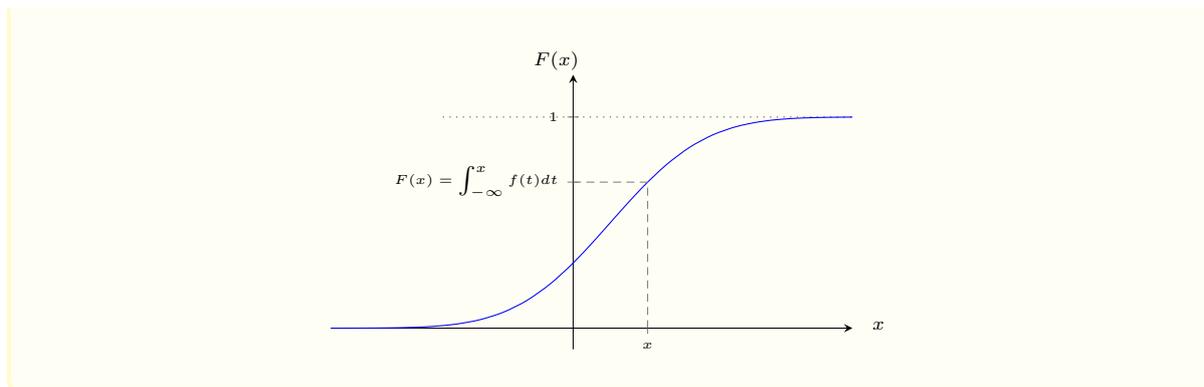
Non è difficile dimostrare (omettiamo la dimostrazione) che ad ogni **funzione di densità** $f(x)$ corrisponde una **funzione di ripartizione** $F(x)$ che è definita come

$$F(x) = \int_{-\infty}^x f(t)dt \quad \text{per ogni } x \in \mathbb{R} \quad (25)$$

(vedi Figura 4.3). Funzioni di ripartizione che possono essere espresse in questo modo vengono chiamate **"assolutamente continue"** e variabili casuali con funzione di ripartizione assolutamente continua vengono a loro volta chiamate assolutamente continue.

Figura 4.3. I due grafici mostrano la relazione tra una funzione di densità e la corrispondente funzione di ripartizione.





Come si evince dalla formula (25), ad ogni **funzione di densità** $f(x)$ corrisponde un'**unica funzione di ripartizione assolutamente continua** $F(x)$, ma è importante tenere presente che ogni funzione di ripartizione assolutamente continua $F(x)$ può essere ottenuta a partire da **infinite** funzioni di densità diverse. Infatti, i valori assunti da una funzione di densità $f(x)$ possono sempre essere modificati in un numero finito di punti $x \in \mathbb{R}$ (o anche su determinati sottoinsiemi infiniti di \mathbb{R}) senza che la funzione di ripartizione $F(x)$ definita nella (25) subisca alcuna variazione.

Tutte le funzioni di densità $f(x)$ che danno luogo alla medesima funzione di ripartizione vengono chiamate **"versioni"** della funzione di densità della funzione di ripartizione $F(x)$ di riferimento. Quando si parla al singolare della funzione di densità di una funzione di ripartizione $F(x)$ che è assolutamente continua, in realtà si intende l'insieme di tutte le funzioni di densità di $F(x)$ e questo insieme forma una cosiddetta **"classe di equivalenza"**.

Siccome esistono sempre infinite **funzioni di densità** $f(x)$ con supporti $S = \{x \in \mathbb{R} : f(x) > 0\}$ diversi che danno luogo alla medesima **funzione di ripartizione assolutamente continua** $F(x)$, **non ha senso parlare del supporto di una distribuzione o di una variabile casuale assolutamente continua** (cosa che invece si poteva fare nel caso discreto). Quando in alcuni libri di testo si parla del supporto di una distribuzione o di una variabile casuale assolutamente continua si intende un altro tipo di supporto di cui in questa dispensa non parleremo.

Si può dimostrare che **tutte le funzioni di ripartizione assolutamente continue sono delle funzioni continue** (ovvero continue in ogni punto $x \in \mathbb{R}$), ma che **non è vero il viceversa**: infatti, esistono delle funzioni di ripartizione che sono continue ma che non possono essere espresse come integrali di una funzione di densità. Tuttavia, nelle applicazioni non si incorre mai in funzioni di ripartizione continue che allo stesso tempo non sono anche assolutamente continue e tutte le funzioni di ripartizione continue che incontreremo in questa dispensa saranno pertanto anche assolutamente continue.

Per ricavare una delle infinite **funzioni di densità** $f(x)$ di una data **funzione di ripartizione** $F(x)$ che è **assolutamente continua** basta porre

$$f(x) = \begin{cases} F'(x) = \frac{dF(x)}{dx} = \lim_{h \rightarrow 0} \frac{F(x+h) - F(x)}{h} & \text{se nel punto } x \text{ la derivata } F'(x) \text{ esiste;} \\ 0 & \text{altrimenti.} \end{cases} \quad (26)$$

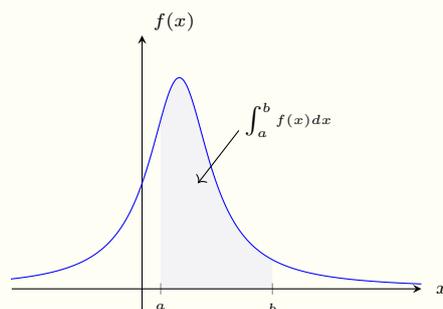
Si noti che se in un numero finito di punti x il calcolo della derivata $F'(x)$ fosse difficile e/o l'esistenza della derivata $F'(x)$ fosse in dubbio, si può comunque sempre ottenere un'altra **versione** della funzione di densità ponendo semplicemente $f(x) = 0$ anche in tali punti.

Di solito, nelle applicazioni si preferisce fare riferimento a **funzioni di densità** $f(x)$ che hanno un andamento "regolare" e che sono continue su un ampio insieme di punti $x \in \mathbb{R}$. Il motivo è molto semplice: se la **funzione di ripartizione** $F(x)$ di una data variabile casuale \tilde{X} è **assolutamente continua**, e se in ogni punto x di un dato intervallo $[a, b]$ una corrispondente funzione di densità $f(x)$ è continua, allora la probabilità

$$\begin{aligned} P(\{a \leq \tilde{X} \leq b\}) &= P(\{a < \tilde{X} \leq b\}) = P(\{a \leq \tilde{X} < b\}) = P(\{a < \tilde{X} < b\}) = \\ &= F(b) - F(a) = \int_{-\infty}^b f(x) dx - \int_{-\infty}^a f(x) dx = \int_a^b f(x) dx \end{aligned}$$

può essere visualizzata come l'area sottesa alla funzione di densità $f(x)$ sopra l'intervallo $[a, b]$ (vedi la Figura 4.4).

Figura 4.4. Il grafico illustra come dall'integrale di una **funzione di densità** continua si ottiene un'area.



Esercizio 4.7. Per quali valori di θ la funzione

$$f_{\theta}(x) = \begin{cases} \theta e^{-\theta x} & \text{se } x > 0 \\ 0 & \text{altrimenti} \end{cases}$$

è una funzione di densità? Si ricavino anche le corrispondenti funzioni di ripartizione assolutamente continue.

Risposta:

Per rispondere al quesito dobbiamo verificare per quali valori di θ la corrispondente funzione $f_\theta(x)$ soddisfa tutte e tre le proprietà **f0** - **f2**. Per quanto concerne la proprietà **f0** è immediato verificare che per qualsiasi valore reale di θ la corrispondente funzione $f_\theta(x)$ ha come dominio l'insieme dei numeri reali. Quindi concludiamo che la proprietà **f0** non implica nessuna restrizione sull'insieme dei possibili valori di θ . Consideriamo dunque le proprietà **f1** e **f2**:

- Siccome $\theta e^{-\theta x} \geq 0$ se e solo se $\theta \geq 0$, possiamo concludere che la proprietà **f1** è soddisfatta se e solo se $\theta \geq 0$.
- Per trovare i valori di $\theta \geq 0$ per i quali risulta soddisfatta anche la proprietà **f2** calcoliamo l'integrale

$$\int_{-\infty}^{\infty} f_\theta(x) dx = \int_0^{\infty} \theta e^{-\theta x} dx.$$

Ovviamente questo integrale è nullo se $\theta = 0$, e quindi possiamo escludere il valore $\theta = 0$. D'altra parte, se $\theta > 0$, si ottiene

$$\begin{aligned} \int_{-\infty}^{\infty} f_\theta(x) dx &= \int_0^{\infty} \theta e^{-\theta x} dx \\ &= \left[\theta \frac{e^{-x}}{(-\theta)} \right]_0^{\infty} = 0 - (-1) = 1 \end{aligned}$$

così come previsto dalla proprietà **f2**. La proprietà **f2** (ovvero la condizione $\int_{-\infty}^{\infty} f_\theta(x) dx = 1$) è quindi soddisfatta se e solo se $\theta > 0$.

Riassumendo i precedenti ragionamenti possiamo dire che la funzione $f_\theta(x)$ è una funzione di densità (ovvero soddisfa tutte e tre le proprietà **f0** - **f2**) se e solo se θ è un numero reale positivo.

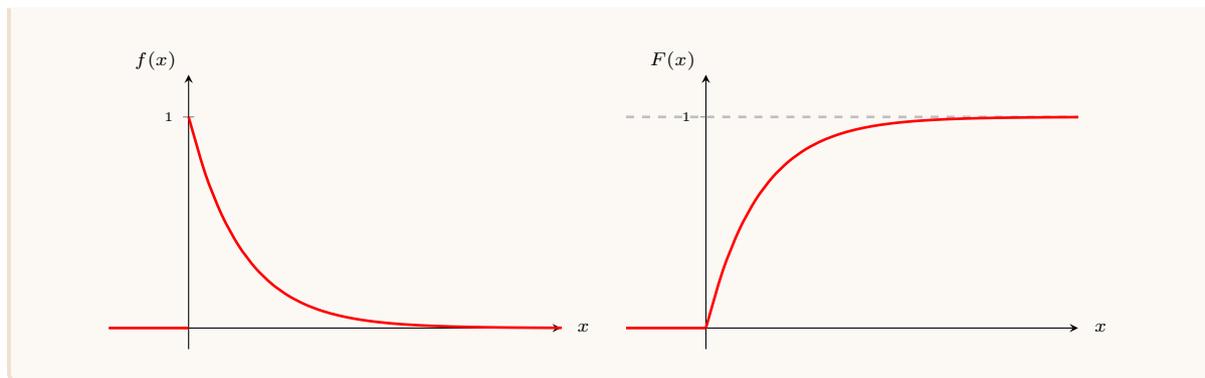
Come abbiamo visto, la funzione di ripartizione che corrisponde ad una data funzione di densità $f(x)$ è definita come

$$F(x) = \int_{-\infty}^x f(t) dt, \quad x \in \mathbb{R}.$$

Le funzioni di ripartizione corrispondenti alle funzioni di densità $f_\theta(x)$ (con $\theta > 0$) sono quindi date da

$$F_\theta(x) = \int_{-\infty}^x f_\theta(t) dt = \begin{cases} \int_{-\infty}^x 0 dt = 0 & \text{per } x < 0 \\ \int_0^x \theta e^{-\theta t} dt = \left[\theta \times \frac{e^{-t}}{(-\theta)} \right]_0^x = 1 - e^{-\theta x} & \text{per } x \geq 0. \end{cases}$$

I due grafici sottostanti mostrano l'andamento della funzione di densità $f_{\theta=1}(x)$ e quello della corrispondente funzione di ripartizione $F_{\theta=1}(x)$.



Esercizio 4.8. Si ricavi una delle infinite funzioni di densità della funzione di ripartizione assolutamente continua

$$F(x) = \begin{cases} 1 - \frac{1}{1+x} & \text{per } x \geq 0 \\ 0 & \text{per } x < 0 \end{cases}$$

Risposta:

Per rispondere ricaveremo la funzione di densità (26) che, ricordiamo, è definita come

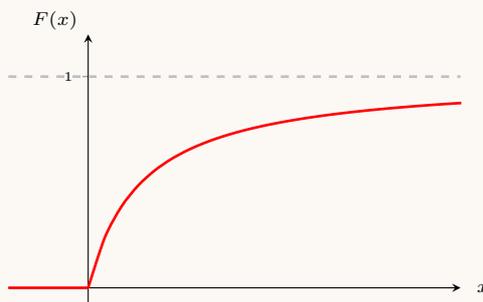
$$f(x) = \begin{cases} F'(x) & \text{se nel punto } x \text{ la derivata } F'(x) \text{ esiste;} \\ 0 & \text{altrimenti.} \end{cases}$$

Usando le solite regole di derivazione non è difficile verificare che

$$F'(x) = \begin{cases} \frac{1}{(1+x)^2} & \text{per } x > 0 \\ 0 & \text{per } x < 0 \end{cases}$$

L'unico punto dove l'esistenza della derivata $F'(x)$ è in dubbio è il punto $x = 0$.

Come si vede nel grafico accanto, $x = 0$ è un **punto angoloso** della funzione di ripartizione $F(x)$ e quindi in tale punto la derivata della funzione $F(x)$ non esiste.



Ponendo $f(x) = F'(x)$ per ogni $x \neq 0$ e ponendo $f(x) = 0$ nel punto $x = 0$ (dove $F(x)$ non è derivabile) vediamo quindi che nel caso in questione la funzione di densità (26) è data da

$$f(x) = \begin{cases} \frac{1}{(1+x)^2} & \text{per } x > 0 \\ 0 & \text{per } x \leq 0. \end{cases}$$

Ovviamente, avremmo potuto procedere anche in modo analitico per stabilire che $x = 0$ sia un punto angoloso di $F(x)$. Infatti, la derivata destra di $F(x)$ nel punto $x = 0$ è data da

$$\lim_{h \rightarrow 0^-} \frac{F(0+h) - F(0)}{h} = \lim_{h \rightarrow 0^-} \frac{0 - 0}{h} = 0,$$

mentre la derivata sinistra è data da

$$\begin{aligned} \lim_{h \rightarrow 0^+} \frac{F(0+h) - F(0)}{h} &= \lim_{h \rightarrow 0^+} \frac{\left(1 - \frac{1}{1+0+h}\right) - 0}{h} \\ [[\cdot \cdot \cdot]] &= \lim_{h \rightarrow 0^+} \frac{1}{1+h} = 1. \end{aligned}$$

Siccome le due derivate direzionali esistono ma non coincidono, possiamo concludere che il punto $x = 0$ sia un punto angoloso della funzione di ripartizione $F(x)$.

Esercizio 4.9 (Distribuzioni uniformi continue). Un esperimento casuale consiste nel selezionare in modo casuale un punto P che giace su un segmento di retta di lunghezza θ . Sia \tilde{X} la variabile casuale che descrive la distanza tra il punto P e il punto iniziale A del segmento, e si assuma che la scelta del punto P avvenga in modo completamente casuale, ovvero in modo tale che per ogni intervallo $[a, b] \subset [0, \theta]$ la probabilità $P(\{a \leq \tilde{X} \leq b\})$ sia proporzionale alla lunghezza dell'intervallo.

- a) Si ricavi la funzione di ripartizione della variabile casuale \tilde{X} .
- b) Si verifichi se la funzione di ripartizione della variabile casuale \tilde{X} è assolutamente continua e, in caso affermativo, si ricavi una corrispondente funzione di densità.

Risposte:

- a) Si ricavi la funzione di ripartizione della variabile casuale \tilde{X} .

Dalle informazioni contenute nella consegna deduciamo che

- i) per ogni $x < 0$ si deve avere $F(x) = P(\{\tilde{X} \leq x\}) = 0$ perché una distanza non può mai essere negativa;
- ii) per ogni $x \in [0, \theta]$ si deve avere

$$\begin{aligned} F(x) &= P(\{\tilde{X} \leq x\}) \\ &= P(\{\tilde{X} < 0\}) + P(\{0 \leq \tilde{X} \leq x\}) \\ &= 0 + c \times (x - 0) = c \times x \end{aligned}$$

dove c è una costante di proporzionalità che dobbiamo determinare;

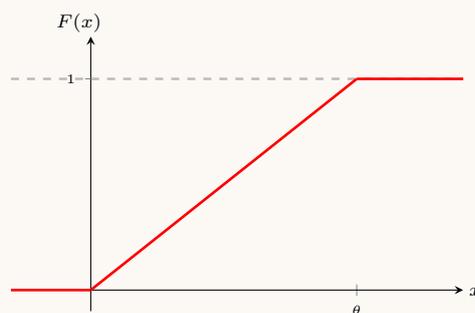
iii) per ogni $x \geq \theta$ si deve avere $F(x) = P(\{\tilde{X} \leq x\}) = 1$ perché per ipotesi la lunghezza del segmento all'interno del quale viene selezionato il punto P è uguale a θ .

Dalle precedenti osservazioni ii) e iii) possiamo dedurre che nel punto $x = \theta$ si deve avere

$$F(\theta) = c \times \theta = 1$$

e che il valore della costante di proporzionalità c deve quindi essere dato da $c = 1/\theta$. La funzione di ripartizione della variabile casuale \tilde{X} è dunque definita come

$$F(x) = P(\{\tilde{X} \leq x\}) = \begin{cases} 0 & \text{per } x < 0 \\ x/\theta & \text{per } 0 \leq x < \theta \\ 1 & \text{per } x \geq \theta. \end{cases} \quad (27)$$



b) Si verifichi se la funzione di ripartizione della variabile casuale \tilde{X} è assolutamente continua e, in caso affermativo, si ricavi una corrispondente funzione di densità.

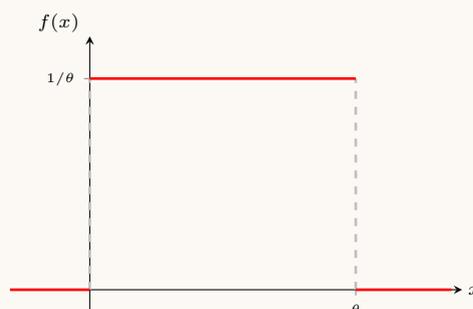
Come abbiamo visto nell'esposizione della teoria sulle [distribuzioni assolutamente continue](#), se una generica [funzione di ripartizione](#) $F(x)$ è assolutamente continua, allora la funzione $f(x)$ che si ottiene attraverso la formula (26) deve essere una sua funzione di densità. Per verificare se la funzione di ripartizione (27) è assolutamente continua, è quindi sufficiente calcolare la corrispondente funzione $f(x)$ definita dalla formula (26), e verificare se integrando questa funzione si ottiene di nuovo la funzione di ripartizione (27). In caso affermativo possiamo concludere che la funzione di ripartizione (27) è assolutamente continua; altrimenti concluderemo che la funzione di ripartizione (27) non è assolutamente continua.

Ora, tenendo presente che la derivata $dF(x)/dx$ esiste in tutti punti $x \in \mathbb{R}$ tranne nei punti $x = 0$ e $x = \theta$ (che sono dei punti angolosi), possiamo

facilmente verificare che nel caso in questione la funzione (26) è data da

$$f(x) = \begin{cases} 1/\theta & \text{se } 0 < x < \theta \\ 0 & \text{altrimenti.} \end{cases}$$

Quindi, come si può anche vedere nel grafico sottostante, la funzione $f(x)$ è costante e pari $1/\theta$ sull'intervallo $(0, \theta)$, e nulla altrove.



Siccome

$$\begin{aligned} \int_{-\infty}^x f(t) dt &= \begin{cases} \int_{-\infty}^x 0 dt = 0 & \text{per } x < 0 \\ \int_{-\infty}^x 0 dt + \int_0^x (1/\theta) dt = 0 + x/\theta & \text{per } 0 \leq x < \theta \\ \int_{-\infty}^x 0 dt + \int_0^x (1/\theta) dt + \int_{\theta}^x 0 dt = 0 + 1 + 0 & \text{per } x \geq \theta. \end{cases} \\ &= \begin{cases} 0 & \text{per } x < 0 \\ x/\theta & \text{per } 0 \leq x < \theta \\ 1 & \text{per } x \geq \theta \end{cases} \\ &= F(x) \quad \text{per ogni } x \in \mathbb{R}, \end{aligned}$$

possiamo concludere che la funzione di ripartizione (27) è assolutamente continua e che la funzione $f(x)$ che abbiamo appena ricavato è una sua funzione di densità. In virtù dell'andamento costante della loro funzione di densità, le distribuzioni $F(x)$ di questo esercizio vengono chiamate **”distribuzioni uniformi” (continue)**.

4.4 Quantili

Con riferimento ad alcune variabili casuali \tilde{X} (che rappresentano per esempio una perdita finanziaria, la magnitudo di un terremoto, la forza del vento, una forza che agisce su determinate strutture, ecc.), si vuole spesso valutare **”il valore x che non viene**

superato con una determinata probabilità p ", ovvero il valore di x tale che

$$F(x) = P(\{\tilde{X} \leq x\}) = p.$$

A seconda dell'andamento della funzione di ripartizione della variabile casuale \tilde{X} , il valore x in questione potrebbe esistere o meno e, se esiste, potrebbe essere univocamente determinato oppure no.

Esempio 4.2 (Quantili - esempio introduttivo). Sia \tilde{X} una variabile casuale con funzione di ripartizione data da

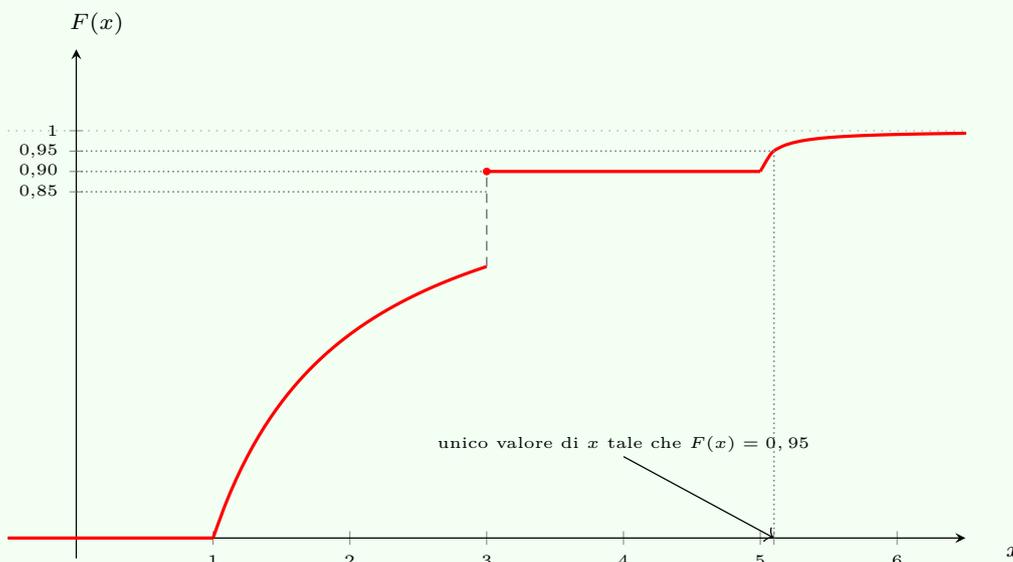
$$F(x) = \begin{cases} 0 & \text{per } x < 1 \\ 1 - \frac{1}{x} & \text{per } 1 \leq x < 3 \\ 0.9 & \text{per } 3 \leq x < 5 \\ 1 - \frac{1}{100x-490} & \text{per } x \geq 5 \end{cases}$$

- Si ricavi il valore x che non viene superato con probabilità pari a 0,85.
- Si ricavi il valore x che non viene superato con probabilità pari a 0,90.
- Si ricavi il valore x che non viene superato con probabilità pari a 0,95.

Risposta:

Per rispondere ai tre quesiti conviene tracciare il grafico della funzione di ripartizione e aggiungere al grafico delle rette orizzontali che intersecano l'asse delle ordinate in corrispondenza di

$$a) F(x) = 0,85; \quad b) F(x) = 0,90; \quad c) F(x) = 0,95.$$



Dal grafico si desume che ...

a) ... non esiste nessun valore x tale che

$$P\left(\{\tilde{X} \leq x\}\right) = F(x) = 0,85$$

b) ... esistono infiniti valori x tali che

$$P\left(\{\tilde{X} \leq x\}\right) = F(x) = 0,90,$$

ovvero tutti i valori x che appartengono all'intervallo $[3, 5]$;

c) ... esiste un unico valore x tale che

$$P\left(\{\tilde{X} \leq x\}\right) = F(x) = 0,95,$$

ovvero il valore che si ottiene risolvendo l'equazione

$$1 - \frac{1}{100x - 490} = 0,95 \quad \Rightarrow \quad x = 5,1.$$

Per superare i problemi di esistenza e unicità evidenziati nel precedente esempio, invece del "valore x che non viene superato con probabilità esattamente pari a p " di solito si calcola "il più piccolo valore x che non viene superato probabilità almeno pari a p ", ovvero il valore

$$x_p = \inf \{x : F(x) \geq p\} \quad (28)$$

(si ricordi che $F(x) = P(\{\tilde{X} \leq x\})$ è la probabilità di non superare il valore x). Infatti, non è difficile dimostrare che per qualsiasi $p \in (0, 1)$ il valore di x_p esiste ed è unico e in ambito probabilistico questo valore viene chiamato **quantile di ordine p** della distribuzione di \tilde{X} o semplicemente quantile di ordine p di \tilde{X} . Di solito la formula (28) viene anche usata per definire il quantile di ordine $p = 1$ (si noti che in questo modo il quantile di ordine $p = 1$ sarà uguale a $+\infty$ a meno che non esista un valore di x tale che $F(x) = 1$), mentre per definire il quantile di ordine $p = 0$ di solito si pone

$$x_0 = \inf \{x : F(x) > p\}$$

con il segno di disuguaglianza stretta al posto del segno di disuguaglianza debole (in questo modo si evita che il quantile di ordine $p = 0$ risulti pari a $-\infty$ quando esistono dei valori di x per i quali $F(x) = 0$).

Come si desume dalla loro definizione,

$$x_p = \begin{cases} \inf \{x : F(x) > p\} & \text{per } p = 0; \\ \inf \{x : F(x) \geq p\} & \text{per } p \in (0, 1] \end{cases}$$

i valori dei quantili di una variabile casuale \tilde{X} dipendono solo dall'andamento della sua funzione di ripartizione $F(x)$ e per questo motivo si può anche parlare dei quantili di una distribuzione senza riferimento ad una specifica variabile casuale.

Come nella statistica descrittiva, anche in ambito probabilistico il valore di un quantile $x_p \dots$

- ... con ordine p che è un multiplo di 0,01 viene chiamato " **p -esimo percentile**";
- ... con ordine p che è un multiplo di 0,10 viene chiamato (a seconda dei casi) "**primo/secondo/.../nono decile**";
- ... con ordine p che è un multiplo di 0,25 viene chiamato (a seconda dei casi) "**primo/secondo/terzo quartile**";
- ... con ordine p pari a 0,5 viene chiamato "**mediana**".

Esercizio 4.10 (Calcolo di quantili). Sia \tilde{X} una variabile casuale con funzione di ripartizione data da

$$F(x) = \begin{cases} 0 & \text{per } x < 1 \\ 1 - \frac{1}{x} & \text{per } 1 \leq x < 3 \\ 0.9 & \text{per } 3 \leq x < 5 \\ 1 - \frac{1}{100x-490} & \text{per } x \geq 5 \end{cases}$$

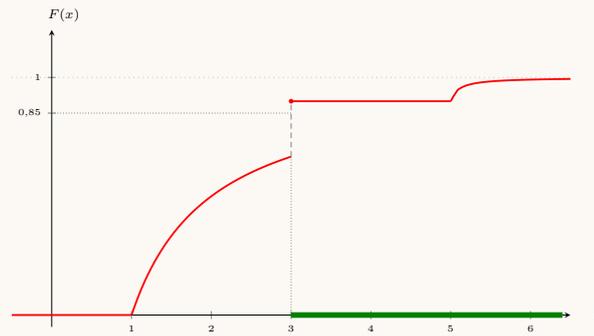
- Si ricavi il valore dell'85-esimo percentile \tilde{X} .
- Si ricavi il valore del nono decile di \tilde{X} .
- Si ricavi il valore del 95-esimo percentile \tilde{X} .

Risposte:

- Si ricavi il valore dell'85-esimo percentile \tilde{X} .

Con l'ausilio del grafico sottostante si vede facilmente che il più piccolo valore di x dove la funzione di ripartizione $F(x)$ è almeno pari a 0,85 è dato da

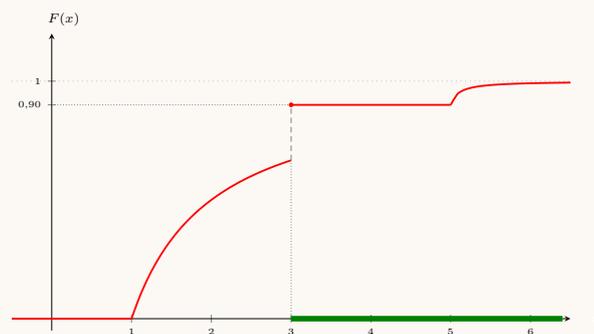
$$x_{0,85} = \inf \{x : F(x) \geq 0,85\} = 3.$$



b) Si ricavi il valore del nono decile di \tilde{X} .

Con l'ausilio del grafico sottostante si vede facilmente che il più piccolo valore di x dove la funzione di ripartizione $F(x)$ è almeno pari a 0,9 è ancora dato da

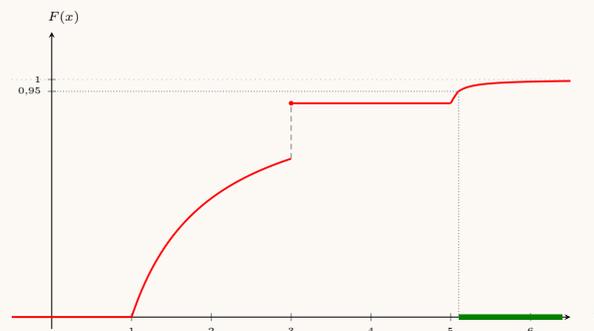
$$x_{0,90} = \inf \{x : F(x) \geq 0,90\} = 3.$$



c) Si ricavi il valore del 95-esimo percentile \tilde{X} .

Con l'ausilio del grafico sottostante si vede facilmente che il più piccolo valore di x dove la funzione di ripartizione $F(x)$ è almeno pari a 0,95 è dato da

$$x_{0,95} = \inf \{x : F(x) \geq 0,95\} = 5,1.$$



Si noti che $x_{0,95} = 5,1$ è l'unico valore di x che soddisfa l'equazione

$$F(x) = 0,95,$$

ovvero l'unico valore di x tale che

$$1 - \frac{1}{100x - 490} = 0,95.$$

Il prossimo esercizio illustra come si ricavano i **quantili** di una **distribuzione discreta** a partire dalla sua **funzione di massa di probabilità**.

Esercizio 4.11 (Quantili di una distribuzione discreta). Si consideri una distribuzione discreta con funzione di probabilità data da

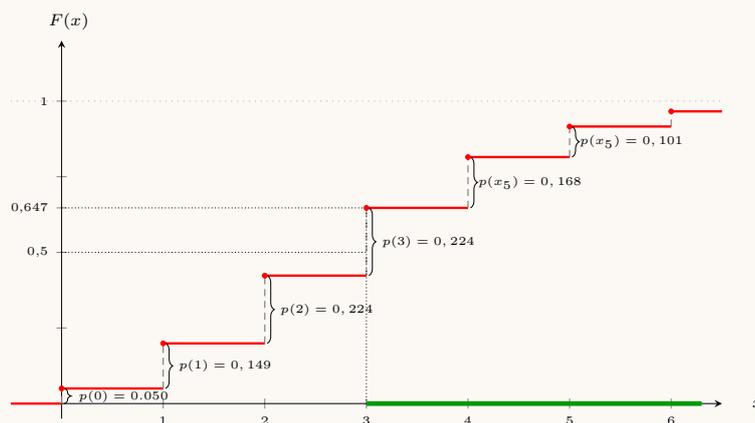
$$p(x) = \begin{cases} \frac{e^{-3} \times 3^x}{x!} & \text{per } x = 0, 1, 2, \dots \\ 0 & \text{altrimenti.} \end{cases}$$

Si ricavino la mediana e il quantile di ordine $p = 0.647$.

Risposta:

Siccome le **funzioni di ripartizione discrete** sono funzioni a *gradini* con andamento orizzontale nei tratti compresi tra i loro punti di discontinuità (ovvero nei tratti compresi tra i punti appartenenti al loro supporto), i **quantili** di una distribuzione discreta possono essere ottenuti **cumulando** i valori che assume la sua **funzione di massa di probabilità** in corrispondenza dei punti del suo supporto.

x	$p(x)$	$F(x)$
0	0,050	0,050
1	0,149	0,199
2	0,224	0,423
3	0,224	0,647
4	0,168	0,815
5	0,101	0,916
6	0,050	0,966
\vdots	\vdots	\vdots



Come si desume dalla tabella (o anche dal grafico), il più piccolo valore di x dove la funzione di ripartizione raggiunge o supera 0,5 è dato da

$$\text{mediana} = x_{0,5} = \inf \{x : F(x) \geq 0,5\} = 3$$

così come è uguale a 3 il valore del quantile di ordine $p = 0,647$, ovvero il valore di

$$x_{0,647} = \inf \{x : F(x) \geq 0,647\} = 3.$$

Esercizio 4.12 (Quantili di una distribuzione continua). Si calcoli il nono decile della distribuzione

$$F(x) = \begin{cases} 0 & \text{per } x < 0 \\ 1 - e^{-x} & \text{per } x \geq 0. \end{cases}$$

Risposta:

Per definizione, il nono decile della distribuzione $F(x)$ è dato da

$$x_{0,9} = \inf\{x : F(x) \geq 0,9\}.$$

Siccome per $x < 0$ si ha $F(x) = 0 < 0,90$, possiamo concludere che i valori di x dove $F(x) \geq 0,9$ sono i valori di x che soddisfano la disuguaglianza

$$1 - e^{-x} \geq 0,90.$$

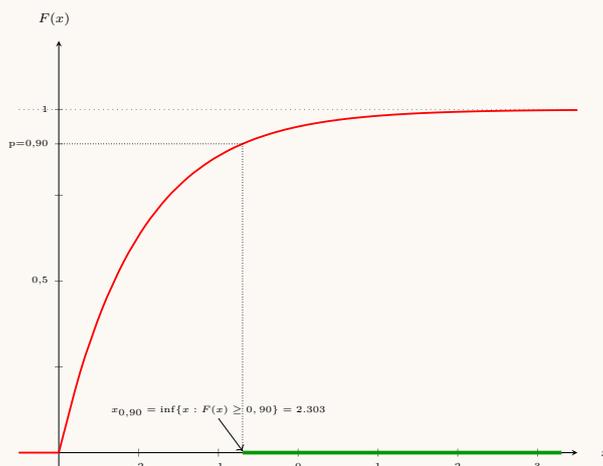
Dopo alcuni semplici passaggi vediamo che questa disuguaglianza è soddisfatta se e solo se

$$x \geq -\ln(1 - 0,9) = 2,303.$$

Il nono decile (ovvero il più piccolo valore di x dove $F(x) \geq 0,9$) è dunque dato da

$$x_{0,90} = 2,303.$$

Il grafico sottostante aiuta a comprendere il ragionamento che ci ha condotto al valore di $x_{0,90} = 2,303$.



Il prossimo esempio illustra un'applicazione dei quantili in ambito finanziario.

Esempio 4.3 (Valore a rischio). In ambito finanziario, il rischio associato ad un portafoglio di attività finanziarie viene spesso misurato attraverso opportuni **quantili** di variabili casuali che descrivono le perdite che potrebbero verificarsi in uno o più istanti temporali futuri. Questi quantili vengono chiamati **”valore a rischio”**. Più precisamente, se \tilde{X} è una variabile casuale che descrive la perdita (espressa in una valuta come il dollaro statunitense, l’euro, ecc.) che potrebbe verificarsi su un determinato portafoglio di attività finanziarie al termine di un determinato **periodo di mantenimento** h , allora il **valore a rischio con livello di confidenza p e periodo di mantenimento h** è definito come il quantile di ordine p della variabile casuale \tilde{X} :

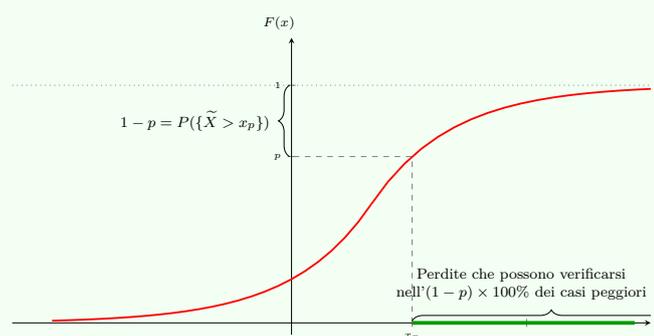
$$VaR_p = x_p = \inf \{x : F(x) \geq p\} \quad (29)$$

(come al solito $F(x)$ indica la funzione di ripartizione di \tilde{X}).

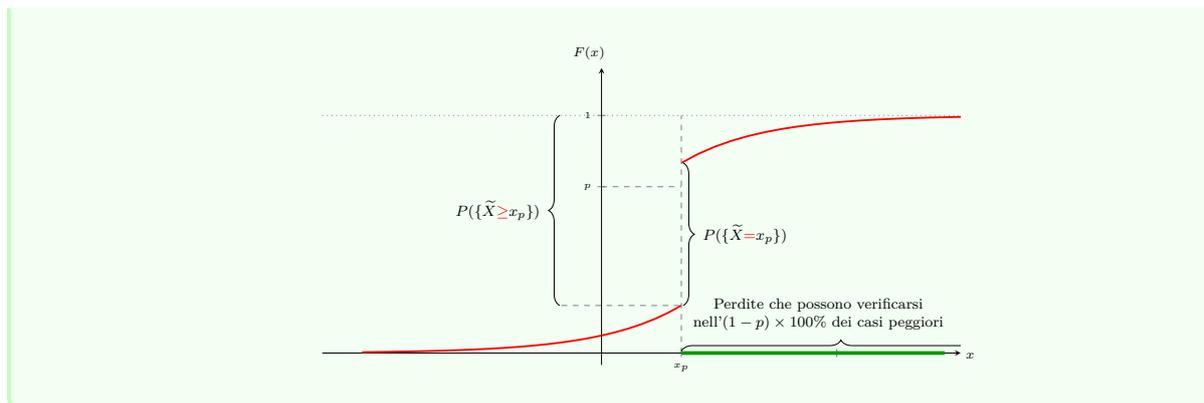
Nelle applicazioni il **livello di confidenza** viene solitamente posto uguale a $p = 0.95$ oppure a $p = 0.99$, e come **periodo di mantenimento** si considera spesso un periodo di $h = 1$ giorno, $h = 10$ giorni o anche di $h = 2$ settimane.

Dalla formula (29) si desume immediatamente che il **valore a rischio con livello di confidenza p** è la **più piccola perdita x che non viene superata con probabilità almeno uguale a p** . Pertanto è importante tenere presente che ...

- ... di solito (vedi il grafico sottostante) nell' $(1 - p) \times 100\%$ dei casi peggiori la perdita effettiva è strettamente maggiore del valore a rischio con livello di confidenza p ...



- ... e che a volte può capitare che la probabilità di incorrere in una perdita che è esattamente **uguale** al valore a rischio x_p sia strettamente positiva e in questi casi la probabilità di incorrere in una perdita che è **maggiore o uguale** al valore a rischio potrebbe essere molto più grande di $1 - p$:



4.5 Distribuzioni congiunte

Finora abbiamo considerato sempre soltanto eventi che sono definiti attraverso una **singola** variabile casuale, ma nelle applicazioni si considerano spesso eventi che sono definiti attraverso **due o più** variabili casuali, come per esempio eventi del tipo

$$\{\omega \in \Omega : (\tilde{X}_1(\omega), \tilde{X}_2(\omega)) \in B\} \quad \text{dove } B \subseteq \mathbb{R}^2.$$

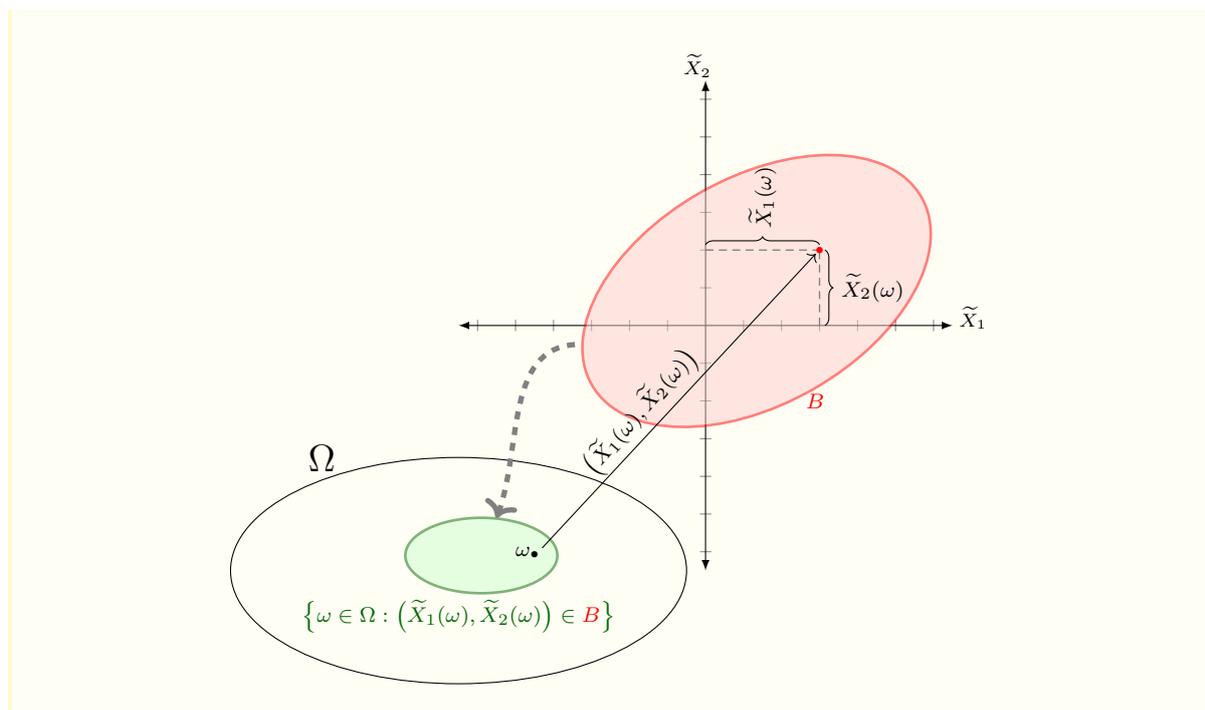
Si noti che un evento di questo tipo è composto da tutti gli eventi elementari $\omega \in \Omega$ per i quali il vettore composto dalle realizzazioni delle due variabili casuali \tilde{X}_1 e \tilde{X}_2 appartiene ad un dato sottoinsieme B del piano \mathbb{R}^2 (vedi la Figura 4.5). Per comodità di notazione, d'ora in poi indicheremo eventi di questo tipo semplicemente con $\{(\tilde{X}_1, \tilde{X}_2) \in B\}$. Ovviamente, con riferimento ad un qualsiasi numero finito k di variabili casuali \tilde{X}_i si possono anche definire eventi del tipo

$$\{\omega \in \Omega : (\tilde{X}_1(\omega), \tilde{X}_2(\omega), \dots, \tilde{X}_k(\omega)) \in B\}, \quad \text{dove } B \subseteq \mathbb{R}^k$$

che, per comodità di notazione, indicheremo semplicemente con

$$\{(\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_k) \in B\}.$$

Figura 4.5. Il grafico illustra come attraverso una coppia di variabili casuali si possono definire degli eventi.



Con riferimento ad una singola variabile casuale \tilde{X} , abbiamo visto che la sua [funzione di ripartizione](#)

$$F(x) = P(\{\tilde{X} \leq x\}),$$

se è definita per ogni $x \in \mathbb{R}$, ha un ruolo molto importante perché

- determina univocamente le probabilità di tutti gli eventi che appartengono alla [\$\sigma\$ -algebra generata dalla variabile casuale \$\tilde{X}\$](#) (Teorema 4.1),
- e perché attraverso la scelta di una opportuna funzione di ripartizione si possono dunque **assegnare** delle probabilità a tutti gli eventi che appartengono alla [\$\sigma\$ -algebra generata dalla variabile casuale \$\tilde{X}\$](#) in modo tale che gli assiomi di Kolmogorov siano soddisfatti (Teorema 4.4).

In questa sezione vedremo che con riferimento ad un generico numero finito k di variabili casuali, la cosiddetta **funzione di ripartizione "congiunta" (o "multivariata")**

$$F(x_1, x_2, \dots, x_k) = P\left(\bigcap_{i=1}^k \{\tilde{X}_i \leq x_i\}\right) \quad (30)$$

ha un ruolo analogo a quello della [funzione di ripartizione "marginale" \(o "univariata"\)](#) di una singola variabile casuale. Prima di procedere dobbiamo però chiarire due questioni preliminari:

- In primo luogo dobbiamo chiarire che la [funzione di ripartizione congiunta di \$k\$ variabili casuali](#) è definita per ogni $(x_1, x_2, \dots, x_k) \in \mathbb{R}^k$ se e solo se tutte e k le variabili casuali sono [misurabili rispetto alla \$\sigma\$ -algebra \$\mathcal{A}\$](#) che è il dominio della

funzione di probabilità $P(\cdot)$ che compare al secondo membro della (30), ovvero se e solo se anche le **funzioni di ripartizione "marginali"** di tutte e k le variabili casuali sono definite per ogni $x \in \mathbb{R}$ (vedi la Definizione 4.2). Questa affermazione può essere facilmente dimostrata facendo riferimento al Lemma 4.1 e alle proprietà delle σ -algebre (omettiamo i dettagli della dimostrazione).

- In secondo luogo dobbiamo chiarire che se la **funzione di ripartizione congiunta di k variabili casuali** è definita per ogni $(x_1, x_2, \dots, x_k) \in \mathbb{R}^k$, allora le probabilità di tutti gli eventi del tipo

$$\{(\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_k) \in B\} \quad \text{con } B \in \sigma(\mathcal{B}_k) \quad (31)$$

devono essere definite e viceversa. In altre parole, la **funzione di ripartizione congiunta di k variabili casuali** è definita per ogni $(x_1, x_2, \dots, x_k) \in \mathbb{R}^k$ se e solo se tutti gli eventi (31) appartengono alla σ -algebra \mathcal{A} che è il dominio della funzione di probabilità $P(\cdot)$ che compare al secondo membro della (30). Ricordiamo che $\sigma(\mathcal{B}_k)$ indica la σ -algebra di Borel su \mathbb{R}^k e che questa σ -algebra contiene tutti i sottoinsiemi di \mathbb{R}^k che potrebbero avere una qualche rilevanza pratica nelle applicazioni (vedi l'Esempio 1.6). La σ -algebra che contiene tutti gli eventi (31) viene chiamata **σ -algebra generata dalle k variabili casuali $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_k$** e viene indicata con il simbolo $\sigma(\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_k)$. Nel linguaggio formale della matematica può essere definita come

$$\sigma(\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_k) = \left\{ \{(\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_k) \in B\} : B \in \sigma(\mathcal{B}_k) \right\}.$$

A questo punto possiamo finalmente descrivere in che senso il ruolo di una **funzione di ripartizione congiunta** è analogo a quello della **funzioni di ripartizione** di una singola variabile casuale. A tal fine enunceremo delle generalizzazioni dei Teoremi 4.1, 4.2 e 4.4 che riguardano la funzione di ripartizione di una singola variabile casuale. Cominciamo con la generalizzazione del Teorema 4.1: essa ci dice che la **funzione di ripartizione congiunta** di k variabili casuali **determina univocamente** le probabilità di tutti gli eventi che appartengono alla **σ -algebra generata dalle k variabili casuali di riferimento**.

Teorema 4.5. Se $(\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_k)$ e $(\tilde{Y}_1, \tilde{Y}_2, \dots, \tilde{Y}_k)$ sono due vettori di variabili casuali che hanno la stessa funzione di ripartizione congiunta,^a allora per ogni $B \in \sigma(\mathcal{B}_k)$ la probabilità dell'evento

$$\{(\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_k) \in B\}$$

è identica alla probabilità dell'evento

$$\{(\tilde{Y}_1, \tilde{Y}_2, \dots, \tilde{Y}_k) \in B\}.$$

In virtù di questo fatto, la funzione di ripartizione congiunta di k variabili casuali viene spesso anche chiamata **distribuzione congiunta**.

^aI due vettori di variabili casuali potrebbero essere anche definiti nell'ambito di due modelli probabilistici diversi.

Come abbiamo anticipato, l'enunciato del Teorema 4.5 può essere riformulato dicendo che la funzione di ripartizione congiunta di k variabili casuali **determina univocamente** le probabilità di **tutti** gli eventi che appartengono alla σ -algebra generata dalle variabili casuali di riferimento. Questo fatto significa che per qualunque $B \in \sigma(\mathcal{B}_k)$ è possibile, almeno in teoria, "ricavare" il corrispondente valore di

$$P\left(\{(\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_k) \in B\}\right) \quad (32)$$

a partire dalla **funzione di ripartizione congiunta** delle variabili casuali $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_k$.

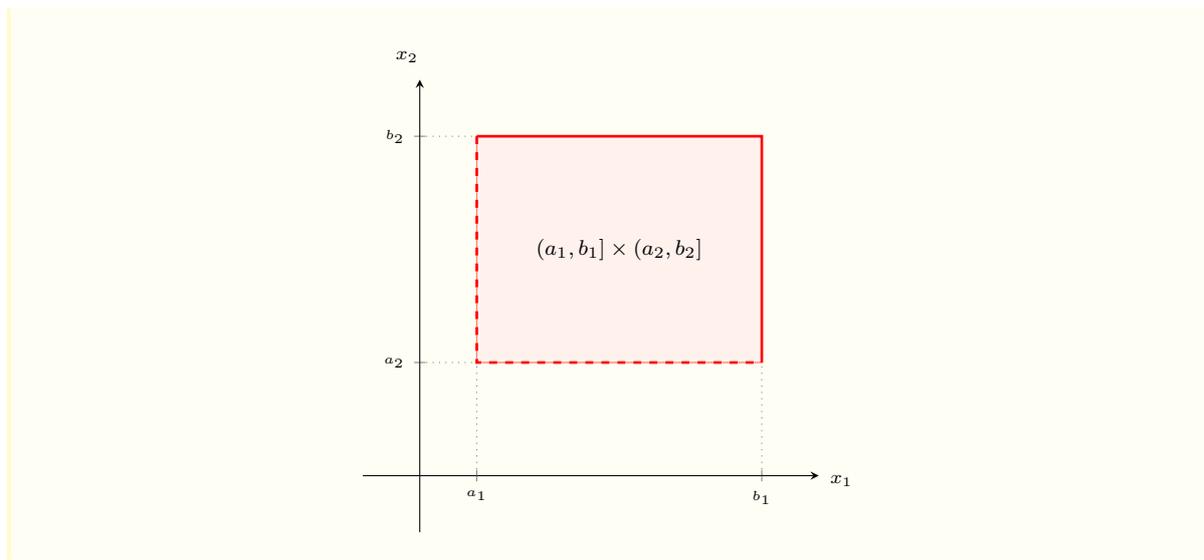
Com'è facile immaginare, a seconda di come è definito l'insieme $B \in \sigma(\mathcal{B}_k)$, potrebbe essere molto difficile ricavare il valore della probabilità (32) a partire dalla **funzione di ripartizione congiunta** delle k variabili casuali coinvolte. Nelle prossime righe mostreremo come si risolve questo problema in alcuni casi particolari molto importanti. Cominceremo con il caso dove $k = 2$ e dove B è un prodotto cartesiano del tipo $(a_1, b_1] \times (a_2, b_2]$ (vedi Figura 4.6). Siccome i punti $(x_1, x_2) \in \mathbb{R}^2$ che appartengono ad un tale prodotto cartesiano sono quelli per i quali $a_1 < x_1 \leq b_1$ e $a_2 < x_2 \leq b_2$, l'evento

$$\{(\tilde{X}_1, \tilde{X}_2) \in B\} \quad \text{con } B = (a_1, b_1] \times (a_2, b_2]$$

può essere espresso come

$$\{a_1 < \tilde{X}_1 \leq b_1\} \cap \{a_2 < \tilde{X}_2 \leq b_2\} = \bigcap_{i=1}^2 \{a_i < \tilde{X}_i \leq b_i\}.$$

Figura 4.6. Il grafico mostra l'area di un rettangolo semiaperto identificato da un prodotto cartesiano del tipo $(a_1, b_1] \times (a_2, b_2]$



Seguendo la sequenza di grafici in Figura 4.7, non è difficile rendersi conto che il valore di

$$P(\{a_1 < \tilde{X}_1 \leq b_1\} \cap \{a_2 < \tilde{X}_1 \leq b_2\}) = P\left(\bigcap_{i=1}^2 \{a_i < \tilde{X}_i \leq b_i\}\right)$$

può essere calcolato come

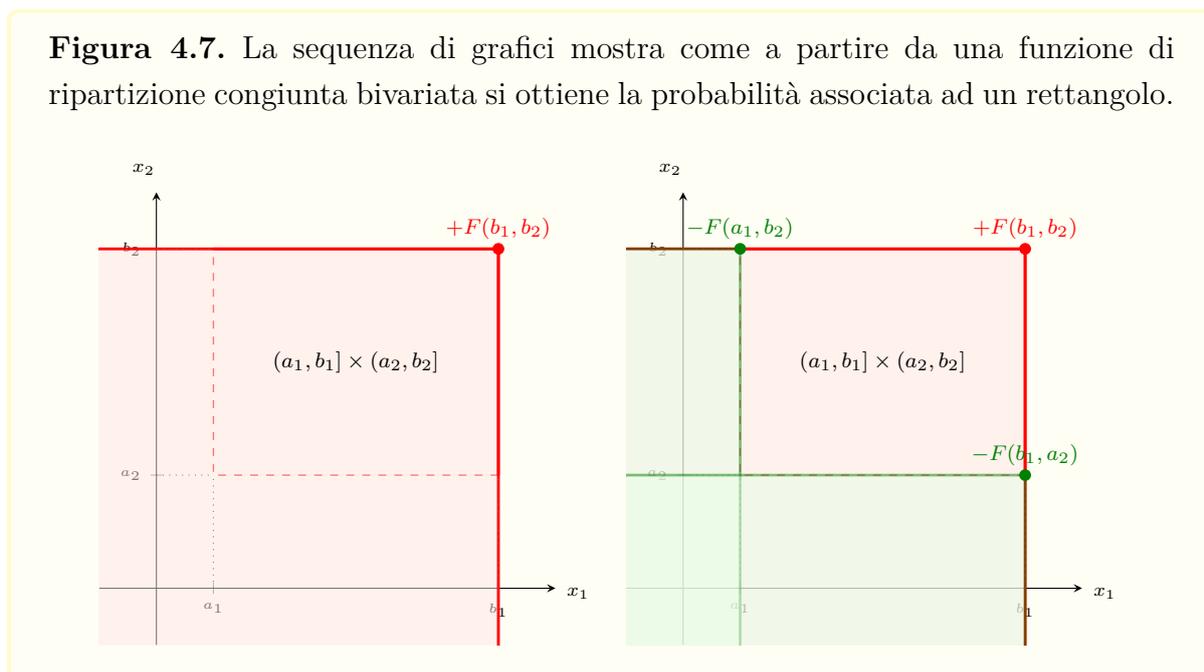
$$P\left(\bigcap_{i=1}^2 \{a_i < \tilde{X}_i \leq b_i\}\right) = F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2) \quad (33)$$

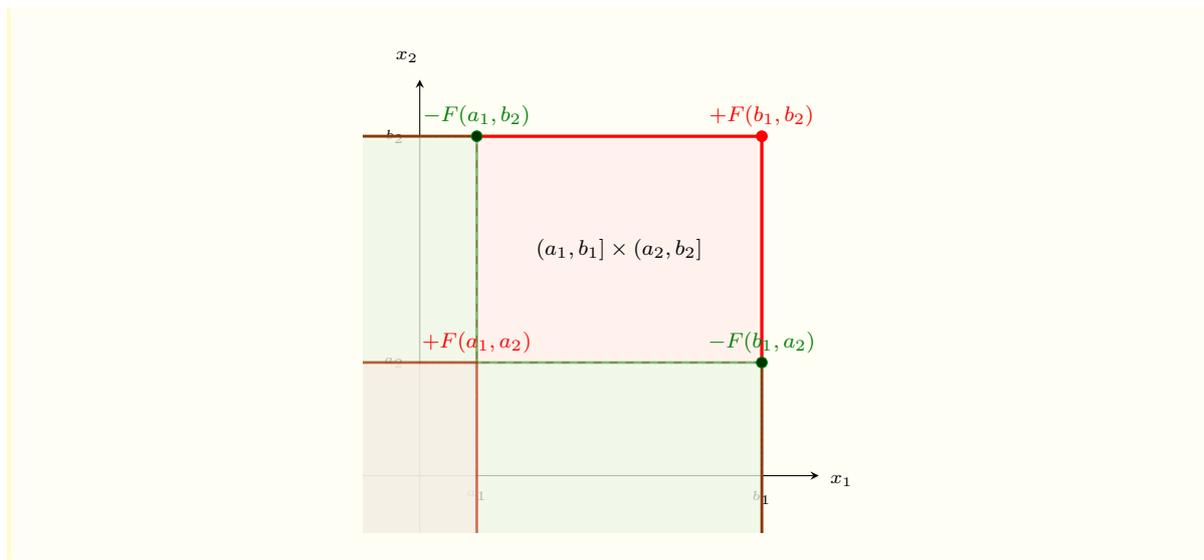
(si notino i segni meno associati a $F(a_1, b_2)$ e a $F(b_1, a_2)$), dove

$$F(x_1, x_2) = P(\{\tilde{X}_1 \leq x_1\} \cap \{\tilde{X}_2 \leq x_2\}), \quad (x_1, x_2) \in \mathbb{R}^2,$$

è la funzione di ripartizione congiunta delle variabili casuali \tilde{X}_1 e \tilde{X}_2 .

Figura 4.7. La sequenza di grafici mostra come a partire da una funzione di ripartizione congiunta bivariata si ottiene la probabilità associata ad un rettangolo.





Consideriamo ora invece un vettore $(\tilde{X}_1, \tilde{X}_2, \tilde{X}_3)$ che è composto da $k = 3$ variabili casuali. Usando la legge L6, si può dimostrare (omettiamo i dettagli della dimostrazione) che nel caso di $k = 3$ variabili casuali, la formula analoga alla (33) è data da

$$P\left(\bigcap_{i=1}^3 \{a_i < \tilde{X}_i \leq b_i\}\right) = +F(b_1, b_2, b_3) - F(a_1, b_2, b_3) - F(b_1, a_2, b_3) - F(b_1, b_2, a_3) + \\ + F(a_1, a_2, b_3) + F(a_1, b_2, a_3) + F(b_1, a_2, a_3) - F(a_1, a_2, a_3),$$

e usando sempre ancora la legge L6, si può dimostrare che la versione generale della formula (33), ovvero la versione che si riferisce ad un generico numero k di variabili casuali, è data da

$$P\left(\bigcap_{i=1}^k \{a_i < \tilde{X}_i \leq b_i\}\right) = \\ = +F(b_1, b_2, \dots, b_k) \\ \quad \text{[[tutti gli argomenti sono del tipo } b_i \text{]]} \\ - F(a_1, b_2, \dots, b_k) - F(b_1, a_2, \dots, b_k) - F(b_1, b_2, \dots, a_k) \\ \quad \text{[[1 argomento è del tipo } a_i \text{ e tutti gli altri sono del tipo } b_i \text{]]} \\ \qquad \qquad \qquad (34) \\ + F(a_1, a_2, \dots, b_k) + F(a_1, b_2, a_3, \dots, b_k) + F(b_1, b_2, \dots, a_{k-1}, a_k) \\ \quad \text{[[2 argomenti sono del tipo } a_i \text{ e tutti gli altri sono del tipo } b_i \text{]]} \\ \vdots \\ + (-1)^k F(a_1, a_2, \dots, a_k) \\ \quad \text{[[tutti gli argomenti sono del tipo } a_i \text{, nessuno è del tipo } b_i \text{]]}$$

(si tenga presente che $(-1)^k$ è uguale a $+1$ oppure a -1 a seconda se il numero di variabili k è pari o dispari). Questa formula generale può essere espressa in modo più conciso come

$$P\left(\bigcap_{i=1}^k \{a_i < \tilde{X}_i \leq b_i\}\right) = \sum_{j=0}^k (-1)^j \times \left[\begin{array}{l} \text{somma di tutti i valori di } F(x_1, x_2, \dots, x_k) \\ \text{per i quali } j \text{ degli argomenti } x_i \text{ sono uguali a } a_i, \\ \text{e tutti gli altri argomenti } x_i \text{ sono uguali a } b_i. \end{array} \right]. \quad (35)$$

La formula (35) è molto importante perché ci permette di identificare un'importante proprietà che tutte le **funzioni di ripartizione congiunte** devono soddisfare: siccome per l'assioma **K1** la probabilità di un evento non può mai essere negativa, l'equazione (35) implica che tutte le **funzioni di ripartizione congiunte** che si riferiscono a k variabili casuali devono necessariamente soddisfare la condizione

F1_k) per ogni insieme di k coppie ordinate (a_i, b_i) tali che $a_i < b_i$ si ha

$$\sum_{j=0}^k (-1)^j \times \left[\begin{array}{l} \text{somma di tutti i valori di } F(x_1, x_2, \dots, x_k) \\ \text{per i quali } j \text{ degli argomenti } x_i \text{ sono uguali a } a_i, \\ \text{e tutti gli altri argomenti } x_i \text{ sono uguali a } b_i. \end{array} \right] \geq 0.$$

Non è difficile verificare che con $k = 1$ la proprietà **F1_k** si riduce alla proprietà **F1** che tutte le funzioni di ripartizione univariate devono soddisfare. L'analogia con le funzioni di ripartizione univariate non finisce qui. Usando gli assiomi di Kolmogorov e le conseguenti leggi del calcolo delle probabilità si può infatti dimostrare che tutte le **funzioni di ripartizione congiunte** che si riferiscono a k variabili casuali devono anche soddisfare le due condizioni ...

F2_k) $F(x_1, x_2, \dots, x_k)$ è continua da destra in ciascuna delle k variabili x_1, x_2, \dots, x_k , ovvero

$$\lim_{h \rightarrow 0^+} F(x_1, x_2, \dots, x_i + h, \dots, x_k) = F(x_1, x_2, \dots, x_i, \dots, x_k)$$

per ogni $i = 1, 2, \dots, k$.

F3_k)

$$\lim_{x_1 \rightarrow \infty} \lim_{x_2 \rightarrow \infty} \cdots \lim_{x_k \rightarrow \infty} F(x_1, x_2, \dots, x_k) = 1$$

e

$$\lim_{x_i \rightarrow -\infty} F(x_1, x_2, \dots, x_i, \dots, x_k) = 0 \quad \text{per ogni } i = 1, 2, \dots, k.$$

Non è difficile verificare che anche queste ultime due proprietà sono delle generalizzazioni delle proprietà **F2** e **F3** che tutte le funzioni di ripartizione univariate devono soddisfare.¹⁹ Per dare il giusto risalto a queste considerazioni, le raccoglieremo in un apposito teorema che generalizza il Teorema 4.2 sulle proprietà delle funzioni di ripartizione univariate.

¹⁹A proposito delle condizioni **F0_k** - **F3_k**, conviene sciogliere un dubbio che potrebbe essere sorto in qualche lettore. Infatti, a prima vista potrebbe sembrare che la condizione **F1_k** sia equivalente alla condizione

Teorema 4.6 (Proprietà delle funzioni di ripartizione congiunte). La funzione di ripartizione congiunta di k variabili casuali è sempre una funzione $F(x_1, x_2, \dots, x_k)$ che ...

$F0_k$) ... ha per dominio l'insieme di tutti i vettori (x_1, x_2, \dots, x_k) che sono composti da k numeri reali (ovvero l'insieme \mathbb{R}^k)

e che soddisfa le tre proprietà $F1_k$, $F2_k$ e $F3_k$.

A questo punto manca soltanto più la generalizzazione del Teorema 4.4:

Teorema 4.7 (Teorema di esistenza per funzioni di ripartizione congiunte). Se k è un qualunque numero intero positivo e $F(x_1, x_2, \dots, x_k)$ è una qualunque funzione che soddisfa le quattro proprietà $F0_k - F3_k$, allora

- a) esiste un'unica funzione di probabilità $P(\cdot)$ che
- ha per dominio la σ -algebra $\sigma(\mathcal{B}_k)$,
 - che soddisfa gli assiomi di Kolmogorov
 - e che inoltre soddisfa la condizione

$$P((-\infty, x_1] \times (-\infty, x_2] \times \dots \times (-\infty, x_k]) = F(x_1, x_2, \dots, x_k)$$

per ogni $(x_1, x_2, \dots, x_k) \in \mathbb{R}^k$.

Di conseguenza,

- b) esistono un modello probabilistico (Ω, \mathcal{A}, P) e k corrispondenti variabili casuali

$$\tilde{X}_1 : \Omega \mapsto \mathbb{R}, \quad \tilde{X}_2 : \Omega \mapsto \mathbb{R}, \quad \dots, \quad \tilde{X}_k : \Omega \mapsto \mathbb{R}$$

tali che $F(x_1, x_2, \dots, x_k)$ è la loro **funzione di ripartizione congiunta**, ovvero tali che

$$P\left(\bigcap_{i=1}^k \{\tilde{X}_i \leq x_i\}\right) = F(x_1, x_2, \dots, x_k) \quad \text{per ogni } (x_1, x_2, \dots, x_k) \in \mathbb{R}^k$$

In virtù di questi risultati, tutte le funzioni $F(x_1, x_2, \dots, x_k)$ che soddisfano le tre proprietà $F0_k - F3_k$ vengono chiamate **"funzione di ripartizione (o distribuzione) congiunta (o multivariata)"** anche senza riferimento a specifiche variabili casuali. Se per chiarezza si vuole specificare anche la dimensione

$F1_k^*$) $F(x_1, x_2, \dots, x_k)$ è una funzione non decrescente di ciascuna delle k variabili x_1, x_2, \dots, x_k ,

o quantomeno che **in presenza delle condizioni $F0_k$, $F2_k$ e $F3_k$** le condizioni $F1_k$ e $F1_k^*$ siano equivalenti. Entrambe queste congetture sono false: infatti, si può dimostrare che la condizione $F1_k$ è più restrittiva della condizione $F1_k^*$ e che esistono delle funzioni $F(x_1, x_2, \dots, x_k)$ che soddisfano le condizioni $F0_k$, $F2_k$ e $F3_k$ e anche la condizione $F1_k^*$, ma non la condizione $F1_k$.

del dominio (ovvero il valore di k), si può dire **”funzione di ripartizione (o distribuzione) univariata/bivariata/trivariata/ecc”**.

Commento: Si noti che anche il Teorema 4.7 può essere interpretato come un **teorema di estensione**. Infatti, la prima parte della conclusione ci dice che le probabilità

$$P((-\infty, x_1] \times (-\infty, x_2] \times \cdots \times (-\infty, x_k]) = F(x_1, x_2, \dots, x_k)$$

possono essere **estese in un unico modo** ad una funzione di probabilità $P(\cdot)$ che soddisfa gli assiomi di Kolmogorov e che ha per dominio la σ -algebra di Borel $\sigma(\mathcal{B}_k)$.

Per completare il quadro conviene osservare che quando consideriamo uno spazio campionario Ω specifico e delle specifiche variabili casuali

$$\tilde{X}_1 : \Omega \mapsto \mathbb{R}, \quad \tilde{X}_2 : \Omega \mapsto \mathbb{R}, \quad \dots, \quad \tilde{X}_k : \Omega \mapsto \mathbb{R},$$

la **funzione di ripartizione congiunta** delle k variabili casuali \tilde{X}_i non può sempre essere una qualunque funzione $F(x_1, x_2, \dots, x_k)$ che soddisfa le quattro proprietà **F0_k - F3_k**, ma a volte potrebbe essere necessario che soddisfi anche delle condizioni aggiuntive che hanno lo scopo di prevenire l’insorgere di **”conflitti di assegnazione”**. Infatti, così come abbiamo già visto nel caso di una singola variabile casuale, se in \mathbb{R}^k esistono due punti (x_1, x_2, \dots, x_k) e (y_1, y_2, \dots, y_k) tali che

$$\bigcap_{i=1}^k \{\omega \in \Omega : \tilde{X}_i(\omega) \leq x_i\} = \bigcap_{i=1}^k \{\omega \in \Omega : \tilde{X}_i(\omega) \leq y_i\},$$

allora, in aggiunta alle proprietà **F0_k - F3_k**, la funzione di ripartizione congiunta delle k variabili casuali \tilde{X}_i deve soddisfare anche la condizione

$$F(x_1, x_2, \dots, x_k) = F(y_1, y_2, \dots, y_k)$$

perché altrimenti si avrebbe un **”conflitto di assegnazione”**. Se, d’altra parte, la corrispondenza

$$(x_1, x_2, \dots, x_k) \mapsto \bigcap_{i=1}^k \{\omega \in \Omega : \tilde{X}_i(\omega) \leq x_i\} \quad (36)$$

è **iniettiva** (ovvero se ad ogni coppia di vettori $(x_1, x_2, \dots, x_k) \neq (y_1, y_2, \dots, y_k)$ corrispondono due eventi $\bigcap_{i=1}^k \{\omega \in \Omega : \tilde{X}_i(\omega) \leq x_i\}$ e $\bigcap_{i=1}^k \{\omega \in \Omega : \tilde{X}_i(\omega) \leq y_i\}$ che sono diversi), allora come **funzione di ripartizione congiunta** delle k variabili casuali \tilde{X}_i si può effettivamente scegliere una qualunque funzione $F(x_1, x_2, \dots, x_k)$ che soddisfa le

tre proprietà $F0_k - F3_k$ senza che correre il rischio di introdurre delle contraddizioni con gli assiomi di Kolmogorov, dei conflitti di assegnazione o qualunque altro tipo di contraddizione.

Come nel caso di una singola variabile casuale, la corrispondenza (36) è iniettiva se e solo se per ogni $(x_1, x_2, \dots, x_k) \in \mathbb{R}^k$ esiste almeno un $\omega \in \Omega$ tale che

$$(\tilde{X}_1(\omega), \tilde{X}_2(\omega), \dots, \tilde{X}_k(\omega)) = (x_1, x_2, \dots, x_k),$$

ovvero se e solo se la corrispondenza

$$\omega \in \Omega \quad \mapsto \quad (\tilde{X}_1(\omega), \tilde{X}_2(\omega), \dots, \tilde{X}_k(\omega)) \in \mathbb{R}^k$$

è **suriettiva** (la dimostrazione della parte "se" di questa affermazione è quasi immediata; la parte "solo se" può essere facilmente dimostrata attraverso un ragionamento per assurdo).

4.6 Distribuzioni marginali associate ad una distribuzione congiunta

Supponiamo ora di conoscere la **funzione di ripartizione congiunta** bivariata

$$F(x_1, x_2) = P\left(\{\tilde{X}_1 \leq x_1\} \cap \{\tilde{X}_2 \leq x_2\}\right) \quad \text{per } (x_1, x_2) \in \mathbb{R}^2$$

e di voler ricavare, a partire da $F(x_1, x_2)$, le **funzioni di ripartizione marginali** delle variabili casuali \tilde{X}_1 e \tilde{X}_2 . Nella dimostrazione sottostante vedremo che queste ultime possono essere ottenute attraverso le formule

$$\begin{aligned} F_1(x_1) &= P\left(\{\tilde{X}_1 \leq x_1\}\right) = \lim_{x_2 \rightarrow \infty} F(x_1, x_2) \quad \text{per ogni } x_1 \in \mathbb{R} \\ F_2(x_2) &= P\left(\{\tilde{X}_2 \leq x_2\}\right) = \lim_{x_1 \rightarrow \infty} F(x_1, x_2) \quad \text{per ogni } x_2 \in \mathbb{R}. \end{aligned} \tag{37}$$

Dimostrazione (Dimostrazione delle formule (37)). Si noti in primo luogo che l'evento $\{\tilde{X}_1 \leq x_1\}$ può essere espresso come

$$\{\tilde{X}_1 \leq x_1\} = \bigcup_{n=1}^{\infty} \left[\{\tilde{X}_1 \leq x_1\} \cap \{\tilde{X}_2 \leq x_{2,n}\} \right]$$

dove $x_{2,n}$ è una qualunque successione di numeri reali che diverge a $+\infty$ (vedi il primo grafico in Figura 4.8). Siccome gli eventi

$$\{\tilde{X}_1 \leq x_1\} \cap \{\tilde{X}_2 \leq x_{2,n}\}, \quad n = 1, 2, \dots$$

formano una successione di eventi non decrescente, possiamo applicare la legge L10 onde concludere che

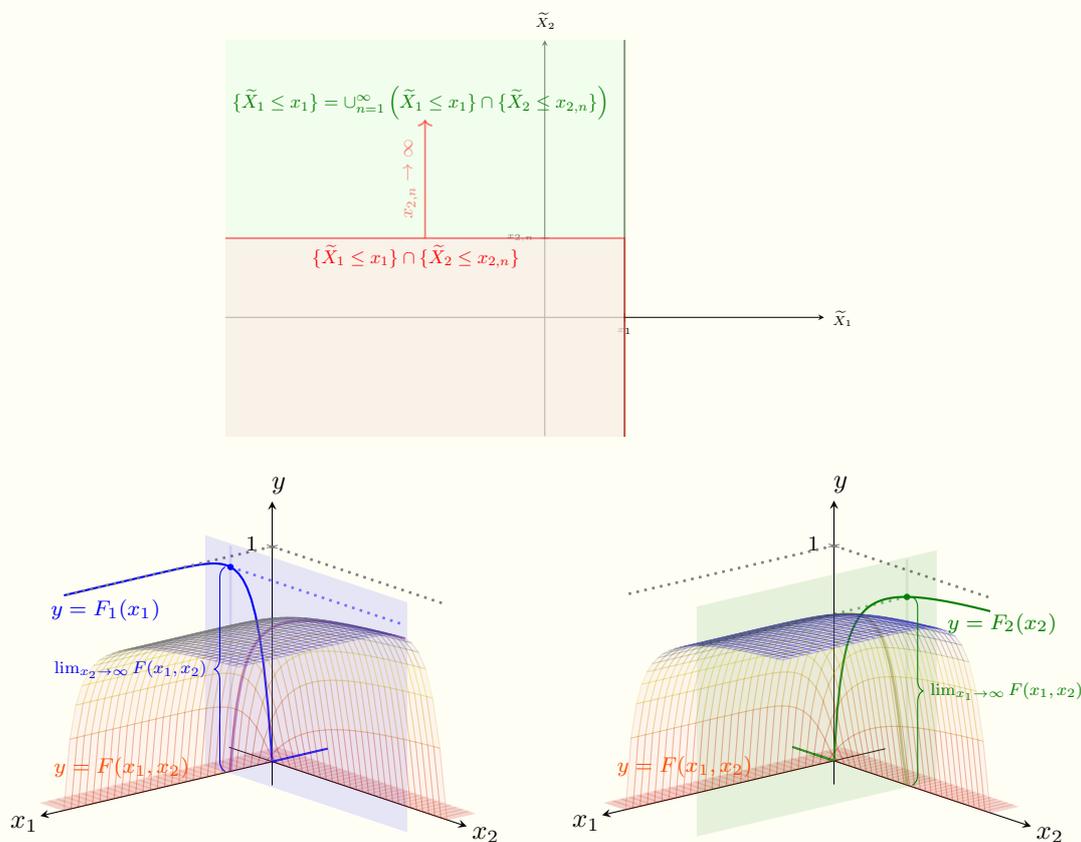
$$\begin{aligned} F_1(x_1) &= P\left(\{\tilde{X}_1 \leq x_1\}\right) \\ &= P\left(\bigcup_{n=1}^{\infty} \left[\{\tilde{X}_1 \leq x_1\} \cap \{\tilde{X}_2 \leq x_{2,n}\} \right]\right) \\ &= \lim_{n \rightarrow \infty} P\left(\{\tilde{X}_1 \leq x_1\} \cap \{\tilde{X}_2 \leq x_{2,n}\}\right) \\ &= \lim_{n \rightarrow \infty} F(x_1, x_{2,n}) = \lim_{x_2 \rightarrow \infty} F(x_1, x_2). \end{aligned}$$

Questo ragionamento dimostra dunque la prima delle due formule (37), ovvero la formula

$$F_1(x_1) = P\left(\{\tilde{X}_1 \leq x_1\}\right) = \lim_{x_2 \rightarrow \infty} F(x_1, x_2) \quad \text{per ogni } x_1 \in \mathbb{R}.$$

Scambiando i ruoli delle variabili casuali \tilde{X}_1 e \tilde{X}_2 si ottiene anche la formula per ricavare la funzione di ripartizione marginale della variabile casuale \tilde{X}_2 . □

Figura 4.8. Il primo grafico mostra che l'evento $\{\tilde{X}_1 \leq x_1\}$ può essere interpretato come unione della successione di eventi $\{\tilde{X}_1 \leq x_1\} \cap \{\tilde{X}_2 \leq x_{2,n}\}$ per $n = 1, 2, \dots$. Gli altri due grafici mostrano come a partire da una funzione di ripartizione congiunta bivariata si ricavano le due corrispondenti funzioni di ripartizione marginali.



Generalizzando il ragionamento che abbiamo visto nella precedente dimostrazione, si può anche dimostrare che se

$$F(x_1, x_2, \dots, x_k) = P\left(\bigcap_{i=1}^k \{\tilde{X}_i \leq x_i\}\right) \quad \text{per } (x_1, x_2, \dots, x_k) \in \mathbb{R}^k$$

è la **funzione di ripartizione congiunta** di k variabili casuali (k può essere qualunque numero maggiore o uguale a 2), allora

$$\begin{aligned} F_{12\dots(k-1)}(x_1, x_2, \dots, x_{k-1}) &= P\left(\bigcap_{i=1}^{k-1} \{\tilde{X}_i \leq x_i\}\right) = \\ &= \lim_{x_k \rightarrow \infty} F(x_1, x_2, \dots, x_{k-1}, x_k) \quad \text{per ogni } (x_1, x_2, \dots, x_{k-1}) \in \mathbb{R}^{k-1} \end{aligned} \quad (38)$$

deve essere la **funzione di ripartizione congiunta** delle prime $k-1$ variabili casuali (ovvero delle variabili casuali \tilde{X}_i con pedice i diverso da k). Scambiando il ruolo della variabile casuale \tilde{X}_k con quello di una qualunque altra variabile casuale \tilde{X}_j con pedice j diverso da k , si vede quindi che la funzione di ripartizione congiunta delle $k-1$ variabili casuali \tilde{X}_i con pedice i diverso da j deve essere data da

$$\begin{aligned} F_{12\dots(j-1)(j+1)\dots k}(x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_k) &= P\left(\bigcap_{i:1 \leq i \leq k \text{ e } i \neq j} \{\tilde{X}_i \leq x_i\}\right) = \\ &= \lim_{x_j \rightarrow \infty} F(x_1, x_2, \dots, x_j, \dots, x_k) \quad \text{per ogni } (x_1, \dots, x_{j-1}, x_{j+1}, x_k) \in \mathbb{R}^{k-1} \end{aligned}$$

(in parole: la funzione di ripartizione congiunta delle k variabili casuali \tilde{X}_i con pedice i diverso da j è uguale a limite per x_j che tende a $+\infty$ della funzione di ripartizione congiunta di tutte le k variabili casuali).

Esercizio 4.13 (Funzioni di ripartizione congiunte). Siano \tilde{X}_1 e \tilde{X}_2 due variabili casuali con funzione di ripartizione congiunta data dalla funzione

$$F(x_1, x_2) = \begin{cases} 0 & \text{se } x_1 \leq 0 \text{ oppure } x_2 \leq 0, \\ 1 - e^{-x_1} + \frac{e^{-x_1(x_2+1)}}{x_2+1} - \frac{1}{x_2+1} & \text{se } x_1 > 0 \text{ e } x_2 > 0 \end{cases}$$

[[Questa funzione di ripartizione congiunta è quella rappresentata negli ultimi due grafici in Figura 4.8.]]

- Si calcoli la probabilità dell'evento $\{1 < \tilde{X}_1 \leq 2\} \cap \{0,5 < \tilde{X}_2 \leq 1\}$.
- Si ricavino le funzioni di ripartizione marginali delle variabili casuali \tilde{X}_1 e \tilde{X}_2 .

Risposte:

- Si calcoli la probabilità dell'evento $\{1 < \tilde{X}_1 \leq 2\} \cap \{0,5 < \tilde{X}_2 \leq 1\}$.

Usando la formula (33) vediamo che la probabilità richiesta è data da

$$\begin{aligned} P\left(\{1 < \tilde{X}_1 \leq 2\} \cap \{0,5 < \tilde{X}_2 \leq 1\}\right) &= \\ &= +F(2, 1) - F(1, 1) - F(2, 0,5) + F(1, 0,5). \end{aligned}$$

Siccome

$$\begin{aligned} F(2, 1) &= 1 - e^{-x_1} + \frac{e^{-x_1(x_2+1)}}{x_2+1} - \frac{1}{x_2+1} \\ &= 1 - e^{-2} + \frac{e^{-2(1+1)}}{1+1} - \frac{1}{1+1} = 0,374, \end{aligned}$$

$$F(1, 1) = \dots = 0,200, \quad F(2, 0,5) = \dots = 0,231,$$

e

$$F(1, 0,5) = \dots = 0,114,$$

possiamo concludere che

$$\begin{aligned} P\left(\{1 < \tilde{X}_1 \leq 2\} \cap \{0,5 < \tilde{X}_2 \leq 1\}\right) &= +0,374 - 0,200 - 0,231 + 0,114 \\ &= 0,057. \end{aligned}$$

b) Si ricavino le funzioni di ripartizione marginali delle variabili casuali \tilde{X}_1 e \tilde{X}_2 .

Usando le due formule (37) vediamo che le funzioni di ripartizione marginali delle variabili casuali \tilde{X}_1 e \tilde{X}_2 sono rispettivamente date da

$$\begin{aligned} F_1(x_1) &= P\left(\{\tilde{X}_1 \leq x_1\}\right) = \lim_{x_2 \rightarrow \infty} F(x_1, x_2) = \\ &= \begin{cases} 0 & \text{se } x_1 \leq 0 \\ \lim_{x_2 \rightarrow \infty} 1 - e^{-x_1} + \frac{e^{-x_1(x_2+1)}}{x_2+1} - \frac{1}{x_2+1} = 1 - e^{-x_1} & \text{se } x_1 > 0. \end{cases} \end{aligned}$$

e da

$$\begin{aligned} F_2(x_2) &= P\left(\{\tilde{X}_2 \leq x_2\}\right) = \lim_{x_1 \rightarrow \infty} F(x_1, x_2) = \\ &= \begin{cases} 0 & \text{se } x_2 \leq 0 \\ \lim_{x_1 \rightarrow \infty} 1 - e^{-x_1} + \frac{e^{-x_1(x_2+1)}}{x_2+1} - \frac{1}{x_2+1} = 1 - \frac{1}{x_2+1} & \text{se } x_2 > 0. \end{cases} \end{aligned}$$

Il secondo e terzo grafico in Figura 4.8 mostrano le due funzioni di ripartizione marginali che abbiamo appena ricavato.

4.7 Distribuzioni congiunte discrete e assolutamente continue

Nella Sezione 4.5 abbiamo visto che la **funzione di ripartizione congiunta** di k variabili casuali **determina univocamente** del probabilità di tutti gli eventi che appartengono alla σ -algebra generata da tali variabili casuali (Teorema 4.5) e che per **assegnare** delle probabilità che soddisfano gli assiomi di Kolmogorov a tutti gli eventi che appartengono alla σ -algebra generata da un numero finito k di variabili casuali è sufficiente specificare la loro **funzione di ripartizione congiunta** (vedi il Teorema 4.7 e le successive osservazio-

ni). Tuttavia, facendo riferimento alla **formula** che definisce una **funzione di ripartizione congiunta bivariata** è difficile capire su quali sottoinsieme di \mathbb{R}^2 viene concentrata molta o poca probabilità (si consideri per esempio la formula che definisce la funzione di ripartizione bivariata dell'Esercizio 4.13) e l'analogo problema si presenta anche con riferimento al **grafico** di una **funzione di ripartizione congiunta** bivariata (si consideri, per esempio, il grafico della funzione di ripartizione congiunta in Figura 4.8). Com'è facile immaginare, il problema di cui stiamo parlando si presenta in forma ancora più grave se consideriamo una **formula** che definisce una **funzione di ripartizione congiunta multivariata** che si riferisce a più di due variabili casuali. Si pensi inoltre alla fase di assegnazione delle probabilità: assegnare delle probabilità pensando al grafico di una **funzione di ripartizione congiunta** bivariata è molto complicato perché, come abbiamo visto, tutte le funzioni di ripartizione congiunte devono soddisfare le quattro proprietà $F0_k - F3_k$ e dal punto di vista grafico non è affatto semplice stabilire se queste proprietà sono soddisfatte o meno. A causa di questi problemi, nelle applicazioni le **funzioni di ripartizione congiunte** vengono di solito descritte e/o specificate soltanto in modo indiretto attraverso ...

- ... una **funzione di massa di probabilità "congiunta"**, ...
- ... una **funzione di densità "congiunta"** ...
- ... oppure una funzione di probabilità/densità **"marginale"** e una o più corrispondenti **funzioni di probabilità/densità "condizionate"**.

Nelle prossime due sezioni vedremo che cosa si intende per una "funzione di massa di probabilità congiunta" e per una "funzione di densità congiunta". Nella sezione intitolata "Distribuzioni condizionate" faremo invece qualche breve cenno sulle "funzioni di probabilità/densità condizionate" e vedremo come attraverso questi due tipi di funzione si può specificare una **distribuzione congiunta** (ricordiamo che i termini "distribuzione congiunta" e "funzione di ripartizione congiunta" sono sinonimi - vedi il Teorema 4.5).

4.7.1 Funzioni di massa di probabilità congiunte

Definizione 4.5 (Funzione di massa di probabilità congiunta). Una **funzione di massa di probabilità congiunta** (o semplicemente **funzione di probabilità congiunta**) è una qualunque funzione $p(x_1, x_2, \dots, x_k)$ che soddisfa le seguenti proprietà:

p0_k) Il dominio di $p(x_1, x_2, \dots, x_k)$ è dato dall'insieme di tutti i vettori (x_1, x_2, \dots, x_k) che sono composti da k numeri reali.

p1_k) $0 \leq p(x_1, x_2, \dots, x_k) \leq 1$ per ogni $(x_1, x_2, \dots, x_k) \in \mathbb{R}^k$ e l'insieme

$$S = \{(x_1, x_2, \dots, x_k) \in \mathbb{R}^k : p(x_1, x_2, \dots, x_k) > 0\}$$

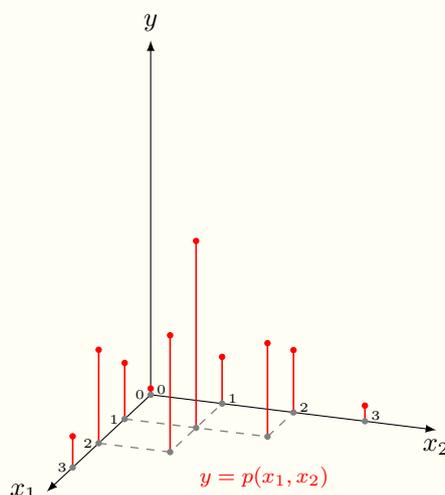
è un sottoinsieme finito o al più infinito numerabile di \mathbb{R}^k .

$$p^{(2)} \sum_{(x_1, x_2, \dots, x_k) \in S} p(x_1, x_2, \dots, x_k) = 1$$

Come nel caso di una **funzione di massa di probabilità** che si riferisce ad una singola variabile casuale, l'insieme S viene chiamato **supporto**.

Nel caso in cui il numero di dimensioni k è uguale a 2 una **funzione di massa di probabilità congiunta** può essere rappresentata come nel grafico in Figura 4.9.

Figura 4.9. Il grafico mostra una **funzione di massa di probabilità congiunta** bivarziata con **supporto finito**. La lunghezza di una asticella rappresenta la probabilità di osservare la coppia di realizzazioni (x_1, x_2) che identifica il corrispondente punto grigio sul piano x_1x_2 .



Le **funzione di massa di probabilità congiunte** hanno un ruolo analogo a quello delle **funzioni di massa di probabilità "univariate"** che abbiamo introdotto nella Definizione 4.3. Infatti, si può dimostrare che:

- 1) Ad ogni **funzione di massa di probabilità congiunta** $p(x_1, x_2, \dots, x_k)$ corrisponde una **funzione di ripartizione congiunta** che è definita come

$$F(x_1, x_2, \dots, x_k) = \sum_{\substack{(t_1, \dots, t_k) \in S : \\ t_1 \leq x_1, \dots, t_k \leq x_k}} p(t_1, t_2, \dots, t_k) \quad \text{per } (x_1, x_2, \dots, x_k) \in \mathbb{R}^k, \quad (39)$$

(la sommatoria è estesa a tutti i valori $p(t_1, t_2, \dots, t_k)$ tali che (t_1, t_2, \dots, t_k) appartiene al supporto S e tali che $t_1 \leq x_1, t_2 \leq x_2, \dots$ e $t_k \leq x_k$). Le **funzioni di ripartizione congiunte** di questo tipo vengono chiamate **"discrete"**.

- 2) Se $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_k$ sono k variabili casuali e se la loro **funzione di ripartizione congiunta** è discreta come quella definita nella (39), allora

a) ciascuna delle k variabili casuali \tilde{X}_i deve essere **discreta** (ovvero la funzione di ripartizione marginale di ciascuna delle k variabili casuali \tilde{X}_i deve essere **discreta**);

b)

$$P\left(\bigcap_{i=1}^k \{\tilde{X}_i = x_i\}\right) = p(x_1, x_2, \dots, x_k) \quad \text{per ogni } (x_1, x_2, \dots, x_k) \in \mathbb{R}^k$$

c) e

$$P\left(\{(\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_k) \in S\}\right) = 1.$$

3) Viceversa, se $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_k$ sono k variabili casuali **discrete** (tutte definite nell'ambito di uno stesso modello probabilistico), allora anche la loro **funzione di ripartizione congiunta** deve essere necessariamente **discreta**.

Dalle precedenti osservazioni si evince che l'interpretazione di una **funzione di massa di probabilità congiunta** è molto semplice e immediata: il valore di $p(x_1, x_2, \dots, x_k)$ ci dice qual è la probabilità di osservare il vettore di realizzazioni (x_1, x_2, \dots, x_k) e questa probabilità può essere strettamente positiva solo su un sottoinsieme finito o al più infinito numerabile di \mathbb{R}^k .

Supponiamo ora che \tilde{X}_1 e \tilde{X}_2 siano due variabili casuali **discrete** (entrambe definite nell'ambito di uno stesso modello probabilistico) e supponiamo di conoscere la loro **funzione di probabilità congiunta**

$$p(x_1, x_2) = P\left(\{\tilde{X}_1 = x_1\} \cap \{\tilde{X}_2 = x_2\}\right) \quad \text{per } (x_1, x_2) \in \mathbb{R}^2.$$

A questo punto ci chiediamo: come possiamo ricavare le **funzioni di massa di probabilità "marginali"** delle due variabili casuali \tilde{X}_1 e \tilde{X}_2 ? Per rispondere a questa domanda conviene esaminare i due grafici in Figura 4.10. Esaminando il primo grafico non è difficile rendersi conto che la funzione di probabilità marginale della variabile casuale \tilde{X}_1 deve essere data da (omettiamo i dettagli di una dimostrazione rigorosa)

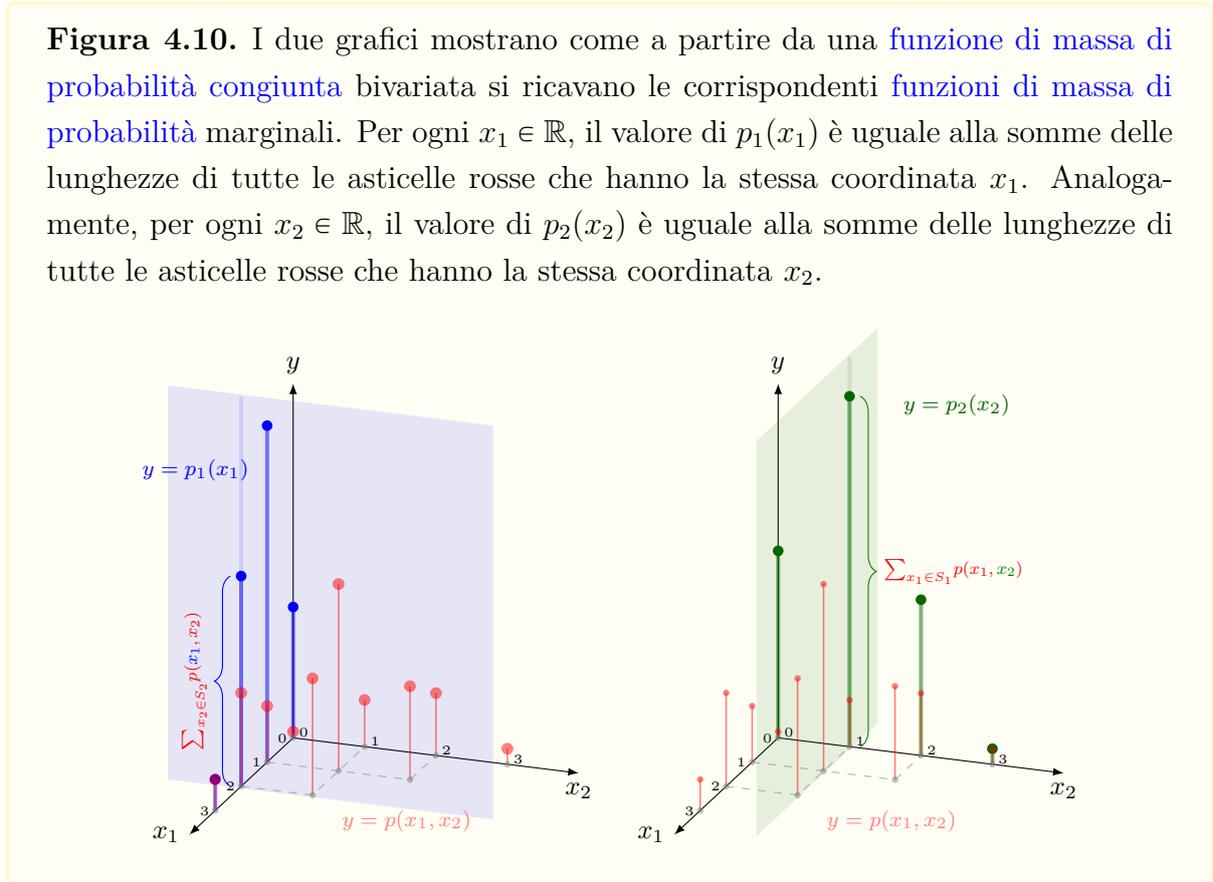
$$p_1(x_1) = P(\{\tilde{X}_1 = x_1\}) = \sum_{x_2 \in S_2} p(x_1, x_2), \quad x_1 \in \mathbb{R}, \quad (40)$$

dove S_2 è il **supporto** della variabile casuale \tilde{X}_2 , ovvero l'insieme di tutti i valori x_2 per i quali la probabilità $P(\{\tilde{X}_2 = x_2\})$ è strettamente positiva (come si vede nel primo grafico in Figura 4.10, alcuni termini della sommatoria $\sum_{x_2 \in S_2} p(x_1, x_2)$ potrebbero essere nulli). Allo stesso modo (si veda il secondo grafico in Figura 4.10), non è difficile rendersi conto che la funzione di probabilità marginale della variabile casuale \tilde{X}_2 deve essere data da

$$p_2(x_2) = P(\{\tilde{X}_2 = x_2\}) = \sum_{x_1 \in S_1} p(x_1, x_2), \quad x_2 \in \mathbb{R}, \quad (41)$$

dove S_1 è il **supporto** della variabile casuale \tilde{X}_1 .

Figura 4.10. I due grafici mostrano come a partire da una **funzione di massa di probabilità congiunta** bivariata si ricavano le corrispondenti **funzioni di massa di probabilità** marginali. Per ogni $x_1 \in \mathbb{R}$, il valore di $p_1(x_1)$ è uguale alla somme delle lunghezze di tutte le asticelle rosse che hanno la stessa coordinata x_1 . Analogamente, per ogni $x_2 \in \mathbb{R}$, il valore di $p_2(x_2)$ è uguale alla somme delle lunghezze di tutte le asticelle rosse che hanno la stessa coordinata x_2 .



Consideriamo ora invece un generico numero $k \geq 2$ di variabili casuali **discrete** e supponiamo che la funzione

$$p(x_1, x_2, \dots, x_k) = P\left(\bigcap_{i=1}^k \{\tilde{X}_i = x_i\}\right) \quad \text{per } (x_1, x_2, \dots, x_k) \in \mathbb{R}^k,$$

sia la loro **funzione di massa di probabilità congiunta**. Anche in questo caso non è difficile dimostrare (omettiamo i dettagli della dimostrazione) che la **funzione di massa di probabilità congiunta** delle variabili casuali \tilde{X}_i con pedice i diverso da un dato pedice j deve essere data da

$$\begin{aligned} p_{12\dots(j-1)(j+1)\dots k}(x_1, x_2, \dots, x_{j-1}, x_{j+1}, \dots, x_k) &= P\left(\bigcap_{1 \leq i \leq k; i \neq j} \{\tilde{X}_i = x_i\}\right) = \\ &= \sum_{x_j \in S_j} p(x_1, x_2, \dots, x_j, \dots, x_k) \quad \text{per } (x_1, x_2, \dots, x_{j-1}, x_{j+1}, \dots, x_k) \in \mathbb{R}^{k-1}, \end{aligned}$$

dove S_j indica il **supporto** della variabile casuale \tilde{X}_j . Applicando ripetutamente questa formula si vede che la **funzione di massa di probabilità** marginale di una qualunque delle k variabili casuali \tilde{X}_i può essere ricavata attraverso la formula

$$\begin{aligned} p_i(x_i) &= P(\{\tilde{X}_i = x_i\}) \\ &= \sum_{x_1 \in S_1} \cdots \sum_{x_{i-1} \in S_{i-1}} \sum_{x_{i+1} \in S_{i+1}} \cdots \sum_{x_k \in S_k} p(x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_k) \quad (42) \\ &\text{per } x_i \in \mathbb{R}. \end{aligned}$$

Esercizio 4.14. Si consideri un'urna che contiene 15 palline di cui 6 sono di colore bianco, 5 di colore grigio e 4 di colore nero. Dall'urna vengono estratte in blocco 3 palline. Sia \tilde{X}_1 la variabile casuale che conta il numero di palline bianche estratte e \tilde{X}_2 la variabile casuale che conta il numero di palline grigie estratte.

- Si ricavi la funzione di massa di probabilità congiunta delle due variabili casuali \tilde{X}_1 e \tilde{X}_2
- Si ricavino le funzioni di massa di probabilità marginali delle variabili casuali \tilde{X}_1 e \tilde{X}_2 .
- Qual è la probabilità che la differenza tra il numero di palline bianche e il numero di palline grigie estratte sia positiva?

Risposte:

Nelle risposte assumeremo che le variabili casuali \tilde{X}_1 e \tilde{X}_2 siano definite nell'ambito del modello probabilistico (Ω, \mathcal{A}, P) con

- spazio campionario Ω uguale all'insieme di tutte le $\#\Omega = \binom{15}{3} = 455$ **combinazioni** di tre palline che si possono ottenere nell'estrazione in blocco,
- $\mathcal{A} = \mathcal{P}(\Omega)$
- e con funzione di probabilità $P(\cdot)$ definita secondo il metodo di assegnazione classico:

$$P(A) = \frac{\#A}{\#\Omega} \quad \text{per ogni } A \subseteq \Omega.$$

- Si ricavi la funzione di massa di probabilità congiunta delle due variabili casuali \tilde{X}_1 e \tilde{X}_2

Ovviamente, le variabili casuali \tilde{X}_1 e \tilde{X}_2 possono assumere soltanto i valori $x = 0$, $x = 1$, $x = 2$ e $x = 3$ e pertanto la loro **funzione di massa di probabilità congiunta**

$$p(x_1, x_2) = P\left(\left\{\tilde{X}_1 = x_1\right\} \cap \left\{\tilde{X}_2 = x_2\right\}\right)$$

deve essere nulla in ogni punto (x_1, x_2) dove x_1 e/o x_2 sono diversi da 0, 1, 2 e 3. Nei rimanenti punti $(x_1, x_2) \in \mathbb{R}^2$ la **funzione di massa di probabilità congiunta** è invece data da

$$p(x_1, x_2) = \frac{\binom{6}{x_1} \times \binom{5}{x_2} \times \binom{4}{3-x_1-x_2}}{\binom{15}{3}},$$

perché per ottenere tutti possibili casi favorevoli all'evento secondo il quale x_1 delle tre palline estratte sono bianche e x_2 delle tre palline estratte sono nere bisogna combinare ...

- ... tutte le $\binom{6}{x_1}$ possibilità per scegliere x_1 palline bianche tra le 6 palline bianche presenti nell'urna ...
- ... con tutte le $\binom{5}{x_2}$ possibilità per scegliere x_2 palline grigie tra le 5 palline grigie presenti nell'urna ...
- ... e ciascuna delle precedenti possibilità per scegliere x_1 palline bianche e x_2 palline grigie deve essere combinata con tutte le $\binom{4}{3-x_1-x_2}$ possibilità per scegliere $3-x_1-x_2$ palline nere tra le quattro palline nere presenti nell'urna (se $3-x_1-x_2 < 0$, conveniamo che $\binom{4}{3-x_1-x_2} = 0$ e in questo caso il valore di $p(x_1, x_2)$ è dunque nullo).

In corrispondenza di $x_1 = 1$ e $x_2 = 2$, per esempio, la funzione di probabilità congiunta è data da

$$\begin{aligned} p(1, 2) &= \frac{\binom{6}{1} \times \binom{5}{2} \times \binom{4}{3-1-2}}{\binom{15}{3}} \\ &= \frac{6 \times 10 \times 1}{455} = \frac{60}{455} = 0,1319. \end{aligned}$$

La tabella sottostante riporta anche i valori della **funzione di massa di probabilità congiunta** che corrispondono a tutti gli altri punti (x_1, x_2) del suo **supporto** (ovviamente, in corrispondenza di tutti gli altri punto è nulla):

$p(x_1, x_2)$	$x_2 = 0$	$x_2 = 1$	$x_2 = 2$	$x_2 = 3$	$p_1(x_1)$
$x_1 = 0$	0,0088	0,0659	0,0879	0,0220	0,1846
$x_1 = 1$	0,0791	0,2637	0,1319	0,0000	0,4747
$x_1 = 2$	0,1319	0,1648	0,0000	0,0000	0,2967
$x_1 = 3$	0,0440	0,0000	0,0000	0,0000	0,0440
$p_2(x_2)$	0,2638	0,4944	0,2198	0,0220	1,0000

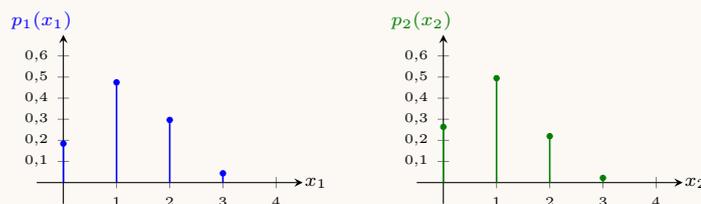
La **funzione di massa di probabilità congiunta** che abbiamo appena ricavato è quella rappresentata nel grafico in Figura 4.9. Non è difficile verificare che questa funzione soddisfa le tre proprietà $p_{0_{k=2}} - p_{2_{k=2}}$ della Definizione 4.5.

- b) Si ricavano le funzioni di massa di probabilità marginali delle variabili casuali \tilde{X}_1 e \tilde{X}_2 .

I valori non nulli delle due **funzioni di massa di probabilità marginali**

$$p(x_1) = P\left(\left\{\tilde{X}_1 = x_1\right\}\right) \quad \text{e} \quad p(x_2) = P\left(\left\{\tilde{X}_2 = x_2\right\}\right)$$

si trovano, rispettivamente, nella riga e nella colonna marginale della precedente tabella. I due grafici in Figura 4.10 mostrano anche le due **funzioni di massa di probabilità marginali**. Ovviamente, le **funzioni di massa di probabilità marginali** possono essere anche rappresentate in modo separato all'interno di un piano cartesiano così come nei due grafici sottostanti.



- c) Qual è la probabilità che la differenza tra il numero di palline bianche e il numero di palline grigie estratte sia positiva?

Per calcolare la probabilità dell'evento $\{\tilde{X}_1 - \tilde{X}_2 > 0\}$ basta sommare i valori delle probabilità congiunte

$$p(x_1, x_2) = P\left(\{\tilde{X}_1 = x_1\} \cap \{\tilde{X}_2 = x_2\}\right)$$

che sono positive e che corrispondono a punti (x_1, x_2) con $x_1 > x_2$ (legge L2). Le probabilità da sommare sono evidenziate in rosso nella tabella sottostante:

$p(x_1, x_2)$	$x_2 = 0$	$x_2 = 1$	$x_2 = 2$	$x_2 = 3$	$p_1(x_1)$
$x_1 = 0$	0,0088	0,0659	0,0879	0,0220	0,1846
$x_1 = 1$	0,0791	0,2637	0,1319	0,0000	0,4747
$x_1 = 2$	0,1319	0,1648	0,0000	0,0000	0,2967
$x_1 = 3$	0,0440	0,0000	0,0000	0,0000	0,0440
$p_2(x_2)$	0,2638	0,4944	0,2198	0,0220	1,0000

La probabilità richiesta è dunque data da

$$\begin{aligned} P\left(\{\tilde{X}_1 - \tilde{X}_2 > 0\}\right) &= \mathbf{0,0791} + \mathbf{0,1319} + \mathbf{0,0440} + \mathbf{0,1648} \\ &= 0,4198. \end{aligned}$$

4.7.2 Funzioni di densità congiunte

Definizione 4.6 (Funzione di densità congiunta). Una **funzione di densità congiunta** è una qualunque funzione $f(x_1, x_2, \dots, x_k)$ che soddisfa tutte e tre le seguenti proprietà:

f0_k) Il dominio di $f(x_1, x_2, \dots, x_k)$ è dato dall'insieme di tutti i vettori (x_1, x_2, \dots, x_k) che sono composti da k numeri reali.

f1_k) $f(x_1, x_2, \dots, x_k) \geq 0$ per ogni $(x_1, x_2, \dots, x_k) \in \mathbb{R}^k$.

f2_k) $f(x_1, x_2, \dots, x_k)$ è una funzione integrabile e

$$\int \int \cdots \int_{\mathbb{R}^k} f(x_1, x_2, \dots, x_k) dx_1 dx_2 \cdots dx_k = 1.^a$$

L'insieme $S = \{(x_1, x_2, \dots, x_k) \in \mathbb{R}^k : f(x_1, x_2, \dots, x_k) > 0\}$ che contiene tutti i punti dove una funzione di densità congiunta è positiva viene chiamato **supporto** della funzione di densità congiunta.

^aTutti gli integrali multipli che compariranno in questa dispensa possono essere interpretati come integrali multipli di Riemann. In questo caso il termine "funzione integrabile" significa "funzione Riemann-integrabile". Altrimenti, volendo interpretare gli integrali multipli come integrali di Lebesgue, il termine "funzione integrabile" può essere interpretato come "funzione misurabile". Siccome le funzioni che incontriamo nelle applicazioni sono sempre Riemann-integrabili (se limitate a sottoinsiemi compatti di \mathbb{R}^k), e a maggior ragione sono quindi anche misurabili, **in questa dispensa non ci preoccuperemo mai di verificare se una data funzione $f(x_1, x_2, \dots, x_k)$ è effettivamente "integrabile" e daremo sempre per scontato che ciò sia vero.**

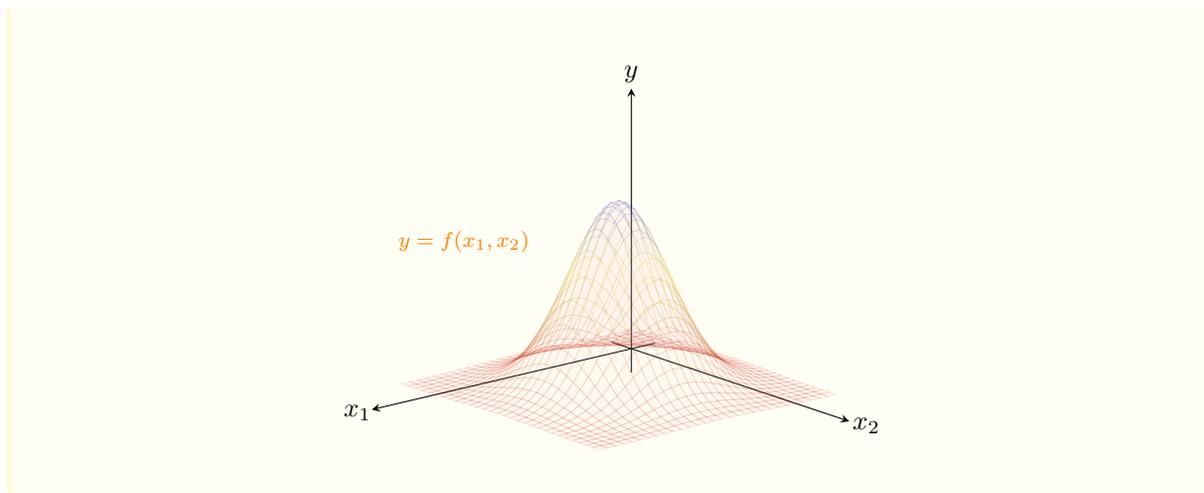
Con riferimento agli integrali multipli che incontreremo in questa dispensa è importante tenere presente che se la funzione integranda $g(x_1, x_2, \dots, x_k)$ è non negativa (ovvero se $g(x_1, x_2, \dots, x_k) \geq 0$ per ogni $(x_1, x_2, \dots, x_k) \in \mathbb{R}^k$), allora un integrale multiplo del tipo

$$\int_{a_k}^{b_k} \cdots \int_{a_2}^{b_2} \int_{a_1}^{b_1} g(x_1, x_2, \dots, x_k) dx_1 dx_2 \cdots dx_k$$

- può sempre essere calcolato come integrale iterato, ovvero risolvendo gli integrali uno alla volta dall'interno verso esterno;
- l'ordine in cui si presentano gli integrali può essere modificato a piacimento senza che ciò incida sul valore dell'integrale.

Nel caso in cui il numero di dimensioni k è uguale a 2, una **funzione di densità congiunta** può essere rappresentata come nel grafico in Figura 4.11.

Figura 4.11. Il grafico mostra una **funzione di densità congiunta** bivariata che è continua in ogni punto $(x_1, x_2) \in \mathbb{R}^2$. Si osservi che in virtù della proprietà f2_k il **volume** compreso tra il piano $x_1 x_2$ e la superficie che rappresenta la funzione di densità congiunta deve essere pari a 1.



Le **funzione di densità congiunte** hanno un ruolo analogo a quello delle **funzioni di densità** "univariate" che si riferiscono ad una singola variabile casuale (vedi la Definizione 4.4). Infatti, si può dimostrare che ad ogni **funzione di densità congiunta** $f(x_1, x_2, \dots, x_k)$ corrisponde una **funzione di ripartizione congiunta** che è definita come

$$F(x_1, x_2, \dots, x_k) = \int_{-\infty}^{x_k} \cdots \int_{-\infty}^{x_2} \int_{-\infty}^{x_1} f(t_1, t_2, \dots, t_k) dt_1 dt_2 \cdots dt_k \quad (43)$$

per $(x_1, x_2, \dots, x_k) \in \mathbb{R}^k$.

Come nel caso univariato, anche le **funzione di ripartizione congiunte** che possono essere rappresentate come integrale di una funzione di densità congiunta vengono chiamate **"assolutamente continue"**. Ovviamente, ad ogni **funzione di densità congiunta** l'equazione (43) associa un'**unica funzione di ripartizione assolutamente continua**, ma è importante tenere presente che anche nel caso multivariato esistono sempre **infinite funzioni di densità congiunte** $f(x_1, x_2, \dots, x_k)$ che attraverso la formula (43) danno luogo esattamente alla stessa funzione di ripartizione $F(x_1, x_2, \dots, x_k)$. Infatti, anche nel caso multivariato è sempre possibile modificare una **funzione di densità congiunta** in un numero finito di punti (o anche su determinati sottoinsiemi infiniti di \mathbb{R}^k) senza che la funzione di ripartizione $F(x_1, x_2, \dots, x_k)$ definita nella (43) subisca alcuna variazione. Come nel caso univariato, tutte le **funzioni di densità congiunte** che danno luogo alla medesima **funzione di ripartizione congiunta** formano una "classe di equivalenza" e le singole **funzioni di densità congiunte** vengono chiamate "versioni" della funzione di densità della **funzione di ripartizione congiunta** di riferimento. Si noti che anche nel caso multivariato ogni classe di equivalenza contiene infinite **funzioni di densità congiunte** che hanno **supporti** diversi, e che quindi non ha senso parlare del supporto di una funzione di ripartizione congiunta che è **assolutamente continua**. Nei libri di testo dove si parla del supporto di una **funzione di ripartizione congiunta assolutamente continua** si fa riferimento ad una definizione di supporto che è diversa da quella che abbiamo dato nella Definizione 4.6.

Consideriamo ora due variabili casuali \tilde{X}_1 e \tilde{X}_2 con **funzione di ripartizione congiunta assolutamente continua**. Assumiamo dunque che

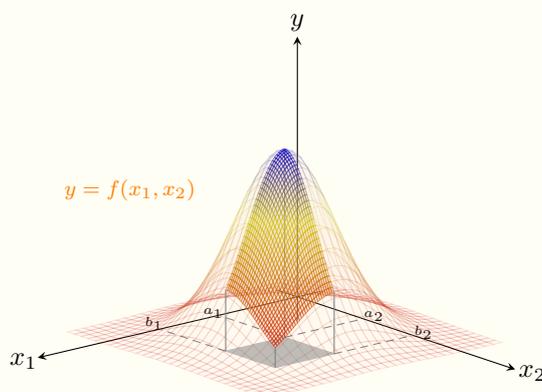
$$\begin{aligned} F(x_1, x_2) &= P\left(\{\tilde{X}_1 \leq x_1\} \cap \{\tilde{X}_2 \leq x_2\}\right) \\ &= \int_{-\infty}^{x_2} \int_{-\infty}^{x_1} f(t_1, t_2) dx_1 dx_2 \quad \text{per ogni } (x_1, x_2) \in \mathbb{R}^2, \end{aligned}$$

dove $f(x_1, x_2)$ è una **funzione di densità congiunta** bivariata, ovvero una funzione che soddisfa le proprietà $f_{0_{k=2}} - f_{2_{k=2}}$. Non è difficile dimostrare (omettiamo i dettagli della dimostrazione) che in questo caso si deve avere

$$P\left(\{a_1 < \tilde{X}_1 \leq b_1\} \cap \{a_2 < \tilde{X}_2 \leq b_2\}\right) = \int_{a_2}^{b_2} \int_{a_1}^{b_1} f(x_1, x_2) dx_1 dx_2, \quad (44)$$

e se la funzione di densità congiunta $f(x_1, x_2)$ è continua in ogni punto del rettangolo $[a_1, b_1] \times [a_2, b_2]$, allora la probabilità al primo membro dell'uguaglianza può essere visualizzata come il **volume** sotteso alla funzione di densità congiunta $f(x_1, x_2)$ sopra il rettangolo $[a_1, b_1] \times [a_2, b_2]$ (vedi il grafico in Figura 4.12).

Figura 4.12. Il grafico visualizza il valore dell'integrale doppio $\int_{a_2}^{b_2} \int_{a_1}^{b_1} f(x_1, x_2) dx_1 dx_2$ come volume sotteso alla **funzione di densità congiunta** $f(x_1, x_2)$.



Come potrebbe sembrare ovvio, nel caso di un generico numero $k \geq 2$ di variabili casuali che hanno **funzione di ripartizione congiunta assolutamente continua**, vale la formula

$$P\left(\bigcap_{i=1}^k \{a_i < \tilde{X}_i \leq b_i\}\right) = \int_{a_k}^{b_k} \cdots \int_{a_2}^{b_2} \int_{a_1}^{b_1} f(x_1, x_2, \dots, x_k) dx_1 dx_2 \cdots dx_k.$$

che generalizza la formula (44).

Consideriamo ora invece le **distribuzioni marginali** associate ad una **distribuzione congiunta** che è **assolutamente continua**. Come al solito, partiamo dal caso più semplice che coinvolge solo due variabili casuali \tilde{X}_1 e \tilde{X}_2 . Assumiamo dunque di nuovo che

$$\begin{aligned} F(x_1, x_2) &= P\left(\{\tilde{X}_1 \leq x_1\} \cap \{\tilde{X}_2 \leq x_2\}\right) \\ &= \int_{-\infty}^{x_2} \int_{-\infty}^{x_1} f(t_1, t_2) dt_1 dt_2 \quad \text{per ogni } (x_1, x_2) \in \mathbb{R}^2, \end{aligned}$$

dove $f(x_1, x_2)$ è una **funzione di densità congiunta** bivariata, ovvero una funzione che soddisfa le proprietà **f0_{k=2} - f2_{k=2}**. Usando la prima delle due formule (37) vediamo che

$$\begin{aligned} F_1(x_1) &= P\left(\{\tilde{X}_1 \leq x_1\}\right) = \lim_{x_2 \rightarrow +\infty} F(x_1, x_2) = \lim_{x_2 \rightarrow +\infty} \int_{-\infty}^{x_2} \int_{-\infty}^{x_1} f(t_1, t_2) dt_1 dt_2 \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{x_1} f(t_1, t_2) dt_1 dt_2 = \int_{-\infty}^{x_1} \left(\int_{-\infty}^{+\infty} f(t_1, t_2) dt_2 \right) dt_1 \quad \text{per ogni } x_1 \in \mathbb{R}, \end{aligned}$$

e da questo risultato deduciamo che \tilde{X}_1 deve essere una variabile casuale **assolutamente continua** e che

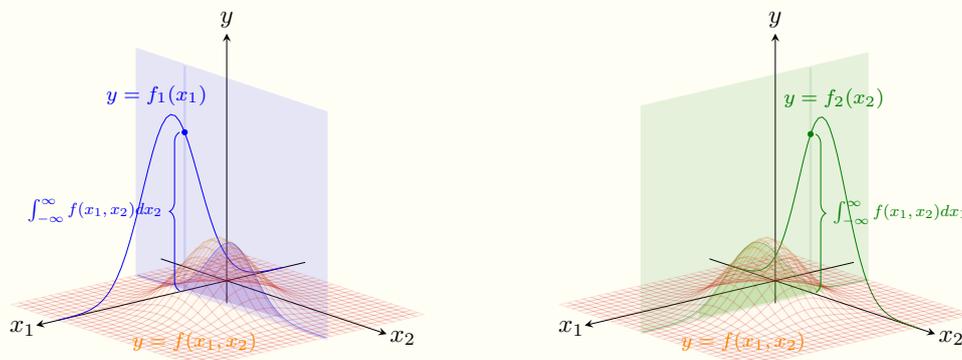
$$f_1(x_1) = \int_{-\infty}^{+\infty} f(t_1 = x_1, t_2) dt_2 = \int_{-\infty}^{+\infty} f(x_1, x_2) dx_2, \quad x_1 \in \mathbb{R} \quad (45)$$

deve essere una sua **funzione di densità** "marginale". Scambiando i ruoli delle variabili casuali \tilde{X}_1 e \tilde{X}_2 vediamo che anche \tilde{X}_2 deve essere una variabile casuale **assolutamente continua** e che

$$f_2(x_2) = \int_{-\infty}^{+\infty} f(x_1, x_2) dx_1, \quad x_2 \in \mathbb{R} \quad (46)$$

deve essere una sua **funzione di densità** marginale. I due grafici in Figura 4.13 aiutano a visualizzare la relazione tra una **funzione di densità congiunta** bivariata e le due corrispondenti **funzioni di densità** marginali.

Figura 4.13. Il due grafici aiutano a visualizzare la relazione tra una **funzione di densità congiunta** bivariata e le due corrispondenti **funzioni di densità** marginali.

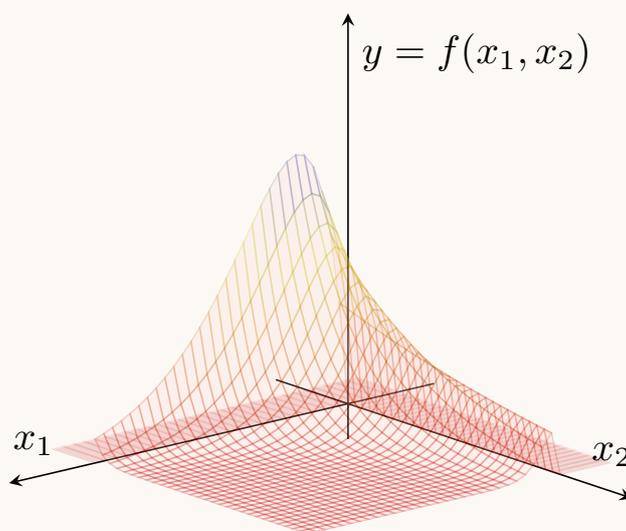


Nel precedente ragionamento abbiamo considerato soltanto $k = 2$ variabili casuali e abbiamo dimostrato che se la loro **distribuzione congiunta** è **assolutamente continua**, allora devono essere **assolutamente continue** anche entrambe le loro **distribuzioni marginali**. Generalizzando il ragionamento, si può dimostrare che se la **funzione di ripartizione congiunta** di $k > 2$ variabili casuali è **assolutamente continua**, allora deve essere **assolutamente continua** anche la **distribuzione congiunta** (o la **distribuzione marginale**) di qualsiasi sottoinsieme delle variabili casuali. Tuttavia, **non è detto che sia vero anche il viceversa**: infatti, non è difficile costruire degli esempi di variabili casuali con **distribuzioni marginali assolutamente continue** che però hanno una **distribuzione congiunta** che **non** è **assolutamente continua**. Si noti che questo fatto differenzia il caso **assolutamente continuo** da quello **discreto**. Infatti, nel caso discreto abbiamo visto che se k variabili casuali hanno **distribuzioni marginali discrete**, allora anche la loro **distribuzione congiunta** deve essere **discreta**.

Esercizio 4.15. Sia

$$f(x_1, x_2) = \begin{cases} x_1 e^{-x_1(x_2+1)} & \text{per } x_1, x_2 > 0 \\ 0 & \text{altrimenti} \end{cases}$$

una funzione di densità congiunta di due variabili casuali \tilde{X}_1 e \tilde{X}_2 .



- Si dimostri che la funzione $f(x_1, x_2)$ soddisfi le tre proprietà $f_{0_{k=2}} - f_{2_{k=2}}$.
- Si ricavi la funzione di ripartizione congiunta di \tilde{X}_1 e \tilde{X}_2 .
- Si calcoli il valore di

$$P\left(\left\{1 < \tilde{X}_1 \leq 2\right\} \cap \left\{0,5 < \tilde{X}_2 \leq 1\right\}\right).$$

- Si ricavino delle funzioni di densità marginali di \tilde{X}_1 e \tilde{X}_2 .

Risposte:

- a) Si dimostri che la funzione $f(x_1, x_2)$ soddisfi le tre proprietà $f0_{k=2}$ - $f2_{k=2}$.

Per dimostrare la proprietà $f0_{k=2}$, basta osservare che la funzione $f(x_1, x_2)$ associa un numero reale ad ogni coppia di numeri reali (x_1, x_2) e che il suo dominio è quindi l'insieme di tutte le coppie ordinate di numeri reali, ovvero \mathbb{R}^2 .

Per dimostrare la proprietà $f1_{k=2}$, è invece sufficiente osservare che $e^z > 0$ per ogni $z \in \mathbb{R}$ e che dunque $x_1 e^{-x_1(x_2+1)} \geq 0$ per ogni $x_1, x_2 > 0$.

Infine, per dimostrare la proprietà $f2_{k=2}$, dobbiamo calcolare il valore dell'integrale doppio

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x_1, x_2) dx_1 dx_2 = \int_0^{\infty} \int_0^{\infty} f(x_1, x_2) dx_1 dx_2.$$

Per rendere più semplice il calcolo di questo integrale conviene invertire l'ordine dei due integrali, ovvero integrare prima rispetto alla variabile x_2 e poi rispetto alla variabile x_1 :

$$\int_0^{\infty} \int_0^{\infty} f(x_1, x_2) dx_1 dx_2 = \int_0^{\infty} \int_0^{\infty} x_1 e^{-x_1(x_2+1)} dx_2 dx_1.$$

Siccome per $x_1 > 0$ il valore dell'integrale interno $\int_0^{\infty} x_1 e^{-x_1(x_2+1)} dx_2$ è dato da

$$\int_0^{\infty} x_1 e^{-x_1(x_2+1)} dx_2 = \left[x_1 \times \frac{e^{-x_1(x_2+1)}}{-x_1} \right]_0^{\infty} = e^{-x_1}$$

e siccome il valore dell'integrale interno nel punto $x_1 = 0$ non incide sul valore dell'integrale doppio (si ricordi che il valore di un integrale di Riemann non cambia se la funzione integranda viene modificata in un singolo punto), possiamo concludere che

$$\begin{aligned} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x_1, x_2) dx_2 dx_1 &= \int_0^{\infty} \int_0^{\infty} x_1 e^{-x_1(x_2+1)} dx_2 dx_1 = \\ &= \int_0^{\infty} e^{-x_1} dx_1 = \dots = 1 \end{aligned}$$

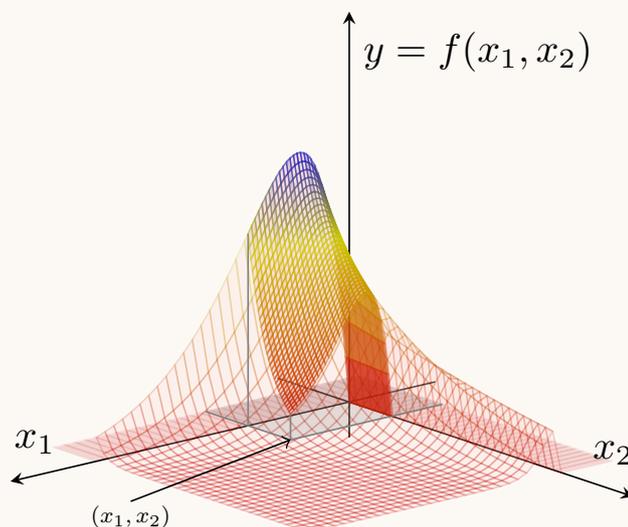
esattamente come previsto dalla proprietà $f2_{k=2}$.

- b) Si ricavi la funzione di ripartizione congiunta di \tilde{X}_1 e \tilde{X}_2 .

La funzione di ripartizione congiunta è data da

$$F(x_1, x_2) = P\left(\left\{\tilde{X}_1 \leq x_1\right\} \cap \left\{\tilde{X}_2 \leq x_2\right\}\right) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} f(t_1, t_2) dt_2 dt_1$$

e il valore di $F(x_1, x_2)$ può dunque essere interpretato come il volume sotteso alla funzione di densità congiunta $f(x_1, x_2)$ sopra l'area infinita del "rettangolo" illimitato che contiene tutti i punti del prodotto cartesiano degli intervalli $(-\infty, x_1]$ e $(-\infty, x_2]$ (vedi il grafico sottostante).



Siccome la funzione di densità congiunta $f(x_1, x_2)$ è nulla se $x_1 \leq 0$ oppure $x_2 \leq 0$, possiamo concludere che

$$F(x_1, x_2) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} f(t_1, t_2) dt_2 dt_1 = 0 \quad \text{se } x_1 \leq 0 \text{ oppure } x_2 \leq 0.$$

Se invece $x_1 > 0$ e $x_2 > 0$, otteniamo

$$\begin{aligned} F(x_1, x_2) &= P\left(\left\{\tilde{X}_1 \leq x_1\right\} \cap \left\{\tilde{X}_2 \leq x_2\right\}\right) \\ &= \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} f(t_1, t_2) dt_2 dt_1 = \int_0^{x_1} \int_0^{x_2} t_1 e^{-t_1(t_2+1)} dt_2 dt_1 \\ &= \int_0^{x_1} \left[t_1 \times \frac{e^{-t_1(t_2+1)}}{-t_1} \right]_0^{x_2} dt_1 \\ &= \int_0^{x_1} (-e^{-t_1(x_2+1)} + e^{-t_1}) dt_1 \\ &= \left[\frac{e^{-t_1(x_2+1)}}{x_2+1} - e^{-t_1} \right]_0^{x_1} = \frac{e^{-x_1(x_2+1)}}{x_2+1} - e^{-x_1} - \frac{1}{x_2+1} + 1. \end{aligned}$$

La funzione di ripartizione congiunta è quindi data da

$$F(x_1, x_2) = \begin{cases} 0 & \text{se } x_1 \leq 0 \text{ oppure } x_2 \leq 0, \\ 1 - e^{-x_1} + \frac{e^{-x_1(x_2+1)}}{x_2+1} - \frac{1}{x_2+1} & \text{se } x_1 > 0 \text{ e } x_2 > 0. \end{cases}$$

Si noti che questa funzione di ripartizione congiunta è ancora quella della consegna dell'Esercizio 4.13.

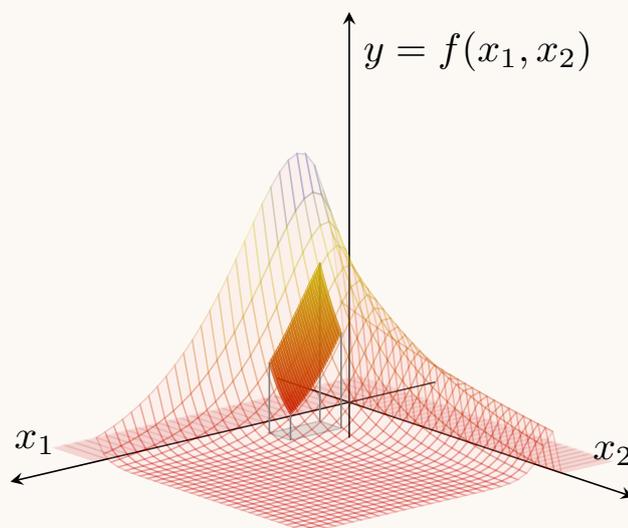
c) Si calcoli il valore di

$$P\left(\left\{1 < \tilde{X}_1 \leq 2\right\} \cap \left\{0,5 < \tilde{X}_2 \leq 1\right\}\right).$$

Nella risposta al quesito a) dell'Esercizio 4.13 abbiamo già calcolato il valore di questa probabilità a partire dalla funzione di ripartizione congiunta, ovvero usando la formula (33). Questa volta calcoleremo la stessa probabilità a partire dalla funzione di densità congiunta, ovvero usando la formula (44):

$$\begin{aligned} P\left(\left\{1 < \tilde{X}_1 \leq 2\right\} \cap \left\{0,5 < \tilde{X}_2 \leq 1\right\}\right) &= \int_{0,5}^1 \int_1^2 f(x_1, x_2) dx_1 dx_2 = \\ & \quad \text{[[inversione dell'ordine degli integrali]]} \\ &= \int_1^2 \int_{0,5}^1 f(x_1, x_2) dx_2 dx_1 = \int_1^2 \int_{0,5}^1 x_1 e^{-x_1(x_2+1)} dx_2 dx_1 = \\ &= \int_1^2 \left[x_1 \times \frac{e^{-x_1(x_2+1)}}{-x_1} \right]_{0,5}^1 dx_1 = \int_1^2 (e^{-1,5x_1} - e^{-2x_1}) dx_1 = \\ &= \int_1^2 e^{-1,5x_1} dx_1 - \int_1^2 e^{-2x_1} dx_1 = \left[\frac{e^{-1,5x_1}}{-1,5} \right]_1^2 - \left[\frac{e^{-2x_1}}{-2} \right]_1^2 \\ &= -\frac{1}{1,5} (e^{-1,5 \times 2} - e^{-1,5 \times 1}) - \frac{1}{2} (e^{-2 \times 2} - e^{-2 \times 1}) = 0,057 \end{aligned}$$

Il grafico sottostante visualizza il valore della probabilità che abbiamo appena calcolato come il volume sotteso alla funzione di densità congiunta sopra l'area del rettangolo con $1 < x_1 \leq 2$ e con $0,5 < x_2 \leq 1$.



d) Si ricavino delle funzioni di densità marginali di \tilde{X}_1 e \tilde{X}_2 .

Usando la formule (45) vediamo che

$$\begin{aligned}
 f_1(x_1) &= \int_{-\infty}^{\infty} f(x_1, x_2) dx_2 \\
 &= \begin{cases} \int_0^{\infty} x_1 e^{-x_1(x_2+1)} dx_2 & \text{per } x_1 > 0 \\ 0 & \text{per } x_1 \leq 0 \end{cases} \\
 &= \begin{cases} \left[x_1 \times \frac{e^{-x_1(x_2+1)}}{-(x_1)} \right]_0^{\infty} & \text{per } x_1 > 0 \\ 0 & \text{per } x_1 \leq 0 \end{cases} \\
 &= \begin{cases} e^{-x_1} & \text{per } x_1 > 0 \\ 0 & \text{per } x_1 \leq 0 \end{cases}
 \end{aligned}$$

è una funzione di densità marginale della variabile casuale \tilde{X}_1 , e che

$$\begin{aligned}
 f_2(x_2) &= \int_{-\infty}^{\infty} f(x_1, x_2) dx_1 \\
 &= \begin{cases} \int_0^{\infty} x_1 e^{-x_1(x_2+1)} dx_1 & \text{per } x_2 > 0 \\ 0 & \text{per } x_2 \leq 0 \end{cases} \\
 &[[\text{integrazione per parti}]] \\
 &= \begin{cases} \left[x_1 \times \frac{e^{-x_1(x_2+1)}}{-(x_2+1)} \right]_0^{\infty} - \int_0^{\infty} \frac{e^{-x_1(x_2+1)}}{-(x_2+1)} dx_1 & \text{per } x_2 > 0 \\ 0 & \text{per } x_2 \leq 0 \end{cases} \\
 &= \begin{cases} 0 - \left[\frac{e^{-x_1(x_2+1)}}{(x_2+1)^2} \right]_0^{\infty} & \text{per } x_2 > 0 \\ 0 & \text{per } x_2 \leq 0 \end{cases} \\
 &= \begin{cases} \frac{1}{(x_2+1)^2} & \text{per } x_2 > 0 \\ 0 & \text{per } x_2 \leq 0 \end{cases}
 \end{aligned}$$

è una funzione di densità marginale della variabile casuale \tilde{X}_2 . Non è difficile verificare che si ottengono esattamente le stesse funzioni di densità marginali se si derivano le due funzioni di ripartizione marginali che abbiamo ricavato nella risposta al quesito b) dell'Esercizio 4.13.

4.8 Distribuzioni condizionate

Come abbiamo già accennato, in molte situazioni pratiche non è affatto semplice specificare direttamente e in modo "realistico" la [funzione di ripartizione congiunta](#) di due o più variabili casuali, soprattutto quando il numero di variabili casuali coinvolte è molto elevato. Tuttavia, in molti casi l'esperimento casuale di riferimento può essere interpre-

tato come un esperimento che si svolge in due o più passi e, come vedremo tra breve, in situazioni di questo genere la distribuzione congiunta di tutte le variabili casuali può a volte essere specificata attraverso passi successivi.

4.8.1 Distribuzioni marginali e condizionate discrete

Per descrivere come la **distribuzione congiunta** di due o più variabili casuali può essere specificata attraverso passi successivi, conviene partire da un esperimento casuale che si svolge in due passi e considerare due variabili casuali discrete \tilde{X}_1 e \tilde{X}_2 , dove il valore assunto da \tilde{X}_1 dipende soltanto dall'esito del primo passo dell'esperimento, mentre quello assunto da \tilde{X}_2 dipende dagli esiti di entrambi i passi. Siccome, per definizione,

$$P(\{\tilde{X}_2 \in x_2\}|\{\tilde{X}_1 = x_1\}) = \frac{P(\{\tilde{X}_1 = x_1\} \cap \{\tilde{X}_2 = x_2\})}{P(\{\tilde{X}_1 = x_1\})}, \quad \text{se } P(\{\tilde{X}_1 = x_1\}) > 0,$$

possiamo scrivere

$$P(\{\tilde{X}_1 = x_1\} \cap \{\tilde{X}_2 = x_2\}) = P(\{\tilde{X}_1 = x_1\}) \times P(\{\tilde{X}_2 \in x_2\}|\{\tilde{X}_1 = x_1\}) \quad (47)$$

e questa uguaglianza può essere ritenuta valida anche se $P(\{\tilde{X}_1 = x_1\}) = 0$ a patto che in questo caso il prodotto al secondo membro venga considerato come nullo: infatti,

$$\{\tilde{X}_1 = x_1\} \cap \{\tilde{X}_2 = x_2\} \subseteq \{\tilde{X}_1 = x_1\}$$

e pertanto (legge L4)

$$P(\{\tilde{X}_1 = x_1\}) = 0 \quad \Rightarrow \quad P(\{\tilde{X}_1 = x_1\} \cap \{\tilde{X}_2 = x_2\}) = 0.$$

Secondo l'uguaglianza (47), possiamo dunque specificare una **funzione di massa di probabilità congiunta**

$$p_{12}(x_1, x_2) = P(\{\tilde{X}_1 = x_1\} \cap \{\tilde{X}_2 = x_2\}), \quad (x_1, x_2) \in \mathbb{R}^2,$$

specificando ...

- ... una **funzione di massa di probabilità marginale**

$$p_1(x_1) = P(\{\tilde{X}_1 = x_1\}), \quad x_1 \in \mathbb{R},$$

che rappresenta le probabilità marginali $P(\{\tilde{X}_1 = x_1\})$, ...

- ... e specificando per ogni $x_1 \in S_1$ (S_1 indica il supporto della variabile casuale \tilde{X}_1 , ovvero l'insieme dei valori x_1 dove $p_1(x_1) > 0$) una corrispondente **funzione di massa di probabilità condizionata della variabile casuale \tilde{X}_2** , ovvero una funzione

$$p_{2|1}(x_2|x_1) = P(\{\tilde{X}_2 \in x_2\}|\{\tilde{X}_1 = x_1\}), \quad x_2 \in \mathbb{R},$$

che, **come funzione di $x_2 \in \mathbb{R}$ con x_1 fisso**, restituisce le probabilità condizionate $P(\{\tilde{X}_2 \in x_2\}|\{\tilde{X}_1 = x_1\})$.

Per mettere in atto questa strategia bisogna ovviamente chiedersi se le **funzioni di massa di probabilità condizionate** devono soddisfare determinate proprietà affinché il prodotto al secondo membro della formula (47) dia effettivamente luogo ad una **funzione di massa di probabilità congiunta**. La risposta a questa domanda è molto semplice: **le funzioni di massa di probabilità condizionate, come funzioni di $x_2 \in \mathbb{R}$ con x_1 fisso, possono essere delle qualunque funzioni di massa di probabilità univariate**, ovvero delle qualunque funzioni che soddisfano le proprietà **p0 - p2**. Infatti, se specifichiamo una **funzione di massa di probabilità** marginale $p_1(x_1)$ (ovvero una qualunque funzione che soddisfa le proprietà **p0 - p2**), e se per ogni x_1 che appartiene al supporto di $p_1(x_1)$ specifichiamo una corrispondente funzione $p_{2|1}(x_2|x_1)$ che (come funzione di $x_2 \in \mathbb{R}$ con x_1 fisso) è anch'essa una funzione che soddisfa le proprietà **p0 - p2**, allora la funzione $p_{12}(x_1, x_2)$ che è definita come

$$p_{12}(x_1, x_2) = \begin{cases} p_1(x_1) \times p_{2|1}(x_2|x_1) & \text{se } p_1(x_1) > 0, \\ 0 & \text{altrimenti} \end{cases} \quad (48)$$

è sempre una **funzione di massa di probabilità** bivariata, ovvero una funzione che soddisfa le proprietà **p0_{k=2} - p2_{k=2}**. Per dimostrare questa affermazione basta osservare che ...

- ... la funzione $p_{12}(x_1, x_2)$ è definita per ogni $(x_1, x_2) \in \mathbb{R}^2$ così come previsto dalla proprietà **p0_{k=2}**;
- la funzione $p_{12}(x_1, x_2)$ soddisfa la proprietà **p1_{k=2}** perché stiamo ipotizzando che la funzione $p_1(x_1)$ e le funzioni $p_{2|1}(x_2|x_1)$ soddisfino la proprietà **p1**;
- la funzione $p_{12}(x_1, x_2)$ soddisfa anche la proprietà **p2_{k=2}** perché assume valori positivi solo se $x_1 \in S_1$ e $x_2 \in S_{2|x_1}$ (S_1 indica il supporto della funzione di probabilità $p_1(x_1)$ e $S_{2|x_1}$ indica il supporto della **funzione di massa di probabilità condizionata** $p_{2|1}(x_2|x_1)$) e perché

$$\begin{aligned} \sum_{x_1 \in S_1} \sum_{x_2 \in S_{2|x_1}} p_1(x_1) \times p_{2|1}(x_2|x_1) &= \sum_{x_1 \in S_1} p_1(x_1) \times \sum_{x_2 \in S_{2|x_1}} p_{2|1}(x_2|x_1) = \\ &= [[\text{le funzioni } p_{2|1}(x_2|x_1) \text{ soddisfano la proprietà } \mathbf{p2}]] = \sum_{x_1 \in S_1} p_1(x_1) \times 1 = \\ &= [[\text{la funzione } p_1(x_1) \text{ soddisfa la proprietà } \mathbf{p2}]] = 1 \times 1 = 1. \end{aligned}$$

D'altra parte, ci si potrebbe anche chiedere se la funzione $p_{12}(x_1, x_2)$ definita nella (48) può essere una **funzione di massa di probabilità congiunta** anche se almeno una delle funzioni di probabilità condizionate $p_{2|1}(x_2|x_1)$ (come funzione di $x_2 \in \mathbb{R}$ con x_1 fisso) **non** è una **funzione di massa di probabilità**, ovvero se almeno una delle **funzioni di massa di probabilità condizionate** $p_{2|1}(x_2|x_1)$ **non** soddisfa le tre le proprietà **p0 - p2**. Non è difficile verificare che la risposta a questa domanda è **negativa!!!**

In sintesi, dal precedente ragionamento deduciamo che per specificare una **funzione di massa di probabilità congiunta** $p(x_1, x_2)$ è sufficiente specificare ...

- ... una **funzione di massa di probabilità** marginale $p_1(x_1)$ che rappresenta le probabilità marginali $P(\{\tilde{X}_1 = x_1\})$...
- ... e specificare, per ogni $x_1 \in S_1$, una **funzione di massa di probabilità condizionata** $p_{2|1}(x_2|x_1)$ che rappresenta le probabilità condizionate $P(\{\tilde{X}_2 \in x_2\}|\{\tilde{X}_1 = x_1\})$. Le **funzioni di massa di probabilità condizionate**, come funzioni di $x_2 \in \mathbb{R}$ con x_1 fisso, devono soddisfare le proprietà p0 - p2, ovvero devono essere anch'esse delle **funzioni di massa di probabilità**. La **funzione di ripartizione discreta** che corrisponde ad una data **funzione di massa di probabilità condizionata** $p_{2|1}(x_2|x_1)$ viene chiamata **funzione di ripartizione (o distribuzione) condizionata di \tilde{X}_2 dato $\tilde{X}_1 = x_1$** .

Il prossimo esercizio illustra un'applicazione del metodo che abbiamo appena descritto.

Esercizio 4.16. Si consideri un esperimento casuale che si svolge in due passi. Al primo passo viene lanciato un dado regolare e al secondo passo vengono estratte in blocco 3 palline da un'urna che ne contiene 6 di cui alcune sono di colore rosso. Si assuma inizialmente che il numero di palline rosse contenute nell'urna sia uguale al punteggio che si ottiene nel lancio del dado. Si indichi con

- \tilde{X}_1 la variabile casuale che descrive il numero di palline rosse presenti nell'urna prima di procedere all'estrazione,
 - e con \tilde{X}_2 la variabile casuale che descrive il numero di palline rosse tra le palline estratte dall'urna.
- a) Si determini la funzione di massa di probabilità congiunta delle variabili casuali \tilde{X}_1 e \tilde{X}_2 .
 - b) Si assuma ora invece che il numero di palline rosse presenti nell'urna sia determinato dall'esito del lancio di una moneta regolare: se esce testa, l'urna contiene 5 palline rosse e una di colore diverso; altrimenti ne contiene soltanto 3 di colore rosso e 3 di colore diverso. Con riferimento a questa nuova ipotesi, come è definita la funzione di massa di probabilità congiunta delle variabili casuali \tilde{X}_1 e \tilde{X}_2 ?
 - c) Si assuma di nuovo che il numero di palline rosse presenti nell'urna sia uguale al punteggio che si ottiene nel lancio di un dado regolare, ma che il numero di palline che vengono estratte (in blocco) dall'urna sia pari a 4 non a 3 come ipotizzato in precedenza. Con riferimento a questa terza ipotesi, come è definita la funzione di massa di probabilità congiunta delle variabili casuali \tilde{X}_1 e \tilde{X}_2 ?

Risposte

- a) Si determini la funzione di massa di probabilità congiunta delle variabili casuali \tilde{X}_1 e \tilde{X}_2 .

Dalle informazioni contenute nella consegna deduciamo che \tilde{X}_1 e \tilde{X}_2 siano due variabili casuali discrete (infatti, entrambe possono assumere solo un numero finito di valori diversi) e che la funzione di massa di probabilità marginale di \tilde{X}_1 (il numero di palline rosse presenti nell'urna) sia data da

$$p_1(x_1) = \begin{cases} 1/6 & \text{per } x_1 = 1, 2, 3, 4, 5, 6; \\ 0 & \text{altrimenti.} \end{cases}$$

Inoltre, deduciamo che per $x_1 = 1, 2, 3, 4, 5, 6$ le funzioni di probabilità condizionate della variabile casuale \tilde{X}_2 siano definite come

$$p_{2|1}(x_2|x_1) = \frac{\binom{x_1}{x_2} \times \binom{6-x_1}{3-x_2}}{\binom{6}{3}} \quad \text{per } x_2 = 0, 1, 2, 3.$$

Infatti, ...

- ... estraendo 3 palline da un'urna che contiene 6 palline si possono ottenere $\binom{6}{3}$ combinazioni di palline diverse ...
- ... e se x_1 delle palline presenti nell'urna sono di colore rosso, allora il numero di combinazioni che contiene x_2 palline rosse deve essere dato da $\binom{x_1}{x_2} \times \binom{6-x_1}{3-x_2}$ perché ci sono $\binom{x_1}{x_2}$ modi per scegliere x_2 palline tra le x_1 palline rosse presenti nell'urna e perché ci sono $\binom{6-x_1}{3-x_2}$ per scegliere $3-x_2$ palline non rosse tra le $6-x_1$ palline non rosse presenti nell'urna. Ovviamente, il coefficiente binomiale $\binom{x_1}{x_2}$ deve considerarsi nullo se $x_2 > x_1$ e il coefficiente binomiale $\binom{6-x_1}{3-x_2}$ deve considerarsi nullo se $3-x_2 > 6-x_1$ (il numero di palline non rosse che si ottengono nelle tre estrazioni non può essere superiore al numero di palline non rosse presenti nell'urna).

La funzione di probabilità congiunta delle variabili casuali \tilde{X}_1 e \tilde{X}_2 è dunque data da

$$p(x_1, x_2) = \begin{cases} p_1(x_1) \times p_{2|1}(x_2|x_1) & \text{se } x_1 \in S_1, \\ 0 & \text{altrimenti} \end{cases}$$

$$= \begin{cases} \frac{1}{6} \times \frac{\binom{x_1}{x_2} \times \binom{6-x_1}{3-x_2}}{\binom{6}{3}} & \text{se } x_1 = 1, 2, 3, 4, 5, 6 \\ & \text{e } x_2 = 0, 1, 2, 3, 4; \\ 0 & \text{altrimenti.} \end{cases}$$

Lasciamo come esercizio il calcolo dei valori numerici delle probabilità congiunte $p(x_1, x_2)$ e la costruzione di una corrispondente tabella a doppia entrata.

Le risposte ai quesiti b) e c) possono essere ottenute in modo analogo. Si noti che per rispondere al quesito b) è sufficiente modificare la funzione di massa di probabilità marginale della variabile casuale \tilde{X}_1 , mentre per rispondere al quesito c) basta modificare le funzioni di massa di probabilità condizionata della variabile casuale \tilde{X}_2 .

Supponiamo ora di aggiungere un terzo passo all'esperimento casuale considerato nel precedente esercizio. Supponiamo, per esempio, che il terzo passo consista nell'estrazione di un certo numero di palline da un'altra urna, e che la composizione di questa ulteriore urna dipenda dagli esiti dei primi due passi dell'esperimento. In questo caso potremmo descrivere l'esito del terzo passo attraverso una terza variabile casuale discreta \tilde{X}_3 che rappresenta, supponiamo, il numero di palline rosse che si ottengono nelle estrazioni dall'urna aggiuntiva. Per specificare la **funzione di massa di probabilità congiunta** di tutte e tre le variabili casuali, ovvero la funzione

$$p_{123}(x_1, x_2, x_3) = P\left(\bigcap_{i=1}^3 \{\tilde{X}_i = x_i\}\right), \quad (x_1, x_2, x_3) \in \mathbb{R}^3,$$

possiamo considerare uno alla volta tutti i punti (x_1, x_2) che appartengono al supporto S_{12} della **funzione di massa di probabilità congiunta**

$$p_{12}(x_1, x_2) = P(\{\tilde{X}_1 = x_1\} \cap \{\tilde{X}_2 = x_2\}), \quad (x_1, x_2) \in \mathbb{R}^2$$

(i punti in questione sono tutte le possibili coppie di realizzazioni delle variabili casuali \tilde{X}_1 e \tilde{X}_2), e specificare per ciascuno di tali punti una opportuna **funzione di massa di probabilità condizionata** $p_{3|12}(x_3|x_1, x_2)$ che, come funzione di $x_3 \in \mathbb{R}$ con x_1 e x_2 fissi, restituisce le probabilità condizionate

$$P\left(\{\tilde{X}_3 = x_3\} \mid \{\tilde{X}_1 = x_1\} \cap \{\tilde{X}_2 = x_2\}\right)$$

Infatti, siccome per ogni $(x_1, x_2) \in S_{12}$ si ha

$$\begin{aligned} P\left(\bigcap_{i=1}^3 \{\tilde{X}_i = x_i\}\right) &= P(\{\tilde{X}_1 = x_1\} \cap \{\tilde{X}_2 = x_2\}) \times \\ &\quad \times P\left(\{\tilde{X}_3 = x_3\} \mid \{\tilde{X}_1 = x_1\} \cap \{\tilde{X}_2 = x_2\}\right), \end{aligned}$$

possiamo comporre le funzioni di massa di probabilità condizionata $p_{3|12}(x_3|x_1, x_2)$ con la **funzione di massa di probabilità congiunta** $p_{12}(x_1, x_2)$ onde ottenere la **funzione di massa di probabilità congiunta trivariata**

$$\begin{aligned} p_{123}(x_1, x_2, x_3) &= P\left(\bigcap_{i=1}^3 \{\tilde{X}_i = x_i\}\right) \\ &= \begin{cases} p_{12}(x_1, x_2) \times p_{3|12}(x_3|x_1, x_2) & \text{se } (x_1, x_2) \in S_{12}, \\ 0 & \text{altrimenti.} \end{cases} \end{aligned}$$

Chiaramente, una volta specificata la **funzione di massa di probabilità congiunta** $p_{123}(x_1, x_2, x_3)$ possiamo andare avanti e aggiungere un quarto, quinto, sesto, ... passo all'esperimento casuale: specificando per ogni punto $(x_1, x_2, \dots, x_{k-1})$ che appartiene al supporto $S_{12\dots(k-1)}$ della **funzione di massa di probabilità congiunta**

$$p_{12\dots(k-1)}(x_1, x_2, \dots, x_{k-1}) = P \left(\bigcap_{i=1}^{k-1} \{\tilde{X}_i = x_i\} \right), \quad (x_1, x_2, \dots, x_{k-1}) \in \mathbb{R}^{k-1},$$

($p_{12\dots(k-1)}(x_1, x_2, \dots, x_{k-1})$ è la **funzione di massa di probabilità congiunta** che è stata ottenuta al passo precedente) una corrispondente **funzione di massa di probabilità condizionata**

$$p_{k|12\dots(k-1)}(x_k | x_1, x_2, \dots, x_{k-1})$$

che restituisce le probabilità condizionate

$$P \left(\{\tilde{X}_k = x_k\} \mid \bigcap_{i=1}^{k-1} \{\tilde{X}_i = x_i\} \right),$$

(la funzione di massa di probabilità condizionata $p_{k|12\dots(k-1)}(x_k | x_1, x_2, \dots, x_{k-1})$, come funzione di $x_k \in \mathbb{R}$ con x_1, x_2, \dots, x_{k-1} fissi, deve essere una funzione che soddisfa le proprietà p0 - p2), si ottiene la **funzione di massa di probabilità congiunta**

$$p_{12\dots k}(x_1, x_2, \dots, x_k) = P \left(\bigcap_{i=1}^k \{\tilde{X}_i = x_i\} \right) = \begin{cases} p_{12\dots(k-1)}(x_1, x_2, \dots, x_{k-1}) \times p_{k|12\dots(k-1)}(x_k | x_1, x_2, \dots, x_{k-1}) \\ \text{se } (x_1, x_2, \dots, x_{k-1}) \in S_{12\dots(k-1)}, \\ 0, \text{ altrimenti} \end{cases} \quad (49)$$

che coinvolge anche la variabile casuale \tilde{X}_k .

Per comodità di notazione, vale la pena osservare che anche nei punti per i quali $(x_1, x_2, \dots, x_{k-1}) \notin S_{12\dots(k-1)}$, ovvero nei punti per i quali non abbiamo definito una **funzione di massa di probabilità condizionata** $p_{k|12\dots(k-1)}(x_k | x_1, x_2, \dots, x_{k-1})$, la suddetta **funzione di massa di probabilità congiunta** $p_{12\dots k}(x_1, x_2, \dots, x_k)$ può comunque essere espressa come

$$p_{12\dots k}(x_1, x_2, \dots, x_k) = p_{12\dots(k-1)}(x_1, x_2, \dots, x_{k-1}) \times p_{k|12\dots(k-1)}(x_k | x_1, x_2, \dots, x_{k-1})$$

perché in tali punti la **funzione di massa di probabilità condizionata**, anche se l'avessimo definita, verrebbe comunque moltiplicata per

$$p_{12\dots(k-1)}(x_1, x_2, \dots, x_{k-1}) = 0.$$

Ipotizzando ora che la **funzione di massa di probabilità congiunta** di k variabili casuali discrete sia stata ottenuta applicando ripetutamente il metodo che abbiamo appena descritto, ovvero ipotizzando che

$$p_{12}(x_1, x_2) = p_1(x_1) \times p_{2|1}(x_2|x_1)$$

$$p_{123}(x_1, x_2, x_3) = p_{12}(x_1, x_2) \times p_{3|12}(x_3|x_1, x_2)$$

$$\vdots$$

$$p_{12\dots k}(x_1, x_2, \dots, x_k) = p_{12\dots(k-1)}(x_1, x_2, \dots, x_{k-1}) \times p_{k|12\dots(k-1)}(x_k|x_1, x_2, \dots, x_{k-1}),$$

vediamo che l'ultima **funzione di massa di probabilità congiunta** di questo elenco può essere espressa come

$$p_{12\dots k}(x_1, x_2, \dots, x_k) = p_1(x_1) \times p_{2|1}(x_2|x_1) \times \dots \times p_{k|12\dots(k-1)}(x_k|x_1, x_2, \dots, x_{k-1}) \quad (50)$$

Come si evince dalla precedente descrizione, la **funzione di massa di probabilità congiunta** di k variabili casuali può sempre essere specificata in più passi, specificando prima la **funzione di massa di probabilità marginale** di una di esse, e specificando successivamente, una dopo l'altra, delle opportune **funzioni di massa di probabilità condizionate**.

A questo punto potrebbe sorgere il dubbio se tutte le **funzioni di probabilità congiunte**

$$p_{12\dots k}(x_1, x_2, \dots, x_k) = P\left(\bigcap_{i=1}^k \{\tilde{X}_i = x_i\}\right), \quad (x_1, x_2, \dots, x_k) \in \mathbb{R}^k,$$

possono essere espresse come nella formula (50). La risposta a questo dubbio è affermativa. Infatti, secondo il teorema della probabilità composta (Teorema 3.1) si ha

$$\begin{aligned} P\left(\bigcap_{i=1}^k \{\tilde{X}_i = x_i\}\right) &= P(\{\tilde{X}_1 = x_1\}) \times P(\{\tilde{X}_2 = x_2\}|\{\tilde{X}_1 = x_1\}) \times \\ &\quad \times P(\{\tilde{X}_3 = x_3\}|\{\tilde{X}_1 = x_1\} \cap \{\tilde{X}_2 = x_2\}) \times \\ &\quad \times \dots \times P\left(\{\tilde{X}_k = x_k\} \middle| \bigcap_{i=1}^{k-1} \{\tilde{X}_i = x_i\}\right) \end{aligned}$$

e le probabilità condizionate al secondo membro di questa uguaglianza sono proprio i valori che restituiscono le **funzioni di massa di probabilità condizionate** che compaiono al secondo membro della (50).

Esercizio 4.17. Si considerino ancora le variabili casuali \tilde{X}_1 e \tilde{X}_2 dell'Esercizio 4.14. Ricordiamo che \tilde{X}_1 e \tilde{X}_2 si riferiscono ad un esperimento casuale che consiste in un'estrazione in blocco di tre palline da un'urna che contiene 6 palline di colore bianco, 5 di colore grigio e 4 di colore nero. Ricordiamo inoltre che \tilde{X}_1 rappresenta

il numero di palline bianche estratte e che \tilde{X}_2 rappresenta il numero di palline grigie estratte.

Si ricavano le funzioni di massa di probabilità condizionate della variabile casuale \tilde{X}_1 e anche quelle della variabile casuale \tilde{X}_2 .

Risposta:

Per ricavare le funzioni di massa di probabilità condizionate conviene partire dalla tabella a doppia entrata che rappresenta la funzione di massa di probabilità congiunta e le due funzioni di massa di probabilità marginali (vedi la risposta al quesito a) dell'Esercizio 4.14):

$p(x_1, x_2)$	$x_2 = 0$	$x_2 = 1$	$x_2 = 2$	$x_2 = 3$	$p_1(x_1)$
$x_1 = 0$	0,0088	0,0659	0,0879	0,0220	0,1846
$x_1 = 1$	0,0791	0,2637	0,1319	0,0000	0,4747
$x_1 = 2$	0,1319	0,1648	0,0000	0,0000	0,2967
$x_1 = 3$	0,0440	0,0000	0,0000	0,0000	0,0440
$p_2(x_2)$	0,2638	0,4944	0,2198	0,0220	1,0000

Siccome il supporto della variabile casuale \tilde{X}_1 contiene quattro valori (ovvero i valori $x_1 = 0$, $x_1 = 1$, $x_1 = 2$ e $x_1 = 3$), le funzioni di massa di probabilità condizionate della variabile casuale \tilde{X}_2 sono quattro. Per ricavarle a partire dalla tabella a doppia entrata, basta dividere i valori di $p(x_1, x_2)$ che si trovano nelle caselle centrali per i corrispondenti totali di riga. Infatti, così facendo otteniamo la tabella

$p_{1 2}(x_1 x_2)$	$x_2 = 0$	$x_2 = 1$	$x_2 = 2$	$x_2 = 3$	Totale
$x_1 = 0$	0,0477	0,3570	0,4762	0,1192	1,0000
$x_1 = 1$	0,1666	0,5555	0,2779	0,0000	1,0000
$x_1 = 2$	0,4446	0,5554	0,0000	0,0000	1,0000
$x_1 = 3$	1,0000	0,0000	0,0000	0,0000	1,0000

e questa tabella contiene tutti i valori non nulli delle quattro funzioni di massa di probabilità condizionate $p_{2|1}(x_2|x_1)$. La funzione di massa di probabilità condizionata con $x_1 = 2$ è per esempio data da

$$p_{2|1}(x_2|x_1 = 2) = \begin{cases} 0,4446 & \text{se } x_2 = 0, \\ 0,5554 & \text{se } x_2 = 1, \\ 0 & \text{altrimenti.} \end{cases}$$

Procedendo in modo analogo (ovvero dividendo le probabilità congiunte $p(x_1, x_2)$ per i totali di colonna) si ottiene anche la tabella che contiene tutti i valori non nulli

delle quattro funzioni di massa di probabilità condizionate della variabile casuale \tilde{X}_1 (si noti che in questo caso anche il supporto della variabile casuale \tilde{X}_2 contiene *quattro* valori e pertanto anche le funzioni di massa di probabilità condizionate della variabile casuale \tilde{X}_1 sono *quattro*):

$p_{1 2}(x_1 x_2)$	$x_2 = 0$	$x_2 = 1$	$x_2 = 2$	$x_2 = 3$	—
$x_1 = 0$	0,0334	0,1333	0,3999	1,0000	—
$x_1 = 1$	0,2998	0,5334	0,6001	0,0000	—
$x_1 = 2$	0,5000	0,3333	0,0000	0,0000	—
$x_1 = 3$	0,1668	0,0000	0,0000	0,0000	—
Totale	1,0000	1,0000	1,0000	1,0000	—

4.8.2 Distribuzioni marginali e condizionate assolutamente continue*

A questo punto ci si potrebbe chiedere se per specificare una funzione di densità congiunta si può utilizzare un metodo analogo a quello che abbiamo descritto nel paragrafo precedente, ovvero se ...

- i) ...specificando una funzione di densità marginale $f_1(x_1)$ (ovvero una funzione che soddisfa le proprietà **f0 - f2**), ...
- ii) ...e specificando per ogni x_1 tale che $f_1(x_1) > 0$ una corrispondente **funzione di densità condizionata** $f_{2|1}(x_2|x_1)$ (ovvero una funzione di $x_2 \in \mathbb{R}$ che soddisfa le proprietà **f0 - f2**),

... si ottiene sempre una funzione

$$f_{12}(x_1, x_2) = \begin{cases} f_1(x_1) \times f_{2|1}(x_2|x_1) & \text{se } f_1(x_1) > 0, \\ 0 & \text{altrimenti,} \end{cases} \quad (51)$$

che è una funzione di densità congiunta (ovvero una funzione che soddisfa le proprietà **f0_{k=2} - f2_{k=2}**). La risposta a questo dubbio è affermativa a patto che la funzione $f(x_1, x_2)$ risulti "integrabile", ma questa condizione aggiuntiva è di interesse puramente teorico perché nelle applicazioni è sempre soddisfatta per qualunque modo "realistico" in cui si potrebbe pensare di definire le funzioni di densità condizionate.

D'ora in poi, per semplificare la notazione, esprimeremo le funzioni di densità congiunte che sono definite come nella (51) semplicemente come

$$f_{12}(x_1, x_2) = f_1(x_1) \times f_{2|1}(x_2|x_1), \quad (x_1, x_2) \in \mathbb{R}^2, \quad (52)$$

senza considerare separatamente il caso dove $f_1(x_1) = 0$. Si noti che questa semplificazione non causa nessuna indeterminazione perché nei punti (x_1, x_2) dove $f_1(x_1) = 0$ la funzione di densità condizionata $f_{2|1}(x_2|x_1)$, anche se fosse definita, verrebbe comunque moltiplicata per $f_1(x_1) = 0$.

Consideriamo ora le implicazioni del suddetto metodo per specificare funzioni di densità congiunte. Consideriamo dunque due variabili casuali \tilde{X}_1 e \tilde{X}_2 con funzione di ripartizione congiunta assolutamente continua e supponiamo che la funzione $f_{12}(x_1, x_2)$ definita nella (52) sia una loro funzione di densità congiunta. Questa ipotesi potrebbe far sorgere vari dubbi. **In primo luogo ci si potrebbe chiedere se la funzione $f_1(x_1)$ è effettivamente una funzione di densità marginale della variabile casuale \tilde{X}_1 .** La risposta a questo dubbio è affermativa. Infatti, siccome stiamo ipotizzando che le

funzioni $f_{2|1}(x_2|x_1)$, come funzioni di $x_2 \in \mathbb{R}$ con x_1 fisso, siano funzioni che soddisfano le proprietà **f0** - **f2**, possiamo concludere

$$\begin{aligned} \int_{-\infty}^{\infty} f_{12}(x_1, x_2) dx_2 &= \int_{-\infty}^{\infty} f_1(x_1) \times f_{2|1}(x_2|x_1) dx_2 = \\ &= f_1(x_1) \times \int_{-\infty}^{\infty} f_{2|1}(x_2|x_1) dx_2 = \text{[[proprietà f2]]} = f_1(x_1) \times 1 = f_1(x_1) \end{aligned}$$

e in virtù della formula (45) possiamo dunque concludere che la funzione $f_1(x_1)$ è effettivamente una funzione di densità marginale della variabile casuale \tilde{X}_1 .

A questo punto potremmo anche interrogarci **sull'interpretazione delle funzioni di densità condizionate** $f_{2|1}(x_2|x_1)$. A prima vista si potrebbe infatti pensare che (analogamente a ciò che accade nel caso discreto) gli integrali del tipo

$$\int_a^b f_{2|1}(x_2|x_1) dx_2 \quad (53)$$

rappresentino le probabilità condizionate

$$P(\{a < \tilde{X}_2 \leq b\} | \{\tilde{X}_1 = x_1\}) = \frac{P(\{a < \tilde{X}_2 \leq b\} \cap \{\tilde{X}_1 = x_1\})}{P(\{\tilde{X}_1 = x_1\})},$$

ma a pensarci meglio ci rendiamo subito conto che questo non può essere vero perché sotto l'ipotesi che stiamo esaminando le suddette probabilità condizionate non sono nemmeno definite (infatti, se la distribuzione congiunta di due variabili casuali è assolutamente continua, allora devono essere assolutamente continue anche entrambe le loro distribuzioni marginali e pertanto il valore di $P(\{\tilde{X}_1 = x_1\})$ deve essere nullo per ogni $x_1 \in \mathbb{R}$). Tuttavia, l'idea che gli integrali nella (53) siano legati a delle probabilità condizionate non è sbagliata. Infatti, se in un dato punto x_1 la funzione di densità marginale $f_1(x_1)$ è continua e positiva, e se in prossimità di x_1 le funzioni di densità condizionate $f_{2|1}(x_2|x_1)$ sono tutte molto "simili" (per non appesantire troppo la trattazione non specificheremo esattamente che cosa intendiamo per "simili"), allora si può dimostrare che

$$\lim_{\epsilon \rightarrow 0^+} P(\{a < \tilde{X}_2 \leq b\} | \{x_1 - \epsilon < \tilde{X}_1 < x_1 + \epsilon\}) = \int_a^b f_{2|1}(x_2|x_1) dx_2 \quad \text{per ogni } a < b.$$

Questa condizione dice che per valori di ϵ prossimi a zero il valore della probabilità condizionate al primo membro della precedente uguaglianza è molto prossimo al valore dell'integrale al secondo membro. Da questa osservazione deduciamo che per ogni prefissato valore di x_1 dove la funzione di densità marginale $f_1(x_1)$ è continua e positiva, la corrispondente funzione di densità condizionate $f_{2|1}(x_2|x_1)$ specifica la distribuzione di \tilde{X}_2 sotto l'ipotesi che la realizzazione di \tilde{X}_1 sia molto prossima a x_1 . Per specificare una funzione di densità condizionate $f_{2|1}(x_2|x_1)$ con riferimento ad una data coppia di variabili casuali, bisogna quindi tener conto dell'ipotesi che \tilde{X}_1 assuma un valore molto prossimo a x_1 .

Nel prossimo esercizio mostreremo come si dovrebbero specificare le funzioni di densità condizionate con riferimento ad una situazione molto semplice.

Esercizio 4.18. Un esperimento che consiste nel selezionare in modo casuale due punti su un segmento di retta AB di lunghezza θ . Si supponga che i due punti vengano selezionati uno dopo l'altro e che dopo aver selezionato il primo punto P_1 , il secondo punto P_2 venga selezionato sul segmento AP_1 .

Si assuma inoltre che ...

- ... per ogni sottosegmento di AB , la probabilità che il punto P_1 si trovi sul sottosegmento

sia proporzionale alla lunghezza del sottosegmento, ...

- ... e che il punto P_2 venga selezionato secondo un metodo analogo a quello utilizzato per selezionare il punto P_1 : per ogni sottosegmento del segmento AP_1 , la probabilità che il punto P_2 si trovi sul sottosegmento è proporzionale alla lunghezza del sottosegmento.

Sia \tilde{X}_1 la variabile casuale che restituisce la distanza tra il punto A e il punto P_1 e sia \tilde{X}_2 la variabile casuale che restituisce la distanza tra il punto A e il punto P_2 . Come dovrebbe essere definita la distribuzione congiunta delle variabili casuali \tilde{X}_1 e \tilde{X}_2 ?

Risposta:

Dalla descrizione del modo in cui viene selezionato il punto P_1 desumiamo che la distribuzione marginale della variabile casuale \tilde{X}_1 sia uniforme e che

$$f_1(x_1) = \begin{cases} 1/\theta & \text{per } 0 < x_1 < \theta, \\ 0 & \text{altrimenti} \end{cases}$$

sia una sua funzione di densità marginale (vedi l'Esercizio 4.9).

Consideriamo ora invece la variabile casuale \tilde{X}_2 . Per specificare le distribuzioni condizionate di \tilde{X}_2 , consideriamo come evento condizionante un evento del tipo $\{x_1 - \epsilon < \tilde{X}_1 < x_1 + \epsilon\}$ con $x_1 \in [0, \theta]$ e con $\epsilon > 0$ arbitrariamente piccolo. Dalla descrizione del modo in cui viene selezionato il punto P_2 , deduciamo che la distribuzione condizionata della variabile casuale \tilde{X}_2 dato l'evento $\{x_1 - \epsilon < \tilde{X}_1 < x_1 + \epsilon\}$ sia approssimativamente uniforme sull'intervallo $[0, x_1]$. Per ogni prefissato valore di x_1 che appartiene all'intervallo $[0, \theta]$, possiamo quindi considerare

$$f_{2|1}(x_2|x_1) = \begin{cases} 1/x_1 & \text{per } 0 < x_2 < x_1, \\ 0 & \text{altrimenti,} \end{cases}$$

come una funzione di densità condizionata di \tilde{X}_2 . Avendo specificato una funzione di densità marginale di \tilde{X}_1 , e avendo specificato per ogni x_1 dove $f_1(x_1) > 0$ anche una corrispondente funzione di densità condizionata $f_{2|1}(x_2|x_1)$, possiamo ora specificare la distribuzione congiunta delle variabili casuali \tilde{X}_1 e \tilde{X}_2 come la distribuzione (congiunta) assolutamente continua che corrisponde alla funzione di densità congiunta

$$\begin{aligned} f_{12}(x_1, x_2) &= f_1(x_1) \times f_{2|1}(x_2|x_1) \\ &= \begin{cases} \frac{1}{\theta x_1} & \text{se } 0 < x_1 < \theta \text{ e } 0 < x_2 < x_1, \\ 0 & \text{altrimenti.} \end{cases} \end{aligned}$$

Si noti che la consegna di questo esercizio ci avrebbe indotto a specificare esattamente le stesse distribuzioni condizionate di \tilde{X}_2 se avessimo "forzato la mano" e considerato eventi condizionanti del tipo $\{\tilde{X}_1 = x_1\}$ (che hanno probabilità nulla) piuttosto che eventi condizionanti del tipo $\{x_1 - \epsilon < \tilde{X}_1 < x_1 + \epsilon\}$ (che invece hanno probabilità prossima a zero ma comunque positiva). Il famoso **paradosso di Borel-Kolmogorov** illustra invece una situazione dove la distribuzione marginale della variabile casuale condizionante \tilde{X}_1 è assolutamente continua come in questo esercizio, ma dove per un'altra variabile casuale \tilde{X}_2 si è indotti a specificare distribuzioni condizionate diverse a seconda se si fa riferimento a eventi condizionanti del tipo $\{\tilde{X}_1 = x_1\}$ (che hanno probabilità nulla) oppure a eventi condizionanti del tipo $\{x_1 - \epsilon < \tilde{X}_1 < x_1 + \epsilon\}$ (che invece hanno probabilità

prossima a zero ma comunque positiva). Nel primo caso si ottiene una distribuzione congiunta completamente irragionevole, mentre nel secondo caso si ottiene la distribuzione congiunta che chiunque giudicherebbe "corretta". Il paradosso di Borel-Kolmogorov ci insegna quindi che non ha senso specificare probabilità condizionate (o distribuzioni condizionate) con riferimento ad un evento condizionante che ha probabilità nulla e che nel caso di una variabile casuale condizionante \tilde{X}_1 che è assolutamente continua, le distribuzioni condizionate di un'altra variabile casuale \tilde{X}_2 dovrebbero sempre essere specificate facendo riferimento a eventi condizionanti del tipo $\{x_1 - \epsilon < \tilde{X}_1 < x_1 + \epsilon\}$.

Chiaramente, anche nel caso assolutamente continuo possiamo aggiungere un terzo, quarto, quinto, ... passo all'esperimento casuale e quindi specificare una funzione di densità congiunta $f_{12\dots k}(x_1, x_2, \dots, x_k)$ come

$$f_{12\dots k}(x_1, x_2, \dots, x_k) = f_1(x_1) \times f_{2|1}(x_2|x_1) \times \dots \times f_{k|12\dots(k-1)}(x_k|x_1, x_2, \dots, x_{k-1}) \quad (54)$$

Si noti che questa formula è analoga alla formula (50) che abbiamo ottenuto nel caso discreto.

Finora abbiamo visto come si può specificare una funzione di densità congiunta attraverso una funzione di densità marginale e delle corrispondenti funzioni di densità condizionate. A questo punto **ci chiediamo se ogni funzione di densità congiunta $f_{12\dots k}(x_1, x_2, \dots, x_k)$ può essere rappresentata in questo modo**, ovvero se ogni funzione di densità congiunta può essere fattorizzata come nella formula (54) in modo tale che

- $f_1(x_1)$ sia una funzione che soddisfa le proprietà f0 - f2,
- per ogni $x_1 \in \mathbb{R}$ la corrispondente funzione $f_{2|1}(x_2|x_1)$, come funzione di $x_2 \in \mathbb{R}$ con x_1 fisso, sia una funzione che soddisfa le proprietà f0 - f2,
- per ogni $(x_1, x_2) \in \mathbb{R}^2$ la corrispondente funzione $f_{3|12}(x_3|x_1, x_2)$, come funzione di $x_3 \in \mathbb{R}$ con x_1 e x_2 fissi, sia una funzione che soddisfa le proprietà f0 - f2,
- ...
- per ogni $(x_1, x_2, \dots, x_{k-1}) \in \mathbb{R}^{k-1}$ la corrispondente funzione $f_{k|12\dots(k-1)}(x_k|x_1, x_2, \dots, x_{k-1})$, come funzione di $x_k \in \mathbb{R}$ con x_1, x_2, \dots, x_{k-1} fissi, sia una funzione che soddisfa le proprietà f0 - f2.

Attraverso la costruzione di opportuni controesempi si può dimostrare che la risposta a questo dubbio è negativa (invitiamo il lettore a costruire un controesempio). Tuttavia, non è difficile dimostrare che **per ogni funzione di densità congiunta $f_{12\dots k}(x_1, x_2, \dots, x_k)$ esiste sempre una versione equivalente $f_{12\dots k}^*(x_1, x_2, \dots, x_k)$ che può essere fattorizzata nel suddetto modo**. Ricordiamo che, per definizione, due funzioni di densità sono versioni (equivalenti) l'una dell'altra se entrambe danno luogo alla medesima funzione di ripartizione.

Esercizio 4.19. Si ricavino le funzioni di densità condizionate $f_{2|1}(x_2|x_1)$ e $f_{1|2}(x_1|x_2)$ della funzione di densità congiunta

$$f_{12}(x_1, x_2) = \begin{cases} x_1 e^{-x_1(x_2+1)} & \text{per } x_1, x_2 > 0 \\ 0 & \text{altrimenti} \end{cases}$$

($f_{12}(x_1, x_2)$ è la funzione di densità congiunta dell'Esercizio 4.15).

La risposta viene lasciata come esercizio.

4.8.3 Distribuzioni marginali e condizionate di tipo diverso

Come abbiamo visto nelle Sezioni 4.8.1 e 4.8.2, per specificare la funzione di ripartizione congiunta di due variabili casuali \tilde{X}_1 e \tilde{X}_2 possiamo specificare una funzione di massa di probabilità marginale $p_1(x_1)$ per la variabile casuale \tilde{X}_1 e delle funzioni di massa di probabilità condizionate $p_{2|1}(x_2|x_1)$ per la variabile casuale \tilde{X}_2 , oppure specificare una funzione di densità marginale $f_1(x_1)$ per la variabile casuale \tilde{X}_1 e delle funzioni di densità condizionate $f_{2|1}(x_2|x_1)$ per la variabile casuale \tilde{X}_2 . Come potrebbe sembrare ovvio, la funzione di ripartizione congiunta di due variabili casuali potrebbe essere anche specificata . . .

- i) . . . combinando una funzione di massa di probabilità marginale $p_1(x_1)$ con delle funzioni di densità condizionate $f_{2|1}(x_2|x_1)$, . . .
- ii) . . . oppure combinando una funzione di densità marginale $f_1(x_1)$ con delle funzioni di massa di probabilità condizionate $p_{2|1}(x_2|x_1)$,

In entrambi i casi la funzione di ripartizione congiunta delle due variabili casuali non è ne discreta ne assolutamente continua, ma nel primo caso possiamo dire per certezza che la funzione di ripartizione marginale della variabile casuale \tilde{X}_2 è assolutamente continua, mentre nel secondo caso la funzione di ripartizione marginale della variabile casuale \tilde{X}_2 potrebbe essere discreta, assolutamente continua oppure di nessuno di questi due tipi.

Più in generale, si può combinare qualsiasi funzione di ripartizione marginale $F_1(x_1)$ con delle funzioni di ripartizione condizionate $F_{2|1}(x_2|x_1)$ che possono essere scelte *quasi* in modo arbitrario onde ottenere una funzione di ripartizione congiunta $F_{12}(x_1, x_2)$. L'unico limite a questo metodo per specificare funzioni di ripartizione congiunte è posto dalla condizione che per ogni prefissato valore di $x_2 \in \mathbb{R}$ la funzione $x_1 \mapsto F_{2|1}(x_2|x_1)$, come funzione di $x_1 \in \mathbb{R}$, deve essere "integrabile" in senso opportuno.

Nel resto di questa sezione parleremo della distribuzione marginale della variabile casuale \tilde{X}_2 che si ottiene quando una funzione di ripartizione discreta $F_1(x_1)$ viene combinata con delle funzioni di ripartizione condizionate $F_{2|1}(x_2|x_1)$. Per concretezza focalizzeremo l'attenzione sul caso dove le funzioni di ripartizione condizionate sono assolutamente continue. Come nella sezione precedente, indicheremo le funzioni di densità condizionate con $f_{2|1}(x_2|x_1)$ in modo tale che

$$F_{2|1}(x_2|x_1) = \int_{-\infty}^{x_2} f_{2|1}(t_2|x_1) dt_2, \quad x_2 \in \mathbb{R}. \quad (55)$$

Consideriamo dunque due variabili casuali \tilde{X}_1 e \tilde{X}_2 , e supponiamo che la funzione di ripartizione marginale della variabile casuale \tilde{X}_1 (che come al solito indichiamo con $F_1(x_1)$) sia discreta con funzione di massa di probabilità

$$p_1(x_1) = P(\{\tilde{X}_1 = x_1\}), \quad x_1 \in \mathbb{R}.$$

Supponiamo inoltre che per ogni $x_1 \in S_1$ (come al solito S_1 indica il supporto della funzione di probabilità $p_1(x_1)$) la corrispondente funzione di ripartizione condizionata della variabile casuale \tilde{X}_2 sia definita come nella (55) e che dunque

$$P(\{\tilde{X}_2 \leq x_2\} | \{\tilde{X}_1 = x_1\}) = \frac{P(\{\tilde{X}_1 = x_1\} \cap \{\tilde{X}_2 \leq x_2\})}{P(\{\tilde{X}_1 = x_1\})} = F_{2|1}(x_2|x_1) = \int_{-\infty}^{x_2} f_{2|1}(t_2|x_1) dt_2$$

per ogni $x_1 \in S_1$ e ogni $x_2 \in \mathbb{R}$. Usando la formula della probabilità totale (vedi il Teorema 3.3) vediamo che sotto queste ipotesi la funzione di ripartizione marginale della variabile casuale \tilde{X}_2 deve essere data

da

$$\begin{aligned}
 F_2(x_2) &= P(\{\tilde{X}_2 \leq x_2\}) \\
 &= \sum_{x_1 \in S_1} P(\{\tilde{X}_1 = x_1\}) \times P(\{\tilde{X}_2 \leq x_2\} | \{\tilde{X}_1 = x_1\}) \\
 &= \sum_{x_1 \in S_1} p_1(x_1) \times F_{2|1}(x_2|x_1) \\
 &= \sum_{x_1 \in S_1} p_1(x_1) \times \int_{-\infty}^{x_2} f_{2|1}(t_2|x_1) dt_2 \\
 &= \int_{-\infty}^{x_2} \left[\sum_{x_1 \in S_1} p_1(x_1) \times f_{2|1}(t_2|x_1) \right] dt_2.
 \end{aligned} \tag{56}$$

La terza uguaglianza mostra che la funzione di ripartizione $F_2(x_2)$ è una media ponderata delle funzioni di ripartizione condizionate $F_{2|1}(x_2|x_1)$, e per questo motivo la funzione di ripartizione $F_2(x_2)$ viene a volte chiamata **mistura di funzioni di ripartizione**. Dall'espressione nell'ultima riga deduciamo che la funzione di ripartizione $F_2(x_2)$ è assolutamente continua e che

$$f_2(x_2) = \sum_{x_1 \in S_1} p_1(x_1) \times f_{2|1}(t_2 = x_2|x_1), \quad x_2 \in \mathbb{R},$$

è una sua funzione di densità. Si noti che $f_2(x_2)$ è la media ponderata delle funzioni di densità condizionate $f_{2|1}(x_2|x_1)$ con pesi dati dalle probabilità marginali $p_1(x_1) = P(\{\tilde{X}_1 = x_1\})$.

Come si evince dalla risposta al quesito a) del prossimo esercizio, in molte applicazioni la funzione di ripartizione marginale di una variabile casuale viene specificata sotto forma di mistura.

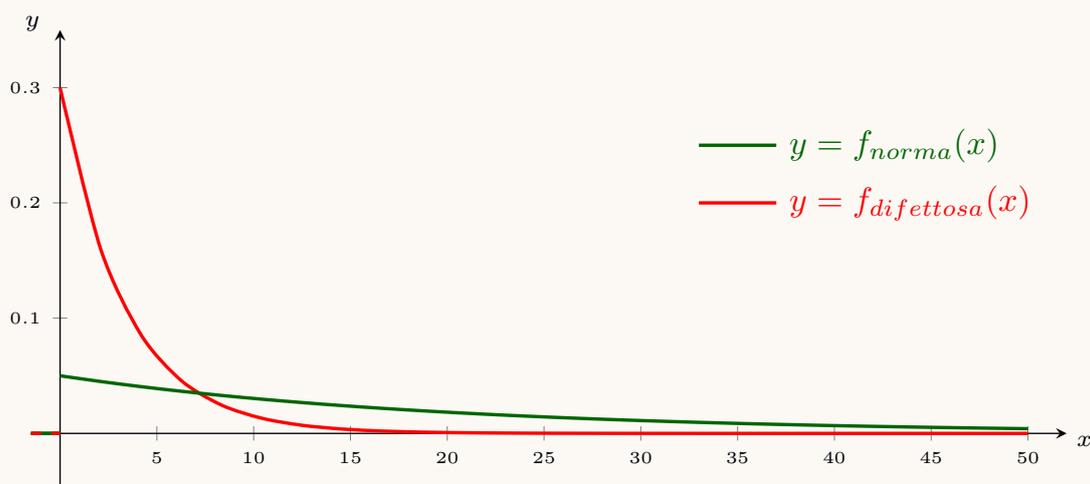
Esercizio 4.20 (Mixture). Si assuma che il 90% delle confezioni di un bene alimentare sia a norma e che in tal caso la distribuzione della durata di conservazione del bene alimentare (espressa in giorni) sia assolutamente continua con funzione di densità

$$f_{norma}(x) = \begin{cases} 0,05 \times e^{-0,05x} & \text{per } x > 0, \\ 0 & \text{altrimenti.} \end{cases}$$

Per il restante 10% di confezioni che sono difettose, si assuma invece che la distribuzione della durata di conservazione (in giorni) sia descritta dalla funzione di densità

$$f_{difettosa}(x) = \begin{cases} 0,3 \times e^{-0,3x} & \text{per } x > 0, \\ 0 & \text{altrimenti.} \end{cases}$$

Il grafico sottostante mostra l'andamento delle due funzioni di densità qui considerate:



- Si ricavi la funzione di ripartizione della variabile casuale che descrive la durata di conservazione del bene alimentare contenuto in una confezione scelta in modo casuale.
- Qual è la probabilità che la durata di conservazione del bene alimentare contenuto in una confezione scelta in modo casuale non sia superiore a 14 giorni?
- Si supponga ora che dopo l'apertura di una confezione non sia più possibile stabilire se la confezione fosse a norma o meno. Si supponga inoltre che il quattordicesimo giorno dopo la data di confezionamento, una confezione scelta in modo casuale venga aperta e che il bene alimentare all'interno della confezione risulti deperito. Si valuti se è più *probabile* che la confezione scelta fosse difettosa piuttosto che a norma.

Risposte:

- Si ricavi la funzione di ripartizione della variabile casuale che descrive la durata di conservazione del bene alimentare contenuto in una confezione scelta in modo casuale.

Per rispondere indicheremo con

- \tilde{X}_1 la variabile casuale che assume il valore 1 se la confezione scelta è difettosa, e che assume il valore 0 altrimenti;
- e indicheremo con \tilde{X}_2 la variabile casuale che descrive la durata di conservazione del bene alimentare contenuto nella confezione scelta.

Dalle ipotesi nella consegna deduciamo immediatamente che la funzione di massa di probabilità della variabile casuale \tilde{X}_1 sia definita come

$$p_1(x_1) = P(\{\tilde{X}_1 = x_1\}) = \begin{cases} 0,9 & \text{se } x_1 = 0, \\ 0,1 & \text{se } x_1 = 1, \\ 0 & \text{altrimenti.} \end{cases}$$

Inoltre, con riferimento all'ipotesi che la confezione scelta sia a norma, possiamo concludere che la corrispondente distribuzione condizionata di \tilde{X}_2 sia definita come

$$\begin{aligned} F_{2|1}(x_2|x_1 = 0) &= P(\{\tilde{X}_2 \leq x_2\}|\{\tilde{X}_1 = 0\}) \\ &= \int_{-\infty}^{x_2} f_{norma}(t) dt \\ &= \begin{cases} \int_0^{x_2} 0,05 \times e^{-0,05 \times t_2} dt_2 = [-e^{-0,05 \times t_2}]_0^{x_2} & \text{se } x_2 \geq 0 \\ 0 & \text{altrimenti.} \end{cases} \\ &= \begin{cases} 1 - e^{-0,05x_2} & \text{se } x_2 \geq 0 \\ 0 & \text{altrimenti.} \end{cases} \end{aligned}$$

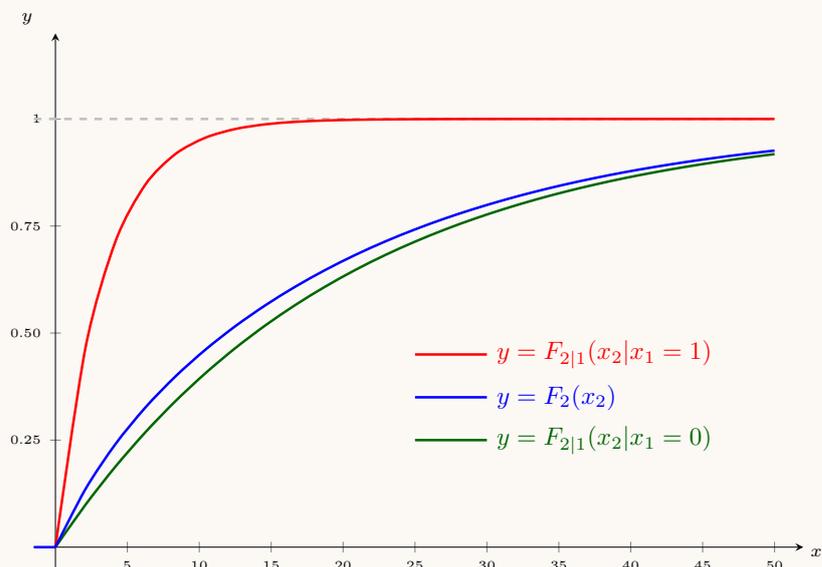
Considerando invece l'ipotesi che la confezione scelta sia difettosa, possiamo concludere che la corrispondente distribuzione condizionata di \tilde{X}_2 sia definita come

$$\begin{aligned} F_{2|1}(x_2|x_1 = 1) &= P(\{\tilde{X}_2 \leq x_2\}|\{\tilde{X}_1 = 1\}) \\ &= \int_{-\infty}^{x_2} f_{difettosa}(t) dt \\ &\quad \vdots \\ &= \begin{cases} 1 - e^{-0,3x_2} & \text{se } x_2 \geq 0 \\ 0 & \text{altrimenti.} \end{cases} \end{aligned}$$

Usando la formula della probabilità totale (così come abbiamo visto nella catena di uguaglianze (56)) vediamo quindi che la distribuzione marginale della durata di conservazione è data da

$$\begin{aligned} F_2(x_2) &= P(\{\tilde{X}_2 \leq x_2\}) \\ &= P(\{\tilde{X}_2 \leq x_2\}|\{\tilde{X}_1 = 0\}) \times P(\{\tilde{X}_1 = 0\}) + \\ &\quad + P(\{\tilde{X}_2 \leq x_2\}|\{\tilde{X}_1 = 1\}) \times P(\{\tilde{X}_1 = 1\}) \\ &= F_{2|1}(x_2|x_1 = 0) \times p_1(x_1 = 0) + F_{2|1}(x_2|x_1 = 1) \times p_1(x_1 = 1) \\ &= \begin{cases} (1 - e^{-0,05 \times x_2}) \times 0,9 + (1 - e^{-0,3 \times x_2}) \times 0,1 & \text{per } x_2 \geq 0, \\ 0 & \text{per } x_2 < 0. \end{cases} \end{aligned}$$

Il grafico sottostante mostra l'andamento delle tre funzioni di ripartizione che abbiamo appena ricavato. Si noti che la curva della distribuzione marginale (curva di colore blu) è sempre compresa tra le due distribuzioni condizionate.



- b) Qual è la probabilità che la durata di conservazione del bene alimentare contenuto in una confezione scelta in modo casuale non sia superiore a 14 giorni?

Usando la funzione di ripartizione marginale trovata al punto precedente vediamo che la probabilità richiesta è data da

$$\begin{aligned} P(\{\tilde{X}_2 \leq 14\}) &= F_2(14) \\ &= (1 - e^{-0,05 \times 14}) \times 0,9 + (1 - e^{-0,3 \times 14}) \times 0,1 \\ &= 0,453 + 0,099 = 0,552. \end{aligned}$$

Si noti che questo valore è una media ponderata dei valori

$$F_{2|1}(x_2 = 14|x_1 = 0) = P(\{\tilde{X}_2 \leq 14\}|\{\tilde{X}_1 = 0\}) = 1 - e^{-0,05 \times 14} = 0,503$$

e

$$F_{2|1}(x_2 = 14|x_1 = 1) = P(\{\tilde{X}_2 \leq 14\}|\{\tilde{X}_1 = 1\}) = 1 - e^{-0,3 \times 14} = 0,985$$

con pesi dati da $P(\{X_1 = 0\}) = 0,9$ e $P(\{X_1 = 1\}) = 0,1$, rispettivamente.

- c) Si supponga ora che dopo l'apertura di una confezione non sia più possibile stabilire se la confezione fosse a norma o meno. Si supponga inoltre che il quattordicesimo giorno dopo la data di confezionamento, una confezione

scelta in modo casuale venga aperta e che il bene alimentare all'interno della confezione risulti deperito. Si valuti se è più *probabile* che la confezione scelta fosse difettosa piuttosto che a norma.

Per rispondere confronteremo i valori delle probabilità condizionate

$$P(\{X_1 = 0\}|\{\tilde{X}_2 \leq 14\})$$

e

$$P(\{X_1 = 1\}|\{\tilde{X}_2 \leq 14\}).$$

Per calcolarle useremo la formula di Bayes:

$$P(\{X_1 = 0\}|\{\tilde{X}_2 \leq 14\}) = \frac{P(\{\tilde{X}_2 \leq 14\}|\{X_1 = 0\}) \times P(\{X_1 = 0\})}{P(\{\tilde{X}_2 \leq 14\})}$$

e

$$P(\{X_1 = 1\}|\{\tilde{X}_2 \leq 14\}) = \frac{P(\{\tilde{X}_2 \leq 14\}|\{X_1 = 1\}) \times P(\{X_1 = 1\})}{P(\{\tilde{X}_2 \leq 14\})}.$$

Nella risposta al quesito b) abbiamo già visto che

$$P(\{\tilde{X}_2 \leq 14\}|\{X_1 = 0\}) \times P(\{X_1 = 0\}) = (1 - e^{-0,05 \times 14}) \times 0,9 = 0,453,$$

$$P(\{\tilde{X}_2 \leq 14\}|\{X_1 = 1\}) \times P(\{X_1 = 1\}) = (1 - e^{-0,3 \times 14}) \times 0,1 = 0,099,$$

e che $P(\{\tilde{X}_2 \leq 14\}) = 0,552$. Sostituendo nella formula di Bayes otteniamo dunque

$$\begin{aligned} P(\{X_1 = 0\}|\{\tilde{X}_2 \leq 14\}) &= \frac{P(\{\tilde{X}_2 \leq 14\}|\{X_1 = 0\}) \times P(\{X_1 = 0\})}{P(\{\tilde{X}_2 \leq 14\})} \\ &= \frac{0,453}{0,552} = 0,821 \end{aligned}$$

e

$$\begin{aligned} P(\{X_1 = 1\}|\{\tilde{X}_2 \leq 14\}) &= \frac{P(\{\tilde{X}_2 \leq 14\}|\{X_1 = 1\}) \times P(\{X_1 = 1\})}{P(\{\tilde{X}_2 \leq 14\})} \\ &= \frac{0,099}{0,552} = 0,179. \end{aligned}$$

In virtù di questi risultati concludiamo che è molto più *probabile* che la confezione scelta fosse a norma.

Si noti che questo risultato è dovuto all'elevato valore $P(\{\tilde{X}_1 = 0\}) = 0,9$. Guardando soltanto ai valori di

$$P(\{\tilde{X}_2 \leq 14\}|\{\tilde{X}_1 = 0\}) = 0,503$$

e

$$P(\{\tilde{X}_2 \leq 14\}|\{\tilde{X}_1 = 1\}) = 0,985$$

si potrebbe essere indotti a pensare che la causa di un periodo di conservazione non superiore a 14 giorni (evento $\{\tilde{X}_2 \leq 14\}$) sia, con un grado di certezza più elevato, riferibile ad un difetto della confezione (evento $\{\tilde{X}_1 = 1\}$) contrariamente a quanto abbiamo appena dimostrato.

4.9 Variabili casuali indipendenti

Nella Sezione 3.3 abbiamo visto che per definizione due eventi A e B sono **indipendenti** se

$$P(A \cap B) = P(A) \times P(B)$$

e nella successiva Sezione 3.4 abbiamo visto che per definizione $k \geq 2$ eventi A_1, A_2, \dots, A_k sono **globalmente indipendenti** se per ogni sottocollezione finita $\{A_{i_1}, A_{i_2}, \dots, A_{i_r}\}$ ($r \leq k$) si ha

$$P\left(\bigcap_{j=1}^r A_{i_j}\right) = P(A_{i_1}) \times P(A_{i_2}) \times \dots \times P(A_{i_r}).$$

Nella prossima definizione introdurremo un concetto di indipendenza (globale) per **variabili casuali** che è analogo a quello riferito a **eventi**.

Definizione 4.7 (Variabili casuali indipendenti). \tilde{X}_1 e \tilde{X}_2 sono due variabili casuali definite nell'ambito di uno stesso modello probabilistico (Ω, \mathcal{A}, P) , allora si dice che \tilde{X}_1 e \tilde{X}_2 sono **indipendenti** se

$$P(\{\tilde{X}_1 \in B_1\} \cap \{\tilde{X}_2 \in B_2\}) = P(\{\tilde{X}_1 \in B_1\}) \times P(\{\tilde{X}_2 \in B_2\})$$

per ogni $B_1, B_2 \in \sigma(\mathcal{B})$.

Se $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_k$ sono $k \geq 2$ variabili casuali definite nell'ambito di uno stesso modello probabilistico (Ω, \mathcal{A}, P) , allora si dice che le k le variabili casuali \tilde{X}_i sono **globalmente indipendenti** se

$$\begin{aligned} P\left(\bigcap_{i=1}^k \{\tilde{X}_i \in B_i\}\right) &= P(\{\tilde{X}_1 \in B_1\}) \times P(\{\tilde{X}_2 \in B_2\}) \times \dots \times P(\{\tilde{X}_k \in B_k\}) = \\ &= \prod_{i=1}^k P(\{\tilde{X}_i \in B_i\}) \quad \text{per ogni } B_1, B_2, \dots, B_k \in \sigma(\mathcal{B}). \end{aligned} \tag{57}$$

Come vedremo nel resto di questa dispensa, variabili casuali **globalmente indipendenti** sono di fondamentale interesse per la teoria della probabilità e pertanto conviene analizza-

re più da vicino la condizione (57) che definisce **variabili casuali globalmente indipendenti**. Innanzitutto conviene osservare che se $\mathcal{X} = \{\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_k\}$ è una collezione finita di variabili casuali **globalmente indipendenti**, allora anche ogni sottocollezione di \mathcal{X} (che contiene almeno due variabili casuali) deve essere una collezione di variabili casuali **globalmente indipendenti**.

Dimostrazione. Siccome $\mathbb{R} \in \sigma(\mathcal{B})$, possiamo porre $B_k = \mathbb{R}$ e in questo modo otteniamo

$$P\left(\bigcap_{i=1}^k \{\tilde{X}_i \in B_i\}\right) = P\left(\bigcap_{i=1}^{k-1} \{\tilde{X}_i \in B_i\}\right) \quad \text{perché } \{\tilde{X}_k \in \mathbb{R}\} = \Omega$$

e

$$\prod_{i=1}^k P(\{\tilde{X}_i \in B_i\}) = \prod_{i=1}^{k-1} P(\{\tilde{X}_i \in B_i\})$$

perché $P(\{\tilde{X}_k \in \mathbb{R}\}) = P(\Omega) = [\text{assioma K1}] = 1$.

Da questo ragionamento deduciamo che se le variabili casuali $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_k$ sono **globalmente indipendenti**, allora devono essere **globalmente indipendenti** anche le prime $k - 1$ variabili casuali. Scambiando il ruolo della variabile casuale \tilde{X}_k con quello di una qualunque altra variabile casuale \tilde{X}_i e/o iterando il precedente ragionamento, si vede dunque che se $\mathcal{X} = \{\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_k\}$ è una collezione finita di variabili casuali **globalmente indipendenti**, allora anche ogni sottocollezione di \mathcal{X} (che contiene almeno due variabili casuali) deve essere una collezione di variabili casuali **globalmente indipendenti**. \square

Sempre con riferimento alla definizione di **variabili casuali globalmente indipendenti**, conviene aggiungere un'altra osservazione: **così come nel caso di eventi l'indipendenza a coppie non implica l'indipendenza globale, lo stesso vale anche con riferimento a variabili casuali** (invitiamo il lettore a costruire un esempio che confermi questa affermazione). Tuttavia, quando nelle applicazioni si considerano **eventi** o **variabili casuali** "indipendenti", si considerano sempre (salvo indicazione contraria) eventi o variabili casuali che sono "globalmente indipendenti" e per questo motivo la parola "globalmente" viene quasi sempre omessa. D'ora in poi seguiremo questa prassi anche in questa dispensa: **d'ora in poi, quando parleremo di eventi o variabili casuali "indipendenti" faremo sempre riferimento a eventi o variabili casuali che sono "globalmente indipendenti"**.

A questo punto ci si potrebbe chiedere **quali fenomeni reali dovrebbero essere descritti attraverso variabili casuali indipendenti**. La risposta a questa domanda è semplice: tutti i fenomeni di cui pensiamo che non si influenzino a vicenda. Nell'ambito di un esperimento casuale che si svolge in più passi dove riteniamo che gli esiti dei singoli passi non si influenzino a vicenda (vedi per esempio l'Esercizio 3.3), le variabili casuali che descrivono gli esiti di passi diversi dovrebbero essere indipendenti. D'altra parte, dovrebbero essere indipendenti anche variabili casuali che si riferiscono a parti "fisicamente separate" di un esperimento casuale. Per esempio, le variabili casuali che descrivono le durate di funzionamento di k macchinari che operano in modo "indipendente" l'uno

dall'altro dovrebbero essere indipendenti. Infine, in molte applicazioni viene ipotizzato che "errori di previsione" di varia natura non si influenzino a vicenda e questi errori di previsione vengono pertanto descritti attraverso variabili casuali indipendenti.

Avendo chiarito quali fenomeni reali dovrebbero essere descritti attraverso variabili casuali indipendenti, conviene anche chiarire come è definita la **funzione di ripartizione congiunta** di k variabili casuali che sono **indipendenti**. Per rispondere a questa domanda osserviamo in primo luogo che in virtù della condizione (57), k variabili casuali possono essere indipendenti **solo se**

$$F_{12\dots k}(x_1, x_2, \dots, x_k) = P\left(\bigcap_{i=1}^k \{\tilde{X}_i \leq x_i\}\right) = \prod_{i=1}^k P\left(\{\tilde{X}_i \leq x_i\}\right) = \prod_{i=1}^k F_i(x_i) \quad (58)$$

per ogni $(x_1, x_2, \dots, x_k) \in \mathbb{R}^k$. Ma è anche vero il viceversa? Ovvero, è sempre vero che la condizione (58) implica l'indipendenza delle k variabili casuali \tilde{X}_i ? Si noti che questa domanda solleva anche un'altra questione: è sempre vero che un prodotto di k funzioni di ripartizione univariate dà luogo ad una funzione di ripartizione congiunta? I prossimi due teoremi (che non dimostreremo) contengono le risposte ad entrambe le domande. Si noti che entrambe le risposte sono affermative.

Teorema 4.8 (Variabili casuali indipendenti). k variabili casuali $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_k$ sono indipendenti se e solo se la loro funzione di ripartizione congiunta è uguale al prodotto delle loro funzioni di ripartizione marginali così come previsto nella condizione (58).

Teorema 4.9 (Teorema di esistenza per variabili casuali indipendenti). Se $F_1(x_1), F_2(x_2), \dots, F_k(x_k)$ sono k funzioni di ripartizione univariate, ovvero k funzioni che soddisfano le proprietà **F0 - F2**, allora esistono un modello probabilistico (Ω, \mathcal{A}, P) e k corrispondenti variabili casuali $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_k$ che sono indipendenti e le cui funzioni di ripartizione marginali sono date dalle k funzioni $F_i(x_i)$.

Se le k funzioni di ripartizione sono tutte identiche, allora si dice che le variabili casuali $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_k$ sono **indipendenti e identicamente distribuite (acronimo i.i.d.)** e per indicare questo fatto si può scrivere

$$\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_k \sim i.i.d.$$

Esercizio 4.21. Si considerino di nuovo le variabili casuali \tilde{X}_1 e \tilde{X}_2 dell'Esercizio

4.13. Ricordiamo che la loro funzione di ripartizione congiunta è definita come

$$F_{12}(x_1, x_2) = \begin{cases} 0 & \text{se } x_1 \leq 0 \text{ oppure } x_2 \leq 0, \\ 1 - e^{-x_1} - \frac{1}{x_2+1} + \frac{e^{-x_1(x_2+1)}}{x_2+1} & \text{se } x_1 > 0 \text{ e } x_2 > 0 \end{cases}$$

e che le corrispondenti funzioni di ripartizione marginali sono rispettivamente date da (vedi la risposta al quesito b))

$$F_1(x_1) = P(\{\tilde{X}_1 \leq x_1\}) = \begin{cases} 0 & \text{se } x_1 \leq 0 \\ 1 - e^{-x_1} & \text{se } x_1 > 0. \end{cases}$$

e da

$$F_2(x_2) = P(\{\tilde{X}_2 \leq x_2\}) = \begin{cases} 0 & \text{se } x_2 \leq 0 \\ 1 - \frac{1}{x_2+1} & \text{se } x_2 > 0. \end{cases}$$

Si verifichi se le variabili casuali \tilde{X}_1 e \tilde{X}_2 sono indipendenti.

Risposta:

Siccome il prodotto delle funzioni di ripartizione marginali

$$\begin{aligned} F_1(x_1) \times F_2(x_2) &= \begin{cases} 0 & \text{se } x_1 \leq 0 \text{ oppure } x_2 \leq 0, \\ (1 - e^{-x_1}) \left(1 - \frac{1}{x_2+1}\right) & \text{se } x_1 > 0 \text{ e } x_2 > 0 \end{cases} \\ &= \begin{cases} 0 & \text{se } x_1 \leq 0 \text{ oppure } x_2 \leq 0, \\ 1 - e^{-x_1} - \frac{1}{x_2+1} + \frac{e^{-x_1}}{x_2+1} & \text{se } x_1 > 0 \text{ e } x_2 > 0 \end{cases} \end{aligned}$$

è diverso dalla funzione di ripartizione congiunta $F_{12}(x_1, x_2)$ (si noti che l'ultimo termine della somma è diverso), possiamo concludere che le variabili casuali \tilde{X}_1 e \tilde{X}_2 non sono indipendenti (si noti che per giungere a questa conclusione abbiamo applicato il Teorema 4.8).

Esercizio 4.22. Siano \tilde{X}_1 e \tilde{X}_2 due variabili casuali indipendenti con funzioni di ripartizione marginali rispettivamente date da

$$F_1(x_1) = P(\{\tilde{X}_1 \leq x_1\}) = \begin{cases} 0 & \text{se } x_1 \leq 0 \\ 1 - e^{-x_1} & \text{se } x_1 > 0. \end{cases}$$

e da

$$F_2(x_2) = P(\{\tilde{X}_2 \leq x_2\}) = \begin{cases} 0 & \text{se } x_2 \leq 0 \\ 1 - \frac{1}{x_2+1} & \text{se } x_2 > 0. \end{cases}$$

- a) Qual è la probabilità che nessuna delle due variabili casuali assuma un valore maggiore di 2?
- b) Qual è la probabilità che almeno una delle due variabili casuali assuma un valore maggiore di 2?
- c) Si calcoli il valore di $P(\{0,5 < \tilde{X}_1 \leq 2\} \cap \{1 < \tilde{X}_2 \leq 3\})$.

La soluzione viene lasciata per esercizio.

Per concludere questa sezione introduttiva sull'indipendenza di variabili casuali, riportiamo un altro teorema che ci servirà in alcuni ragionamenti che faremo più avanti. L'enunciato del teorema è perfettamente ovvio, ma una sua dimostrazione sarebbe piuttosto lunga e tecnica e quindi non lo dimostreremo.

Teorema 4.10. Se $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_k$ sono k variabili casuali indipendenti e se

- \tilde{Y}_1 è una funzione (misurabile) delle prime k_1 variabili casuali \tilde{X}_i :

$$\tilde{Y}_1 = g_1(\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_{k_1});$$

- \tilde{Y}_2 è una funzione (misurabile) delle successive $k_2 - k_1$ variabili casuali \tilde{X}_i :

$$\tilde{Y}_2 = g_2(\tilde{X}_{k_1+1}, \tilde{X}_{k_1+2}, \dots, \tilde{X}_{k_2});$$

- ...

- \tilde{Y}_r è una funzione delle ultime $k - k_1 - k_2 - \dots - k_{r-1}$ variabili casuali \tilde{X}_i :

$$\tilde{Y}_r = g_r(\tilde{X}_{k_{r-1}+1}, \tilde{X}_{k_{r-1}+2}, \dots, \tilde{X}_k)$$

(si noti che nessuna delle k variabili casuali \tilde{X}_i concorre alla definizione di più di una delle r variabili casuali \tilde{Y}_j), allora anche le variabili casuali $\tilde{Y}_1, \tilde{Y}_2, \dots, \tilde{Y}_r$ devono essere indipendenti.

4.9.1 Variabili casuali indipendenti: il caso discreto

Supponiamo ora che $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_k$ siano k **variabili casuali discrete** che sono **indipendenti**. Dalla Definizione 4.7 discende immediatamente che in questo caso la loro **funzione di massa di probabilità congiunta** deve essere data da

$$p_{12\dots k}(x_1, x_2, \dots, x_k) = P\left(\bigcap_{i=1}^k \{\tilde{X}_i = x_i\}\right) = \prod_{i=1}^k P(\{\tilde{X}_i = x_i\}) = \prod_{i=1}^k p_i(x_i) \quad (59)$$

per ogni $(x_1, x_2, \dots, x_k) \in \mathbb{R}^k$. Viceversa, si può anche dimostrare che se la **funzione di massa di probabilità congiunta** di k variabili casuali discrete (di cui a priori non sappiamo se sono indipendenti o meno) è uguale al prodotto delle corrispondenti **funzioni di massa di probabilità marginali**, allora le k variabili casuali devono essere **indipendenti**. Il prossimo teorema riassume quanto abbiamo appena detto:

Teorema 4.11 (Variabili casuali discrete indipendenti). k variabili casuali discrete $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_k$ sono indipendenti se e solo se la loro funzione di massa di probabilità congiunta è uguale al prodotto delle loro funzioni di massa di probabilità marginali così come previsto nella condizione (59).

Esercizio 4.23. Si consideri di nuovo l'esperimento casuale dell'Esercizio 3.3 che, ricordiamo, consiste nel

- (passo 1) lancio di una moneta,
- (passo 2) con successivo lancio di un tetraedro ...
- (passo 3) e termina con il lancio di un dado.

Si indichi con

- \tilde{X}_1 la variabile casuale che assume il valore 1 se l'esito del lancio della moneta è testa, e che assume il valore 0 altrimenti;
- \tilde{X}_2 la variabile casuale che descrive il numero ottenuto nel lancio del tetraedro;
- \tilde{X}_3 la variabile casuale che descrive il numero ottenuto nel lancio dado.

Si verifichi se le variabili casuali \tilde{X}_1, \tilde{X}_2 e \tilde{X}_3 sono indipendenti.

Risposta:

Per rispondere al quesito faremo riferimento al modello probabilistico (Ω, \mathcal{A}, P) che abbiamo definito nella soluzione dell'Esercizio 3.3. Quindi consideriamo come ...

- **spazio campionario** Ω l'insieme di tutte le terne $\omega = (\omega_1, \omega_2, \omega_3)$ dove

- ω_1 può essere T (come testa) oppure C (come croce);
- ω_2 può essere uno dei numeri naturali 1, 2, 3 oppure 4;
- e dove ω_3 può essere uno dei numeri naturali 1, 2, 3, 4, 5 oppure 6;
- come **spazio di eventi** \mathcal{A} l'insieme delle parti di Ω ;
- e come **funzione di probabilità** $P(\cdot)$ la funzione di probabilità corrispondente al metodo di assegnazione classico:

$$P(A) = \frac{\#A}{\#\Omega} = \frac{\#A}{2 \times 4 \times 6}, \quad \text{per ogni } A \subseteq \Omega.$$

Non è difficile verificare che con riferimento a questo modello probabilistico le funzioni di massa di probabilità marginali delle variabili casuali \tilde{X}_1 , \tilde{X}_2 e \tilde{X}_3 sono rispettivamente date da

$$p_1(x_1) = P(\{\tilde{X}_1 = x_1\}) = \begin{cases} 1/2 & \text{per } x_1 = 0, 1; \\ 0 & \text{altrimenti,} \end{cases}$$

$$p_2(x_2) = P(\{\tilde{X}_2 = x_2\}) = \begin{cases} 1/4 & \text{per } x_2 = 1, 2, 3, 4; \\ 0 & \text{altrimenti,} \end{cases}$$

$$p_3(x_3) = P(\{\tilde{X}_3 = x_3\}) = \begin{cases} 1/6 & \text{per } x_3 = 0, 1, 2, 4, 5, 6; \\ 0 & \text{altrimenti.} \end{cases}$$

e che la loro funzione di massa di probabilità congiunta è data da

$$\begin{aligned} p_{123}(x_1, x_2, x_3) &= P\left(\left\{\tilde{X}_1 = x_1\right\} \cap \left\{\tilde{X}_2 = x_2\right\} \cap \left\{\tilde{X}_3 = x_3\right\}\right) \\ &= \frac{\#\left(\left\{\tilde{X}_1 = x_1\right\} \cap \left\{\tilde{X}_2 = x_2\right\} \cap \left\{\tilde{X}_3 = x_3\right\}\right)}{\#\Omega} \\ &= \begin{cases} \frac{1}{2 \times 4 \times 6} & \text{se } x_1 = 1, 2, \quad x_2 = 1, 2, 3, 4 \quad \text{e} \quad x_3 = 1, 2, 3, 4, 5, 6; \\ 0 & \text{altrimenti.} \end{cases} \end{aligned}$$

Siccome (la verifica è semplice)

$$p(x_1, x_2, x_3) = p_1(x_1) \times p_2(x_2) \times p_3(x_3) \quad \text{per ogni } (x_1, x_2, x_3) \in \mathbb{R}^3,$$

la condizione necessaria e sufficiente del Teorema 4.11 è soddisfatta e quindi possiamo concludere che le tre variabili casuali \tilde{X}_1 , \tilde{X}_2 e \tilde{X}_3 sono indipendenti.

Esercizio 4.24. Si consideri una successione di quattro lanci indipendenti di una moneta.

- Qual è la probabilità di ottenere testa in tutti e quattro i lanci?
- Qual è la probabilità di ottenere testa in entrambi i primi due lanci e di ottenere croce negli ultimi due lanci?
- Qual è la probabilità di ottenere due volte testa e due volte croce indipendentemente dall'ordine in cui si presentano questi esiti?

Risposte:

Nelle risposte indicheremo con \tilde{X}_i la variabile casuale che assume il valore 1 se l'esito dell' i -esimo lancio è testa, e che assume il valore zero altrimenti. Siccome per ipotesi i cinque lanci sono "indipendenti", assumeremo che le 4 variabili casuali \tilde{X}_i siano i.i.d. e che la funzione di probabilità comune ad esse sia data da

$$p(x) = P(\{\tilde{X}_i = x\}) = \begin{cases} 0,5 & \text{se } x = 0, 1 \\ 0 & \text{altrimenti.} \end{cases}$$

- Qual è la probabilità di ottenere testa in tutti e quattro i lanci?

Dall'ipotesi di indipendenza discende che la probabilità di ottenere sempre testa è data da

$$P\left(\bigcap_{i=1}^4 \{\tilde{X}_i = 1\}\right) = \prod_{i=1}^4 P(\{\tilde{X}_i = 1\}) = \prod_{i=1}^4 p(1) = [p(1)]^4 = 0,5^4 = 0,0625.$$

- Qual è la probabilità di ottenere testa in entrambi i primi due lanci e di ottenere croce negli ultimi due lanci?

Usando ancora l'ipotesi di indipendenza vediamo che la probabilità richiesta è data da

$$\begin{aligned} & P\left(\{\tilde{X}_1 = 1\} \cap \{\tilde{X}_2 = 1\} \cap \{\tilde{X}_3 = 0\} \cap \{\tilde{X}_4 = 0\}\right) = \\ & = P(\{\tilde{X}_1 = 1\}) \times P(\{\tilde{X}_2 = 1\}) \times P(\{\tilde{X}_3 = 0\}) \times P(\{\tilde{X}_4 = 0\}) \\ & = 0,5 \times 0,5 \times 0,5 \times 0,5 = 0,5^4 = 0,0625. \end{aligned}$$

- Qual è la probabilità di ottenere due volte testa e due volte croce indipendentemente dall'ordine in cui si presentano questi esiti?

Scambiando i pedici delle variabili casuali \tilde{X}_i nella risposta al quesito precedente ci rendiamo facilmente conto che la probabilità di ottenere due volte

testa e due volte croce **in un determinato ordine** è sempre uguale a 0,0625 qualunque sia l'ordine in cui si presentano le teste e le croci. Siccome ci sono $\binom{4}{2} = 4!/(2! \times 2!) = 6$ modi per ottenere due volte testa e due volte croce ($\binom{4}{2} = 6$ è infatti il numero di modi per scegliere due dei quattro eventi $\{\tilde{X}_i = x\}$ dove porre $x = 1$), e siccome ciascuno di questi modi è incompatibile con tutti gli altri, possiamo concludere che

$$\begin{aligned} P(\{\text{si ottiene 2 volte testa e 2 volte croce}\}) &= \\ &= \binom{4}{2} \times 0,5^4 = 6 \times 0,0625 = 0,375. \end{aligned}$$

Si noti che l'evento

{si ottiene 2 volte testa e 2 volte croce}

può essere anche indicato come

$$\left\{ \sum_{i=1}^4 \tilde{X}_i = 2 \right\}.$$

4.9.2 Variabili casuali indipendenti: il caso assolutamente continuo

Il seguente teorema (che non dimostreremo) contiene condizioni necessarie e sufficienti affinché k variabili **assolutamente continue** siano **indipendenti**.

Teorema 4.12 (Variabili casuali assolutamente continue indipendenti). k variabili casuali assolutamente continue sono indipendenti **se e solo se**

- la loro distribuzione congiunta è assolutamente continua
- e per qualsiasi scelta di k funzioni di densità marginali $f_1(x_1), f_2(x_2), \dots, f_k(x_k)$, la funzione

$$f_{12\dots k}(x_1, x_2, \dots, x_k) = \prod_{i=1}^k f_i(x_i), \quad (x_1, x_2, \dots, x_k) \in \mathbb{R}^k,$$

è una funzione di densità congiunta delle k variabili casuali.

Esercizio 4.25. Si considerino due variabili casuali \tilde{X}_1 e \tilde{X}_2 che sono assolutamente

continue e anche indipendenti. Si assuma che

$$f_1(x_1) = \begin{cases} 0 & \text{se } x_1 \leq 0 \\ e^{-x_1} & \text{se } x_1 > 0 \end{cases}$$

sia una funzione di densità della variabile casuale \tilde{X}_1 , e che

$$f_2(x_2) = \begin{cases} 0 & \text{se } x_2 \leq 0 \\ \frac{1}{(x_2+1)^2} & \text{se } x_2 > 0 \end{cases}$$

sia una funzione di densità della variabile casuale \tilde{X}_2 .

- Qual è la probabilità che nessuna delle due variabili casuali assuma un valore maggiore di 2?
- Qual è la probabilità che almeno una delle due variabili casuali assuma un valore maggiore di 2?
- Si calcoli il valore di $P(\{0,5 < \tilde{X}_1 \leq 2\} \cap \{1 < \tilde{X}_2 \leq 3\})$.
- Si ricavi l'espressione di una funzione di densità congiunta.

La soluzione viene lasciata per esercizio.

4.10 Successioni infinite di variabili casuali indipendenti

Anche se nel mondo reale un esperimento casuale non può essere ripetuto **infinite** volte, alcuni esperimenti casuali possono essere ripetuti un **numero arbitrario di volte** e quindi potrebbe essere interessante analizzare dal punto di vista teorico che cosa accade se il numero di repliche può essere aumentato a piacimento. Per fare alcuni esempi di esperimenti casuali che possono essere ripetuti un numero arbitrario di volte, basti pensare ad una successione di lanci di una moneta o di un dado, oppure ad una successione di misurazioni di una determinata grandezza fisica. Non è difficile intuire che per analizzare esperimenti di questo tipo bisogna fare riferimento ad un modello probabilistico (Ω, \mathcal{A}, P) che può **"ospitare"** infinite variabili casuali che sono indipendenti. Ma che cosa si intende per **infinite** variabili casuali indipendenti? (si ricordi che per ora abbiamo soltanto definito che cosa si intende per un numero **finito** di variabili casuali indipendenti - vedi la Definizione 4.7). La risposta questa domanda è quasi ovvia:

Definizione 4.8 (Collezioni infinite di variabili casuali indipendenti). Se \mathcal{X} è una **collezione infinita** di variabili casuali che sono definite nell'ambito di uno stesso

modello probabilistico, allora si dice che \mathcal{X} è una **collezione di variabili casuali indipendenti** se ogni sottocollezione finita di \mathcal{X} è una collezione di variabili casuali globalmente indipendenti.

Avendo chiarito che cosa si intende per una collezione infinita di variabili casuali indipendenti, si apre un altro dubbio: esistono modelli probabilistici che possono "ospitare" una collezione infinita di variabili casuali indipendenti? ... oppure l'ipotesi di infinite variabili casuali indipendenti introduce una contraddizione con gli assiomi di Kolmogorov? (si ricordi che in questo caso l'ipotesi di infinite variabili casuali indipendenti potrebbe essere usata per dimostrare qualsiasi affermazione). Le risposte a queste domande possono essere dedotte dal famoso **teorema di esistenza di Kolmogorov** di cui il prossimo teorema (che si riferisce soltanto a collezioni formate da un'infinità **numerabile** di variabili casuali) può essere considerato come un corollario.

Teorema 4.13 (Teorema di esistenza per successioni infinite di variabili casuali indipendenti). Se $F_1(\cdot), F_2(\cdot), \dots$ è una successione infinita di funzioni di ripartizione univariate, allora esistono un modello probabilistico (Ω, \mathcal{A}, P) e una successione infinita di variabili casuali

$$\tilde{X}_1 : \Omega \mapsto \mathbb{R}, \quad \tilde{X}_2 : \Omega \mapsto \mathbb{R}, \quad \dots$$

che sono indipendenti e le cui funzioni di ripartizione marginali sono date dalle funzioni $F_1(x), F_2(x), \dots$

Se le funzioni di ripartizione $F_i(\cdot)$ sono tutte identiche, allora si dice che $\tilde{X}_1, \tilde{X}_2, \dots$ è una **successione di variabili casuali indipendenti e identicamente distribuite (i.i.d)**. Per indicare che le variabili casuali di una determinata successione sono i.i.d. si può scrivere

$$\tilde{X}_1, \tilde{X}_2, \dots \sim i.i.d.$$

In questa dispensa non faremo spesso riferimento al precedente teorema: di fatto ci servirà soltanto per confermare l'esistenza di determinati modelli probabilistici a partire dai quali ricaveremo delle distribuzioni notevoli.

5 Valore atteso

Il **valore atteso** è la controparte della media aritmetica nella statistica descrittiva. Infatti, così come la media aritmetica di un carattere quantitativo Y può essere interpretata come baricentro della distribuzione di frequenze di Y , il valore atteso di una variabile

casuale $\tilde{Y} : \Omega \mapsto \mathbb{R}$ può, come vedremo tra breve, essere interpretato come il baricentro della **distribuzione di probabilità** di \tilde{Y} .

Solitamente, nei libri di testo avanzati, la definizione generale di valore atteso è basata sull'**integrale di Lebesgue**, ma siccome in questa dispensa non vogliamo fare riferimento a questo tipo di integrale, forniremo come **definizione** di valore atteso soltanto delle **formule** per calcolare il valore atteso nei casi in cui la variabile casuale \tilde{Y} di riferimento è

- una variabile casuale **discreta** o una funzione di un numero finito di variabili casuali discrete,
- una variabile casuale **assolutamente continua** oppure una funzione di un numero finito di variabili casuali con **distribuzione congiunta assolutamente continua**.

5.1 Il valore atteso nel caso discreto

Definizione 5.1 (Valore atteso di una variabile casuale discreta). Se \tilde{X} è una variabile casuale discreta con funzione di massa di probabilità $p(x)$ e supporto S , allora diciamo che il **valore atteso** $E(\tilde{X})$ "esiste" se

$$\sum_{x \in S} |x| \times p(x) < \infty, \quad (60)$$

e in tal caso definiamo il valore numerico di $E(\tilde{X})$ come

$$E(\tilde{X}) = \sum_{x \in S} x \times p(x). \quad (61)$$

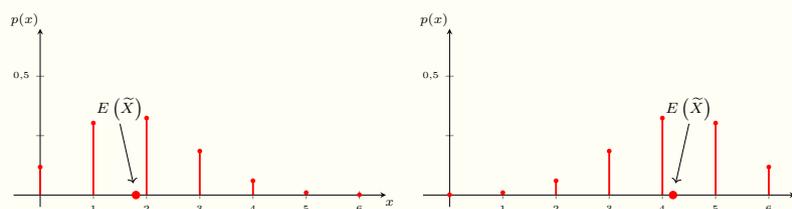
Se invece la sommatoria nella (60) non converge, allora diciamo che il valore atteso $E(\tilde{Y})$ "non esiste".

Si noti che l'*esistenza* e il *valore numerico* del valore atteso di una variabile casuale **discreta** dipendono soltanto dalla sua **funzione di massa di probabilità**. Come vedremo tra breve, lo stesso vale anche per il valore atteso di una variabile casuale **assolutamente continua** e si può dimostrare che anche **in generale l'esistenza e il valore numerico del valore atteso di una qualunque variabile casuale (discreta, assolutamente continua oppure di altro altro tipo)** dipendono soltanto dall'**andamento della funzione di ripartizione della variabile casuale**. Pertanto sarebbe più corretto parlare del valore atteso di una **funzione di ripartizione** o di una **distribuzione** piuttosto che del valore atteso di una variabile casuale. Ma siccome si incontrano spesso delle situazioni dove la distribuzione di una variabile casuale è ignota, spesso risulta più naturale parlare del valore atteso di una variabile casuale piuttosto che del valore atteso della sua

distribuzione. Nel prosieguo di questa dispensa useremo, a seconda della convenienza, entrambi i modi di dire.

Come si desume dalla formula (61), il **valore atteso** di una variabile casuale **discreta**, se esiste, è la media ponderata dei valori x che appartengono al suo supporto dove i pesi sono dati dalle probabilità $p(x)$ associate a tali valori. Dal punto di vista grafico il valore atteso di una variabile casuale discreta può quindi essere interpretato come baricentro della sua **funzione di massa di probabilità** (vedi Figura 5.1).

Figura 5.1. I due grafici mostrano i valori attesi di due distribuzioni discrete.



Con riferimento all'esistenza del valore atteso di una variabile casuale **discreta**, conviene fare alcune osservazioni:

- Innanzitutto osserviamo che la condizione di esistenza (60), che per comodità riscriviamo

$$\sum_{x \in S} |x| \times p(x) < +\infty,$$

è soddisfatta se e solo se sono soddisfatte entrambe le disuguaglianze

$$\sum_{x \in S: x < 0} x \times p(x) > -\infty \quad \text{e} \quad \sum_{x \in S: x \geq 0} x \times p(x) < +\infty. \quad (62)$$

Se $E(\tilde{X})$ esiste, il valore numerico di $E(\tilde{X})$ può dunque essere calcolato come²⁰

$$E(\tilde{X}) = \sum_{x \in S} x \times p(x) = \sum_{x \in S: x < 0} x \times p(x) + \sum_{x \in S: x \geq 0} x \times p(x).$$

D'altra parte, se almeno una delle due disuguaglianze (62) non è soddisfatta, allora il valore numerico di $E(\tilde{X})$, calcolato secondo la formula

$$E(\tilde{X}) = \sum_{x \in S: x < 0} x \times p(x) + \sum_{x \in S: x \geq 0} x \times p(x),$$

sarebbe, a seconda dei casi, dato da $E(\tilde{X}) = +\infty$, $E(\tilde{X}) = -\infty$ oppure dalla forma indeterminata $E(\tilde{X}) = +\infty - \infty$. Tuttavia, la Definizione 5.1 esclude la possibilità che il valore numerico sia infinito oppure indeterminato prevedendo la possibilità

²⁰Si ricordi che il limite di una serie assolutamente convergente non dipende dall'ordine in cui si presentano i suoi termini.

che il valore atteso non esista. A proposito dell'esistenza conviene aggiungere che in alcuni libri di testo la definizione del valore atteso è estesa anche al caso dove soltanto una delle due serie nella (62) è convergente, e in tal caso il valore numerico del valore atteso viene definito come $E(\tilde{X}) = \pm\infty$ a seconda di quale delle due serie nella (62) è quella che converge. Salvo esplicita indicazione contraria, in questa dispensa non seguiremo questa impostazione e rimarremo fedeli alla Definizione 5.1.

- Non è difficile rendersi conto che la condizione di esistenza

$$\sum_{x \in S} |x| \times p(x) < \infty$$

è sempre soddisfatta se il supporto S è un sottoinsieme **limitato** di \mathbb{R} . Infatti, se $|x| < c < \infty$ per ogni $x \in S$, si ottiene

$$\sum_{x \in S} |x| \times p(x) \leq c \times \sum_{x \in S} p(x) = c \times 1 < \infty.$$

Siccome ogni insieme finito è anche limitato, la condizione di esistenza

$$\sum_{x \in S} |x| \times p(x) < \infty$$

è dunque sempre soddisfatta se il supporto S è un insieme finito.

Esercizio 5.1. Sia \tilde{X} una variabile casuale discreta con funzione di probabilità data da

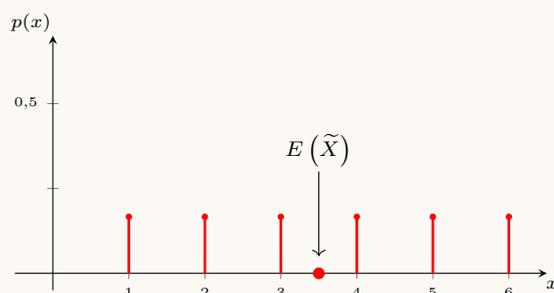
$$p(x) = \begin{cases} \frac{1}{6} & \text{per } x = 1, 2, 3, 4, 5, 6, \\ 0 & \text{altrimenti.} \end{cases}$$

Si determini il valore atteso di \tilde{X} (se esiste).

Risposta:

Siccome il supporto $S = \{1, 2, 3, 4, 5, 6\}$ è finito, possiamo concludere che il valore atteso $E(\tilde{X})$ esiste. Il valore numerico di $E(\tilde{X})$ è dato da

$$E(\tilde{X}) = \sum_{x=1}^6 \left(x \times \frac{1}{6} \right) = \frac{1}{6} \times \sum_{x=1}^6 x = \frac{1}{6} (1 + 2 + 3 + 4 + 5 + 6) = 3,5.$$



Esercizio 5.2. Sia \tilde{X} una variabile casuale discreta con funzione di probabilità data da

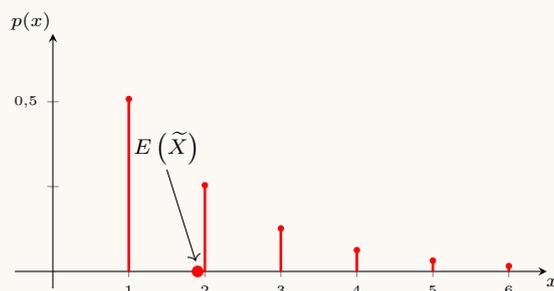
$$p(x) = \begin{cases} \frac{2^{6-x}}{63} & \text{per } x = 1, 2, 3, 4, 5, 6, \\ 0 & \text{altrimenti.} \end{cases}$$

Si determini il valore atteso di \tilde{X} .

Risposta:

Anche in questo esempio il supporto $S = \{1, 2, 3, 4, 5, 6\}$ è finito e quindi possiamo concludere che valore atteso $E(\tilde{X})$ esiste. Il valore numerico di $E(\tilde{X})$ è dato da

$$\begin{aligned} E(\tilde{X}) &= \sum_{x=1}^6 x \times \frac{2^{6-x}}{63} \\ &= 1 \times \frac{32}{63} + 2 \times \frac{16}{63} + 3 \times \frac{8}{63} + 4 \times \frac{4}{63} + 5 \times \frac{2}{63} + 6 \times \frac{1}{63} \\ &= \frac{120}{63} = 1,905. \end{aligned}$$



Nei due precedenti esempi abbiamo considerato solo funzioni di probabilità con supporto S finito e pertanto il problema dell'esistenza del valore atteso non si è mai posto. Per ottenere un caso di mancata esistenza del valore atteso basta considerare la funzione

di probabilità

$$p(x) = \begin{cases} \frac{1}{x(x+1)} = \frac{1}{x} - \frac{1}{x+1} & \text{per } x = 1, 2, 3, \dots, \\ 0 & \text{altrimenti.} \end{cases}$$

Infatti, con questa funzione di probabilità si ottiene

$$\sum_{x \in S} |x| \times p(x) = \sum_{x=1}^{\infty} x \times \frac{1}{x(x+1)} = \sum_{x=1}^{\infty} \frac{1}{x+1}$$

e, com'è noto, la serie $\sum_{x=1}^{\infty} \frac{1}{x+1}$ è divergente (non dimostreremo questo fatto).

Per completare il quadro sul caso discreto dobbiamo ancora dare la definizione di valore atteso per una funzione di una o più variabili casuali discrete:

Definizione 5.2 (Valore atteso di una funzione di variabili casuali discrete). Se $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_k$ sono k (un numero finito) variabili casuali discrete con funzione di massa di probabilità congiunta $p(x_1, x_2, \dots, x_k)$ e supporto (della funzione di massa di probabilità congiunta) S , e se

$$\tilde{Y} = g(\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_k)$$

è una funzione di queste k variabili casuali, allora diciamo che il **valore atteso** $E(\tilde{Y})$ "esiste" se

$$\sum_{(x_1, x_2, \dots, x_k) \in S} |g(x_1, x_2, \dots, x_k)| \times p(x_1, x_2, \dots, x_k) < \infty,$$

e nel caso in cui esiste definiamo il valore numerico di $E(\tilde{Y})$ come

$$E(\tilde{Y}) = \sum_{(x_1, x_2, \dots, x_k) \in S} g(x_1, x_2, \dots, x_k) \times p(x_1, x_2, \dots, x_k).$$

Se d'altra parte la condizione di esistenza non è soddisfatta, allora diciamo che il **valore atteso** $E(\tilde{Y})$ "non esiste".

Il valore atteso $E(\tilde{Y})$ può essere anche indicato con $E\left(g(\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_k)\right)$.

Come il lettore potrà facilmente verificare rispondendo al quesito del prossimo esercizio, verificare l'esistenza o calcolare il valore numerico di valori attesi di funzioni di un'unica variabile casuale discreta non è molto diverso dallo stesso problema riferito ad una variabile casuale discreta alla quale non viene applicata nessuna funzione. D'altra parte, in questa dispensa la definizione di valore atteso di funzioni di più di una variabile casuale (discreta e non) ci servirà solo per il calcolo di covarianze e quindi non presenteremo degli esercizi generici che riguardano il valore atteso di una funzione di più di una singola variabile casuale.

Si considerino ancora le variabili casuali \tilde{X} degli Esercizi 5.1 e 5.2. Si verifichi, per ciascuna di tali variabili casuali, se il valore atteso $E(\tilde{X}^2)$ esiste. Si calcoli il valore numerico di $E(\tilde{X}^2)$ in caso affermativo.

La risposta viene lasciata per esercizio.

5.2 Il valore atteso nel caso assolutamente continuo

Definizione 5.3 (Valore atteso di una variabile casuale assolutamente continua). Se \tilde{X} è una variabile casuale con distribuzione assolutamente continua e se $f(\cdot)$ è una sua funzione di densità, allora diciamo che il **valore atteso** $E(\tilde{X})$ **"esiste"** se

$$\int_{-\infty}^{\infty} |x| \times f(x) dx < \infty, \quad (63)$$

e in tal caso definiamo il valore numerico di $E(\tilde{X})$ come

$$E(\tilde{X}) = \int_{-\infty}^{\infty} x \times f(x) dx. \quad (64)$$

Se invece l'integrale nella (63) non converge, allora diciamo che il valore atteso $E(\tilde{X})$ **"non esiste"**.

Si noti che, anche se solo implicitamente, questa definizione conferma quanto avevamo già osservato, ovvero che l'esistenza e il valore numerico del valore atteso di una qualsunque variabile casuale dipendono soltanto dalla distribuzione della variabile casuale: **per verificare l'esistenza e/o per calcolare il valore numerico del valore atteso di una variabile casuale assolutamente continua possiamo quindi utilizzare una qualsiasi delle sue infinite funzioni di densità e otterremo sempre lo stesso risultato!!!**

Consideriamo ora la condizione di esistenza del valore atteso. Le osservazioni che avevamo fatto nel caso discreto devono essere adattate al caso assolutamente continuo come segue:

- La condizione di esistenza (63) è soddisfatta se e solo se sono soddisfatte entrambe le disuguaglianze

$$\int_{-\infty}^0 x \times f(x) dx > -\infty \quad \text{e} \quad \int_0^{\infty} x \times f(x) dx < +\infty. \quad (65)$$

Se $E(\tilde{X})$ esiste, il valore numerico di $E(\tilde{X})$ può dunque essere calcolato come²¹

$$E(\tilde{X}) = \int_{-\infty}^0 x \times f(x)dx + \int_0^{\infty} x \times f(x)dx.$$

Anche nel caso assolutamente continuo alcuni libri di testo estendono la definizione del valore atteso $E(\tilde{X})$ al caso dove solo uno dei due integrali nella (65) converge, ma in questa dispensa non seguiremo questo tipo di impostazione.

- La condizione di esistenza

$$\int_{-\infty}^{\infty} |x| \times f(x)dx < \infty$$

è sempre soddisfatta se il **supporto della funzione di densità** $f(x)$ è un sottoinsieme limitato di \mathbb{R} . Infatti, se $f(x) = 0$ per ogni $|x| > c$, possiamo scrivere

$$\int_{-\infty}^{\infty} |x| \times f(x)dx = \int_{-c}^c |x| \times f(x)dx \leq c \times \int_{-c}^c f(x)dx = c < \infty.$$

Esercizio 5.3. Sia \tilde{X} una variabile casuale assolutamente continua con funzione di densità data da

$$f(x) = \begin{cases} e^{-x} & \text{per } x > 0 \\ 0 & \text{altrimenti.} \end{cases}$$

Si ricavi il valore atteso di \tilde{X} (se esiste).

Risposta:

Siccome la funzione di densità $f(x)$ è positiva solo per $x > 0$, possiamo concludere che

$$\int_{-\infty}^{\infty} |x| \times f(x)dx = \int_0^{\infty} x \times f(x)dx.$$

Per verificare l'esistenza di $E(\tilde{X})$ è quindi sufficiente verificare se

$$\int_0^{\infty} x \times f(x)dx = \int_0^{\infty} x \times e^{-x}dx < \infty,$$

e se questa condizione è soddisfatta, il valore dell'integrale $\int_0^{\infty} x \times e^{-x}dx$ fornirà

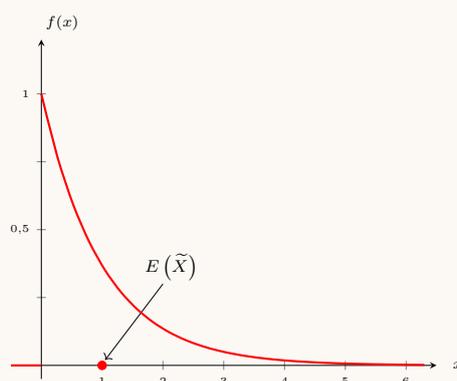
²¹Si ricordi che un integrale assolutamente convergente può essere "spezzato" in due o più parti modo arbitrario.

direttamente anche il valore numerico di $E(\tilde{X})$. Siccome

$$\begin{aligned} \int_0^{\infty} x \times e^{-x} dx &= [[\text{integrazione per parti}]] \\ &= \left[x \times \frac{e^{-x}}{-1} \right]_0^{\infty} - \int_0^{\infty} \frac{e^{-x}}{-1} dx \\ &= 0 - 0 + \int_0^{\infty} e^{-x} dx = \left[\frac{e^{-x}}{-1} \right]_0^{\infty} = 0 - (-1) = 1 < \infty, \end{aligned}$$

possiamo concludere che il valore atteso $E(\tilde{X})$ esiste e che il suo valore numerico è dato da $E(\tilde{X}) = 1$.

Nel grafico sottostante si vede che anche nel caso assolutamente continuo il valore atteso fornisce il baricentro della funzione di densità che è stata usata per calcolarlo. Come abbiamo già osservato, il valore atteso (e dunque il baricentro) è lo stesso per qualsiasi funzione di densità che è equivalente.



Esercizio 5.4. Sia \tilde{X} una variabile casuale assolutamente continua con funzione di densità data da

$$f(x) = \begin{cases} x^{-2} & \text{per } x > 1, \\ 0 & \text{altrimenti.} \end{cases}$$

Si verifichi che il valore atteso di \tilde{X} non esiste.

Risposta:

Nel caso in questione l'integrale per verificare la condizione di esistenza è dato da

$$\begin{aligned} \int_{-\infty}^{\infty} |x| \times f(x) dx &= \int_1^{\infty} x \times x^{-2} dx \\ &= \int_1^{\infty} x^{-1} dx = [\ln x]_1^{\infty} = \infty - 0 = \infty \end{aligned}$$

e siccome questo integrale è divergente, possiamo concludere che il valore atteso di \tilde{X} non esiste.

A questo punto, per completezza, daremo ancora la "definizione" di valore atteso di una funzione **di una o più** variabili casuali con distribuzione congiunta assolutamente continua anche se in questa dispensa questa definizione ci servirà soltanto per calcolare valori attesi di funzioni di una **singola** variabile casuale assolutamente continua. La definizione è perfettamente analoga a quella che abbiamo già visto nel caso discreto:

Definizione 5.4 (Valore atteso di una funzione di variabili casuali con distribuzione congiunta assolutamente continua). Se $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_k$ sono k (un numero finito) variabili casuali con distribuzione congiunta assolutamente continua e $f(x_1, x_2, \dots, x_k)$ è una loro funzione di densità congiunta, e se $g(x_1, x_2, \dots, x_k)$ è una funzione "integrabile" in senso opportuno,^a allora si dice che il valore atteso della variabile casuale $\tilde{Y} = g(\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_k)$ "esiste" se

$$\int \int \cdots \int_{\mathbb{R}^k} |g(x_1, x_2, \dots, x_k)| \times f(x_1, x_2, \dots, x_k) dx_1 dx_2 \cdots dx_k < \infty,$$

e in tal caso definiamo il valore numerico di $E(\tilde{Y})$ come

$$E(\tilde{Y}) = \int \int \cdots \int_{\mathbb{R}^k} g(x_1, x_2, \dots, x_k) \times f(x_1, x_2, \dots, x_k) dx_1 dx_2 \cdots dx_k.$$

Se d'altra parte la condizione di esistenza non è soddisfatta, allora si dice che il **valore atteso** $E(\tilde{Y})$ "non esiste".

Il valore atteso $E(\tilde{Y})$ può essere anche indicato con $E(g(\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_k))$.

^a $g(x_1, x_2, \dots, x_k)$ deve essere una funzione Riemann-integrabile se gli integrali vengono interpretati come integrali di Riemann, altrimenti, se gli integrali vengono interpretati come integrali di Lebesgue, allora $g(x_1, x_2, \dots, x_k)$ può essere una qualunque funzione misurabile. Ribadiamo che tutte le funzioni $g(x_1, x_2, \dots, x_k)$ che si incontrano nelle applicazioni sono Riemann-integrabili e quindi anche misurabili.

5.3 Proprietà del valore atteso

In questa sezione forniremo un elenco delle principali proprietà del valore atteso. Queste proprietà sono del tutto generali, ovvero non sono subordinate alla condizione che la distribuzione marginale o congiunta delle variabili casuali coinvolte sia discreta oppure assolutamente continua.

Prima di enunciare le proprietà del valore atteso conviene chiarire una piccola sottigliezza sulla quale fino ad ora abbiamo sorvolato. Fino ad ora abbiamo **definito** funzioni

di variabili casuali scrivendo semplicemente

$$\tilde{Y} = g(\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_k). \quad (66)$$

Abbiamo sempre dato per scontato che questo modo di scrivere indicasse il fatto che

$$\tilde{Y}(\omega) = g(\tilde{X}_1(\omega), \tilde{X}_2(\omega), \dots, \tilde{X}_k(\omega)) \quad \text{per ogni } \omega \in \Omega$$

e che quindi

$$P(\{\tilde{Y} = g(\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_k)\}) = P(\Omega) = [[\text{assioma K2}]] = 1.$$

Tuttavia, per interpretare in modo più estensivo i risultati fin qui esposti (ci stiamo riferendo al Teorema 4.10 e alle Definizioni 5.2 e 5.4), conviene osservare che essi rimangono validi anche se

$$\tilde{Y}(\omega) \neq g(\tilde{X}_1(\omega), \tilde{X}_2(\omega), \dots, \tilde{X}_k(\omega)) \quad \text{per qualche } \omega \in \Omega,$$

a patto che

$$P(\{\tilde{Y} \neq g(\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_k)\}) = 0.$$

Anche nel prosieguo di questa dispensa ci imbattemo in varie relazioni che coinvolgono una o più variabili casuali e quando parleremo di una tale relazione, salvo esplicita indicazione contraria, daremo sempre per scontato che essa sia soddisfatta con probabilità 1 e quindi non escluderemo la possibilità che per qualche $\omega \in \Omega$ la relazione non sia soddisfatta. Per esempio, quando scriveremo

$$\tilde{X}_1 \geq \tilde{X}_2$$

intenderemo dire che $P(\{\tilde{X}_1 \geq \tilde{X}_2\}) = 1$. Si noti dunque che la relazione $\tilde{X}_1 \geq \tilde{X}_2$ non esclude la possibilità che

$$\tilde{X}_1(\omega) < \tilde{X}_2(\omega) \quad \text{per qualche } \omega \in \Omega,$$

ma implica in ogni caso che

$$P(\{\tilde{X}_1 < \tilde{X}_2\}) = 0.$$

Allo stesso modo, quando diremo che **tra due o più variabili casuali \tilde{X}_i esiste una relazione affine (o lineare)**, intenderemo dire che esistono delle costanti a, b_1, b_2, \dots, b_k con almeno una costante b_i diversa da zero (nel caso di una relazione lineare la costante a deve essere nulla), tali che

$$P(\{a + b_1\tilde{X}_1 + b_2\tilde{X}_2 + \dots + b_k\tilde{X}_k = 0\}) = 1.$$

Si noti che anche questa condizione non esclude la possibilità che

$$a + b_1\tilde{X}_1 + b_2\tilde{X}_2 + \dots + b_k\tilde{X}_k \neq 0 \quad \text{per qualche } \omega \in \Omega,$$

ma prevede in ogni caso che

$$P(\{a + b_1\tilde{X}_1 + b_2\tilde{X}_2 + \cdots + b_k\tilde{X}_k \neq 0\}) = 0.$$

Avendo chiarito come si devono interpretare relazioni che coinvolgono variabili casuali, possiamo finalmente enunciare le principali proprietà del valore atteso. Per non appesantire troppo la trattazione non forniremo le dimostrazioni.

E0) Se a è una costante reale e $\tilde{X} = a$ (in questo caso si dice che \tilde{X} è una variabile casuale **"degenere"**), allora il valore atteso $E(\tilde{X})$ esiste e il suo valore numerico è dato da $E(\tilde{X}) = a$.

E1) (Proprietà di monotonia). Se entrambi i valori attesi $E(\tilde{X}_1)$ e $E(\tilde{X}_2)$ esistono e se $\tilde{X}_1 \geq \tilde{X}_2$, allora $E(\tilde{X}_1) \geq E(\tilde{X}_2)$. Questa proprietà rimane valida se al posto delle disuguaglianze deboli " \geq " si sostituiscono delle disuguaglianze forti " $>$ ".

E2) (Valore atteso di una trasformazione affine). Se $E(\tilde{X})$ esiste e $\tilde{Y} = a + b\tilde{X}$, allora deve esistere anche $E(\tilde{Y})$ e

$$E(\tilde{Y}) = a + bE(\tilde{X}).$$

E3) (Disuguaglianza di Jensen). Se $g(\cdot)$ è una funzione **convessa** e se entrambi i valori attesi $E(\tilde{X})$ e $E(g(\tilde{X}))$ esistono, allora si deve avere

$$g(E(\tilde{X})) \leq E(g(\tilde{X})).$$

D'altra parte, se $g(x)$ è una funzione **concava** e se entrambi i valori attesi $E(\tilde{X})$ e $E(g(\tilde{X}))$ esistono, allora si deve avere

$$g(E(\tilde{X})) \geq E(g(\tilde{X})).$$

Entrambe le disuguaglianze sono **strette** se \tilde{X} non è una variabile casuale degenere e la funzione $g(\cdot)$ è **strettamente convessa** ovvero **strettamente concava**.

Applicando la disuguaglianza di Jensen (proprietà E3) con $g(x) = |x|$ e $g(x) = x^2$ si ottengono le due disuguaglianze

$$|E(\tilde{X})| \leq E(|\tilde{X}|) \quad \text{e} \quad [E(\tilde{X})]^2 \leq E(\tilde{X}^2)$$

(a patto che i valori attesi coinvolti esistano).

E4) (Valore atteso di somme). Se entrambi i valori attesi $E(\tilde{X}_1)$ e $E(\tilde{X}_2)$ esistono e se $\tilde{Y} = \tilde{X}_1 + \tilde{X}_2$, allora deve esistere anche $E(\tilde{Y})$ e

$$E(\tilde{Y}) = E(\tilde{X}_1) + E(\tilde{X}_2).$$

Più in generale, se $E(\tilde{X}_i)$ esiste per ogni $i = 1, 2, \dots, k$ e se

$$\tilde{Y} = \tilde{X}_1 + \tilde{X}_2 + \dots + \tilde{X}_k,$$

allora deve esistere anche $E(\tilde{Y})$ e

$$E(\tilde{Y}) = E(\tilde{X}_1) + E(\tilde{X}_2) + \dots + E(\tilde{X}_k).$$

E' importante tenere presente che il valore atteso $E\left(\sum_{i=1}^k \tilde{X}_i\right)$ potrebbe esistere anche se almeno uno dei k valori attesi $E(\tilde{X}_i)$ non esiste. Per fare esempio banale dove questo accade basta considerare il caso dove $\tilde{X}_1 = -\tilde{X}_2$ e dove il valore atteso di \tilde{X}_1 non esiste (e quindi non esiste nemmeno quello di \tilde{X}_2). In questo caso si ha $\tilde{X}_1 + \tilde{X}_2 = 0$ e quindi (proprietà E0) il valore atteso $E(\tilde{X}_1 + \tilde{X}_2)$ esiste e il suo valore numerico è nullo (proprietà E0).

Combinando le proprietà E2 e E4 si ottiene la proprietà seguente:

E4*) (Valore atteso di trasformazioni affini). Se $E(\tilde{X}_i)$ esiste per ogni $i = 1, 2, \dots, k$ e

$$\tilde{Y} = a + b_1\tilde{X}_1 + b_2\tilde{X}_2 + \dots + b_k\tilde{X}_k,$$

allora deve esistere anche $E(\tilde{Y})$ e

$$E(\tilde{Y}) = a + b_1E(\tilde{X}_1) + b_2E(\tilde{X}_2) + \dots + b_kE(\tilde{X}_k).$$

E5) (Valore atteso del prodotto di variabili casuali indipendenti) Se \tilde{X}_1 e \tilde{X}_2 sono due variabili casuali **indipendenti**, allora il valore atteso $E(\tilde{X}_1 \times \tilde{X}_2)$ esiste **se e solo se** esistono anche entrambi i valori attesi $E(\tilde{X}_1)$ e $E(\tilde{X}_2)$ e in questo

caso si ha

$$E(\tilde{X}_1 \times \tilde{X}_2) = E(\tilde{X}_1) \times E(\tilde{X}_2).$$

Più in generale, se $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_k$ sono k variabili casuali **indipendenti**, allora il valore atteso $E\left(\prod_{i=1}^k \tilde{X}_i\right)$ esiste **se e solo se** esistono anche i valori attesi di tutte e k le variabili casuali \tilde{X}_i e in tal caso si ha

$$E\left(\prod_{i=1}^k \tilde{X}_i\right) = \prod_{i=1}^k E(\tilde{X}_i).$$

Anche sulla proprietà **E5** conviene fare qualche osservazione:

- Se due o più variabili casuale **non** sono indipendenti, allora il valore atteso del loro prodotto potrebbe esistere anche se i valori attesi delle singole variabili casuali non esistono, e viceversa.
- Se due o più variabili casuali **non** sono indipendenti e se il valore atteso del loro prodotto e i valori attesi delle singole variabili casuali esistono, allora di solito si ottiene

$$E\left(\prod_{i=1}^k \tilde{X}_i\right) \neq \prod_{i=1}^k E(\tilde{X}_i).$$

E6) (Disuguaglianza di Markov). Se $E(\tilde{X})$ esiste (si noti che questa condizione equivale ad assumere l'esistenza di $E(|\tilde{X}|)$), si ha

$$P\left(\{|\tilde{X} - E(\tilde{X})| > \epsilon\}\right) \leq \frac{E(|\tilde{X}|)}{\epsilon} \quad \text{per ogni } \epsilon > 0.$$

E7) (Disuguaglianza di Cauchy-Schwarz per valori attesi). Se $E(\tilde{X}_1^2)$ e $E(\tilde{X}_2^2)$ esistono, allora esiste anche $E(\tilde{X}_1 \times \tilde{X}_2)$ e

$$[E(\tilde{X}_1 \times \tilde{X}_2)]^2 \leq E(\tilde{X}_1^2) \times E(\tilde{X}_2^2).$$

La disuguaglianza diventa un'uguaglianza **se e solo se** tra le variabili casuali \tilde{X}_1 e \tilde{X}_2 esiste una **relazione lineare**.

Con riferimento alla disuguaglianza di Cauchy-Schwarz è importante tenere presente che il valore atteso $E(\tilde{X}_1 \times \tilde{X}_2)$ potrebbe esistere anche se almeno uno dei due valori attesi $E(\tilde{X}_1^2)$ e $E(\tilde{X}_2^2)$ non esiste.

Come si desume da questo elenco, **le proprietà del valore atteso sono perfettamente analoghe quelle della media aritmetica nella statistica descrittiva.** L'unica differenza consiste nella complicazione legata alla possibilità che un valore atteso non esista.

5.4 Applicazioni del valore atteso

5.4.1 La definizione di gioco equo e il paradosso di San Pietroburgo

Secondo quanto abbiamo visto finora, il valore atteso può essere interpretato come un "operatore" che a determinate variabili casuali (o meglio, a determinate distribuzioni) associa un numero reale. In virtù dell'analogia con la meda aritmetica della statistica descrittiva, non è una sorpresa che per la teoria della probabilità il valore atteso sia un concetto di fondamentale importanza. Infatti, le definizioni di molti altri concetti di cui si occupa la teoria della probabilità sono basate sul valore atteso. Nelle prossime pagine introdurremo due di questi concetti: quello di **gioco equo** e quello dell'**utilità attesa**.

Definizione 5.5 (Gioco equo). Sia \tilde{X} la variabile casuale che rappresenta la vincita (lorda) associata ad un gioco (o ad una scommessa), e sia p il prezzo che si deve pagare per partecipare al gioco. Allora, per definizione,

- si dice che il gioco è **"equo"** se $E(\tilde{X}) = p$;
- si dice che il gioco è **"vantaggioso"** se $E(\tilde{X}) > p$;
- si dice che il gioco è **"in perdita"** se $E(\tilde{X}) < p$.

Si noti che le definizioni di gioco "equo", "vantaggioso" o "in perdita" non dipendono soltanto dal prezzo p del gioco, ma che dipendono anche in modo cruciale dalla distribuzione della vincita lorda \tilde{X} , ovvero dal modo in cui sono state assegnate le probabilità agli eventi che possono verificarsi al termine del gioco!!!! Per chiarire questa osservazione conviene considerare un gioco dove la vincita lorda \tilde{X} può assumere soltanto un numero finito di valori diversi che indicheremo con

$$x_1, \quad x_2, \quad x_3, \quad \dots, \quad x_k.$$

Chiaramente, in questo caso \tilde{X} deve essere una variabile casuale discreta e il valore atteso di \tilde{X} (che secondo la Definizione 5.5 rappresenta il prezzo "equo" del gioco) può dunque essere espresso come

$$E(\tilde{X}) = \sum_{i=1}^k x_i P(\{\tilde{X} = x_i\}) = x_1 P(\{\tilde{X} = x_1\}) + \dots + x_k P(\{\tilde{X} = x_k\}).$$

Si noti ora che il prezzo "equo" del gioco (ovvero $E(\tilde{X})$) ha un significato diverso a seconda del modo in cui sono state assegnate le probabilità

$$P(A_i) = P(\{\tilde{X} = x_i\}), \quad i = 1, 2, \dots, k.$$

- 1) Se le probabilità $P(A_i) = P(\{\tilde{X} = x_i\})$ sono state assegnate secondo il **metodo classico**, ovvero se ciascuna delle suddette probabilità è uguale al rapporto tra il corrispondente numero di casi favorevoli e il numero complessivo di casi possibili che possono verificarsi al termine del gioco, allora quello che secondo la Definizione 5.5 è il prezzo equo del gioco, ovvero il valore di

$$p = E(\tilde{X}) = x_1 P(\{\tilde{X} = x_1\}) + \dots + x_k P(\{\tilde{X} = x_k\}),$$

è la **media ponderata delle vincite lorde \tilde{X} con pesi dati dalle probabilità definite come rapporti tra casi favorevoli e casi possibili.**

- 2) Se le probabilità $P(A_i) = P(\{\tilde{X} = x_i\})$ sono state assegnate secondo il **metodo frequentista**, ovvero se ciascuna delle suddette probabilità rappresenta la frequenza relativa con cui l'evento $A_i = \{\tilde{X} = x_i\}$ di riferimento si è verificato in un (elevato) numero n di repliche del gioco, allora quello che secondo la Definizione 5.5 è il prezzo equo del gioco, ovvero il valore di

$$p = E(\tilde{X}) = x_1 P(\{\tilde{X} = x_1\}) + \dots + x_k P(\{\tilde{X} = x_k\}),$$

è la **media delle vincite lorde \tilde{X} che si sarebbero realizzate nelle n repliche del gioco.**

- 3) Se invece le probabilità $P(A_i) = P(\{\tilde{X} = x_i\})$ sono state assegnate secondo il **metodo soggettivo**, ovvero se ciascuna delle suddette probabilità rappresenta il prezzo che un dato individuo ritiene "equo" per il titolo di una scommessa dove si vince 1 euro se l'evento $A_i = \{\tilde{X} = x_i\}$ di riferimento si verifica e dove non si riceve nulla in caso contrario, allora quello che secondo la Definizione 5.5 è il prezzo equo del gioco, ovvero il valore di

$$E(\tilde{X}) = x_1 P(\{\tilde{X} = x_1\}) + \dots + x_k P(\{\tilde{X} = x_k\}),$$

rappresenta, **per l'individuo che ha assegnato le probabilità, il prezzo "equo" di un portafoglio** che è composto da

- x_1 titoli di scommesse sull'evento $A_1 = \{\tilde{X} = x_1\}$,
- x_2 titoli di scommesse sull'evento $A_2 = \{\tilde{X} = x_2\}$,
- ...
- e di x_k titoli di scommesse sull'evento $A_k = \{\tilde{X} = x_k\}$.

Si noti che il valore finale di questo portafoglio è uguale alla realizzazione della variabile casuale \tilde{X} che rappresenta la vincita lorda del gioco!!!

Esercizio 5.5. Per definizione, la "quota" offerta da un'agenzia di scommesse per un evento incerto è l'importo (espresso in euro) che l'agenzia si impegna a pagare ad uno scommettitore che ha puntato 1 euro su tale evento nel caso in cui l'evento in questione si verificasse.

Un'agenzia di scommessa offre a quota $Q = 2,5$ il verificarsi di un determinato evento A .

- Si supponga che si tratti di un gioco "equo". Qual è la probabilità che si verifichi l'evento A ?
- Si supponga ora invece che per uno scommettitore che acquista un tagliando della scommessa sull'evento A il gioco sia "in perdita". Che cosa si può dire sulla probabilità dell'evento A ?
- Si supponga ora che un'agenzia di scommesse offra a quota $Q_1 = 2,5$ il verificarsi di un dato evento A e che contemporaneamente offra a quota $Q_2 = 1,5$ il verificarsi dell'evento complementare \bar{A} . Può trattarsi di due giochi che sono entrambi "equi"?

Risposte:

- Si supponga che si tratti di un gioco "equo". Qual è la probabilità che si verifichi l'evento A ?

Per rispondere indichiamo con \tilde{X} la variabile casuale che descrive la vincita lorda di uno scommettitore che punta 1 euro sull'evento A .

Non è difficile rendersi conto che \tilde{X} è una variabile casuale discreta con funzione di massa di probabilità data da

$$p(x) = \begin{cases} P(A) & \text{se } x = Q = 2,5 \\ 1 - P(A) & \text{se } x = 0 \\ 0 & \text{altrimenti,} \end{cases}$$

dove $P(A) \in [0, 1]$ è l'ignota probabilità dell'evento A . Si noti che qualunque sia l'ignoto valore di $P(A)$, la vincita lorda \tilde{X} può assumere soltanto i valori $x = 0$ e $x = 2,5$. Dunque, qualunque sia l'ignoto valore di $P(A)$, il valore atteso di \tilde{X} esiste.

Siccome uno scommettitore deve pagare 1 euro per assicurarsi di ricevere $Q = 2,5$ euro nel caso in cui si verificasse l'evento A , possiamo concludere che il gioco in questione è "equo" se e solo se

$$E(\tilde{X}) = Q \times P(A) + 0 \times [1 - P(A)] = 1,$$

ovvero se e solo se

$$P(A) = \frac{1}{Q} = \frac{1}{2,5} = 0,4.$$

- b) Si supponga ora invece che per uno scommettitore che acquista un tagliando della scommessa sull'evento A il gioco sia "in perdita". Che cosa si può dire sulla probabilità dell'evento A ?

Per uno scommettitore il gioco è "in perdita" se e solo se

$$\begin{aligned} E(\tilde{X}) &= Q \times P(A) + 0 \times [1 - P(A)] < 1 \quad \Leftrightarrow \\ \Leftrightarrow \quad P(A) &< \frac{1}{Q} = \frac{1}{2,5} = 0,4. \end{aligned}$$

- c) Si supponga ora che un'agenzia di scommesse offra a quota $Q_1 = 2,5$ il verificarsi di un dato evento A e che contemporaneamente offra a quota $Q_2 = 1,5$ il verificarsi dell'evento complementare \bar{A} . Può trattarsi di due giochi che sono entrambi "equi"?

Per rispondere a questa domanda dobbiamo innanzitutto **estendere la nostra definizione di gioco equo al caso in cui vi sia la possibilità di partecipare a più di un singolo gioco**: per definizione, due o più giochi "contemporanei" (ovvero giochi i cui esiti potrebbero non essere indipendenti) sono (congiuntamente) equi se per qualsiasi sistema di scommesse il valore atteso della vincita lorda è uguale al prezzo che si deve pagare per acquistarlo.

A questo punto, possiamo rispondere al quesito c) osservando che nel caso di **due giochi contemporanei** la suddetta definizione di gioco equo può essere soddisfatta **solo se** esiste un **unico** modello probabilistico (Ω, \mathcal{A}, P) che può "ospitare" due variabili casuali \tilde{X}_1 e \tilde{X}_2 che rappresentano le vincite lorde associate ai due giochi, e che assicuri che i valori attesi di queste due variabili casuali siano entrambi uguali ai prezzi che si devono pagare per partecipare ai rispettivi giochi (non è difficile dimostrare che oltre ad essere **necessaria**, questa condizione è anche **sufficiente**: per farlo, basta sfruttare il fatto che la vincita lorda associata ad un sistema di scommesse è sempre una combinazione lineare delle vincite lorde associate alle singole scommesse che compongono il sistema). Infatti, non è difficile rendersi conto che nel caso contemplato dal quesito c) questa **condizione necessaria** (e sufficiente) non è soddisfatta: infatti, nella risposta al quesito a) abbiamo già visto che la scommessa sull'evento A è equa se e solo se

$$P(A) = \frac{1}{Q_1} = \frac{1}{2,5} = 0,4.$$

Ragionando allo stesso modo, vediamo che la scommessa associata all'evento complementare \bar{A} è equa se e solo se

$$P(\bar{A}) = \frac{1}{Q_2} = \frac{1}{1,5} = 0,666.$$

Siccome la legge **L3** prevede che

$$P(\bar{A}) = 1 - P(A) \quad \Rightarrow \quad \frac{1}{Q_2} = 1 - \frac{1}{Q_1} \quad \Rightarrow \quad \frac{1}{Q_1} + \frac{1}{Q_2} = 1,$$

ma con le quote $Q_1 = 2,5$ e $Q_2 = 1,5$ si ottiene

$$\frac{1}{Q_1} + \frac{1}{Q_2} = 0,4 + 0,666 > 1,$$

possiamo concludere che con le quote Q_1 e Q_2 la **condizione necessaria (e sufficiente) di cui sopra non può essere soddisfatta**: questo significa che, considerati congiuntamente, **i due giochi non possono essere "equi"!!!**

Facendo riferimento alla disuguaglianza

$$P(A) + P(\bar{A}) = \frac{1}{Q_1} + \frac{1}{Q_2} = 0,4 + 0,666 > 1$$

possiamo anche valutare come si deve intervenire per rendere i due giochi in questione congiuntamente equi: bisogna aumentare il valore di almeno una delle due quote Q_1 e/o Q_2 .

Prima di terminare questo esempio vale la pena osservare che con le quote

$$Q_1 = 2,5 \text{ per l'evento } A \quad \text{e} \quad Q_2 = 1,5 \text{ per l'evento } \bar{A}$$

l'agenzia di scommesse può incamerare una vincita certa se riesce a vendere i tagliandi delle due scommesse nella giusta proporzione: infatti, **vendendo**

- n (un numero positivo qualunque) tagliandi di scommesse sull'evento A
- e $n \times \frac{Q_1}{Q_2}$ tagliandi di scommesse sull'evento \bar{A} ,

l'agenzia di scommessa otterrebbe un **incasso immediato di**

$$n + n \times \frac{Q_1}{Q_2} \text{ euro,}$$

e al termine del gioco **pagherebbe**

- $n \times Q_1$ euro nel caso in cui si verificasse l'evento A
- e $(n \times \frac{Q_1}{Q_2}) \times Q_2 = n \times Q_1$ euro (esattamente la stessa cifra) nel caso in cui si verificasse l'evento \bar{A} .

Siccome

$$n + n \times \frac{Q_1}{Q_2} - n \times Q_1 = nQ_1 \left(\frac{1}{Q_1} + \frac{1}{Q_2} - 1 \right) = n \times 0,165 > 0,$$

l'agenzia di scommesse conseguirebbe dunque un guadagno certo di $n \times 0,165$ euro!!!

A prima vista potrebbe sembrare che tutti i giochi con prezzo inferiore al valore atteso della vincita lorda siano particolarmente convenienti. Il **paradosso di San Pietroburgo**²² chiarisce che questo non è affatto vero: spesso il prezzo massimo che siamo disposti a pagare è molto più piccolo del prezzo che secondo la Definizione 5.5 sarebbe "equo". Il paradosso di San Pietroburgo rende evidente questo fatto attraverso una scommessa che verte su una **successione di lanci consecutivi di una moneta**. La successione di

²² Il paradosso di San Pietroburgo è nato da un problema inviato per lettera da Nicolas Bernoulli (1687 - 1759) a Pierre Raymond de Montmort (1678 - 1719) nell'anno 1713 e deve il suo nome al fatto che è stato inserito in una memoria redatta da Daniel Bernoulli (1700 - 1782; Daniel Bernoulli era il cugino di Nicolas Bernoulli) pubblicata nell'anno 1738 in una raccolta di testi scientifici dell'Accademia delle scienze di San Pietroburgo. La storia di questo paradosso è riassunta in *Todhunter (1865) "History of the theory of probability", textually unaltered reprint, Chelsea Publishing Company, New York, 1965.*

lanci viene interrotta nel momento in cui si ottiene per la prima volta il lato testa, oppure se in (per esempio) $n = 10^6$ (un milione) di lanci non si ottiene mai il lato testa. Nel momento in cui si interrompono i lanci, allo scommettitore viene riconosciuto un premio che dipende dal numero di lanci che sono stati effettuati:

- se si ottiene testa al primo lancio, lo scommettitore vince $2^1 = 2$ euro (e il gioco termina);
- se si ottiene per la prima volta testa al secondo lancio, lo scommettitore vince $2^2 = 4$ euro (e il gioco termina);
- se si ottiene testa per la prima volta al terzo lancio, lo scommettitore vince $2^3 = 8$ euro (e il gioco termina);
- ecc..
- se dopo il lancio numero $n = 10^6$ non si è ancora mai ottenuto il lato testa, lo scommettitore non vince nulla (e il gioco termina).

Quanto saremmo disposti a pagare per partecipare a questa scommessa???

Probabilmente poco: infatti, se assumiamo che in ogni singolo lancio la probabilità di ottenere testa sia pari a 0,5 e che gli esiti dei lanci siano indipendenti (un'ipotesi sulla quale ben pochi avrebbero da obiettare),²³ allora la probabilità di ottenere per la prima volta testa ...

- ... in un lancio successivo al quinto (se ciò non accadesse si vincerebbero al più **32** euro) è data da

$$P\left(\bigcap_{i=1}^5 \{\tilde{X}_i = 0\}\right) = 0,5^5 = 0,03125 \quad \Rightarrow$$

$$\Rightarrow P(\{\tilde{X} > \mathbf{32}\}) = P(\{\tilde{X} \geq 64\}) = 0,03125 \dots$$

- ... in un lancio successivo al sesto (se ciò non accadesse si vincerebbero al più **64** euro) è data da

$$P\left(\bigcap_{i=1}^6 \{\tilde{X}_i = 0\}\right) = 0,5^6 = 0,015625 \quad \Rightarrow$$

$$\Rightarrow P(\{\tilde{X} > \mathbf{64}\}) = P(\{\tilde{X} \geq 128\}) = 0,015625$$

- ...

²³Per dimostrare l'esistenza di un modello probabilistico che rispecchia queste ipotesi possiamo fare riferimento al teorema di esistenza per successioni infinite di variabili casuali indipendenti, ovvero al Teorema 4.13: basta definire \tilde{X}_i come la variabile casuale che assume il valore uno oppure zero a seconda se l'esito dell' i -esimo lancio è testa oppure croce, e definire la funzione di massa di probabilità di \tilde{X}_i ponendo $p(0) = P(\{\tilde{X}_i = 0\}) = P(\{\tilde{X}_i = 1\}) = p(1) = 0,5$.

- ... in un lancio successivo al decimo (se ciò non accadesse si vincerebbero al più **1024** euro) è data da

$$P\left(\bigcap_{i=1}^{10}\{\tilde{X}_i = 0\}\right) = 0,5^{10} = 0,00098 \quad \Rightarrow$$

$$\Rightarrow P(\{\tilde{X} > 1024\}) = P(\{\tilde{X} \geq 2048\}) = 0,00098$$

- ...

Ora, il paradosso di San Pietroburgo verte sul fatto che nonostante agli occhi di una persona comune questo tipo di scommessa è poco allettante, **il valore atteso della vincita lorda \tilde{X} (e quindi il prezzo che secondo la Definizione 5.5 renderebbe questa scommessa un gioco equo) è uguale $n = 10^6$ euro, ovvero a 1 milione di euro!!!** Per dimostrare che il valore atteso della vincita lorda è così elevato basta osservare che ...

- la probabilità di ottenere per la prima volta testa al **primo lancio** è data da

$$P(\{\tilde{X}_1 = 1\}) = \frac{1}{2^1} \quad \Rightarrow \quad P(\{\tilde{X} = 2^1\}) = \frac{1}{2^1}$$

- la probabilità di ottenere per la prima volta testa al **secondo lancio** è data da

$$P(\{\tilde{X}_1 = 0\} \cap \{\tilde{X}_2 = 1\}) = \frac{1}{2^2} \quad \Rightarrow \quad P(\{\tilde{X} = 2^2\}) = \frac{1}{2^2};$$

- la probabilità di ottenere per la prima volta testa al **terzo lancio** è data da

$$P\left(\bigcap_{i=1}^{3-1}\{\tilde{X}_i = 0\} \cap \{\tilde{X}_3 = 1\}\right) = \frac{1}{2^3} \quad \Rightarrow \quad P(\{\tilde{X} = 2^3\}) = \frac{1}{2^3};$$

- e che, in generale, la probabilità di ottenere per la prima volta testa al **k -esimo lancio** (per qualunque $k = 1, 2, \dots$) è data da

$$P\left(\bigcap_{i=1}^{k-1}\{\tilde{X}_i = 0\} \cap \{\tilde{X}_k = 1\}\right) = \frac{1}{2^k} \quad \Rightarrow \quad P(\{\tilde{X} = 2^k\}) = \frac{1}{2^k}.$$

Dal risultato generale all'ultimo punto deduciamo che il valore atteso della vincita lorda \tilde{X} sia effettivamente dato da

$$E(\tilde{X}) = \sum_{k=1}^{10^6} 2^k \times \frac{1}{2^k} = \sum_{k=1}^{10^6} 1 = 10^6$$

come ci eravamo proposti di dimostrare. **Aumentando il numero massimo di lanci di lanci $n = 10^6$, possiamo dunque portare il "prezzo equo" $E(\tilde{X})$ ad una cifra arbitrariamente elevata!!!** Tuttavia, agli occhi di una persona comune, oltre ad

un certo limite, il valore della scommessa rimane praticamente immutato se il numero massimo di lanci viene ulteriormente aumentato.

Come abbiamo appena visto, **in alcuni tipi di scommesse il prezzo p che rende un gioco equo è molto più elevato del prezzo che una persona "ragionevole" sarebbe disposta a pagare.**

Come si possono caratterizzare queste scommesse??? La risposta è semplice: si tratta di scommesse dove con probabilità prossima a 1 si vincono importi modesti e dove con probabilità molto piccola si vincono importi stratosferici.

5.4.2 La teoria dell'aspettativa morale di Daniel Bernoulli

Visto che nessuno sarebbe disposto a pagare il prezzo "equo" per partecipare alla scommessa del paradosso di San Pietroburgo, conviene chiedersi come si potrebbe determinare un prezzo più "ragionevole". Per risolvere questo problema, Daniel Bernoulli, nella sua memoria (vedi la nota [22]), propone di applicare la cosiddetta **teoria della "aspettativa morale"** da lui appositamente sviluppata.

La teoria dell'aspettativa morale è basata sull'ipotesi che **l'utilità marginale del denaro sia inversamente proporzionale alla quantità di denaro che si possiede.** Ovviamente, questa ipotesi implica che l'utilità del denaro sia descritta da una curva logaritmica: $u(x) = \ln x$.²⁴ Se \tilde{X} è la variabile casuale che rappresenta la vincita lorda di una scommessa e x_0 è il patrimonio iniziale di un soggetto che deve decidere se partecipare o meno alla scommessa, allora, secondo la teoria dell'aspettativa morale, il prezzo massimo che il soggetto è disposto a pagare per assicurarsi il diritto di riscuotere la vincita lorda \tilde{X} sarebbe il valore $p = p_{acq}^*$ che rende **l'aspettativa morale**

$$m(p; x_0) = E \left(u(x_0 + \tilde{X} - p) \right) \quad (67)$$

uguale a **livello di utilità certo** che corrisponde al suo patrimonio iniziale x_0 , ovvero al livello di utilità $u(x_0) = \ln x_0$ (vedi il grafico in Figura 5.2). Secondo questo criterio, il prezzo massimo che un soggetto sarebbe disposto a pagare per la scommessa del paradosso di San Pietroburgo deve quindi essere una soluzione $p = p_{acq}^*$ dell'equazione

$$\sum_{i=1}^n \ln(x_0 + 2^i - p) \times \frac{1}{2^i} = \ln x_0 \quad (68)$$

(ricordiamo che n è il numero massimo di lanci). Vedendo questa equazione potrebbe sorgere il dubbio che non esista nessuna soluzione $p = p_{acq}^*$ che la soddisfi. Non è difficile

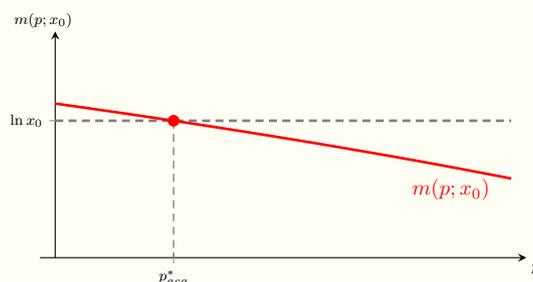
²⁴Se x rappresenta il denaro che si possiede e $u(x)$ rappresenta la sua "utilità", allora "l'utilità marginale del denaro è inversamente proporzionale a x " significa che $u'(x) = b/x$ per qualche $b > 0$. Com'è facile verificare, questa condizione è soddisfatta se e solo se $u(x) = \int u'(x)dx = \int (b/x)dx = b \ln x + a$ dove anche a è una costante. Nel testo considereremo sempre il caso in cui $a = 0$ e $b = 1$. Si può dimostrare che tutti i risultati riportati nel testo non cambiano se al posto di questi due valori usassimo una qualunque altra coppia di valori reali con $b > 0$.

dimostrare che questo non è vero e che la suddetta equazione ha sempre un'unica soluzione $p = p_{acq}^*$. Tuttavia, la soluzione p_{acq}^* non può essere espressa in termini di funzioni elementari e per questo motivo, nella sua memoria, Daniel Bernoulli fornisce alcuni suggerimenti pratici per ottenere un'approssimazione alla soluzione p_{acq}^* . Per non appesantire troppo la trattazione tralasciamo questi suggerimenti e ci limitiamo a fornire una tabella con le soluzioni p_{acq}^* che corrispondono ad alcuni valori del patrimonio iniziale x_0 e al caso in cui il numero massimo di lanci è $n = 100$ (piuttosto che $n = 10^6$ come in precedenza: quanto saremmo disposti a pagare per passare da $n = 100$ a $n = 10^6$? ... nulla? ... un centesimo? ... un euro?). Per ottenere i valori di $p = p_{acq}^*$ riportati nella tabella abbiamo utilizzato il metodo delle tangenti (vedi le soluzioni dell'Esercizio 5.6).

x_0	p_{acq}^*	x_0	p_{acq}^*
100	7,79	100.000	17,55
1.000	10,95	1.000.000	20,87
10.000	14,24	10.000.000	24,20

Si noti che secondo il criterio dell'aspettativa morale un individuo con un patrimonio iniziale di 10 milioni di euro sarebbe disposto a pagare al più $p_{acq}^* = 24,20$ euro per assicurarsi la vincita lorda \tilde{X} della scommessa del paradosso di San Pietroburgo con $n = 100$ (si ricordi che con questa scommessa si possono vincere fino a $2^{100} \simeq 1,268 \times 10^{30}$ euro - una cifra stratosferica)!!!

Figura 5.2. Il grafico mostra come si determina il prezzo $p = p_{acq}^*$ che rende l'aspettativa morale $m(p; x_0)$ uguale al livello di utilità certo che corrisponde al patrimonio iniziale x_0 . Anche se dal punto di vista grafico l'aspettativa morale $m(p; x_0)$ sembra una funzione lineare di p , l'equazione (67) mostra che ciò non è vero.



Ovviamente, l'aspettativa morale

$$m(p; x_0) = E \left(u(x_0 + \tilde{X} - p) \right)$$

può essere calcolata anche con funzioni di utilità $u(x)$ diverse da quella logaritmica e così facendo si ottengono, se esistono, altri valori di $p = p_{acq}^*$ che rendono l'aspettativa

morale uguale al livello di utilità iniziale $u(x_0)$. Abbiamo detto "se esistono", perché non è sempre vero che l'equazione

$$E\left(u(x_0 + \tilde{X} - p)\right) = u(x_0) \quad (69)$$

abbia una soluzione. Tuttavia, si può dimostrare che se una tale soluzione p_{acq}^* esiste, e se la funzione di utilità $u(x)$ è strettamente crescente, allora la soluzione p_{acq}^* deve essere unica (questa affermazione può essere facilmente dimostrata usando la proprietà E1). Siccome in questa dispensa assumeremo sempre che la variabile indipendente x di una funzione di utilità $u(x)$ rappresenti il denaro posseduto da un soggetto di riferimento, e siccome non ha senso considerare l'ipotesi che il livello di utilità di un individuo rimanga costante oppure che diminuisca all'aumentare del denaro da lui posseduto, **in questa dispensa daremo sempre per scontato che tutte le funzioni di utilità $u(x)$ siano strettamente crescenti** e non ripeteremo esplicitamente questa ipotesi. Inoltre, onde evitare complicazioni tecniche che poco aggiungono alla sostanza di ciò che diremo, assumeremo sempre che l'equazione (69) abbia una soluzione $p = p_{acq}^*$ e che questa soluzione sia dunque unica. Come si desume dai ragionamenti che ci hanno condotti all'equazione (69), il valore della soluzione $p = p_{acq}^*$ può essere interpretato come il **prezzo massimo che un individuo con funzione di utilità $u(x)$ è disposto a pagare per partecipare ad una scommessa con vincita lorda \tilde{X}** . Si noti che non stiamo escludendo il caso che \tilde{X} possa assumere valori negativi: se la realizzazione di \tilde{X} è negativa, colui che ha pagato per partecipare alla scommessa deve pagare la controparte. Com'è facile intuire, se \tilde{X} può assumere anche (o soltanto) valori negativi, il prezzo $p = p_{acq}^*$ che si ottiene dall'equazione (69) potrebbe a sua volta essere negativo. In questo caso il modulo di p_{acq}^* rappresenta il prezzo minimo che l'individuo vuole ricevere dalla controparte per partecipare alla scommessa.

Nel prossimo esercizio considereremo una scommessa più semplice di quella del paradosso di San Pietroburgo e calcoleremo i prezzi p_{acq}^* che corrispondono a diverse ipotesi sull'andamento della funzione di utilità $u(x)$.

Esercizio 5.6. Tizio dispone di un patrimonio iniziale di $x_0 = 5$ milioni di euro e deve decidere se partecipare ad una scommessa dove può vincere 3 milioni di euro con probabilità 0,6, oppure perdere 2 milioni di euro con probabilità 0,4.

- a) Si supponga che la funzione di utilità di Tizio sia definita come $u(x) = \ln x$ così come previsto nella versione originale del criterio dell'aspettativa morale di Daniel Bernoulli. Si applichi il criterio dell'aspettativa morale per determinare il prezzo massimo p_{acq}^* che Tizio è disposto a pagare per partecipare alla scommessa.
- b) Si supponga ora invece che la funzione di utilità di Tizio sia definita co-

me $u(x) = x$. Si applichi il criterio dell'aspettativa morale per determinare il prezzo massimo p_{acq}^* che Tizio è disposto a pagare per partecipare alla scommessa.

- c) Si supponga infine che la funzione di utilità di Tizio sia definita come $u(x) = e^{x-5}$. Si applichi il criterio dell'aspettativa morale per determinare il prezzo massimo p_{acq}^* che Tizio è disposto a pagare per partecipare alla scommessa.

Risposte:

- a) Si supponga che la funzione di utilità di Tizio sia definita come $u(x) = \ln x$ così come previsto nella versione originale del criterio dell'aspettativa morale di Daniel Bernoulli. Si applichi il criterio dell'aspettativa morale per determinare il prezzo massimo p_{acq}^* che Tizio è disposto a pagare per partecipare alla scommessa.

Sia \tilde{X} la variabile casuale che rappresenta la vincita lorda di Tizio. Come abbiamo visto, secondo la teoria dell'aspettativa morale di Daniel Bernoulli il prezzo massimo è la soluzione $p = p_{acq}^*$ dell'equazione che eguaglia l'utilità attesa di Tizio

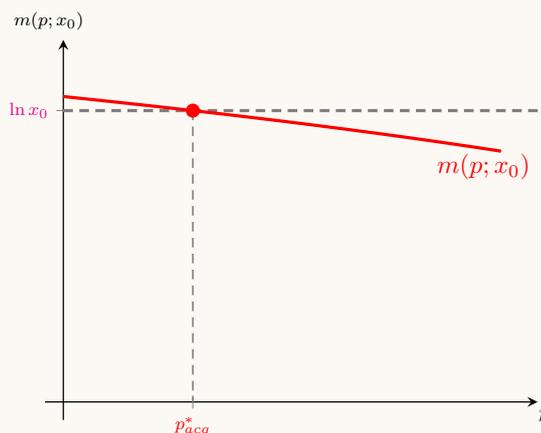
$$\begin{aligned} m(p; x_0) &= E(u(x_0 + \tilde{X} - p)) \\ &= E(\ln(5 + \tilde{X} - p)) \\ &= 0,6 \times \ln(5 + 3 - p) + 0,4 \times \ln(5 - 2 - p) \\ &= 0,6 \times \ln(8 - p) + 0,4 \times \ln(3 - p) \end{aligned} \quad (70)$$

al suo livello di utilità attuale $u(x_0) = \ln x_0 = \ln 5 = 1,609$.

Per rispondere al quesito dobbiamo quindi trovare una soluzione $p = p_{acq}^*$ all'equazione

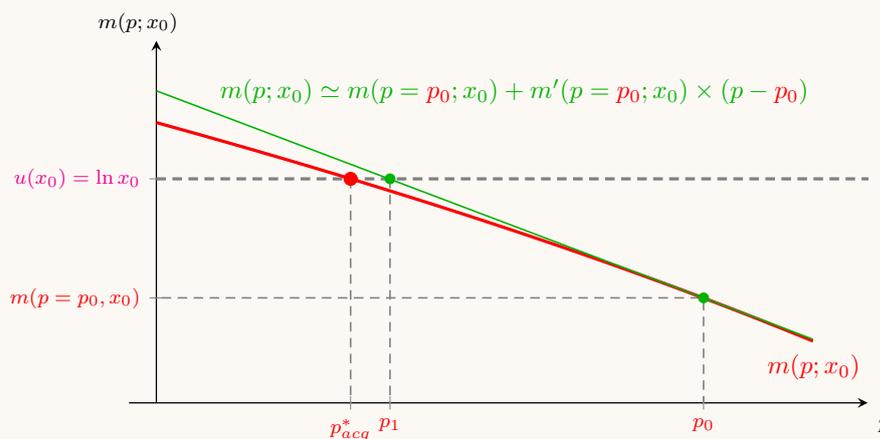
$$m(p; x_0) = 0,6 \times \ln(8 - p) + 0,4 \times \ln(3 - p) = 1,609$$

Il grafico sottostante mostra il valore di p_{acq}^* che dobbiamo trovare:



Siccome il valore di $p = p^*_{acq}$ non può essere calcolato attraverso una formula esplicita, tenteremo di trovare una sua approssimazione attraverso il metodo delle tangenti:

- 1) Il grafico sottostante mostra l'esito del primo passo del metodo delle tangenti:



Per avviare la procedura iterativa del metodo delle tangenti abbiamo scelto come punto iniziale p_0 l'aspettativa matematica

$$p = p_0 = E(\tilde{X}) = 3 \times 0,6 - 2 \times 0,4 = 1$$

e abbiamo calcolato la retta tangente (la retta di colore verde) all'aspettativa morale $m(p; x_0)$ che corrisponde al punto $p = p_0$. Come si vede nel grafico, per valori di p che sono prossimi a p_0 questa retta fornisce un'approssimazione all'andamento dell'aspettativa morale $m(p; x_0)$:

$$m(p; x_0) \simeq m(p = p_0; x_0) + m'(p = p_0; x_0) \times (p - p_0).$$

Per trovare un'approssimazione al valore di $p = p^*$ abbiamo poi calcolato il punto p_1 dove l'ordinata della retta tangente è uguale al livello di utilità iniziale $u(x_0) = \ln x_0 = \ln 5 = 1,609$:

$$m(p = p_0; x_0) + m'(p = p_0; x_0) \times (p - p_0) = u(x_0) \quad \Rightarrow$$

$$\Rightarrow p = p_1 = p_0 + \frac{u(x_0) - m(p = p_0; x_0)}{m'(p = p_0; x_0)} \quad (71)$$

Per calcolare il valore numerico di p_1 , oltre al valore di $p_0 = 1$ e al valore $u(x_0) = \ln x_0 = \ln 5 = 1,609$, ci servono anche ...

* ... il valore di

$$m(p = p_0; x_0) = E(\ln(5 + \tilde{X} - p_0))$$

$$= 0,6 \times \ln(5 + 3 - p_0) + 0,4 \times \ln(5 - 2 - p_0)$$

$$= 0,6 \times \ln(5 + 3 - 1) + 0,4 \times \ln(5 - 2 - 1) = 1,445,$$

* ... e il valore di $m'(p = p_0; x_0)$:

$$m'(p; x_0) = \frac{dm(p; x_0)}{dp} = -0,6 \times \frac{1}{8 - p} - 0,4 \times \frac{1}{3 - p} \quad (72)$$

$$\Rightarrow m'(p = p_0; x_0) = -0,6 \times \frac{1}{8 - 1} - 0,4 \times \frac{1}{3 - 1} = -0,286.$$

Sostituendo questi valori nella (71) otteniamo

$$\Rightarrow p = p_1 = 1 + \frac{1,609 - 1,445}{(-0,286)} = 0,426.$$

Per valutare se il valore di p_1 è sufficientemente prossimo al valore di p_{acq}^* che stiamo cercando, calcoliamo il valore dell'aspettativa morale nel punto $p = p_1$:

$$m(p = p_1; x_0) = 0,6 \times \ln(8 - 0,427) + 0,4 \times \ln(3 - 0,427) = 1,593.$$

Siccome questo valore non ci sembra abbastanza vicino al valore obiettivo $u(x_0) = 1,609$, procediamo ad un'altra iterazione del metodo delle tangenti, ovvero ripetiamo la suddetta procedura con p_1 al posto di p_0 .

2) Nella seconda iterazione del metodo delle tangenti dobbiamo quindi calcolare il valore di

$$\Rightarrow p = p_2 = p_1 + \frac{u(x_0) - m(p = p_1; x_0)}{m'(p = p_1; x_0)}$$

(questa formula si ottiene a partire dalla formula (71) aumentando i pedici delle p di una unità).

Usando la formula (72) per calcolare la derivata dell'aspettativa morale nel punto $p = p_1 = 0,426$, otteniamo

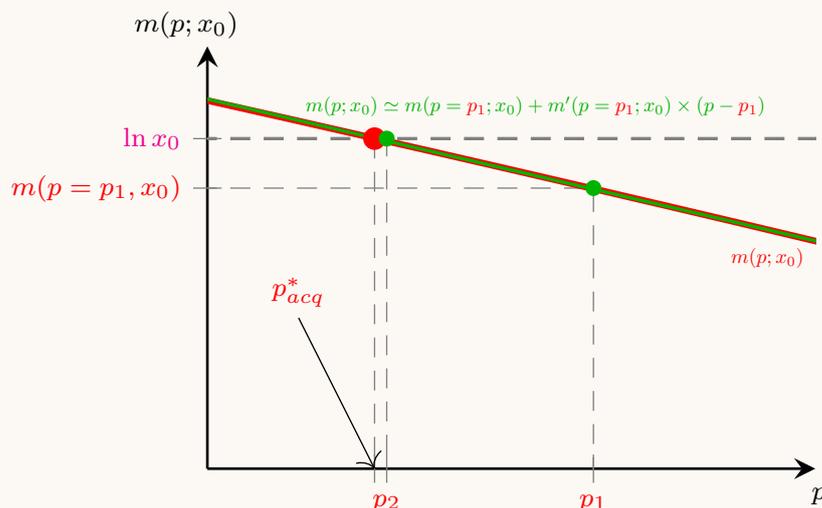
$$m'(p = p_1; x_0) = -0,6 \times \frac{1}{8 - 0,427} - 0,4 \times \frac{1}{3 - 0,427} = -0,235,$$

e usando questo risultato vediamo che il valore di p_2 è dato da

$$p_2 = 0,426 + \frac{1,609 - 1,593}{(-0,235)} = 0,359$$

(si ricordi che il valore di $m(p = p_1; x_0) = 1,593$ lo avevamo già calcolato al termine della prima iterazione per valutare se in corrispondenza di p_1 l'aspettativa morale fosse sufficientemente prossima al valore obiettivo $u(x_0) = \ln x_0 = \ln 5 = 1,609$).

Il grafico sottostante mostra come abbiamo determinato il valore di $p_2 = 0,359$



Per valutare se il valore di $p_2 = 0,359$ è sufficientemente prossimo al valore di p_{acq}^* che stiamo cercando, confrontiamo l'aspettativa morale che corrisponde a $p_2 = 0,359$ con il valore obiettivo $u(x_0) = 1,609$. Siccome

$$m(p = p_2; x_0) = 0,6 \times \ln(8 - 0,359) + 0,4 \times \ln(3 - 0,359) = 1,609$$

e siccome questo valore coincide con il valore obiettivo fino alla terza cifra decimale, possiamo ritenerci soddisfatti e interrompere le iterazioni del metodo delle tangenti.

Il prezzo massimo p_{acq}^* che Tizio è disposto a pagare per partecipare alla scommessa è dunque all'incirca uguale a $p_2 = 0,359$ milioni di euro. Si noti che anche questo prezzo (come quello della scommessa del paradosso di San Pietroburgo) è più piccolo dell'aspettativa matematica $p_0 = E(\tilde{X}) = 1$ euro.

- b) Si supponga ora invece che la funzione di utilità di Tizio sia definita come $u(x) = x$. Si applichi il criterio dell'aspettativa morale per determinare il prezzo massimo p_{acq}^* che Tizio è disposto a pagare per partecipare alla scommessa.

Con la funzione di utilità $u(x) = x$, l'aspettativa morale di Tizio è data da

$$\begin{aligned} m(p; x_0) &= E(u(x_0 + \tilde{X} - p)) \\ &= E(x_0 + \tilde{X} - p) \\ \text{[[proprietà E2]]} &= x_0 + E(\tilde{X}) - p. \end{aligned}$$

Com'è facile verificare, il valore $p = p_{acq}^*$ che rende questa aspettativa morale uguale al corrispondente livello di utilità iniziale

$$u(x_0) = x_0$$

è dato dall'aspettativa matematica $p_{acq}^* = E(\tilde{X}) = 1$ euro che abbiamo già calcolato nella risposta al quesito a). Il prezzo massimo che Tizio è disposto a pagare è quindi dato dall'aspettativa matematica $p_{acq}^* = E(\tilde{X}) = 1$.

Si noti che saremmo giunti alla stessa conclusione se al posto della funzione di utilità $u(x) = x$ avessimo considerato una qualunque trasformazione affine crescente, ovvero una qualunque funzione del tipo $u(x) = a + bx$ con $b > 0$.

- c) Si supponga infine che la funzione di utilità di Tizio sia definita come $u(x) = e^{x-5}$. Si applichi il criterio dell'aspettativa morale per determinare il prezzo massimo p_{acq}^* che Tizio è disposto a pagare per partecipare alla scommessa.

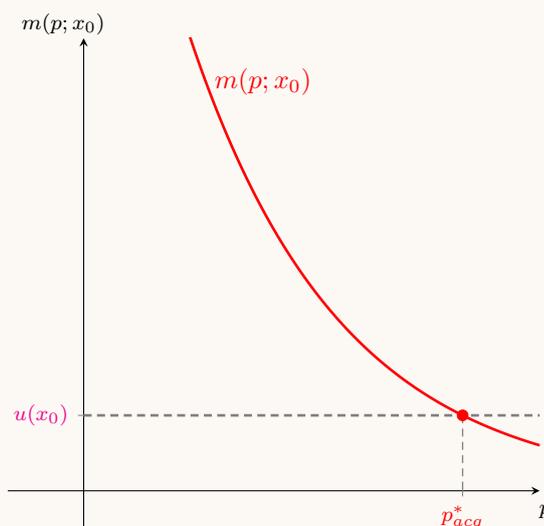
L'aspettativa morale che corrisponde alla funzione di utilità $u(x) = e^{x-5}$ è data da

$$\begin{aligned} m(p; x_0 = 5) &= E(u(x_0 + \tilde{X} - p)) = E\left(e^{x_0 + \tilde{X} - p - 5}\right) \\ E\left(e^{x_0 + \tilde{X} - p - 5}\right) &= 0,6 \times e^{5+3-p-5} + 0,4 \times e^{5-2-p-5} \quad (73) \\ &= 0,6 \times e^{3-p} + 0,4 \times e^{-2-p}, \end{aligned}$$

e per determinare il prezzo $p = p_{acq}^*$ richiesto dobbiamo porre $m(p; x_0 = 5) = 1$:

$$0,6 \times e^{3-p} + 0,4 \times e^{-2-p} = 1. \quad (74)$$

Il grafico sottostante mostra il prezzo $p = p_{acq}^*$ che dobbiamo determinare:



Anche in questo caso il valore di $p = p_{acq}^*$ non può essere calcolato attraverso una formula esplicita e deve quindi essere determinato attraverso un metodo numerico iterativo come per esempio quello delle tangenti che abbiamo già applicato nella risposta al quesito a).

Applicando il metodo delle tangenti e avviando le iterazioni con l'aspettativa matematica $p = p_0 = 1$, si perviene, dopo quattro iterazioni, al valore $p = p_4 = 2,494$ in corrispondenza del quale l'aspettativa morale risulta uguale a

$$m(p = p_4; x_0) = 0,999656 \simeq 1,000.$$

La tabella sottostante mostra i risultati delle quattro le iterazioni (invitiamo il lettore a verificare questi risultati):

iterazione	p_i	$m(p = p_i; x_0)$	$m'(p = p_i; x_0)$
0	1	4,453	-4,453
1	1,775	2,052	-2,052
2	2,288	1,228	-1,228
3	2,474	1,020	-1,020
4	2,494	1,000	—

Per verificare i risultati delle iterazioni basta applicare la formula

$$p = p_{i+1} = p_i + \frac{u(x_0) - m(p = p_i; x_0)}{m'(p = p_i; x_0)}$$

che abbiamo già ricavato nella risposta al quesito a) (vedi la (71)). Nel caso in questione la verifica è facilitata dal fatto che la

$$m'(p; x_0 = 5) = \frac{dm(p; x_0 = 5)}{dp} = -0,6 \times e^{3-p} - 0,4 \times e^{-2-p}$$

coincide con l'aspettativa morale moltiplicata per -1 (vedi la (73)).

In virtù del risultato ottenuto con il metodo delle tangenti possiamo dunque concludere che il prezzo massimo p_{acq}^* che Tizio è disposto a pagare per partecipare alla scommessa è all'incirca uguale a $p_4 = 2,474$ euro. Si noti che questo prezzo è maggiore dell'aspettativa matematica $p_0 = E(\tilde{X}) = 1$ euro.

Nella soluzione dell'Esercizio 5.6 abbiamo ottenuto i seguenti risultati:

- Con $u(x) = \ln x$ (una funzione strettamente concava), il prezzo massimo p_{acq}^* che Tizio è disposto a pagare per partecipare alla scommessa è minore dell'aspettativa matematica: $p_{acq}^* < p_0 = E(\tilde{X})$.
- Con $u(x) = x$ (una funzione lineare), il prezzo massimo p_{acq}^* che Tizio è disposto a pagare per partecipare alla scommessa è uguale all'aspettativa matematica: $p_{acq}^* = p_0 = E(\tilde{X})$.
- Con $u(x) = e^{x-5}$ (una funzione convessa), il prezzo massimo p_{acq}^* che Tizio è disposto a pagare per partecipare alla scommessa p_{acq}^* è maggiore dell'aspettativa matematica: $p_{acq}^* > p_0 = E(\tilde{X})$.

A priori si potrebbe pensare che questi risultati dipendano soltanto dal particolare tipo di scommessa che abbiamo considerato (ovvero dalla distribuzione della vincita lorda \tilde{X}) e/o dalle particolari funzioni di utilità che abbiamo considerato. Tuttavia, come si desume dal prossimo teorema, questi risultati possono essere giustificati in un modo molto elegante.

Teorema 5.1 (Teorema sul criterio dell'aspettativa morale - parte 1). Sia \tilde{X} la variabile casuale che rappresenta la vincita lorda associata ad una scommessa (non stiamo escludendo che \tilde{X} possa assumere valori negativi), sia x_0 il valore del patrimonio iniziale di un dato individuo e sia $u(x)$ la sua funzione di utilità (strettamente crescente). Si assuma che

- \tilde{X} non sia una variabile casuale degenere,
- che l'aspettativa matematica $p_0 = E(\tilde{X})$ esista e
- che esista un unico prezzo p_{acq}^* che soddisfa il criterio dell'aspettativa morale, ovvero che l'equazione

$$E(u(x_0 + \tilde{X} - p)) = u(x_0)$$

abbia un'unica soluzione $p = p_{acq}^*$.

Sotto queste ipotesi si può dimostrare che:

- a) p_{acq}^* è (strettamente) minore di $p_0 = E(\tilde{X})$ se $u(x)$ è una funzione (strettamente) concava;
- b) p_{acq}^* è uguale a $p_0 = E(\tilde{X})$ se $u(x)$ è una funzione lineare;
- c) p_{acq}^* è (strettamente) maggiore di $p_0 = E(\tilde{X})$ se $u(x)$ è una funzione (strettamente) convessa.

Dimostrazione. La dimostrazione del Teorema 5.1 è basata sulla disuguaglianza di Jensen (proprietà E3) e sulla proprietà di linearità del valore atteso (proprietà E2). Dimostreremo soltanto la conclusione al punto a).

Siccome al punto a) viene ipotizzato che la funzione di utilità $u(x)$ sia concava, possiamo concludere che

$$\begin{aligned} u(x_0) &= E\left(u(x_0 + \tilde{X} - p_{acq}^*)\right) \\ [[\text{proprietà E3}]] &\leq u\left(E(x_0 + \tilde{X} - p_{acq}^*)\right) \\ [[\text{proprietà E2}]] &= u(x_0 + E(\tilde{X}) - p_{acq}^*). \end{aligned}$$

Tenendo presente che

$$u(x_0) = u(x_0 + E(\tilde{X}) - E(\tilde{X})) = u(x_0 + E(\tilde{X}) - p_0),$$

vediamo dunque che

$$\begin{aligned} u(x_0 + E(\tilde{X}) - p_0) &\leq u(x_0 + E(\tilde{X}) - p_{acq}^*) \Leftrightarrow \\ \Leftrightarrow [[u(x) \text{ è strettamente crescente}]] & \quad x_0 + E(\tilde{X}) - p_0 \leq x_0 + E(\tilde{X}) - p_{acq}^* \Leftrightarrow \\ & \Leftrightarrow p_{acq}^* \leq p_0 = E(\tilde{X}). \end{aligned}$$

Fin qui abbiamo quindi dimostrato che $p_{acq}^* \leq p_0$ se $u(x)$ è una **qualunque** funzione concava. Per dimostrare che $p_{acq}^* < p_0$ se $u(x)$ è **strettamente** concava, basta osservare che in questo caso la disuguaglianza di Jensen (proprietà E3) è anche soddisfatta in senso stretto e che tutte le disuguaglianze di questa dimostrazione possono quindi essere sostituite con delle disuguaglianze strette. \square

Supponiamo ora che \tilde{X} sia la vincita lorda associata ad una scommessa il cui prezzo "equo" secondo la Definizione 5.5 è nullo: $E(\tilde{X}) = 0$. Come si desume dalla conclusione del precedente Teorema 5.1

- a) un soggetto con **funzione di utilità strettamente concava** associa un **prezzo massimo** p_{acq}^* **negativo** a questa scommessa e quindi vuole essere pagato partecipare alla scommessa;

- b) un soggetto con **funzione di utilità lineare** associa un **prezzo massimo** p_{acq}^* **nullo** a questa scommessa e quindi non è disposto a pagare nulla per partecipare alla scommessa, ma è allo stesso tempo indifferente rispetto alla possibilità di partecipare alla scommessa (senza pagare nulla) o meno;
- c) un soggetto con **funzione di utilità strettamente convessa** associa un **prezzo massimo** p_{acq}^* **positivo** a questa scommessa e quindi è disposto a pagare pur di partecipare alla scommessa.

In virtù di queste considerazioni,

- a) i soggetti con **funzione di utilità strettamente concava** vengono chiamati **"avversi al rischio"**,
- b) i soggetti con **funzione di utilità lineare** vengono chiamati **"neutrali al rischio"**,
- c) e i soggetti con **funzione di utilità strettamente convessa** vengono chiamati **"amanti del rischio"**.

Consideriamo ora un problema leggermente diverso da quello di trovare il prezzo massimo che un soggetto è disposto a pagare per **acquistare** un titolo di una scommessa (che, non è difficile rendersi conto, è identico al problema di determinare il prezzo al quale un soggetto è disposto ad **emettere** un titolo di una scommessa). **Consideriamo dunque un soggetto, lo chiameremo Tizio, che è già impegnato in una scommessa e indichiamo ancora con \tilde{X} la variabile casuale che rappresenta la sua vincita lorda** (anche qui non escludiamo il caso che la vincita lorda possa assumere valori negativi e che al termine della scommessa Tizio debba dunque pagare la controparte). Supponiamo che il patrimonio iniziale di Tizio ammonti a x_0 euro e che Tizio voglia **vendere** i diritti e gli eventuali oneri legati alla scommessa. **Qual è il prezzo minimo che Tizio vuole ricevere per trasferire i diritti e gli eventuali oneri legati alla scommessa ad un soggetto terzo?** Secondo il criterio dell'aspettativa morale il prezzo minimo è dato dalla soluzione $p = p_{ven}^*$ dell'equazione che eguaglia la sua aspettativa morale al suo livello di utilità certo corrispondente alla somma tra il suo patrimonio iniziale x_0 e il compenso p che riceve per il trasferire la scommessa:

$$E(u(x_0 + \tilde{X})) = u(x_0 + p). \quad (75)$$

Siccome stiamo assumendo che la funzione di utilità $u(x)$ sia strettamente crescente, questa equazione ha un'unica soluzione che è data da

$$p = p_{ven}^* = u^{-1} \left(E(u(x_0 + \tilde{X})) \right) - x_0, \quad (76)$$

dove $u^{-1}(\cdot)$ è la funzione inversa della funzione $u(x)$, ovvero l'unica funzione tale che $u^{-1}(u(x)) = x$ per ogni x che appartiene al dominio della funzione $u(x)$. **Si noti che il**

prezzo minimo p_{ven}^* potrebbe anche essere negativo. In questo caso p_{ven}^* rappresenterebbe il prezzo massimo che Tizio è disposto a pagare per trasferire i suoi diritti e oneri legati alla scommessa ad un soggetto terzo.

Esercizio 5.7. Tizio dispone di un patrimonio iniziale di $x_0 = 5$ milioni di euro e si è impegnato in una scommessa dove può vincere 3 milioni di euro con probabilità 0,6, oppure perdere 2 milioni di euro con probabilità 0,4 (si noti che questa scommessa è identica a quella che abbiamo già considerato nell'Esercizio 5.6).

- Si supponga che la funzione di utilità di Tizio sia definita come $u(x) = \ln x$. Si applichi il criterio dell'aspettativa morale per determinare il prezzo minimo p_{ven}^* che Tizio vuole ricevere per trasferire i diritti e gli oneri legati alla scommessa ad un soggetto terzo.
- Si supponga ora invece che la funzione di utilità di Tizio sia definita come $u(x) = x$. Si applichi il criterio dell'aspettativa morale per determinare il prezzo minimo p_{ven}^* che Tizio vuole ricevere per trasferire i diritti e gli oneri legati alla scommessa ad un soggetto terzo.
- Si supponga infine che la funzione di utilità di Tizio sia definita come $u(x) = e^{x-5}$. Si applichi il criterio dell'aspettativa morale per determinare il prezzo minimo p_{ven}^* che Tizio vuole ricevere per trasferire i diritti e gli oneri legati alla scommessa ad un soggetto terzo.

Risposte:

- Si supponga che la funzione di utilità di Tizio sia definita come $u(x) = \ln x$. Si applichi il criterio dell'aspettativa morale per determinare il prezzo minimo p_{ven}^* che Tizio vuole ricevere per trasferire i diritti e gli oneri legati alla scommessa ad un soggetto terzo.

Indichiamo sempre ancora con \tilde{X} la variabile casuale che rappresenta la vincita lorda legata alla scommessa (ovvero l'importo che Tizio riceve dalla controparte qualora la realizzazione di \tilde{X} fosse positiva, e l'importo che Tizio deve pagare alla controparte qualora la realizzazione di \tilde{X} fosse negativa). Per rispondere al quesito dobbiamo calcolare il prezzo p_{ven}^* nella formula (76), ovvero dobbiamo risolvere l'equazione (75).

A tal fine osserviamo che l'aspettativa morale di Tizio è data da

$$E(u(x_0 + \tilde{X})) = 0,6 \times \ln(5 + 3) + 0,4 \times \ln(5 - 2) = 1,687$$

e ponendo questo valore uguale al livello di utilità

$$u(x_0 + p) = \ln(x_0 + p) = \ln(5 + p),$$

ovvero al livello di utilità che Tizio raggiungerebbe nel caso in cui vendesse i diritti e gli oneri legati alla scommessa e incassasse il compenso p , si ottiene

$$\ln(5 + p) = 1,687 \quad \Leftrightarrow \quad p = p_{ven}^* = e^{1,687} - 5 = 0,403.$$

Si noti che anche questo prezzo, così come quello che abbiamo determinato nella risposta al quesito a) dell'Esercizio 5.6, è strettamente minore dell'aspettativa matematica $p = E(\tilde{X})$.

- b) Si supponga ora invece che la funzione di utilità di Tizio sia definita come $u(x) = x$. Si applichi il criterio dell'aspettativa morale per determinare il prezzo minimo p_{ven}^* che Tizio vuole ricevere per trasferire i diritti e gli oneri legati alla scommessa ad un soggetto terzo.

Se consideriamo la funzione di utilità $u(x) = x$, l'aspettativa morale di Tizio è data da

$$E(u(x_0 + \tilde{X})) = 0,6 \times (5 + 3) + 0,4 \times (5 - 2) = 6,$$

e ponendo questo valore uguale al livello di utilità

$$u(x_0 + p) = x_0 + p = 5 + p,$$

ovvero al livello di utilità che Tizio raggiungerebbe nel caso in cui vendesse i diritti e gli oneri legati alla scommessa e incassasse il compenso p , otteniamo

$$5 + p = 6 \quad \Rightarrow \quad p = 1.$$

Si noti che questo prezzo, così come quello che abbiamo determinato nella risposta al quesito b) dell'Esercizio 5.6, coincide esattamente con l'aspettativa matematica $p = E(\tilde{X}) = 1$.

- c) Si supponga infine che la funzione di utilità di Tizio sia definita come $u(x) = e^{x-5}$. Si applichi il criterio dell'aspettativa morale per determinare il prezzo minimo p_{ven}^* che Tizio vuole ricevere per trasferire i diritti e gli oneri legati alla scommessa ad un soggetto terzo.

Se consideriamo la funzione di utilità $u(x) = e^{x-5}$, l'aspettativa morale di Tizio è data da

$$E(u(x_0 + \tilde{X})) = 0,6 \times e^{5+3-5} + 0,4 \times e^{5-2-5} = 12,105,$$

e ponendo questo valore uguale al livello di utilità

$$u(x_0 + p) = e^{x_0 + p - 5} = e^p,$$

ovvero al livello di utilità che Tizio raggiungerebbe nel caso in cui vendesse i diritti e gli oneri legati alla scommessa e incassasse il compenso p , otteniamo

$$e^p = 12,105 \quad \Rightarrow \quad p = p_{ven}^* = \ln 12,105 = 2,494.$$

Si noti che anche questo prezzo, così come quello che abbiamo determinato nella risposta al quesito c) dell'Esercizio 5.6, è strettamente maggiore dell'aspettativa matematica $p = E(\tilde{X})$.

Come abbiamo visto nelle soluzioni del precedente esercizio, anche il prezzo minimo di vendita p_{ven}^* risulta (strettamente) minore o maggiore dell'aspettativa matematica (ovvero del valore atteso) a seconda se la funzione di utilità $u(x)$ è concava oppure convessa. Anche qui non si tratta di un caso particolare legato alle particolari funzioni di utilità che abbiamo considerato e/o alla particolare distribuzione della vincita lorda \tilde{X} , ma di una conseguenza del verso della concavità della funzione di utilità, ovvero di una conseguenza dell'attitudine nei confronti del rischio del soggetto che valuta il prezzo minimo di vendita p_{ven}^* . Il prossimo teorema formalizza questa osservazione.

Teorema 5.2 (Teorema sul criterio dell'aspettativa morale - parte 2). Sia \tilde{X} la variabile casuale che rappresenta la vincita lorda associata ad una scommessa nella quale un dato soggetto si è impegnato (non escludiamo il caso che \tilde{X} possa assumere valori negativi), sia x_0 il valore del suo patrimonio iniziale prima di conoscere l'esito della scommessa e sia $u(x)$ la sua funzione di utilità (strettamente crescente).

Si assuma che

- \tilde{X} non sia una variabile casuale degenere,
- che l'aspettativa matematica $p_0 = E(\tilde{X})$ esista,
- e che esista anche l'aspettativa morale $E(u(x_0 + \tilde{X}))$.

Allora, secondo il criterio dell'aspettativa morale,

$$p = p_{ven}^* = u^{-1} \left(E(u(x_0 + \tilde{X})) \right) - x_0$$

è il **prezzo minimo** al quale il soggetto è disposto a cedere i diritti e gli oneri legati alla scommessa, e questo prezzo è (strettamente) minore, uguale oppure (strettamente) maggiore dell'aspettativa matematica $p_0 = E(\tilde{X})$ a seconda se la sua funzione di utilità $u(x)$ è (strettamente) concava, lineare oppure (strettamente) convessa.

Dimostrazione. Anche la dimostrazione di questo teorema (come quella del Teorema 5.1) è basata sulla disuguaglianza di Jensen (proprietà E3) e sulla proprietà E2. Anche questa volta dimostreremo soltanto la parte della conclusione che si riferisce a funzioni di utilità concave.

La dimostrazione è semplice: siccome stiamo ipotizzando che $u(x)$ sia una funzione concava, la sua funzione inversa $u^{-1}(y)$ deve essere una funzione convessa (per ottenere il grafico della funzione inversa di una qualunque funzione strettamente monotona, basta considerare la retta bisettrice del primo e terzo quadrante come asse di simmetria e tracciare la curva che rispetto a questo asse di simmetria è simmetrica alla curva che rappresenta la funzione di partenza) e usando la disuguaglianza di Jensen vediamo quindi che

$$\begin{aligned} p_{ven}^* &= u^{-1}\left(E(u(x_0 + \tilde{X}))\right) - x_0 \\ [[\text{proprietà E3}]] &\leq E\left(u^{-1}(u(x_0 + \tilde{X}))\right) - x_0 \\ &= E(x_0 + \tilde{X}) - x_0 \\ [[\text{proprietà E2}]] &= x_0 + E(\tilde{X}) - x_0 = E(\tilde{X}). \end{aligned}$$

Fin qui abbiamo quindi dimostrato che $p_{ven}^* \leq p_0$ se $u(x)$ è una **qualunque** funzione concava. Per dimostrare che $p_{ven}^* < p_0$ se $u(x)$ è **strettamente** concava, basta osservare che in questo caso deve essere strettamente concava anche la funzione inversa $u^{-1}(y)$, e che la disuguaglianza di Jensen (proprietà E3) è dunque anche soddisfatta in senso stretto: tutte le disuguaglianze di questa dimostrazione possono quindi essere sostituite con delle disuguaglianze strette. \square

Per interpretare le conclusioni del precedente teorema possiamo dire che:

- un individuo avverso al rischio (funzione di utilità concava) è disposto a vendere una "vincita" incerta \tilde{X} anche ad un prezzo p_{ven}^* minore del prezzo $p_0 = E(\tilde{X})$ che secondo la Definizione 5.5 sarebbe quello "equo";
- un individuo neutrale al rischio (funzione di utilità concava) vuole invece ricevere almeno il prezzo "equo";
- un individuo amante del rischio (funzione di utilità convessa) vuole ricevere più del prezzo equo "equo".

5.4.3 Osservazione sul metodo di assegnazione soggettivo

A prima vista il **criterio dell'aspettativa morale potrebbe sembrare in contraddizione con il metodo di assegnazione soggettivo di de Finetti**. Infatti,

- se $P(A) \in (0, 1)$ è la probabilità che un soggetto, chiamiamolo Tizio, assegna ad un determinato evento A , ...

- e se Tizio è avverso al rischio, ...

allora (secondo il criterio dell'aspettativa morale) il prezzo massimo p_{acq}^* che Tizio è disposto a pagare per acquistare un titolo che incorpora il diritto di ricevere 1 euro al verificarsi dell'evento A è sempre strettamente minore di $P(A)$.

Dimostrazione. Per dimostrare questa affermazione basta considerare la variabile casuale \tilde{X} che rappresenta la vincita lorda di un soggetto che ha acquistato il suddetto titolo e applicare il Teorema 5.1. Siccome la funzione di massa di probabilità di \tilde{X} è data da

$$p(x) = P(\{\tilde{X} = x\}) = \begin{cases} P(A) & \text{se } x = 1 \\ 1 - P(A) & \text{se } x = 0 \\ 0 & \text{altrimenti} \end{cases}$$

e siccome

$$E(\tilde{X}) = 1 \times P(A) + 0 \times [1 - P(A)] = P(A),$$

dobbiamo avere $p_{acq}^* < P(A)$ (vedi la prima conclusione del Teorema 5.1). \square

Quindi, se Tizio valuta titoli di scommesse secondo il criterio dell'aspettativa morale e Tizio è avverso al rischio, allora $P(A)$ non può essere il prezzo al quale Tizio è disposto ad acquistare un titolo che incorpora il diritto di ricevere 1 euro al verificarsi dell'evento A . A prima vista potrebbe sembrare che questo ragionamento introduca un grave dilemma:

- o si rinuncia all'idea che i prezzi di scommesse vengano determinati attraverso il criterio dell'aspettativa morale,
- oppure si accetta l'idea che individui avversi al rischio non siano disposti ad assegnare probabilità $P(A)$ che sono strettamente comprese tra zero e uno (si ricordi infatti che nel Teorema 5.1 viene ipotizzato che la vincita lorda non sia una variabile casuale degenera).

Chiaramente nessuna di queste soluzioni è desiderabile perché

- il criterio dell'aspettativa morale è molto utile per risolvere paradossi come quello di San Pietroburgo,
- e perché nelle nostre scelte quotidiane siamo tutti avversi al rischio e nella nostra vita quotidiana abbiamo spesso a che fare con eventi di cui non siamo certi se si verificheranno o meno.

De Finetti risolve questo dilemma introducendo una piccola **finzione**: secondo il suo metodo di assegnazione soggettivo la probabilità $P(A)$ di un evento rappresenterebbe il prezzo al quale un individuo è disposto ad acquistare ed emettere titoli di scommesse sull'evento A **sotto l'ipotesi che l'individuo sia costretto ad accettare qualsiasi proposta di vendita e emissione che gli viene recapitata**. Si noti che il criterio dell'aspettativa morale fornisce invece il prezzo al quale un individuo è disposto ad acquistare o emettere un titolo in modo **volontario** e non in modo **costrittivo** come ipotizzato da de Finetti.

5.4.4 Teorie dell'utilità attesa*

Come abbiamo visto nella Sezione 5.4.2, attraverso il criterio dell'aspettativa morale si perviene a valutazioni di scommesse che sono più "ragionevoli" di quelle che si ottengono attraverso il criterio del prezzo "equo". Tuttavia, il criterio dell'aspettativa morale può sembrare, e certamente lo è, arbitrario:

- Perché dare così tanta importanza al valore atteso di una funzione di utilità?
- Come si dovrebbe scegliere la funzione di utilità?

Le teorie dell'utilità attesa di von Neumann Morgenstern e di Savage forniscono, ciascuna a modo suo, delle risposte a queste domande. Anche se si tratta di teorie molto interessanti, queste teorie esulano dal contenuto di un corso introduttivo sui modelli probabilistici e quindi non le tratteremo.

6 Varianza e covarianza

6.1 Definizioni di varianza e covarianza

In questa sezione definiremo la varianza e la covarianza con riferimento a [variabili casuali](#). Le definizioni di questi due indici sono perfettamente analoghe alle definizioni degli omonimi indici della statistica descrittiva. L'unica complicazione aggiuntiva è legata al fatto che con riferimento a variabili casuali la varianza e/o la covarianza potrebbero non esistere.

Definizione 6.1 (Varianza e covarianza). La **varianza** di una variabile casuale \tilde{X} è definita come

$$\text{var}(\tilde{X}) = E\left(\left[\tilde{X} - E(\tilde{X})\right]^2\right)$$

se entrambi i valori attesi al secondo membro esistono. In caso contrario si dice che la varianza di \tilde{X} non esiste.

La **covarianza** tra \tilde{X}_1 e \tilde{X}_2 è definita come

$$\text{cov}(\tilde{X}_1, \tilde{X}_2) = E\left(\left[\tilde{X}_1 - E(\tilde{X}_1)\right] \times \left[\tilde{X}_2 - E(\tilde{X}_2)\right]\right).$$

se tutti e tre i valori attesi al secondo membro esistono. In caso contrario si dice che la covarianza tra \tilde{X}_1 e \tilde{X}_2 non esiste.

Ovviamente, non solo le **definizioni**, ma anche le **interpretazioni** della varianza e covarianza sono perfettamente analoghe a quelle degli omonimi indici della statistica

descrittiva: la varianza, o meglio, lo **scarto quadratico medio**

$$\sigma(\tilde{X}) = \sqrt{\text{var}(\tilde{X})},$$

fornisce una **misura per l'incertezza sulla realizzazione di una variabile casuale**, mentre la covarianza fornisce una **misura per la concordanza tra le realizzazioni di due variabili casuali**.

Calcolare il valore numerico di una varianza o covarianza attraverso le formule definitorie può essere un compito molto laborioso. Infatti, ...

- per calcolare la varianza di una variabile casuale discreta attraverso la formula definitoria, bisogna prima calcolare il valore atteso

$$\mu = E(\tilde{X}) = \sum_{x \in S} xp(x)$$

e poi sostituirlo nella formula

$$\text{var}(\tilde{X}) = E\left(\left[\tilde{X} - E(\tilde{X})\right]^2\right) = \sum_{x \in S} (x - \mu)^2 p(x).$$

Come vedremo tra breve (proprietà **V4**), se $\mu = E(\tilde{X}) \neq 0$, la varianza di una variabile casuale discreta può essere calcolata in modo più agevole attraverso la **formula indiretta**

$$\text{var}(\tilde{X}) = E(\tilde{X}^2) - [E(\tilde{X})]^2 = \sum_{x \in S} x^2 p(x) - \left[\sum_{x \in S} xp(x)\right]^2$$

(si noti che questa formula è perfettamente analoga alla formula indiretta della statistica descrittiva).

- Chiaramente, lo stesso discorso vale anche per la varianza di una variabile casuale assolutamente continua: volendo applicare la formula definitoria bisogna prima calcolare il valore di

$$\mu = E(\tilde{X}) = \int_{-\infty}^{\infty} xf(x)dx$$

e poi sostituirlo nella formula

$$\text{var}(\tilde{X}) = E\left(\left[\tilde{X} - E(\tilde{X})\right]^2\right) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)dx.$$

Anche in questo caso è di solito più conveniente fare riferimento alla formula indiretta (proprietà **V4**)

$$\text{var}(\tilde{X}) = E(\tilde{X}^2) - [E(\tilde{X})]^2 = \int_{-\infty}^{\infty} x^2 f(x)dx - \left[\int_{-\infty}^{\infty} xf(x)dx\right]^2.$$

Vediamo ora invece la covarianza.

- Volendo applicare la formula definitoria nel caso discreto, bisogna prima calcolare entrambi i valori attesi

$$\mu_1 = E(\tilde{X}_1) = \sum_{x_1 \in S_1} x_1 p_1(x_1) \quad \text{e} \quad \mu_2 = E(\tilde{X}_2) = \sum_{x_2 \in S_2} x_2 p_2(x_2),$$

e poi sostituirli nella formula

$$\begin{aligned} cov(\tilde{X}_1, \tilde{X}_2) &= E\left(\left[\tilde{X}_1 - E(\tilde{X}_1)\right] \times \left[\tilde{X}_2 - E(\tilde{X}_2)\right]\right) \\ &= \sum_{(x_1, x_2) \in S} (x_1 - \mu_1)(x_2 - \mu_2) p(x_1, x_2). \end{aligned}$$

In queste formule abbiamo indicato con $p(x_1, x_2)$ la funzione di probabilità congiunta delle due variabili casuali, con S il supporto di $p(x_1, x_2)$, con $p_1(x)$ e $p_2(x)$ le funzioni di probabilità marginali delle due variabili casuali e con S_1 e S_2 i supporti delle due funzioni di probabilità marginali.

Si noti che se almeno uno dei due valori attesi $\mu_1 = E(\tilde{X}_1)$ e $\mu_2 = E(\tilde{X}_2)$ è diverso da zero, è più conveniente calcolare la covarianza attraverso la formula indiretta (proprietà V5)

$$\begin{aligned} cov(\tilde{X}_1, \tilde{X}_2) &= E(\tilde{X}_1 \tilde{X}_2) - E(\tilde{X}_1)E(\tilde{X}_2) \\ &= \sum_{(x_1, x_2) \in S} x_1 x_2 p(x_1, x_2) - \left(\sum_{x_1 \in S_1} x_1 p_1(x_1)\right) \times \left(\sum_{x_2 \in S_2} x_2 p_2(x_2)\right). \end{aligned}$$

- Volendo invece applicare la formula definitoria per calcolare una covarianza nel caso assolutamente continuo, bisogna prima calcolare entrambi i valori attesi attraverso le formule

$$\mu_1 = E(\tilde{X}_1) = \int_{-\infty}^{\infty} x_1 f_1(x_1) dx_1 \quad \text{e} \quad \mu_2 = E(\tilde{X}_2) = \int_{-\infty}^{\infty} x_2 f_2(x_2) dx_2,$$

e poi sostituirli nella formula

$$\begin{aligned} cov(\tilde{X}_1, \tilde{X}_2) &= E\left(\left[\tilde{X}_1 - E(\tilde{X}_1)\right] \times \left[\tilde{X}_2 - E(\tilde{X}_2)\right]\right) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_1 - \mu_1)(x_2 - \mu_2) f(x_1, x_2) dx_2 dx_1. \end{aligned}$$

In queste formule abbiamo indicato con $f(x_1, x_2)$ una funzione di densità congiunta delle due variabili casuali, e con $f_1(x_1)$ e $f_2(x_2)$ delle corrispondenti funzioni di densità marginali.

Anche nel caso assolutamente continuo, se almeno uno dei due valori attesi $\mu_1 = E(\tilde{X}_1)$ e $\mu_2 = E(\tilde{X}_2)$ è diverso da zero, è più conveniente calcolare la covarianza attraverso la formula indiretta (proprietà V5)

$$\begin{aligned} cov(\tilde{X}_1, \tilde{X}_2) &= E(\tilde{X}_1 \tilde{X}_2) - E(\tilde{X}_1)E(\tilde{X}_2) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 f(x_1, x_2) dx_2 dx_1 - \left(\int_{-\infty}^{\infty} x_1 f_1(x_1) dx_1\right) \times \left(\int_{-\infty}^{\infty} x_2 f_2(x_2) dx_2\right). \end{aligned}$$

Ovviamente, le suddette formule per il calcolo della varianza e covarianza sono valide se e solo se tutte le sommatorie/tutti gli integrali sono assolutamente convergenti (altrimenti la varianza e/o la covarianza non esistono).

Esercizio 6.1 (Calcolo della varianza e della covarianza). Si considerino ancora le due variabili casuali \tilde{X}_1 e \tilde{X}_2 dell'Esercizio 4.14. Ricordiamo che la loro funzione di probabilità congiunta può essere rappresentata attraverso la seguente tabella a doppia entrata:

$p(x_1, x_2)$	$x_2 = 0$	$x_2 = 1$	$x_2 = 2$	$x_2 = 3$	$p(x_1)$
$x_1 = 0$	0,0088	0,0659	0,0879	0,0220	0,1846
$x_1 = 1$	0,0791	0,2637	0,1319	0,0000	0,4747
$x_1 = 2$	0,1319	0,1648	0,0000	0,0000	0,2967
$x_1 = 3$	0,0440	0,0000	0,0000	0,0000	0,0440
$p(x_2)$	0,2638	0,4944	0,2198	0,0220	1,0000

Si ricavino i valori di $E(\tilde{X}_1)$, $E(\tilde{X}_2)$, $E(\tilde{X}_1\tilde{X}_2)$, $var(\tilde{X}_1)$, $var(\tilde{X}_2)$ e anche quello di $cov(\tilde{X}_1, \tilde{X}_2)$.

[Soluzione: $E(\tilde{X}_1) = 1,2001$; $E(\tilde{X}_2) = 1$; $E(\tilde{X}_1\tilde{X}_2) = 0,8571$; $var(\tilde{X}_1) = 0,6173$; $var(\tilde{X}_2) = 0,5716$; $cov(\tilde{X}_1, \tilde{X}_2) = -0,3430$.]

6.2 Proprietà della varianza e covarianza

V0) Il valore numerico di $var(\tilde{X})$ non può mai essere negativo.

$var(\tilde{X})$ esiste ed è uguale a 0 se e solo se \tilde{X} è una variabile casuale **degenere**.

V1) Il valore numerico di $cov(\tilde{X}_1, \tilde{X}_2)$ può essere positivo, negativo o anche nullo.

Se almeno una delle due variabili casuali \tilde{X}_1 e \tilde{X}_2 è **degenere** e $cov(\tilde{X}_1, \tilde{X}_2)$ esiste, allora $cov(\tilde{X}_1, \tilde{X}_2) = 0$, ma la covarianza $cov(\tilde{X}_1, \tilde{X}_2)$ può essere nulla anche in altri casi.

V2) (Simmetria della covarianza). $cov(\tilde{X}_1, \tilde{X}_2)$ esiste se e solo se anche $cov(\tilde{X}_2, \tilde{X}_1)$ esiste e in tal caso i valori numerici delle due covarianze sono identici:

$$cov(\tilde{X}_1, \tilde{X}_2) = cov(\tilde{X}_2, \tilde{X}_1).$$

V3) Se $\tilde{X}_1 = \tilde{X}_2$, allora la covarianza $cov(\tilde{X}_1, \tilde{X}_2)$ esiste se e solo se esistono anche le varianze di entrambe le variabili casuali e in questo caso si ha

$$cov(\tilde{X}_1, \tilde{X}_2) = var(\tilde{X}_1) = var(\tilde{X}_2).$$

V4) (Formula indiretta della varianza). La varianza $var(\tilde{X})$ esiste se e solo se $E(\tilde{X}^2)$ esiste e in questo caso si ha

$$var(\tilde{X}) = E(\tilde{X}^2) - E(\tilde{X})^2.$$

V5) (Formula indiretta della covarianza). La covarianza $cov(\tilde{X}_1, \tilde{X}_2)$ esiste se e solo se tutti e tre i valori attesi $E(\tilde{X}_1)$, $E(\tilde{X}_2)$ e $E(\tilde{X}_1\tilde{X}_2)$ esistono e in questo caso si ha

$$cov(\tilde{X}_1, \tilde{X}_2) = E(\tilde{X}_1\tilde{X}_2) - E(\tilde{X}_1) \times E(\tilde{X}_2).$$

V6) (Indipendenza). Se \tilde{X}_1 e \tilde{X}_2 sono due variabili casuali indipendenti e se la loro covarianza esiste, allora si deve avere $cov(\tilde{X}_1, \tilde{X}_2) = 0$.

V7) (Disuguaglianza di Cauchy - Schwarz per varianze e covarianze). Se $var(\tilde{X}_1)$ e $var(\tilde{X}_2)$ esistono, allora esiste anche $cov(\tilde{X}_1, \tilde{X}_2)$ e

$$\left[cov(\tilde{X}_1, \tilde{X}_2) \right]^2 \leq var(\tilde{X}_1) \times var(\tilde{X}_2).$$

La precedente disuguaglianza è un'uguaglianza se e solo se **tra le due variabili casuali esiste una relazione affine**. In particolare, la disuguaglianza di Cauchy-Schwartz è un'uguaglianza se una o entrambe le variabili casuali sono **degeneri**.

Con riferimento alla disuguaglianza di Cauchy-Schwartz conviene osservare che l'esistenza di $cov(\tilde{X}_1, \tilde{X}_2)$ non implica l'esistenza di nessuna delle due varianze $var(\tilde{X}_1)$ e $var(\tilde{X}_2)$. Si possono infatti costruire degli esempi dove $cov(\tilde{X}_1, \tilde{X}_2)$ esiste e dove una o entrambe le suddette varianze non esistono (invitiamo il lettore a costruire un tale esempio).

Come nella statistica descrittiva, anche nella teoria della probabilità si può definire un **coefficiente di correlazione lineare** ponendo

$$r(\tilde{X}_1, \tilde{X}_2) = \frac{\text{cov}(\tilde{X}_1, \tilde{X}_2)}{\sqrt{\text{var}(\tilde{X}_1) \times \text{var}(\tilde{X}_2)}} = \frac{\text{cov}(\tilde{X}_1, \tilde{X}_2)}{\sigma(\tilde{X}_1) \times \sigma(\tilde{X}_2)}.$$

Usando la disuguaglianza di Cauchy - Schwarz si vede che

- il coefficiente di correlazione lineare $r(\tilde{X}_1, \tilde{X}_2)$ è ben definito se e solo se entrambe le varianze $\text{var}(\tilde{X}_1)$ e $\text{var}(\tilde{X}_2)$ esistono e sono positive
- e che $r(\tilde{X}_1, \tilde{X}_2)$ può assumere soltanto valori che appartengono all'intervallo $[-1, 1]$. In particolare, $r(\tilde{X}_1, \tilde{X}_2) = \pm 1$ se e solo se esistono due costanti reali a e b , con b dello stesso segno di $r(\tilde{X}_1, \tilde{X}_2)$ (quindi la costante b non può essere nulla), tali che

$$\tilde{X}_1 = a + b \times \tilde{X}_2.$$

V8) (Varianza di trasformazioni affini di una singola variabile casuale). Se $\text{var}(\tilde{X})$ esiste e $\tilde{Y} = a + b\tilde{X}$, allora deve esistere anche $\text{var}(\tilde{Y})$ e

$$\text{var}(\tilde{Y}) = \text{var}(a + b\tilde{X}) = b^2 \text{var}(\tilde{X}).$$

V8*) (Varianza di trasformazioni affini - caso generale). Se tutte le varianze $\text{var}(\tilde{X}_1), \text{var}(\tilde{X}_2), \dots, \text{var}(\tilde{X}_k)$ esistono e

$$\tilde{Y} = a + b_1\tilde{X}_1 + b_2\tilde{X}_2 + \dots + b_k\tilde{X}_k,$$

allora deve esistere anche $\text{var}(\tilde{Y})$ e

$$\text{var}(\tilde{Y}) = \text{var}\left(a + \sum_{i=1}^k b_i\tilde{X}_i\right) = \sum_{i=1}^k \sum_{j=1}^k b_i b_j \sigma_{ij} \quad (77)$$

dove

$$\sigma_{ij} = \begin{cases} \text{cov}(\tilde{X}_i, \tilde{X}_j) & \text{se } i \neq j, \\ \text{var}(\tilde{X}_i) & \text{se } i = j. \end{cases}$$

In particolare, se le variabili casuali $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_k$ sono indipendenti, si ottiene

$$\text{var}(\tilde{Y}) = \text{var}\left(a + \sum_{i=1}^k b_i\tilde{X}_i\right) = \sum_{i=1}^k b_i^2 \sigma_{ii} = \sum_{i=1}^k b_i^2 \text{var}(\tilde{X}_i)$$

perché in questo caso le covarianze sono tutte nulle (proprietà V6).

Come abbiamo visto nella proprietà V8, ...

- nel caso di una trasformazione affine di una singola variabile casuale, la formula (77) per calcolare la sua varianza si riduce a

$$\text{var}(\tilde{Y}) = \text{var}(a + b_1\tilde{X}_1) = b^2\sigma_{11} = b^2\text{var}(\tilde{X}_1).$$

- Nel caso di una trasformazione affine di $k = 2$ variabili casuali (proprietà V8*), i termini della doppia sommatoria nella formula (77) danno invece luogo alla formula

$$\text{var}(\tilde{Y}) = \text{var}(a + b_1\tilde{X}_1 + b_2\tilde{X}_2) = \sum_{i=1}^2 \sum_{j=1}^2 b_i b_j \sigma_{ij} = b_1^2\sigma_{11} + b_2^2\sigma_{22} + 2b_1b_2\sigma_{12} \quad (78)$$

perché $\sigma_{ij} = [[\text{proprietà V7}]] = \sigma_{ji}$.

- Nel caso di una trasformazione affine di $k = 3$ variabili casuali, i termini della doppia sommatoria nella formula (77) danno luogo alla formula

$$\begin{aligned} \text{var}(\tilde{Y}) &= \text{var}(a + b_1\tilde{X}_1 + b_2\tilde{X}_2 + b_3\tilde{X}_3) = \sum_{i=1}^3 \sum_{j=1}^3 b_i b_j \sigma_{ij} \\ &= b_1^2\sigma_{11} + b_2^2\sigma_{22} + b_3^2\sigma_{33} + 2b_1b_2\sigma_{12} + 2b_1b_3\sigma_{13} + 2b_2b_3\sigma_{23}. \end{aligned} \quad (79)$$

- ecc..

Come si evince dalle formule (78) e (79), nella doppia sommatoria

$$\sum_{i=1}^k \sum_{j=1}^k b_i b_j \sigma_{ij}$$

tutti i termini $b_i b_j \sigma_{ij}$ che hanno indici i e j diversi compaiono due volte.

Si noti che la formula (77) può essere espressa anche in forma matriciale. Infatti, ponendo

$$\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix} \quad \text{e} \quad \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1k} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{k1} & \sigma_{k2} & \cdots & \sigma_{kk} \end{bmatrix},$$

non è difficile verificare che

$$\mathbf{b}^\top \Sigma \mathbf{b} = \sum_{i=1}^k \sum_{j=1}^k b_i b_j \sigma_{ij}, \quad (80)$$

dove \mathbf{b}^\top è il vettore trasposto del vettore \mathbf{b} .

V9) (Covarianza tra trasformazioni affini di una singola variabile casuale). Se $\tilde{Y} = a + b\tilde{X}_1$ e $cov(\tilde{X}_1, \tilde{X}_2)$ esiste, allora deve esistere anche $cov(\tilde{Y}, \tilde{X}_2)$ e

$$cov(\tilde{Y}, \tilde{X}_2) = cov(a + b\tilde{X}_1, \tilde{X}_2) = b cov(\tilde{X}_1, \tilde{X}_2).$$

Più in generale, se

$$\tilde{Y}_1 = a + b\tilde{X}_1 \quad \text{e} \quad \tilde{Y}_2 = c + d\tilde{X}_2$$

e $cov(\tilde{X}_1, \tilde{X}_2)$ esiste, allora deve esistere anche $cov(\tilde{Y}_1, \tilde{Y}_2)$ e

$$cov(\tilde{Y}_1, \tilde{Y}_2) = cov(a + b\tilde{X}_1, c + d\tilde{X}_2) = bd cov(\tilde{X}_1, \tilde{X}_2). \quad (81)$$

V9*) (Covarianza tra trasformazioni affini - caso generale). Se

$$\tilde{Y}_1 = a + \sum_{i=1}^k b_i \tilde{X}_i \quad \text{e} \quad \tilde{Y}_2 = c + \sum_{i=1}^k d_i \tilde{X}_i$$

e se le **varianze** di tutte e k le variabili casuali \tilde{X}_i esistono, allora deve esistere anche $cov(\tilde{Y}_1, \tilde{Y}_2)$ e

$$cov(\tilde{Y}_1, \tilde{Y}_2) = \sum_{i=1}^k \sum_{j=1}^k b_i d_j \sigma_{ij} \quad (82)$$

dove

$$\sigma_{ij} = \begin{cases} cov(\tilde{X}_i, \tilde{X}_j) & \text{se } i \neq j, \\ var(\tilde{X}_i) & \text{se } i = j. \end{cases}$$

Si noti che la formula (81), che qui per comodità riscriviamo

$$cov(\tilde{Y}_1, \tilde{Y}_2) = cov(a + b\tilde{X}_1, c + d\tilde{X}_2) = bd cov(\tilde{X}_1, \tilde{X}_2),$$

è il caso particolare della formula (82) che si ottiene se $k = 2$, $b_1 = b$, $d_2 = d$ e $b_2 = d_1 = 0$. Infatti, in questo caso particolare si ha

$$\tilde{Y}_1 = a + b_1 \tilde{X}_1 \quad \text{e} \quad \tilde{Y}_2 = c + d_2 \tilde{X}_2,$$

e applicando la formula (82) si ottiene

$$cov(\tilde{Y}_1, \tilde{Y}_2) = \sum_{i=1}^2 \sum_{j=1}^2 b_i d_j \sigma_{ij} = b_1 d_2 \sigma_{12} = b_1 d_2 cov(\tilde{X}_1, \tilde{X}_2).$$

Si noti che sul lato destro di quest'ultima formula mancano i tre termini

$$b_1 d_1 \sigma_{11}, \quad b_2 d_1 \sigma_{21} = b_2 d_1 \sigma_{12}, \quad b_2 d_2 \sigma_{22}$$

perché in tutti questi tre termini uno dei due coefficienti b_i o d_j è nullo.

Anche la formula (82) può essere espressa in forma matriciale. Infatti, ponendo

$$\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix}, \quad \mathbf{d} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_k \end{bmatrix} \quad \text{e} \quad \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1k} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{k1} & \sigma_{k2} & \cdots & \sigma_{kk} \end{bmatrix},$$

non è difficile verificare che

$$\mathbf{b}^\top \Sigma \mathbf{d} = \sum_{i=1}^k \sum_{j=1}^k b_i d_j \sigma_{ij}. \quad (83)$$

V10) (Disuguaglianza di Cebicef). Se $\text{var}(\tilde{X})$ esiste, allora si ha

$$P\left(\left\{\left|\tilde{X} - E(\tilde{X})\right| > \epsilon\right\}\right) \leq \frac{\text{var}(\tilde{X})}{\epsilon^2} \quad \text{per ogni } \epsilon > 0.$$

6.3 Applicazioni

Come si evince dalle proprietà E4*, V8* e V9*, a partire dai valori attesi e dalla **matrice di varianza e covarianza** si possono calcolare valori attesi, varianze e covarianze che si riferiscono a **trasformazioni affini** di un insieme di variabili casuali. In questa sezione presenteremo due trasformazioni affini che sono molto importanti in ambito finanziario e illustreremo alcune applicazioni delle suddette proprietà.

6.3.1 Valore atteso e varianza del rendimento di un portafoglio titoli

Consideriamo un portafoglio con C_1 euro investiti in un primo titolo, C_2 euro investiti in un secondo titolo, ecc.. Supponiamo che il portafoglio sia composto da k titoli e indichiamo con

$$C_P = C_1 + C_2 + \cdots + C_k$$

il valore complessivo del portafoglio. In quanto segue assumeremo che $C_P > 0$.

Indichiamo ora con $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_k$ le variabili casuali che rappresentano i **rendimenti aritmetici** dei k titoli. Per definizione, il rendimento aritmetico di un titolo è la variazione relativa del valore del titolo durante un certo periodo di riferimento (giorno, settimana, mese, anno). Se consideriamo un periodo di investimento durante il quale il possessore del portafoglio non effettua compravendite di titoli e indichiamo con \tilde{C}_1 ,

$\tilde{C}_2, \dots, \tilde{C}_k$ i valori finali dei k titoli inclusi nel portafoglio, allora possiamo calcolare i rendimenti aritmetici dei k titoli come

$$\tilde{X}_i = \frac{\tilde{C}_i - C_i}{C_i} = \frac{\tilde{C}_i}{C_i} - 1, \quad i = 1, 2, \dots, k,$$

e il valore finale dell'intero portafoglio può quindi essere espresso come

$$\begin{aligned} \tilde{C}_P &= \tilde{C}_1 + \tilde{C}_2 + \dots + \tilde{C}_k \\ &= C_1 \times (1 + \tilde{X}_1) + C_2 \times (1 + \tilde{X}_2) + \dots + C_k \times (1 + \tilde{X}_k). \end{aligned}$$

Usando questa formula vediamo che il rendimento aritmetico dell'intero portafoglio deve essere dato da

$$\tilde{Y}_P = \frac{\tilde{C}_P - C_P}{C_P} = \dots = \frac{C_1}{C_P} \times \tilde{X}_1 + \frac{C_2}{C_P} \times \tilde{X}_2 + \dots + \frac{C_k}{C_P} \times \tilde{X}_k$$

e indicando con $w_i = C_i/C_P$ la quota di capitale iniziale investita nell' i -esimo titolo, possiamo esprimere il rendimento aritmetico dell'intero portafoglio come una media ponderata dei rendimenti aritmetici dei singoli titoli:

$$\tilde{Y}_P = \frac{\tilde{C}_P - C_P}{C_P} = w_1 \times \tilde{X}_1 + w_2 \times \tilde{X}_2 + \dots + w_k \times \tilde{X}_k. \quad (84)$$

Per derivare questa formula abbiamo usato soltanto l'ipotesi che il valore iniziale C_P del portafoglio sia positivo. La formula (84) rimane dunque valida anche se alcuni pesi sono negativi a patto che la somma di tutti i pesi sia pari a 1. Si noti che il caso di uno o più pesi negativi si verifica se e solo se uno o più titoli vengono venduti allo scoperto e i proventi delle vendite vengono investiti negli altri titoli.

La formula (84) esprime il rendimento aritmetico di un portafoglio come una trasformazione lineare (e quindi affine) dei rendimenti aritmetici dei singoli titoli che compongono il portafoglio. Quindi possiamo applicare la proprietà E4* del valore atteso e la proprietà V8* della varianza onde concludere che

$$\mu_P = E(\tilde{Y}_P) = [[\text{proprietà E4*}]] = \sum_{i=1}^k w_i \times E(\tilde{X}_i) = \sum_{i=1}^k w_i \mu_i \quad (85)$$

e che

$$\sigma_P^2 = var(\tilde{Y}_P) = [[\text{proprietà V8*}]] = \sum_{i=1}^k \sum_{j=1}^k w_i w_j \sigma_{ij} = [[\text{formula (80)}]] = \mathbf{w}^T \mathbf{\Sigma} \mathbf{w}. \quad (86)$$

In quest'ultima formula abbiamo indicato con \mathbf{w} il vettore i cui elementi sono dati dai pesi w_1, w_2, \dots, w_k .

Esercizio 6.2. Un investitore ha investito $C_1 = 100$ mila euro nel titolo A e $C_2 = 150$ mila euro nel titolo B .

Si supponga che i rendimenti annuali dei due titoli siano descritti da due variabili casuali \tilde{X}_1 (rendimento annuale del titolo A) e \tilde{X}_2 (rendimento annuale del titolo B) con

$$\mu_1 = E(\tilde{X}_1) = 0,06, \quad \mu_2 = E(\tilde{X}_2) = 0,03$$

e con matrice di varianza e covarianza

$$\Sigma = \begin{bmatrix} 0,0009 & -0,0001 \\ -0,0001 & 0,0002 \end{bmatrix}$$

Si calcolino il rendimento annuale atteso e la varianza del rendimento annuale del portafoglio titoli dell'investitore.

Risposta

In questa risposta interpreteremo tutti i rendimenti come rendimenti **aritmetici**. Così facendo possiamo infatti esprimere il rendimento del portafoglio come

$$\begin{aligned} \tilde{Y}_P &= [[\text{formula (84)}]] = \frac{100}{250} \times \tilde{X}_1 + \frac{150}{250} \times \tilde{X}_2 \\ &= 0,4 \times \tilde{X}_1 + 0,6 \times \tilde{X}_2 \end{aligned}$$

e possiamo sfruttare le formule (85) e (86) per calcolare il rendimento atteso $\mu_P = E(\tilde{Y}_P)$ e la varianza $\sigma_P^2 = var(\tilde{Y}_P)$:

$$\begin{aligned} \mu_P &= E(\tilde{Y}_P) = E(0,4 \times \tilde{X}_1 + 0,6 \times \tilde{X}_2) \\ [[\text{formula (85)}]] &= 0,4 \times E(\tilde{X}_1) + 0,6 \times E(\tilde{X}_2) \\ &= 0,4 \times 0,06 + 0,6 \times 0,03 = 0,042; \end{aligned}$$

$$\begin{aligned} \sigma_P^2 &= var(\tilde{Y}_P) = var(0,4 \times \tilde{X}_1 + 0,6 \times \tilde{X}_2) \\ [[\text{formula (86)}]] & \\ &= 0,4^2 \times var(\tilde{X}_1) + 0,6^2 \times var(\tilde{X}_2) + 2 \times 0,4 \times 0,6 \times cov(\tilde{X}_1, \tilde{X}_2) \\ &= 0,4^2 \times 0,0009 + 0,6^2 \times 0,0002 + 2 \times 0,4 \times 0,6 \times (-0,0001) \\ &= 0,000168. \end{aligned}$$

Nel prossimo esercizio prenderemo spunto dall'**Arbitrage Pricing Theory (APT)** di Stephen A. Ross²⁵ per presentare un'applicazione della proprietà V9* della covarianza. L'APT è una nota teoria che descrive la relazione tra il rischio e il rendimento di

²⁵Vedi Stephen A. Ross (1976), "The arbitrage theory of capital asset pricing", *Journal of Economic Theory*. 13 (3): 341-360

attività finanziarie. In questa dispensa non presenteremo i dettagli di questa teoria, ma ci limiteremo ad esporre alcuni tratti essenziali della teoria del portafoglio di Markowitz che può essere considerata come precursore dell'APT.

Esercizio 6.3. Si supponga che i rendimenti aritmetici di due titoli azionari siano descritti da due variabili casuali \tilde{Y}_1 e \tilde{Y}_2 che sono definite come segue:

$$\begin{aligned}\tilde{Y}_1 &= 0,01 + 2\tilde{X}_1 - \tilde{X}_2 + \tilde{X}_3 \\ \tilde{Y}_2 &= 0,06 - 3\tilde{X}_1 + 4\tilde{X}_2 + \tilde{X}_4\end{aligned}$$

dove

- \tilde{X}_1 e \tilde{X}_2 sono due variabili casuali che rappresentano **fattori di rischio sistemici** che influenzano i rendimenti di entrambi i titoli,
- e dove \tilde{X}_3 e \tilde{X}_4 sono due variabili casuali che rappresentano **fattori di rischio idiosincratici** che influenzano il rendimento di soltanto un singolo titolo.

Si assuma che i valori attesi delle variabili casuali che rappresentano i fattori di rischio siano tutti nulli e che la loro matrice di varianza e covarianza sia data da

$$\Sigma = \begin{bmatrix} 0,0004 & & & & \\ 0,0001 & 0,0008 & & & \\ 0 & 0 & 0,0002 & & \\ 0 & 0 & 0 & 0,0001 & \\ & & & & \end{bmatrix}$$

(si tenga presente che le matrici di varianza e covarianza sono sempre simmetriche).

- Si calcolino il valore atteso e la varianza delle variabili casuali che rappresentano i rendimenti dei due titoli.
- Si calcoli la covarianza tra le variabili casuali che rappresentano i rendimenti dei due titoli.
- Si consideri un portafoglio con il 30% del capitale investito nel primo titolo e il restante 70% del capitale investito nel secondo titolo. Si calcoli il valore atteso e la varianza del rendimento aritmetico del portafoglio.
- Si consideri un portafoglio che è composto soltanto dai due titoli in questione. Si determini la composizione del portafoglio che minimizza la varianza del rendimento aritmetico.

Risposte:

- a) Si calcolino il valore atteso e la varianza delle variabili casuali che rappresentano i rendimenti dei due titoli.

Tenendo presente che i valori attesi delle variabili casuali che rappresentano i fattori di rischio sono tutti nulli e applicando la proprietà E4* vediamo che

$$\begin{aligned} E(\tilde{Y}_1) &= E(0,01 + 2\tilde{X}_1 - \tilde{X}_2 + \tilde{X}_3) \\ &= 0,01 + 2E(\tilde{X}_1) - E(\tilde{X}_2) + E(\tilde{X}_3) = 0,01 \end{aligned}$$

e che

$$\begin{aligned} E(\tilde{Y}_2) &= E(0,06 - 3\tilde{X}_1 + 4\tilde{X}_2 + \tilde{X}_4) \\ &= 0,06 - 3E(\tilde{X}_1) + 4E(\tilde{X}_2) + E(\tilde{X}_4) = 0,06. \end{aligned}$$

Usando la proprietà V8* della varianza vediamo invece che le varianze dei rendimenti aritmetici dei due titoli sono date da

$$\begin{aligned} var(\tilde{Y}_1) &= var(0,01 + 2\tilde{X}_1 - \tilde{X}_2 + \tilde{X}_3) \\ &[[\text{proprietà V8* \& formula (80)}]] \\ &= \begin{bmatrix} 2 & -1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0,0004 & 0,0001 & 0 & 0 \\ 0,0001 & 0,0008 & 0 & 0 \\ 0 & 0 & 0,0002 & 0 \\ 0 & 0 & 0 & 0,0001 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \\ 1 \\ 0 \end{bmatrix} \\ &= 2 \times 2 \times 0,0004 + 2 \times (-1) \times 0,0001 + (-1) \times 2 \times 0,0001 + \\ &\quad + (-1) \times (-1) \times 0,0008 + 1 \times 1 \times 0,0002 \\ &= 0,0022 \end{aligned}$$

e

$$\begin{aligned} var(\tilde{Y}_2) &= var(0,06 - 3\tilde{X}_1 + 4\tilde{X}_2 + \tilde{X}_4) \\ &[[\text{proprietà V8* \& formula (80)}]] \\ &= \begin{bmatrix} -3 & 4 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0,0004 & 0,0001 & 0 & 0 \\ 0,0001 & 0,0008 & 0 & 0 \\ 0 & 0 & 0,0002 & 0 \\ 0 & 0 & 0 & 0,0001 \end{bmatrix} \begin{bmatrix} -3 \\ 4 \\ 0 \\ 1 \end{bmatrix} \\ &= (-3) \times (-3) \times 0,0004 + (-3) \times 4 \times 0,0001 + 4 \times (-3) \times 0,0001 + \\ &\quad + 4 \times 4 \times 0,0008 + 1 \times 1 \times 0,0001 \\ &= 0,0141. \end{aligned}$$

- b) Si calcoli la covarianza tra le variabili casuali che rappresentano i rendimenti dei due titoli.

Usando la proprietà V9* della covarianza otteniamo

$$\begin{aligned}
 cov(\tilde{Y}_1, \tilde{Y}_2) &= cov(0,01 + 2\tilde{X}_1 - \tilde{X}_2 + \tilde{X}_3, 0,06 - 3\tilde{X}_1 + 4\tilde{X}_2 + \tilde{X}_4) \\
 &[[\text{proprietà V9* \& formula (83)}]] \\
 &= \begin{bmatrix} 2 & -1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0,0004 & 0,0001 & 0 & 0 \\ 0,0001 & 0,0008 & 0 & 0 \\ 0 & 0 & 0,0002 & 0 \\ 0 & 0 & 0 & 0,0001 \end{bmatrix} \begin{bmatrix} -3 \\ 4 \\ 0 \\ 1 \end{bmatrix} \\
 &= 2 \times (-3) \times 0,0004 + 2 \times 4 \times 0,0001 + (-1) \times (-3) \times 0,0001 + \\
 &\quad + (-1) \times 4 \times 0,0008 \\
 &= -0,0045.
 \end{aligned}$$

- c) Si consideri un portafoglio con il 30% del capitale investito nel primo titolo e il restante 70% del capitale investito nel secondo titolo. Si calcoli il valore atteso e la varianza del rendimento aritmetico del portafoglio.

Come abbiamo visto nella formula (84), il rendimento aritmetico del portafoglio può essere espresso come

$$\tilde{Y}_P = w_1 \times \tilde{X}_1 + w_2 \times \tilde{X}_2 = 0,30 \times \tilde{X}_1 + 0,70 \times \tilde{X}_2,$$

e il valore atteso e la varianza del rendimento aritmetico del portafoglio possono essere calcolati come

$$\begin{aligned}
 \mu_P &= E(\tilde{Y}_P) \\
 &[[\text{formula (85)}]] \\
 &= w_1 \times E(\tilde{X}_1) + w_2 \times E(\tilde{X}_2) \\
 &[[\text{risposta al quesito a)}]] \\
 &= 0,30 \times 0,01 + 0,70 \times 0,06 = 0,045.
 \end{aligned}$$

$$\begin{aligned}
 \sigma_P^2 &= var(\tilde{Y}_P) \\
 &[[\text{formula (86) \& risposte ai quesiti a) e b)}]] \\
 &= \begin{bmatrix} 0,30 & 0,70 \end{bmatrix} \begin{bmatrix} 0,0022 & -0,0045 \\ -0,0045 & 0,0141 \end{bmatrix} \begin{bmatrix} 0,30 \\ 0,70 \end{bmatrix} \\
 &= 0,30 \times 0,30 \times 0,022 + 0,30 \times 0,70 \times (-0,0045) + \\
 &\quad + 0,70 \times 0,30 \times (-0,0045) + 0,7 \times 0,7 \times 0,0141 \\
 &= 0,30^2 \times 0,0022 + 0,70^2 \times 0,0141 + 2 \times 0,30 \times 0,70 \times (-0,0045) \\
 &= 0,0052.
 \end{aligned}$$

- d) Si consideri un portafoglio che è composto soltanto dai due titoli in questione. Si determini la composizione del portafoglio che minimizza la varianza del rendimento aritmetico.

Come si desume dalla formula (86), la relazione tra la varianza $\sigma_P^2 = \text{var}(\tilde{Y}_P)$ e i pesi è descritta dall'equazione

$$\sigma_P^2 = \text{var}(\tilde{Y}_P) = w_1^2\sigma_{11} + w_2^2\sigma_{22} + 2w_1w_2\sigma_{12}$$

dove (vedi le risposte ai quesiti a) e b))

$$\sigma_{11} = \text{var}(\tilde{Y}_1) = 0,0022,$$

$$\sigma_{22} = \text{var}(\tilde{Y}_2) = 0,0141,$$

$$\sigma_{12} = \text{cov}(\tilde{Y}_1, \tilde{Y}_2) = -0,0045.$$

Siccome i pesi w_1 e w_2 devono sommare a 1 (si ricordi che i pesi rappresentano le quote di capitale investite nei due titoli), la varianza del rendimento del portafoglio può essere espressa come funzione di soltanto uno dei due pesi, diciamo di w_1 :

$$\begin{aligned} w_1 + w_2 = 1 &\Rightarrow w_2 = 1 - w_1 \Rightarrow \\ \Rightarrow \sigma_P^2 = \text{var}(\tilde{Y}_P) &= w_1^2\sigma_{11} + (1 - w_1)^2\sigma_{22} + 2w_1(1 - w_1)\sigma_{12} \Rightarrow \end{aligned}$$

$$\Rightarrow \sigma_P^2 = \text{var}(\tilde{Y}_P) = (\sigma_{11} + \sigma_{22} - 2\sigma_{12})w_1^2 - 2(\sigma_{22} - \sigma_{12})w_1 + \sigma_{22}. \quad (87)$$

Sostituendo i valori numerici delle varianze e covarianze otteniamo

$$\begin{aligned} \sigma_P^2 &= (0,0022 + 0,0141 + 2 \times 0,0045)w_1^2 - 2(0,0141 + 0,0045)w_1 + 0,0141 \\ &= 0,0253w_1^2 - 0,0372w_1 + 0,0141 \end{aligned}$$

e la relazione che lega la varianza del portafoglio al peso w_1 è dunque descritta dall'equazione

$$\sigma_P^2 = 0,0253w_1^2 - 0,0372w_1 + 0,0141.$$

Si noti che questa equazione può essere rappresentata attraverso una parabola convessa. Per trovare il peso w_1 che minimizza la varianza del rendimento aritmetico del portafoglio dobbiamo quindi trovare il punto di minimo di questa parabola, ovvero il valore di $w_1 = w_1^*$ che annulla la derivata

$$\frac{d\sigma_P^2}{dw_1} = 2 \times 0,0253w_1 - 0,0372.$$

Ponendo

$$\frac{d\sigma_P^2}{dw_1} = 0$$

otteniamo

$$2 \times 0,0253w_1 - 0,0372 = 0 \quad \Rightarrow \quad w_1 = \frac{0,0372}{2 \times 0,0253} = 0,7352.$$

Il portafoglio che minimizza la varianza è quindi composto per il $w_1 = 73,52\%$ del suo valore dal primo titolo e per il restante $w_2 = 1 - w_1 = 26,48\%$ del suo valore dal secondo titolo.

Prima di concludere questa sezione conviene analizzare anche la relazione tra il **rendimento logaritmico** di un portafoglio e i rendimenti logaritmici dei singoli titoli che lo compongono. A tal fine osserviamo che il rendimento logaritmico dell' i -esimo titolo del portafoglio è dato da

$$\tilde{X}'_i = \ln \tilde{C}_i - \ln C_i = \ln \frac{\tilde{C}_i}{C_i}$$

e che il valore finale del portafoglio può quindi essere espresso come

$$\begin{aligned} \tilde{C}_P &= \tilde{C}_1 + \tilde{C}_2 + \dots + \tilde{C}_k \\ &= C_1 \times e^{\tilde{X}'_1} + C_2 \times e^{\tilde{X}'_2} + \dots + C_k \times e^{\tilde{X}'_k} \end{aligned} \quad (88)$$

Usando questa espressione vediamo che il rendimento logaritmico del portafoglio è dato da

$$\begin{aligned} \tilde{Y}'_P &= \ln \tilde{C}_P - \ln C_P = \ln \frac{\tilde{C}_P}{C_P} \\ &= \ln \left(\frac{C_1}{C_P} \times e^{\tilde{X}'_1} + \frac{C_2}{C_P} \times e^{\tilde{X}'_2} + \dots + \frac{C_k}{C_P} \times e^{\tilde{X}'_k} \right) \\ &= \ln \left(w_1 \times e^{\tilde{X}'_1} + w_2 \times e^{\tilde{X}'_2} + \dots + w_k \times e^{\tilde{X}'_k} \right). \end{aligned}$$

La relazione tra il rendimento logaritmico del portafoglio e i rendimenti logaritmici dei singoli titoli che lo compongono è quindi descritta dall'equazione

$$\tilde{Y}'_P = \ln \left(w_1 \times e^{\tilde{X}'_1} + w_2 \times e^{\tilde{X}'_2} + \dots + w_k \times e^{\tilde{X}'_k} \right).$$

Si noti che questa relazione non è affine e quindi non possiamo sfruttare le proprietà **E4*** e **V8*** per esprimere il valore atteso e la varianza del rendimento logaritmico di un portafoglio come funzioni dei valori attesi e della matrice di varianza e covarianza dei rendimenti logaritmici dei singoli titoli che lo compongono.

6.3.2 Valore atteso e varianza del rendimento complessivo riferito a più periodi di investimento

Come abbiamo visto nella precedente sezione, per analizzare la relazione tra il rendimento di un portafoglio e i rendimenti dei singoli titoli che lo compongono conviene fare

riferimento ai **rendimenti aritmetici** e non a quelli **logaritmici** perché nel primo caso la relazione è lineare, mentre nel secondo caso la relazione non è né lineare né affine. In questa sezione vedremo invece che per analizzare il rendimento di un titolo durante un arco temporale che copre più periodi di investimento conviene, almeno da un certo punto di vista, fare riferimento ai rendimenti logaritmici e non a quelli aritmetici.

Supponiamo quindi che \tilde{P}_0 sia il prezzo all'inizio del primo periodo di investimento, e che $\tilde{P}_1, \tilde{P}_2, \dots$ siano i prezzi dello stesso titolo alla fine del primo periodo di investimento, alla fine del secondo periodo di investimento, ecc.. Si noti che secondo la notazione qui adottata \tilde{P}_0 è una variabile casuale, ma che in molte applicazioni il prezzo all'inizio del primo periodo di investimento è noto. In tal caso basta considerare \tilde{P}_0 come una variabile casuale degenera. Prima di procedere conviene inoltre chiarire che i periodi di investimento potrebbero anche avere durate diverse. In questa sezione non imponiamo dunque nessuna ipotesi sulle durate dei singoli periodi di investimento, ma assumeremo che i periodi di investimento siano consecutivi, ovvero che il prezzo finale di ciascun periodo di investimento coincida con il prezzo iniziale del periodo di investimento successivo. In virtù di quest'ultima ipotesi possiamo definire i rendimenti logaritmici che si riferiscono ai singoli periodi di investimento come

$$\tilde{X}_i = \ln \tilde{P}_i - \tilde{P}_{i-1} = \ln \frac{\tilde{P}_i}{\tilde{P}_{i-1}}, \quad i = 1, 2, \dots$$

e possiamo calcolare il rendimento logaritmico complessivo dei primi k periodi di investimento come

$$\begin{aligned} \tilde{Y}_k &= \ln \tilde{P}_k - \ln \tilde{P}_0 \\ &= (\ln \tilde{P}_k - \ln \tilde{P}_{k-1}) + (\ln \tilde{P}_{k-1} - \ln \tilde{P}_{k-2}) + \dots + (\ln \tilde{P}_1 - \ln \tilde{P}_0) \\ &= \tilde{X}_k + \tilde{X}_{k-1} + \dots + \tilde{X}_1 = \tilde{X}_1 + \tilde{X}_2 + \dots + \tilde{X}_k. \end{aligned}$$

Il rendimento logaritmico complessivo dei primi k periodi di investimento è quindi la somma dei rendimenti logaritmici dei singoli periodi:

$$\tilde{Y}_k = \ln \tilde{P}_k - \ln \tilde{P}_0 = \tilde{X}_1 + \tilde{X}_2 + \dots + \tilde{X}_k. \quad (89)$$

Trattandosi di una trasformazione affine possiamo quindi applicare la proprietà **E4*** del valore atteso e la proprietà **V8*** della varianza onde ottenere le formule

$$E(\tilde{Y}_k) = E(\tilde{X}_1) + E(\tilde{X}_2) + \dots + E(\tilde{X}_k) \quad (90)$$

e

$$var(\tilde{Y}_k) = \sum_{i=1}^k \sum_{j=1}^k \sigma_{ij}. \quad (91)$$

Se ora consideriamo il caso particolare in cui i rendimenti logaritmici \tilde{X}_i sono variabili casuali i.i.d. con valore atteso $\mu = E(\tilde{X}_i)$ e varianza $\sigma^2 = \sigma_{ii} = var(\tilde{X}_i)$ (si noti che in

questo caso tutte le covarianze $\sigma_{ij} = \text{cov}(\tilde{X}_i, \tilde{X}_j)$ con $i \neq j$ devono essere nulle - proprietà V1), le formule (90) e (91) si riducono a

$$E(\tilde{Y}_k) = k \times \mu \quad (92)$$

e

$$\text{var}(\tilde{Y}_k) = k \times \sigma^2. \quad (93)$$

Dalle formule (92) e (93) deduciamo che

- se i rendimenti logaritmici \tilde{X}_i sono rappresentati da variabili casuali i.i.d. che sono dotate di varianza (si ricordi che per definizione in questo caso deve esistere anche il loro valore atteso)
- e se i singoli periodi di investimento sono tutti della stessa durata,

allora il valore atteso e la varianza del rendimento logaritmico complessivo riferito ad un generico numero k di periodi di investimento sono entrambi proporzionali a k .

Esercizio 6.4. Secondo un investitore, i rendimenti logaritmici mensili di un titolo sono descritti da variabili casuali i.i.d. con valore atteso $\mu = 0,01$ e scarto quadratico medio $\sigma = 0,004$.

Si calcoli il valore atteso e la varianza del rendimento logaritmico annuale.

Risposta:

Usando le formule (92) e (93) vediamo che il valore atteso e la varianza del rendimento logaritmico annuale sono rispettivamente dati da

$$E(\tilde{Y}_k) = k \times \mu \quad \Rightarrow \quad E(\tilde{Y}_{12}) = 12 \times 0,01 = 0,12$$

e da

$$\text{var}(\tilde{Y}_k) = k \times \sigma^2 \quad \Rightarrow \quad \text{var}(\tilde{Y}_{12}) = 12 \times 0,004^2 = 0,000192.$$

Consideriamo ora invece la relazione tra il **rendimento aritmetico** complessivo riferito a più periodi di investimento e i **rendimenti aritmetici** dei singoli periodi. Ovviamente, i rendimenti aritmetici riferiti ai singoli periodi sono dati dalle variabili casuali

$$\tilde{X}_i = \frac{\tilde{P}_i - \tilde{P}_{i-1}}{\tilde{P}_{i-1}} = \frac{\tilde{P}_i}{\tilde{P}_{i-1}} - 1 \quad \text{per } i = 1, 2, \dots$$

e il rendimento aritmetico complessivo riferito ai primi k periodi di investimento è dunque dato da

$$\begin{aligned}\tilde{Y}_k &= \frac{\tilde{P}_k - \tilde{P}_0}{\tilde{P}_0} = \frac{\tilde{P}_k}{\tilde{P}_0} - 1 \\ &= \frac{\tilde{P}_k}{\tilde{P}_{k-1}} \times \frac{\tilde{P}_{k-1}}{\tilde{P}_{k-2}} \times \cdots \times \frac{\tilde{P}_1}{\tilde{P}_0} - 1 \\ &= \left(1 + \tilde{X}_k\right) \times \left(1 + \tilde{X}_{k-1}\right) \times \cdots \times \left(1 + \tilde{X}_1\right) - 1.\end{aligned}$$

La relazione tra il rendimento aritmetico complessivo \tilde{Y}_k e i rendimenti aritmetici \tilde{X}_i che si riferiscono ai singoli periodi è dunque descritta dall'equazione

$$\tilde{Y}_k = \left(1 + \tilde{X}_k\right) \times \left(1 + \tilde{X}_{k-1}\right) \times \cdots \times \left(1 + \tilde{X}_1\right) - 1.$$

Si noti che questa relazione non è affine e quindi non possiamo sfruttare le proprietà E4* e V8* per esprimere il valore atteso e la varianza del rendimento aritmetico complessivo come funzioni del valore atteso e della matrice di varianza e covarianza dei rendimenti aritmetici che si riferiscono ai singoli periodi di investimento.

6.3.3 Teoria del portafoglio di Markowitz

In questa sezione presenteremo alcuni tratti essenziali della teoria del portafoglio di Markowitz anche nota come "Modern Portfolio theory" (MPT). Questa teoria prese le mosse da un articolo scientifico di Harry Markowitz.²⁶ La MPT è basata su due ipotesi:

- 1) L'ipotesi che nel momento in cui un investitore sceglie la composizione del suo portafoglio titoli egli consideri soltanto un **singolo periodo di investimento** (un giorno, una settimana, un mese, ...).
- 2) L'ipotesi che l'investitore applichi il **criterio media-varianza**: questa ipotesi prevede che l'investitore confronti tutti i possibili portafogli P soltanto sulla base dei valori attesi μ_P e delle varianze σ_P^2 delle variabili casuali \tilde{Y}_P che descrivono i loro rendimenti e che
 - ... tra due portafogli P e P' con $\mu_P > \mu_{P'}$ e $\sigma_P^2 < \sigma_{P'}^2$, l'investitore preferisca sempre il portafoglio P ,
 - ... che a parità di rendimento atteso μ_P l'investitore preferisca portafogli con varianza minore σ_P^2 ,
 - ... e che a parità di varianza σ_P^2 l'investitore preferisca portafogli con rendimento atteso μ_P maggiore.

²⁶Vedi Harry Markowitz (1952), "Portfolio selection", *The Journal of Finance*, Vol. 7, No. 1, pp. 77-91.

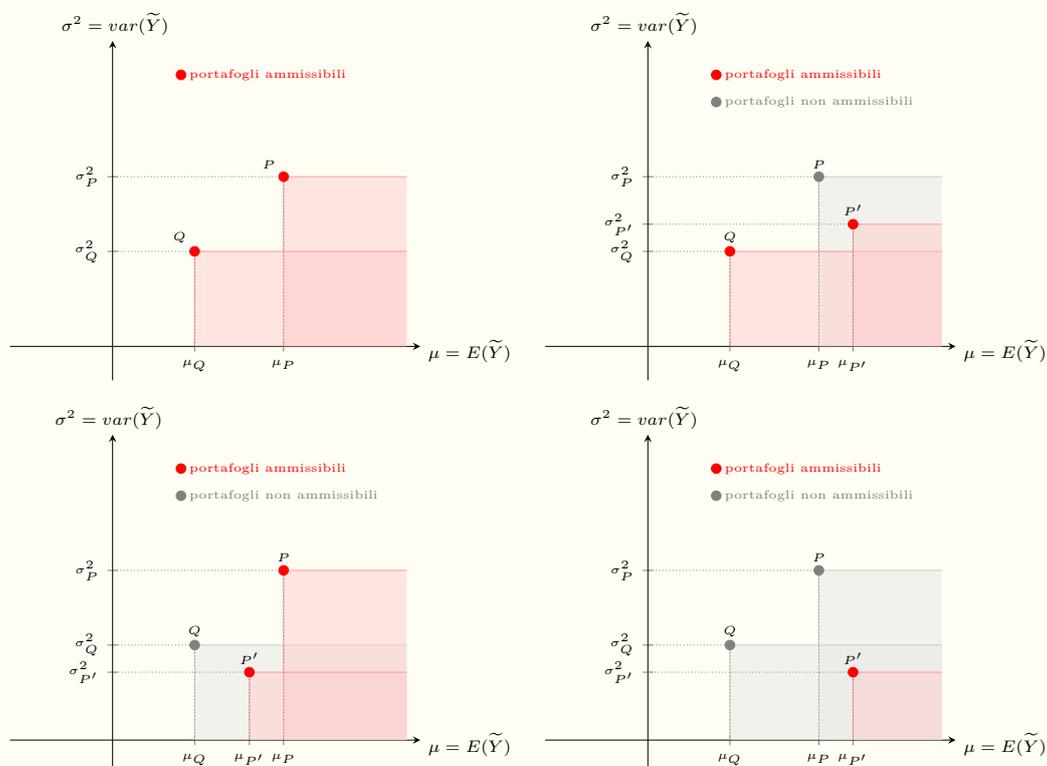
Nella teoria del portafoglio di Markowitz la varianza viene quindi considerata come una misura per il rischio di un portafoglio.

Il criterio media-varianza porta a distinguere tra portafogli che sono **ammissibili** e portafogli che non lo sono. I portafogli ammissibili sono i portafogli P rispetto ai quali non esistono altri portafogli P' tali che

- $\sigma_{P'}^2 \leq \sigma_P^2$ e $\mu_{P'} > \mu_P$ (ovvero portafogli P' che per un rischio minore o uguale a quello di P offrono un rendimento atteso strettamente maggiore)
- oppure tali che $\sigma_{P'}^2 < \sigma_P^2$ e $\mu_{P'} \geq \mu_P$ (ovvero portafogli P' che per un rischio strettamente minore di quello di P offrono una rendimento atteso maggiore o uguale a quello del portafoglio P).

Rappresentando tutti i possibili portafogli P attraverso dei punti $P = (\mu_P, \sigma_P^2)$ all'interno di un piano cartesiano, possiamo riformulare la definizione di portafoglio ammissibile in modo più intuitivo: un portafoglio P è ammissibile se e solo se non esistono altri portafogli P' che nel piano media-varianza si trovano a sud-est rispetto al punto $P = (\mu_P, \sigma_P^2)$ che rappresenta il portafoglio P (vedi Figura 6.1).

Figura 6.1. I grafici illustrano come si identificano i portafogli ammissibili nel piano media-varianza.



Nel prosieguo di questa sezione vedremo come a partire dai valori attesi e dalle varianze e covarianze dei rendimenti dei titoli che sono disponibili sul mercato si ricava

l'insieme di tutti portafogli P che secondo il criterio media-varianza di Markowitz sono ammissibili. Per semplicità **considereremo solo il caso dove sul mercato sono disponibili soltanto due titoli**. In questo caso possiamo indicare con \tilde{X}_1 la variabile casuale che rappresenta il rendimento del primo titolo, con \tilde{X}_2 la variabile casuale che rappresenta il rendimento del secondo titolo e possiamo esprimere il rendimento di un generico portafoglio P come

$$\tilde{Y}_P = \text{[[formula (84)]]} = w_1 \times \tilde{X}_1 + w_2 \times \tilde{X}_2, \quad (94)$$

dove $w_1 = C_1/C_P$ rappresenta la quota di capitale investita nel primo titolo e $w_2 = C_2/C_P$ rappresenta la quota di capitale investita nel secondo titolo. Come nell'articolo²⁶ di Markowitz considereremo soltanto portafogli P con valore iniziale positivo, ovvero soltanto portafogli P con $C_P = C_1 + C_2 > 0$, ma per evitare inutili complicazioni, ammetteremo anche la possibilità che uno dei due pesi w_1 oppure w_2 sia negativo. Si noti che il caso di un peso negativo si verifica se e solo se uno dei due titoli viene venduto allo scoperto e i proventi della vendita vengono investiti nell'altro titolo.

Siccome $w_1 + w_2 = 1$, possiamo esprimere il peso w_2 come $w_2 = 1 - w_1$, e sostituendo nella (94) possiamo dunque esprimere il rendimento di un generico portafoglio P soltanto in funzione del peso w_1 :

$$\tilde{Y}_P = w_1 \times \tilde{X}_1 + (1 - w_1) \times \tilde{X}_2.$$

Applicando la formula (85) vediamo dunque che il valore atteso del rendimento di un generico portafoglio P può essere espresso come

$$\begin{aligned} \mu_P &= E(\tilde{Y}_P) = E\left(w_1 \times \tilde{X}_1 + (1 - w_1) \times \tilde{X}_2\right) = \\ &\text{[[formula (85)]]} \\ &= w_1 E(\tilde{X}_1) + (1 - w_1) E(\tilde{X}_2) \\ &= w_1 \mu_1 + (1 - w_1) \mu_2 \\ &= \mu_2 + (\mu_1 - \mu_2) w_1 \end{aligned}$$

e applicando la formula (86) otteniamo anche un'espressione per la varianza:

$$\begin{aligned} \sigma_P^2 &= \text{var}(\tilde{Y}_P) = \text{var}\left(w_1 \tilde{X}_1 + (1 - w_1) \tilde{X}_2\right) = \\ &\text{[[formula (86)]]} \\ &= w_1^2 \sigma_{11} + (1 - w_1)^2 \sigma_{22} + 2w_1(1 - w_1) \sigma_{12} \\ &= w_1^2 (\sigma_{11} + \sigma_{22} - 2\sigma_{12}) - 2w_1(\sigma_{22} - \sigma_{12}) + \sigma_{22} \\ &= w_1^2 (\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}) - 2w_1(\sigma_2^2 - \sigma_{12}) + \sigma_2^2. \end{aligned}$$

Come si desume dalle espressioni

$$\mu_P = E(\tilde{Y}_P) = \mu_2 + (\mu_1 - \mu_2) w_1 \quad (95)$$

e

$$\sigma_P^2 = \text{var}(\tilde{Y}_P) = w_1^2(\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}) - 2w_1(\sigma_2^2 - \sigma_{12}) + \sigma_2^2 \quad (96)$$

che abbiamo appena ricavato, il rendimento atteso e la varianza dei rendimenti di un generico portafoglio P dipendono

- dai valori attesi $\mu_1 = E(\tilde{X}_1)$ e $\mu_2 = E(\tilde{X}_2)$,
- dalle varianze $\sigma_1^2 = \text{var}(\tilde{X}_1)$ e $\sigma_2^2 = \text{var}(\tilde{X}_2)$,
- dalla covarianza $\sigma_{12} = \text{cov}(\tilde{X}_1, \tilde{X}_2)$,
- e dal peso w_1 del primo titolo (si ricordi che $w_2 = 1 - w_1$).

Per determinare i pesi w_1 che identificano i portafogli P che sono ammissibili bisogna quindi "conoscere" i valori di

$$\begin{aligned} \mu_1 &= E(\tilde{X}_1), & \mu_2 &= E(\tilde{X}_2) \\ \sigma_1^2 &= \text{var}(\tilde{X}_1), & \sigma_2^2 &= \text{var}(\tilde{X}_2), & \sigma_{12} &= \text{cov}(\tilde{X}_1, \tilde{X}_2). \end{aligned} \quad (97)$$

Ovviamente, per qualsiasi coppia di titoli che si possono acquistare nel mondo reale questi valori sono ignoti. Per superare questo problema Markowitz suggerisce di stimarli a partire da dati storici e di applicare eventualmente degli aggiustamenti suggeriti da esperti. Chiaramente, quando si assegnano dei valori a questi parametri bisogna prestare attenzione non solo al fatto di scegliere dei valori "realistici", ma prima ancora al fatto di considerare soltanto valori che siano tra di loro "coerenti", ovvero di considerare soltanto valori $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$ e σ_{12} per i quali esistono un modello probabilistico (Ω, \mathcal{A}, P) e due variabili casuali $\tilde{X}_1 : \Omega \mapsto \mathbb{R}$ e $\tilde{X}_2 : \Omega \mapsto \mathbb{R}$ che soddisfano tutte e cinque le uguaglianze (97). Non è difficile dimostrare che questa condizione di esistenza è soddisfatta se e solo se i valori di σ_1^2, σ_2^2 e σ_{12} soddisfano la disuguaglianza di Cauchy-Schwarz (proprietà V7), ovvero se e solo se

$$\sigma_{12}^2 \leq \sigma_1^2 \sigma_2^2.$$

Assumendo che i valori dei parametri (97) siano noti e tra di loro coerenti, vedremo ora come si determina l'insieme di tutti i portafogli ammissibili, ovvero vedremo quali valori di w_1 corrispondono ai portafogli P che secondo il criterio media-varianza di Markowitz sono ammissibili. A tal fine conviene partire dall'equazione (96) che qui, per comodità, riscriviamo:

$$\sigma_P^2 = \text{var}(\tilde{Y}_P) = w_1^2(\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}) - 2w_1(\sigma_2^2 - \sigma_{12}) + \sigma_2^2$$

Come si vede da questa equazione, $\sigma_P^2 = \text{var}(\tilde{Y}_P)$ come funzione di $w_1 \in \mathbb{R}$ può essere soltanto una parabola oppure una retta. Siccome

$$\begin{aligned} \sigma_1^2 + \sigma_2^2 - 2\sigma_{12} &= \text{var}(\tilde{X}_1) + \text{var}(\tilde{X}_2) - 2\text{cov}(\tilde{X}_1, \tilde{X}_2) \\ [[\text{proprietà V8* e V0}]] &= \text{var}(\tilde{X}_1 - \tilde{X}_2) \geq 0, \end{aligned}$$

possiamo anche dire che ...

- se si tratta di una parabola, ovvero se $\sigma_1^2 + \sigma_2^2 - 2\sigma_{12} > 0$, allora la parabola deve essere convessa;
- se si tratta di una retta, ovvero se $\text{var}(\tilde{X}_1 - \tilde{X}_2) = 0$, allora $\tilde{X}_1 - \tilde{X}_2$ deve essere una variabile casuale **degenere**, ovvero deve esistere una costante reale a tale che $P(\{\tilde{X}_1 - \tilde{X}_2 = a\}) = 1$ (proprietà V0).

Il caso della retta. Consideriamo prima il caso in cui

$$\sigma_P^2 = \text{var}(\tilde{Y}_P) = w_1^2(\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}) - 2w_1(\sigma_2^2 - \sigma_{12}) + \sigma_2^2$$

come funzione di $w_1 \in \mathbb{R}$ è una retta. Come abbiamo appena visto, in questo caso deve esistere una costante reale a tale che

$$P(\{\tilde{X}_1 - \tilde{X}_2 = a\}) = 1 \quad \Leftrightarrow \quad P(\{\tilde{X}_1 = \tilde{X}_2 + a\}) = 1.$$

Usando questa condizione non è difficile dimostrare (tra breve vedremo la dimostrazione) che nel caso della retta possono verificarsi soltanto due situazioni:

- R1) **Non esiste nessun portafoglio ammissibile.** Questa situazione si verifica se il valore di a è diverso da zero. Infatti, se $a > 0$ possiamo migliorare con probabilità 1 il rendimento di qualsiasi portafoglio vendendo allo scoperto il secondo titolo per un importo aggiuntivo e investendo i proventi della vendita nel primo titolo che con probabilità uguale a 1 ha un rendimento maggiore. Con $a < 0$ possiamo mettere in atto l'operazione opposta. In altre parole, se $a \neq 0$ possiamo mettere in atto un'operazione di arbitraggio.
- R2) **Tutti i portafogli sono ammissibili.** Questa situazione si verifica se il valore di a è invece nullo. Infatti, in questo caso i rendimenti di entrambi i titoli sono uguali con probabilità 1 e pertanto qualsiasi portafoglio ha lo stesso rendimento con probabilità 1 (questo caso corrisponde dunque al caso in cui sul mercato esistono soltanto titoli di un unico tipo).

Dimostrazione. Per dimostrare in modo rigoroso che R1 e R2 sono le uniche due situazioni che possono verificarsi nel caso della retta, ovvero nel caso in cui

$$\sigma_P^2 = \text{var}(\tilde{Y}_P) = w_1^2(\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}) - 2w_1(\sigma_2^2 - \sigma_{12}) + \sigma_2^2$$

come funzione di $w_1 \in \mathbb{R}$ è una retta, partiamo dall'osservazione che nel caso di una retta si deve avere

$$\sigma_1^2 + \sigma_2^2 - 2\sigma_{12} = 0 \quad \Rightarrow \quad \sigma_{12} = \frac{\sigma_1^2 + \sigma_2^2}{2} \quad \Rightarrow \quad (98)$$

[[disuguaglianza di Cauchy-Schwarz - proprietà V7]]

$$\begin{aligned} \sigma_{12}^2 &= \left(\frac{\sigma_1^2 + \sigma_2^2}{2} \right)^2 \leq \sigma_1^2 \sigma_2^2 \quad \Rightarrow \quad \sigma_1^4 + \sigma_2^4 - 2\sigma_1^2 \sigma_2^2 \leq 0 \quad \Rightarrow \\ &\Rightarrow (\sigma_1^2 - \sigma_2^2)^2 \leq 0 \quad \Rightarrow \quad \sigma_1^2 = \sigma_2^2 \quad \Rightarrow \\ &\Rightarrow \sigma_1^2 = \sigma_2^2 = \text{[[seconda equazione nella riga (98)]]} = \sigma_{12} \end{aligned}$$

Quindi, se

$$\sigma_P^2 = \text{var}(\tilde{Y}_P) = w_1^2(\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}) - 2w_1(\sigma_2^2 - \sigma_{12}) + \sigma_2^2$$

è un retta, allora si deve avere $\sigma_1^2 = \sigma_2^2 = \sigma_{12}$ e di conseguenza si deve anche avere

$$\sigma_P^2 = \text{var}(\tilde{Y}_P) = \sigma_2^2 = \sigma_1^2 = \sigma_{12}$$

per qualsiasi scelta del peso w_1 . Questo ragionamento dimostra dunque che nel caso della retta **le varianze dei rendimenti di entrambi i titoli devono coincidere e si possono costruire solo portafogli P con varianza identica a quella dei due titoli che sono disponibili sul mercato.** Ricordando l'equazione (96) del rendimento atteso di un generico portafoglio, che qui per comodità riscriviamo

$$\mu_P = E(\tilde{Y}_P) = \mu_2 + (\mu_1 - \mu_2)w_1,$$

possiamo ora facilmente dimostrare che nel caso della retta possono effettivamente verificarsi soltanto le due situazioni che abbiamo già descritto in precedenza:

- **Non esiste nessun portafoglio ammissibile.** Questa situazione si verifica se $\mu_1 \neq \mu_2$: infatti, se $\mu_1 \neq \mu_2$ allora il rendimento atteso di qualsiasi portafoglio può essere aumentato aumentando (se $\mu_1 > \mu_2$) o diminuendo (se $\mu_1 < \mu_2$) il valore del peso w_1 . Siccome al variare di w_1 la varianza $\sigma_P^2 = \text{var}(\tilde{Y}_P)$ rimane costante, questa operazione conduce sempre ad un portafoglio preferibile a quello di partenza. Pertanto tutti i portafoglio sono "migliorabili" e nessun portafoglio è quindi ammissibile.
- **Tutti i portafogli sono ammissibili.** Questa situazione si verifica se $\mu_1 = \mu_2$. In questo caso al variare di w_1 rimane costante anche il rendimento atteso $\mu_P = E(\tilde{Y}_P)$ (vedi l'equazione (96)) e tutti i portafogli hanno quindi lo stesso rendimento atteso $\mu_P = E(\tilde{Y}_P)$ e lo stesso rischio $\sigma_P^2 = \text{var}(\tilde{Y}_P)$.

□

Il caso della parabola. Consideriamo ora invece il caso in cui

$$\sigma_P^2 = \text{var}(\tilde{Y}_P) = w_1^2(\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}) - 2w_1(\sigma_2^2 - \sigma_{12}) + \sigma_2^2$$

come funzione di $w_1 \in \mathbb{R}$ è una **parabola**. Come abbiamo già osservato, in questo caso si deve avere

$$\text{var}(\tilde{X}_1 - \tilde{X}_2) = [[\text{proprietà V8*}]] = \sigma_1^2 + \sigma_2^2 - 2\sigma_{12} > 0$$

e pertanto la parabola deve essere convessa. Per determinare il punto di minimo della parabola calcoliamo dunque la derivata

$$\frac{d\sigma_P^2}{dw_1} = \frac{d \text{var}(\tilde{Y}_P)}{dw_1} = 2(\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}) - 2(\sigma_2^2 - \sigma_{12})$$

e la poniamo uguale a zero. In questo modo vediamo che il minimo assoluto di $\sigma_P^2 = \text{var}(\tilde{Y}_P)$ deve trovarsi nel punto

$$w_1 = w_1^* = \frac{\sigma_2^2 - \sigma_{12}}{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}}.$$

A questo punto distinguiamo due casi:

- 1) Il caso in cui $\mu_1 = \mu_2$, ovvero il caso in cui i rendimenti attesi di entrambi i titoli coincidono. In questo caso, per qualsiasi scelta del peso w_1 , il rendimento atteso

$$\mu_P = E(\tilde{Y}_P) = \mu_2 + (\mu_1 - \mu_2)w_1$$

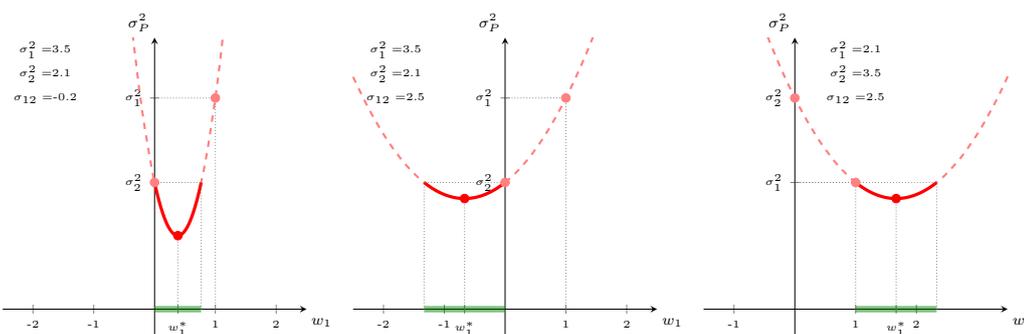
si riduce a

$$\mu_P = E(\tilde{Y}_P) = \mu_2 = \mu_1$$

e pertanto si possono costruire soltanto portafogli P con rendimento atteso identico a quello comune ai due titoli sono disponibili sul mercato. Tra tutti questi portafogli il portafoglio con peso

$$w_1 = w_1^* = \frac{\sigma_2^2 - \sigma_{12}}{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}}$$

è ovviamente l'unico portafoglio ammissibile in quanto è l'unico portafoglio che minimizza il valore di $\sigma_P^2 = \text{var}(\tilde{Y}_P)$. Come si desume dai tre grafici sottostanti (i tre grafici si riferiscono a tre situazioni diverse che si possono ottenere variando il valore di σ_{12}), ...



... se $\sigma_P^2 = \text{var}(\tilde{Y}_P)$ come funzione di w_1 è una parabola e $\mu_1 = \mu_2$, allora ...

- ... esistono infiniti portafogli portafogli P con rischio $\sigma_P^2 = \text{var}(\tilde{Y}_P)$ più piccolo di entrambi i portafogli che investono soltanto in uno dei due titoli presenti sul mercato, ovvero dei portafogli con $w_1 = 0$ e $w_1 = 1$ (striscia verde sull'asse delle ascisse)
- ... e, a seconda del valore di w_1^* , tutti questi infiniti portafogli devono avere
 - * peso w_1 compreso tra 0 e 1, se $0 < w_1^* < 1$ (primo grafico),
 - * peso $w_1 < 0$, se $w_1^* < 0$ (secondo grafico),
 - * oppure peso $w_1 > 1$, se $w_1^* > 1$ (terzo grafico).

- 2) Consideriamo ora invece il caso dove $\mu_1 \neq \mu_2$ (e dove $\sigma_P^2 = \text{var}(\tilde{Y}_P)$ come funzione di w_1 è una parabola).

In questo caso il rendimento atteso

$$\mu_P = E(\tilde{Y}_P) = \mu_2 + (\mu_1 - \mu_2)w_1$$

come funzione di $w_1 \in \mathbb{R}$ è una retta strettamente crescente oppure strettamente decrescente, e scegliendo opportunamente il valore di w_1 possiamo dunque ottenere portafogli P con qualunque rendimento atteso. Siccome

$$\mu_P = E(\tilde{Y}_P) = \mu_2 + (\mu_1 - \mu_2)w_1 \quad \Leftrightarrow \quad w_1 = \frac{\mu_P - \mu_2}{\mu_1 - \mu_2},$$

possiamo esprimere

$$\sigma_P^2 = \text{var}(\tilde{Y}_P) = w_1^2(\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}) - 2w_1(\sigma_2^2 - \sigma_{12}) + \sigma_2^2$$

come funzione di μ_P piuttosto che come funzione di w_1 : sostituendo

$$w_1 = \frac{\mu_P - \mu_2}{\mu_1 - \mu_2}$$

nell'espressione

$$\sigma_P^2 = \text{var}(\tilde{Y}_P) = w_1^2(\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}) - 2w_1(\sigma_2^2 - \sigma_{12}) + \sigma_2^2$$

otteniamo (invitiamo il lettore a verificare i passaggi algebrici)

$$\sigma_P^2 = \text{var}(\tilde{Y}_P) = A\mu_P^2 - 2B\mu_P + C$$

con

$$A = \frac{\sigma_{11} + \sigma_{22} - 2\sigma_{12}}{(\mu_1 - \mu_2)^2}, \quad B = \frac{\mu_1(\sigma_{22} - \sigma_{12}) + \mu_2(\sigma_{11} - \sigma_{12})}{(\mu_1 - \mu_2)^2},$$

$$C = \frac{\mu_1^2\sigma_{22} + \mu_2^2\sigma_{11} - 2\mu_1\mu_2\sigma_{12}}{(\mu_1 - \mu_2)^2}.$$

Attraverso la funzione

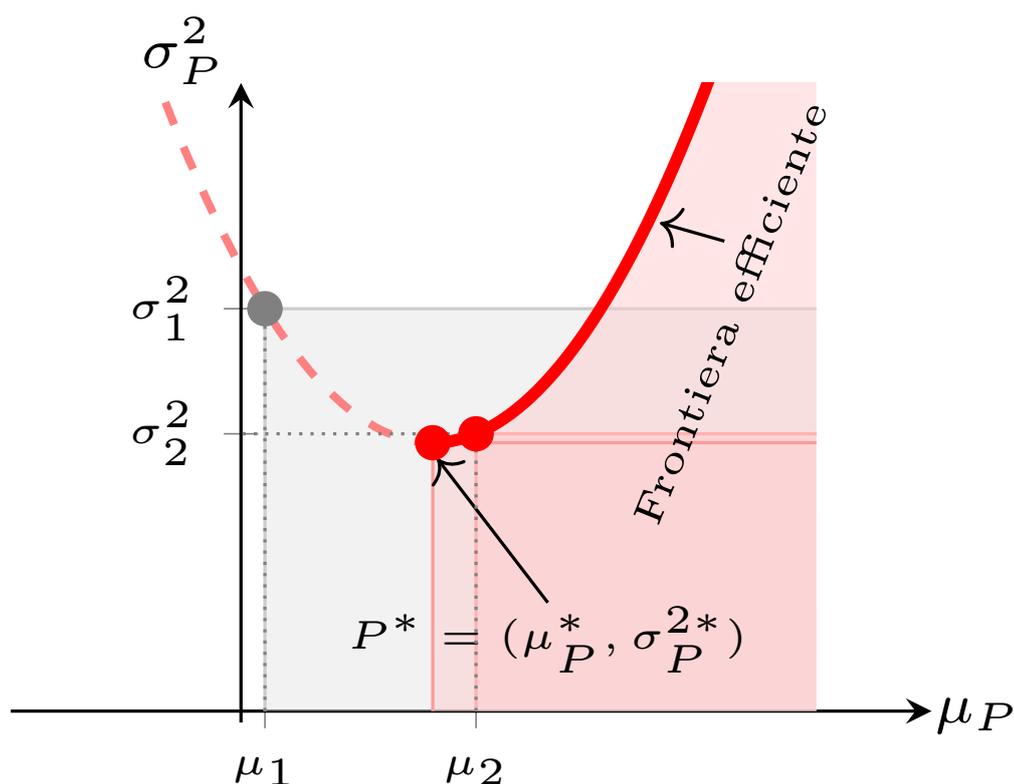
$$\sigma_P^2 = \text{var}(\tilde{Y}_P) = A\mu_P^2 - 2B\mu_P + C \quad (99)$$

possiamo ora valutare la posizione di tutti i possibili portafogli P nel piano media-varianza (si ricordi che la relazione tra μ_P e w_1 è biunivoca).

Siccome

$$A = \frac{\sigma_{11} + \sigma_{22} - 2\sigma_{12}}{(\mu_1 - \mu_2)^2} > 0 \quad \dots$$

(si ricordi che stiamo analizzando il caso $\text{var}(\tilde{X}_1 - \tilde{X}_2) = \sigma_{11} + \sigma_{22} - 2\sigma_{12} > 0$), possiamo immediatamente concludere che $\sigma_P^2 = \text{var}(\tilde{Y}_P)$ come funzione di $\mu_P = E(\tilde{Y}_P)$ deve essere una parabola convessa. Quindi possiamo concludere che **i portafogli P ammissibili (secondo il criterio media-varianza) sono quelli che si trovano sul ramo crescente di questa parabola**. Questo ramo viene chiamato **frontiera efficiente** ed è rappresentato nel grafico sottostante:



Per determinare il punto di minimo della parabola

$$\sigma_P^2 = \text{var}(\tilde{Y}_P) = A\mu_P^2 - 2B\mu_P + C$$

basta porre uguale a zero la derivata

$$\frac{d\sigma_P^2}{d\mu_P} = \frac{d\text{var}(\tilde{Y}_P)}{d\mu_P} = 2A - 2B.$$

In questo modo vediamo che il punto di minimo si trova nel punto

$$\mu_P^* = \frac{B}{A} = \frac{\mu_1(\sigma_{22} - \sigma_{12}) + \mu_2(\sigma_{11} - \sigma_{12})}{\sigma_{11} + \sigma_{22} - 2\sigma_{12}}$$

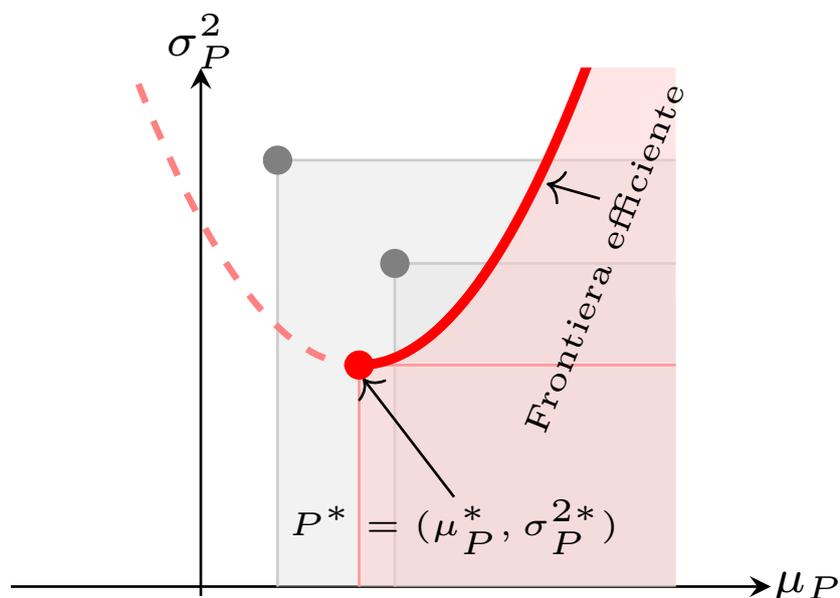
(si ricordi che stiamo analizzando il caso $\text{var}(\tilde{X}_1 - \tilde{X}_2) = \sigma_{11} + \sigma_{22} - 2\sigma_{12} > 0$).

Il grafico della frontiera efficiente che abbiamo visto poco sopra è stato ottenuto stimando i valori di μ_1 , μ_2 , σ_1^2 , σ_2^2 e σ_{12} sulla base dei rendimenti giornalieri (espressi in termini percentuali) dei titoli AAPL e AMZN del periodo 01/01/2021-30/04/2021 (fonte Bloomberg):

- Per il titolo AAPL si ottiene $\mu_1 = 0,01\%$ e $\sigma_1 = \sqrt{3,52} \%$;
- Per il titolo AMZN si ottiene $\mu_2 = 0,10\%$ e $\sigma_2 = \sqrt{2,43} \%$;
- e per la covarianza dei rendimenti (espressi come percentuali) si ottiene $\sigma_{12} = 2,05$.

Esaminando con attenzione il grafico della frontiera efficiente notiamo che anche se il portafoglio che è composto soltanto dal titolo AAPL (ovvero il punto (μ_1, σ_1^2)) non è ammissibile, esistono comunque infiniti portafogli ammissibili che contengono una quota positiva w_1 del titolo AAPL, ovvero tutti i portafogli sulla frontiera efficiente il cui rendimento è compreso tra μ_P^* e μ_2 .²⁷

Prima di concludere questa sezione conviene aggiungere qualche commento sul caso più realistico dove sul mercato sono disponibili tre o più titoli. Ovviamente, anche in questo caso tutti i portafogli che possono essere costruiti attraverso i titoli disponibili sul mercato possono essere rappresentati attraverso dei punti (μ_P, σ_P^2) nel piano media-varianza e si può dimostrare che (salvo casi eccezionali) tutti i punti che rappresentano dei portafogli ottenibili con i titoli presenti sul mercato devono trovarsi "all'interno" oppure "sul bordo" dell'area delimitata da una parabola convessa (vedi il grafico sottostante)



Anche nel caso in cui sono disponibili più di due titoli, l'insieme dei portafogli ammissibili si trova dunque sul ramo crescente di una parabola convessa!!! Siccome il ramo crescente di questa parabola separa l'insieme delle combinazioni (μ_P, σ_P^2) che sono "raggiungibili" dall'insieme di tutte le combinazioni $(\mu_{P'}, \sigma_{P'}^2)$ che secondo il criterio media-varianza sarebbero preferibili ma che non sono raggiungibili attraverso i titoli disponibili sul mercato, questo ramo della parabola viene chiamato "**frontiera efficiente**". Si noti che nel caso particolare che abbiamo esaminato in questa sezione, ovvero nel caso in cui sono disponibili soltanto due titoli, tutti i portafogli $P = (\mu_P, \sigma_P^2)$ raggiungibili devono

²⁷Infatti, i portafogli che contengono una quota positiva del titolo AAPL sono quelli che hanno un rendimento atteso minore di μ_2 e i punti sulla frontiera efficiente con $\mu_P < \mu_2$ rappresentano portafogli ammissibili con rendimento atteso minore di μ_2 .

trovarsi allineati lungo una parabola convessa e che in questo caso particolare non esistono dunque altri portafogli raggiungibili che si trovano "all'interno" della parabola.

7 Momenti e funzione generatrice dei momenti

7.1 I momenti

Come vedremo tra breve, la successione dei valori attesi

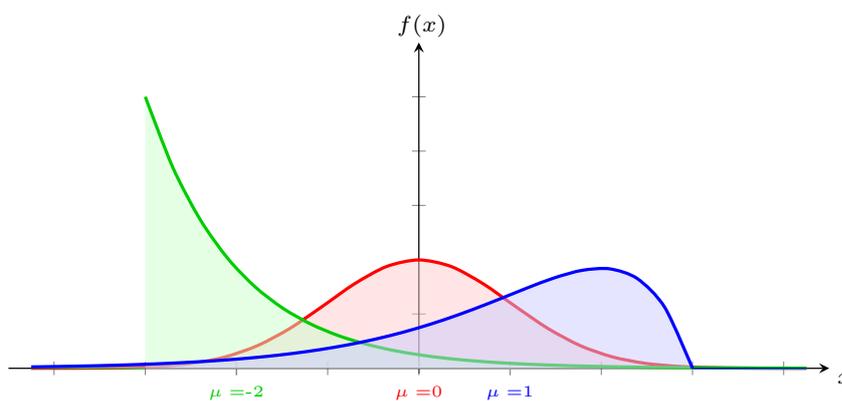
$$\mu_1 = E(\tilde{X}), \quad \mu_2 = E(\tilde{X}^2), \quad \mu_3 = E(\tilde{X}^3), \quad \dots,$$

se esistono, contiene molte informazioni sulla distribuzione della variabile casuale \tilde{X} e in alcuni casi può anche identificarla in modo univoco.

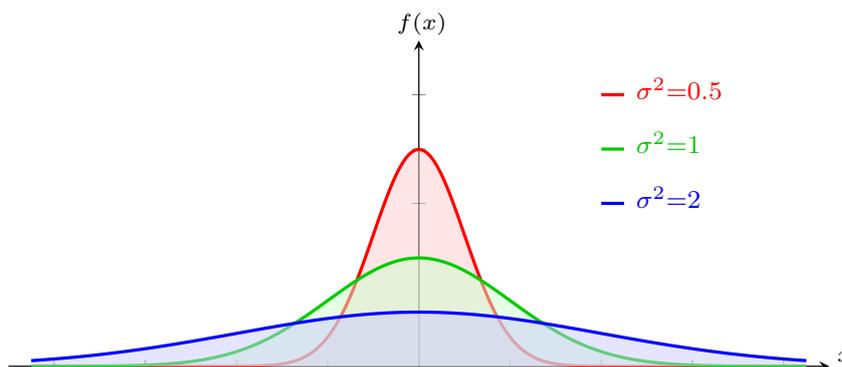
I valori attesi in questione vengono chiamati **momenti** della variabile casuale (o della sua distribuzione). In particolare, se k è un numero naturale, il corrispondente valore atteso $\mu_k = E(\tilde{X}^k)$, se esiste, viene chiamato **momento di ordine k** .

Per rendersi conto dell'importanza dei momenti basta osservare che ...

- $\mu_1 = E(\tilde{X})$ è il **valore atteso** della distribuzione di \tilde{X} . Il momento primo viene di solito indicato solo con la lettera greca μ senza il pedice. Il grafico sottostante mostra i momenti primi di alcune distribuzioni assolutamente continue. Si ricordi che per distribuzioni assolutamente continue il momento primo può essere interpretato baricentro delle corrispondenti funzioni di densità (vedi l'Esercizio 5.3).



- Il valore di $\mu_2 = E(\tilde{X}^2)$ è legato alla variabilità delle possibili realizzazioni di \tilde{X} . Infatti, per variabili casuali con valore atteso nullo si ha $\sigma^2 = var(\tilde{X}) = \mu_2$, mentre per variabili casuali con valore atteso $\mu = \mu_1$ qualunque si ha $\sigma^2 = var(\tilde{X}) = E([\tilde{X} - \mu]^2) = \mu_2 - \mu_1^2$ (formula indiretta per il calcolo della varianza - proprietà V4). Il grafico sottostante mostra alcune funzioni di densità con valore atteso nullo e con diversi valori del momento secondo $\mu_2 = E(\tilde{X}^2)$ (che coincide con la varianza visto che i valori attesi delle funzioni di densità rappresentate sono tutti e tre nulli):



- Il momento $\mu_3 = E(\tilde{X}^3)$ è invece legato all'**indice di asimmetria di Pearson**. Per una variabile casuale con valore atteso nullo questo indice è definito come

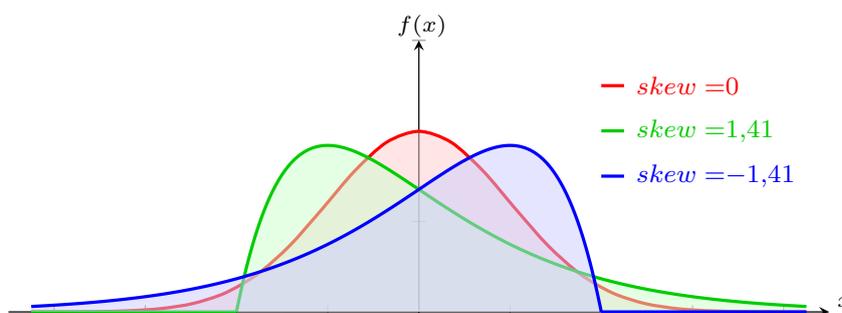
$$skew(\tilde{X}) = \frac{\mu_3}{\mu_2^{3/2}} = \frac{E(\tilde{X}^3)}{[E(\tilde{X}^2)]^{3/2}}.$$

La definizione generale dell'indice $skew(\cdot)$ di Pearson, ovvero quella che può essere applicata anche nel caso in cui il valore atteso $\mu = \mu_1 = E(\tilde{X})$ è diverso da zero, è invece data da

$$skew(\tilde{X}) = \frac{E([\tilde{X} - \mu]^3)}{[E([\tilde{X} - \mu]^2)]^{3/2}}.$$

Si può dimostrare che questo indice è ben definito se e solo se \tilde{X} non è una variabile casuale **degenere** e tutti e tre i momenti $\mu_1 = \mu = E(\tilde{X})$, $\mu_2 = E(\tilde{X}^2)$ e $\mu_3 = E(\tilde{X}^3)$ esistono. In tal caso può essere calcolato a partire da questi ultimi attraverso una formula che qui non riportiamo.

Come si vede nel grafico sottostante, l'indice di asimmetria $skew(\cdot)$ di Pearson è nullo per funzioni di densità (o di massa di probabilità) simmetriche; negativo per funzioni di densità (o di massa di probabilità) che "si allungano" verso sinistra e positivo per funzioni di densità (o di massa di probabilità) che "si allungano" verso destra.



- Il momento $\mu_4 = E(\tilde{X}^4)$ è legato all'indice di "curtosi" di Pearson, che per una variabile casuale con valore atteso nullo è definito come

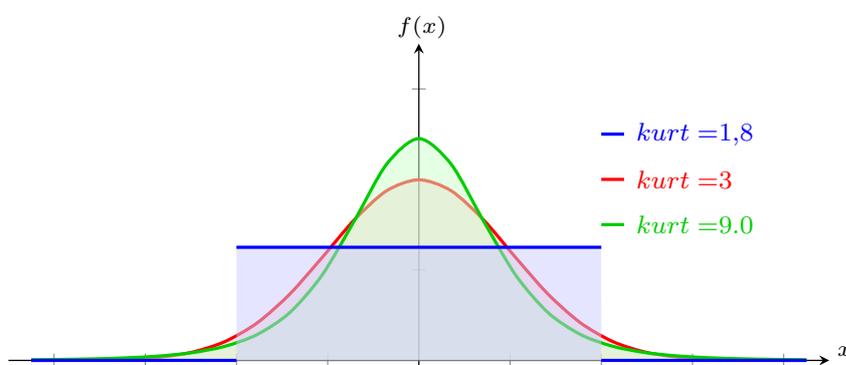
$$kurt(\tilde{X}) = \frac{\mu_4}{\mu_2^2} = \frac{E(\tilde{X}^4)}{[E(\tilde{X}^2)]^2}.$$

In generale, per una variabile casuale con valore atteso $\mu = \mu_1 = E(\tilde{X})$ qualunque, l'indice di curtosi di Pearson è invece definito come

$$kurt(\tilde{X}) = \frac{E([\tilde{X} - \mu]^4)}{[E([\tilde{X} - \mu]^2)]^2}.$$

Si può dimostrare che l'indice $kurt(\tilde{X})$ è ben definito se e solo se \tilde{X} non è una variabile casuale **degenere** e tutti e quattro i momenti $\mu_1 = \mu = E(\tilde{X})$, $\mu_2 = E(\tilde{X}^2)$, $\mu_3 = E(\tilde{X}^3)$ e $\mu_4 = E(\tilde{X}^4)$ esistono. In tal caso il valore di $kurt(\tilde{X})$ può essere calcolato a partire da questi ultimi attraverso una formula che qui non riportiamo.

L'interpretazione dell'indice di curtosi è una questione controversa: alcuni autori sostengono che misuri contemporaneamente il grado di appiattimento e anche lo spessore delle code di una funzione di densità (o di massa di probabilità), altri sostengono che l'indice di curtosi non dica nulla sull'appiattimento ma che riguardi soltanto lo spessore delle code. Il grafico sottostante mostra alcune funzioni di densità simmetriche con diversi valori dell'indice di curtosi.



Sulla base del valore dell'indice $kurt(\cdot)$, le distribuzioni possono essere classificate in tre categorie:

- la categoria delle distribuzioni **platicurtiche** (se $kurt(\cdot) < 3$),
- la categoria delle distribuzioni **mesocurtiche** (se $kurt(\cdot) = 3$)
- la categoria delle distribuzioni **leptocurtiche** (se $kurt(\cdot) > 3$)

Il valore $kurt(\cdot) = 3$ che separa le distribuzioni platicurtiche da quelle leptocurtiche è il valore dell'indice di curtosi comune a tutte le distribuzioni "normali" che introdurremo più avanti.

7.2 La funzione generatrice dei momenti (fgm)

Come abbiamo accennato nella sezione precedente, non è detto che tutti i momenti di una distribuzione esistano, ma si può dimostrare che **se per un dato ordine $k > 0$ il corrispondente momento $\mu_k = E(\tilde{X}^k)$ esiste, allora devono esistere anche tutti i momenti di ordine inferiore a k !!!**

Quando tra breve analizzeremo le proprietà di alcune famiglie di distribuzioni notevoli, vedremo che **per molte distribuzioni esiste l'intera successione dei momenti** (ovvero esistono i momenti di ogni ordine k finito). Se tutti i momenti della distribuzione di una variabile casuale \tilde{X} esistono, allora in molti casi (ma non sempre!!!) esiste anche un valore reale $\epsilon > 0$ tale che

$$\text{il valore atteso } E\left(e^{t\tilde{X}}\right) \text{ esiste per ogni } t \in (-\epsilon, +\epsilon).$$

Se questa condizione è soddisfatta (si ricordi che l'esistenza del valore atteso di una variabile casuale o di una sua funzione dipende solo dalla distribuzione della variabile casuale), allora si dice che la distribuzione della variabile casuale \tilde{X} è dotata di **funzione generatrice dei momenti (fgm)** e quest'ultima è definita come

$$m(t) = E\left(e^{t\tilde{X}}\right)$$

per ogni $t \in \mathbb{R}$ dove il valore atteso al secondo membro esiste. Invece di dire che la distribuzione di una variabile casuale "è dotata" di fgm, spesso si dice anche che la fgm della distribuzione "esiste"; e invece di parlare della fgm della "distribuzione" di una variabile casuale spesso si parla semplicemente della fgm di una "variabile casuale", anche se quest'ultimo modo di dire è fuorviante visto che tutte le variabili casuali con la stessa distribuzione hanno la stessa fgm (se esiste), ovvero visto che l'esistenza e il modo in cui eventualmente è definita una fgm dipendono soltanto dalla distribuzione della variabile casuale di riferimento.

Applicando la definizione di fgm al caso di una distribuzione discreta con funzione di massa di probabilità $p(x)$ possiamo facilmente verificare che la fgm di $p(x)$ (o della distribuzione che corrisponde a $p(x)$) esiste se e solo se esiste un $\epsilon > 0$ tale che

$$\sum_{x \in S} e^{tx} \times p(x) < \infty \quad \text{per ogni } t \in (-\epsilon, +\epsilon)$$

(si ricordi che la funzione esponenziale può assumere solo valori positivi), e che in tal caso la fgm è definita come

$$m(t) = \sum_{x \in S} e^{tx} \times p(x)$$

per ogni $t \in \mathbb{R}$ tale che la sommatoria al secondo membro converge. Il dominio della fgm $m(t)$ è dunque l'insieme dei numeri reali t per i quali la suddetta sommatoria è convergente. Si noti che la fgm di una funzione di massa di probabilità $p(x)$ (o della

corrispondente distribuzione) esiste sempre se il supporto di $p(x)$ è finito o, più in generale, limitato. In entrambi questi casi la fgm ha come dominio l'insieme di tutti i numeri reali.

Applicando invece la definizione di fgm al caso di una distribuzione assolutamente continua della quale $f(x)$ è una delle infinite funzioni di densità, possiamo facilmente verificare che la fgm di $f(x)$ (o della distribuzione che corrisponde a $f(x)$) esiste se e solo se esiste un $\epsilon > 0$ tale che

$$\int_{-\infty}^{\infty} e^{tx} \times f(x) dx < \infty \quad \text{per ogni } t \in (-\epsilon, +\epsilon)$$

(si ricordi che la funzione esponenziale può assumere solo valori positivi) e che in tal caso la fgm è definita come

$$m(t) = \int_{-\infty}^{\infty} e^{tx} \times f(x) dx$$

per ogni $t \in \mathbb{R}$ tale che l'integrale al secondo membro converge. Il dominio della fgm $m(t)$ è dunque l'insieme dei numeri reali t per i quali il suddetto integrale è convergente. Si noti che anche la fgm di una funzione di densità $f(x)$ (o della corrispondente distribuzione) esiste sempre se il supporto di $f(x)$ è limitato. In quest'ultimo caso la fgm ha come dominio l'insieme di tutti i numeri reali.

La fgm è un concetto di interesse prettamente teorico grazie al quale si possono snellire molte dimostrazioni. Infatti, facendo riferimento alla fgm e ad alcuni teoremi che la riguardano, si possono dedurre molti risultati che altrimenti sarebbero ostici da dimostrare (la parte difficile della dimostrazione viene quindi "scaricata" sui teoremi che riguardano la fgm - più avanti, quando studieremo le proprietà di alcune famiglie di distribuzioni notevoli vedremo alcune dimostrazioni di questo genere). Per mettere a frutto il concetto di fgm dobbiamo quindi conoscere i teoremi che la riguardano. Nelle prossime righe vedremo quattro di questi teoremi. Tuttavia, dimostreremo soltanto il secondo e il terzo di essi perché gli altri due richiederebbero dimostrazioni troppo tecniche.

Teorema 7.1 (Teorema sulla relazione tra la fgm e i momenti). Se \tilde{X} è una variabile casuale dotata di fgm e $m(t)$ è la sua fgm, allora ...

- ... esistono i momenti $\mu_k = E(\tilde{X}^k)$ di ogni ordine $k > 0$;
- esiste un numero reale $\epsilon > 0$ tale che

$$m(t) = E\left(e^{t\tilde{X}}\right) = \sum_{k=0}^{\infty} \frac{t^k}{k!} E(\tilde{X}^k) = \sum_{k=0}^{\infty} \frac{t^k}{k!} \mu_k \quad \text{per ogni } t \in (-\epsilon, \epsilon);$$

- i momenti della variabile casuale \tilde{X} possono essere ricavati attraverso la formula

$$\mu_k = E(\tilde{X}^k) = \left. \frac{d^k m(t)}{dt^k} \right|_{t=0} \quad \text{per ogni } k = 1, 2, \dots$$

Si noti che la conclusione al primo punto del precedente teorema ci dice che **l'esistenza della fgm implica l'esistenza di tutti i momenti, ma si ricordi che non è vero il viceversa** (infatti, esistono distribuzioni che non sono dotate di fgm per le quali, tuttavia, l'intera successione dei momenti esiste).

Teorema 7.2 (fgm di una trasformazione affine). Se \tilde{X} è una variabile casuale dotata di fgm e $\tilde{Y} = a + b\tilde{X}$ è una trasformazione affine della variabile casuale \tilde{X} , allora anche \tilde{Y} è dotata di fgm e in questo caso la fgm di \tilde{Y} deve essere data da

$$m_{\tilde{Y}}(t) = E\left(e^{t\tilde{Y}}\right) = e^{at} \times E\left(e^{tb\tilde{X}}\right) = e^{at} \times m_{\tilde{X}}(tb)$$

per ogni t dove il valore atteso $m_{\tilde{X}}(tb) = E\left(e^{tb\tilde{X}}\right)$ esiste ($m_{\tilde{X}}(tb)$ indica la fgm di \tilde{X} nel punto tb).

Dimostrazione. Siccome stiamo ipotizzando che \tilde{X} sia una variabile casuale dotata di fgm e che quindi esista un $\epsilon > 0$ tale che

$$\text{il valore atteso } E\left(e^{t\tilde{X}}\right) \text{ esiste per ogni } t \in (-\epsilon, +\epsilon),$$

possiamo concludere che

$$\begin{aligned} E\left(e^{t\tilde{Y}}\right) &= E\left(e^{t[a+b\tilde{X}]}\right) = E\left(e^{ta} \times e^{tb\tilde{X}}\right) = \\ &= \text{[[proprietà E2]]} = e^{ta} \times E\left(e^{tb\tilde{X}}\right) \end{aligned}$$

per ogni $t \in (-\epsilon/b, \epsilon/b)$. Da questo risultato deduciamo che anche la variabile casuale $\tilde{Y} = a + b\tilde{X}$ è dotata di fgm e che

$$m_{\tilde{Y}}(t) = E\left(e^{t\tilde{Y}}\right) = e^{ta} \times E\left(e^{tb\tilde{X}}\right) = e^{at} \times m_{\tilde{X}}(tb)$$

in ogni punto $t \in \mathbb{R}$ dove il valore atteso $m_{\tilde{X}}(tb) = E\left(e^{tb\tilde{X}}\right)$ esiste. \square

Teorema 7.3 (Fgm di una somma di variabili casuali indipendenti). Se $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_k$ sono k variabili casuali indipendenti e tutte e k le variabili casuali \tilde{X}_i sono dotate di fgm, allora anche la loro somma

$$\tilde{X} = \tilde{X}_1 + \tilde{X}_2 + \dots + \tilde{X}_k$$

è dotata di fgm e quest'ultima è data dal prodotto delle fgm delle k variabili casuali \tilde{X}_i :

$$m_{\tilde{X}}(t) = m_{\tilde{X}_1}(t) \times m_{\tilde{X}_2}(t) \times \dots \times m_{\tilde{X}_k}(t) = \prod_{i=1}^k m_{\tilde{X}_i}(t)$$

per ogni t dove il prodotto delle fgm delle k variabili casuali \tilde{X}_i è definito.

Dimostrazione. Come si deduce dal Teorema 4.10, se $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_k$ sono variabili casuali indipendenti, allora, per qualsiasi $t \in \mathbb{R}$, devono essere indipendenti anche le variabili casuali

$$\tilde{Y}_1 = e^{t\tilde{X}_1}, \quad \tilde{Y}_2 = e^{t\tilde{X}_2}, \quad \dots, \quad \tilde{Y}_k = e^{t\tilde{X}_k}.$$

Se per un dato $t \in \mathbb{R}$ tutti i valori attesi

$$E(e^{t\tilde{X}_1}), \quad E(e^{t\tilde{X}_2}), \quad \dots, \quad E(e^{t\tilde{X}_k})$$

esistono, possiamo dunque scrivere

$$\begin{aligned} E(e^{t\tilde{X}}) &= E(e^{t(\tilde{X}_1 + \tilde{X}_2 + \dots + \tilde{X}_k)}) \\ &= E(e^{t\tilde{X}_1} \times e^{t\tilde{X}_2} \times \dots \times e^{t\tilde{X}_k}) \\ \text{[[proprietà E5]]} &= E(e^{t\tilde{X}_1}) \times E(e^{t\tilde{X}_2}) \times \dots \times E(e^{t\tilde{X}_k}). \end{aligned}$$

Siccome nell'enunciato del teorema viene ipotizzato che tutte le variabili casuali \tilde{X}_i siano dotate di fgm, ovvero che per ciascuna delle variabili casuali \tilde{X}_i esista un intervallo aperto che contiene l'origine tale che per ogni t appartenente a questo intervallo il corrispondente valore atteso $E(e^{t\tilde{X}_i})$ esiste, possiamo concludere che anche il valore atteso $E(e^{t\tilde{X}})$ esista per ogni t in un intervallo aperto che contiene l'origine (si noti infatti che l'intersezione di un numero finito di intervalli aperti che contengono l'origine è a sua volta un intervallo aperto che contiene l'origine). Da questa osservazione deduciamo che anche la variabile casuale $\tilde{X} = \tilde{X}_1 + \tilde{X}_2 + \dots + \tilde{X}_k$ è dotata di fgm e che, in tutti i punti dove tutte le k fgm $m_{\tilde{X}_i}(t) = E(e^{t\tilde{X}_i})$ sono definite, la fgm della variabile casuale \tilde{X} è uguale al prodotto delle k fgm $m_{\tilde{X}_i}(t) = E(e^{t\tilde{X}_i})$. \square

Teorema 7.4 (Teorema di unicità della fgm). Se \tilde{X} e \tilde{Y} sono due variabili casuali che sono entrambe dotate di fgm e se le fgm di \tilde{X} e \tilde{Y} sono uguali, allora \tilde{X} e \tilde{Y} hanno anche la stessa distribuzione.

8 Distribuzioni notevoli

8.1 Il concetto di distribuzione notevole

In questa sezione analizzeremo le caratteristiche di alcune famiglie di distribuzioni notevoli, ovvero di alcune famiglie di distribuzioni che spesso si incontrano nelle applicazioni della teoria della probabilità. Il motivo per cui si incorre così spesso in queste distribuzioni risiede nel fatto che esse sorgono come distribuzioni di variabili casuali che sono legate a **esperimenti casuali "ideali"**. Per esperimento casuale ideale non intendiamo un esperimento casuale che si svolge nel mondo reale, ma un esperimento casuale astratto che è descritto da un modello probabilistico (Ω, \mathcal{A}, P) con delle caratteristiche ben precise. Nell'ambito della teoria della probabilità gli esperimenti casuali ideali possono dunque essere considerati come le controparti delle figure geometriche ideali (quadrati, rettangoli, cerchi, ecc.) nella geometria: gli esperimenti casuali ideali sono delle descrizioni astratte di esperimenti casuali che si svolgono nel mondo reale così come le figure geometriche ideali sono delle descrizioni astratte di figure che incontriamo nel mondo reale.

In questa dispensa abbiamo già incontrato vari esperimenti casuali ideali senza chiamarli con questo nome. Nell'Esercizio 2.2, per esempio, abbiamo considerato estrazioni "casuali" da un'urna (con o senza reimmissione) e abbiamo visto che questo tipo di esperimento casuale può essere descritto attraverso il modello probabilistico (Ω, \mathcal{A}, P) dove lo spazio campionario Ω è l'insieme di tutte le disposizioni (con o senza ripetizione) che si possono formare con le palline dell'urna, e dove la funzione di probabilità $P(\cdot)$ è quella che corrisponde al metodo di assegnazione classico. Chiaramente, quando nel mondo reale estraiamo delle palline da un'urna non possiamo mai essere sicuri che il suddetto modello probabilistico fornisca la migliore descrizione possibile del modo in cui avvengono le estrazioni, ma in molte circostanze fornisce comunque una descrizione che molti giudicherebbero "ragionevole". Per questo motivo, nella teoria della probabilità vengono studiate le proprietà del suddetto modello probabilistico e l'esperimento casuale di riferimento viene dunque considerato come un esperimento casuale "ideale".

Allo stesso modo può essere considerato come "ideale" un esperimento casuale che consiste in una successione infinita di lanci di una moneta, oppure in una successione finita o anche infinita di prove indipendenti che vengono eseguite sotto le medesime condizioni e che possono dar luogo a due esiti diversi: un "successo" o un "insuccesso". Le singole prove di un tale esperimento vengono chiamate **prove binomiali** oppure **prove bernoulliane** e un esperimento casuale ideale che consiste in un numero finito o infinito di prove binomiali (o bernoulliane) indipendenti che vengono eseguite sotto le medesime condizioni (ovvero con la medesima probabilità di successo) viene chiamato **esperimento binomiale** o **processo di Bernoulli** (o anche processo "bernoulliano"). Come si desume dal Teorema 4.13 (vedi anche l'Esercizio 4.24), un processo bernoulliano può sempre essere descritto attraverso un qualunque modello probabilistico (Ω, \mathcal{A}, P) che può ospitare una successione infinita (o a volte anche soltanto finita) di variabili casuali i.i.d. che possono assumere soltanto i valori 0 (in caso di "insuccesso") o 1 (in caso di "successo"). Le variabili casuali in questione vengono chiamate "indicatrici" e, come vedremo tra breve, a partire dalle variabili casuali indicatrici che descrivono l'esito di un processo di Bernoulli si possono definire molte altre variabili casuali come, per esempio, una variabile casuale che conta il numero complessivo di successi in un determinato insieme finito di prove, la variabile casuale che conta il numero di prove necessarie per ottenere il primo successo, ecc.. Nelle prossime sezioni ricaveremo le distribuzioni di queste (e altre) variabili casuali che sono legate ad un processo di Bernoulli.

Oltre a estrazioni da urne e a processi di Bernoulli, considereremo anche un'altra importante classe di esperimenti casuali ideali, i cosiddetti **processi di Poisson**. Questi ultimi sono delle descrizioni astratte di esperimenti casuali reali che vengono chiamati **processi puntuali**. Un processo puntuale è un esperimento casuale che consiste nella rilevazione degli istanti temporali in cui si verificano determinati eventi che, per evitare ambiguità, chiameremo "successi". Spesso questi esperimenti si svolgono sotto condizioni

che li rendono molto simili ad una successione di prove indipendenti che si svolgono a distanza temporale molto ridotta e che hanno probabilità di successo costante e prossima a zero. Come vedremo tra breve, processi puntuali di questo tipo possono essere descritti attraverso modelli probabilistici con delle caratteristiche ben precise e pertanto possono essere considerati come particolari esperimenti casuali ideali che vengono chiamati processi di Poisson.

8.2 La famiglia delle distribuzioni di Bernoulli

Una distribuzione di Bernoulli o "bernoulliana" è una distribuzione discreta con supporto $S = \{0, 1\}$, ovvero una distribuzione discreta che concentra tutta la probabilità sui numeri naturali $x = 0$ e $x = 1$. La funzione di massa di probabilità di una distribuzione bernoulliana è quindi definita come

$$p(x) = \begin{cases} 1 - p & \text{se } x = 0, \\ p & \text{se } x = 1, \\ 0 & \text{altrimenti} \end{cases}$$

dove p è un parametro che deve essere un numero reale strettamente compreso tra 0 e 1 (se $p = 0$ o $p = 1$ si ottengono distribuzioni degeneri che concentrano tutta la probabilità in un solo punto).

Distribuzioni bernoulliane possono sorgere nell'ambito di qualunque esperimento casuale, o meglio, nell'ambito di qualunque modello probabilistico. Infatti, se A è un qualunque evento che appartiene alla σ -algebra \mathcal{A} di un modello probabilistico (Ω, \mathcal{A}, P) , possiamo sempre definire una variabile casuale $\tilde{X} : \Omega \mapsto \mathbb{R}$ ponendo

$$\tilde{X}(\omega) = \begin{cases} 1 & \text{se } \omega \in A; \\ 0 & \text{altrimenti} \end{cases}$$

e la distribuzione di questa variabile casuale sarà ovviamente bernoulliana con parametro $p = P(A)$. Variabili casuali con distribuzione bernoulliana vengono a loro volta chiamate bernoulliane oppure spesso anche "**indicatrici**" o "**indicatore**" perché "indicano" se un determinato evento A si è verificato o meno. Per indicare che una determinata variabile casuale \tilde{X} ha distribuzione bernoulliana di parametro p scriveremo

$$\tilde{X} \sim \text{Bernoulli}(p).$$

Valore atteso, varianza e fgm. Le formule per il valore atteso, la varianza e la fgm di una distribuzione bernoulliana sono molto semplici. Infatti, se $\tilde{X} \sim \text{Bernoulli}(p)$, si hanno i seguenti risultati:

- Per qualsiasi valore di $p \in (0, 1)$ il **valore atteso** $E(\tilde{X})$ esiste ed è dato da

$$\mu = E(\tilde{X}) = 1 \times p + 0 \times (1 - p) = p. \quad (100)$$

- Per qualsiasi valore di $p \in (0, 1)$ la **varianza** $var(\tilde{X})$ esiste ed è data da

$$\sigma^2 = var(\tilde{X}) = (1 - p)^2 \times p + (0 - p)^2 \times (1 - p) = \dots = p(1 - p). \quad (101)$$

- Per qualsiasi valore di $p \in (0, 1)$ la **fgm** di \tilde{X} esiste ed è data da

$$m(t) = E\left(e^{t \times \tilde{X}}\right) = e^{t \times 1} \times p + e^{t \times 0} \times (1 - p) = pe^t + (1 - p), \quad t \in \mathbb{R}. \quad (102)$$

Esercizio 8.1 (Asimmetria e curtosi delle distribuzioni bernoulliane).

- Si ricavino delle espressioni per l'indice di asimmetria $skew(\tilde{X})$ e l'indice di curtosi $kurt(\tilde{X})$ di una distribuzione di Bernoulli con generico parametro p .
- Si traccino dei grafici che mostrano l'andamento di $skew(\tilde{X})$ e di $kurt(\tilde{X})$ al variare del valore di p .

Esercizio 8.2. Nel gioco della roulette francese un giocatore d'azzardo può puntare un importo C su uno o più settori di una ruota che è suddivisa in 37 settori numerati. Al termine del gioco, se la pallina si ferma in uno dei settori scelti dal giocatore, gli verrà corrisposto un importo pari a 36 volte il valore di C diviso per il numero di settori scelti. In caso contrario il giocatore perderà l'importo C che ha puntato.

- Si consideri un giocatore d'azzardo che punta un importo di $C = 1$ euro su un singolo settore. Si stabilisca se si tratta di un gioco "equo".
- Si consideri un giocatore d'azzardo che punta complessivamente un importo di $C = 1$ euro su n settori (dove n è un numero intero positivo non superiore a 37). Si stabilisca se si tratta di un gioco "equo".
- Con riferimento alla situazione descritta al punto precedente, come varia la varianza del guadagno netto del giocatore d'azzardo all'aumentare di n ?
- Un giocatore d'azzardo che possiede un capitale di 2 euro è in dubbio tra due possibili strategie di gioco:
 - fare due giocate puntando entrambe le volte un euro sui numeri dispari e poi smettere di giocare (18 settori su 37 sono contrassegnati con numeri dispari),

- oppure fare un'unica giocata puntando direttamente entrambi gli euro sui numeri dispari e poi smettere di giocare.

Si confrontino, con riferimento a queste due strategie, i valori attesi e le varianze delle variabili casuali che descrivono il guadagno netto del giocatore d'azzardo nel momento in cui egli smette di giocare. Quale di queste due strategie è preferibile secondo il criterio media-varianza di Markowitz?

Risposte:

- a) Si consideri un giocatore d'azzardo che punta un importo di $C = 1$ euro su un singolo settore. Si stabilisca se si tratta di un gioco "equo".

Per rispondere al quesito conviene definire \tilde{X} come la variabile casuale bernoulliana che assume il valore 1 se la pallina si ferma nel settore scelto dal giocatore, e che assume il valore 0 altrimenti.

Dalla descrizione del gioco della roulette deduciamo che il valore del parametro p della distribuzione bernoulliana di \tilde{X} deve essere $p = 1/37$.

Il guadagno lordo del giocatore d'azzardo può quindi essere definito come

$$\tilde{G} = 36\tilde{X} - 1$$

e per verificare se la scommessa del giocatore d'azzardo è equa dobbiamo verificare se il valore atteso di \tilde{G} (che ovviamente esiste) è nullo o meno.

Nel caso in questione otteniamo

$$\begin{aligned} E(\tilde{G}) &= E(36\tilde{X} - 1) \\ [[\text{proprietà E2}]] &= 36 \times E(\tilde{X}) - 1 \\ &[[\text{valore atteso di distribuzioni bernoulliane}]] \\ &= 36 \times \frac{1}{37} - 1 = -\frac{1}{37} \end{aligned}$$

e quindi possiamo concludere che per il giocatore d'azzardo il gioco è svantaggioso.

- b) Si consideri un giocatore d'azzardo che punta complessivamente un importo di $C = 1$ euro su n settori (dove n è un numero intero positivo non superiore a 37). Si stabilisca se si tratta di un gioco "equo".

Anche qui possiamo considerare una variabile casuale bernoulliana, che questa volta indicheremo con \tilde{X}_n , che assume il valore 1 se la pallina si ferma in uno degli n settori scelti dal giocatore, e che assume il valore 0 altrimenti. Questa volta il parametro p della distribuzione bernoulliana è pari a $p = n/37$.

Per verificare se il gioco in questione è equo dobbiamo sempre ancora verificare se il valore atteso del guadagno netto

$$\tilde{G}_n = \frac{36}{n} \times \tilde{X}_n - 1$$

è nullo.

Ragionando come nella risposta del quesito precedente vediamo che

$$\begin{aligned} E(\tilde{G}_n) &= E\left(\frac{36}{n} \times \tilde{X}_n - 1\right) \\ [[\text{proprietà E2}]] &= \frac{36}{n} \times E(\tilde{X}_n) - 1 \\ &[[\text{valore atteso di distribuzioni bernoulliane}]] \\ &= \frac{36}{n} \times \frac{n}{37} - 1 = -\frac{1}{37} \end{aligned}$$

e quindi possiamo concludere che anche in questo caso per il giocatore d'azzardo il gioco è svantaggioso.

- c) Con riferimento alla situazione descritta al punto precedente, come varia la varianza del guadagno netto del giocatore d'azzardo all'aumentare di n ?

La varianza del guadagno netto è data da

$$\begin{aligned} \text{var}(\tilde{G}_n) &= \text{var}\left(\frac{36}{n} \times \tilde{X}_n - 1\right) \\ [[\text{proprietà V8}]] &= \left(\frac{36}{n}\right)^2 \times \text{var}(\tilde{X}_n) \\ &[[\text{varianza di distribuzioni bernoulliane}]] \\ &= \left(\frac{36}{n}\right)^2 \times \frac{n}{37} \times \left(1 - \frac{n}{37}\right) = \frac{36^2}{37} \left(\frac{1}{n} - \frac{1}{37}\right). \end{aligned}$$

Ovviamente, la varianza del guadagno netto \tilde{G}_n è nulla se il giocatore punta su tutti gli $n = 37$ settori.

- d) Un giocatore d'azzardo che possiede un capitale di 2 euro è in dubbio tra due possibili strategie di gioco:
- fare due giocate puntando entrambe le volte un euro sui numeri dispari e poi smettere di giocare (18 settori su 37 sono contrassegnati con numeri dispari),
 - oppure fare un'unica giocata puntando direttamente entrambi gli euro sui numeri dispari e poi smettere di giocare.

Si confrontino, con riferimento a queste due strategie, i valori attesi e le varianze delle variabili casuali che descrivono il guadagno netto del giocatore d'azzardo nel momento in cui egli smette di giocare. Quale di queste due strategie è preferibile secondo il criterio media-varianza di Markowitz?

Consideriamo la prima delle due strategie: con questa strategia il guadagno netto del giocatore d'azzardo può essere definito come

$$\tilde{G}_{Strategia\ 1} = \tilde{G}_{18} + \tilde{G}'_{18}$$

dove \tilde{G}_{18} e \tilde{G}'_{18} sono le due variabili casuali che descrivono i guadagni netti nelle due giocate. Siccome le variabili casuali \tilde{G}_{18} e \tilde{G}'_{18} si riferiscono a due giocate diverse, e siccome non c'è motivo di ritenere che gli esiti delle due giocate si influenzino a vicenda, assumeremo che \tilde{G}_{18} e \tilde{G}'_{18} siano variabili casuali indipendenti.

Inoltre, siccome \tilde{G}_{18} e \tilde{G}'_{18} si riferiscono allo stesso tipo di scommessa, assumeremo che le loro distribuzioni (e quindi i loro valori attesi e le loro varianze) siano identiche (si ricordi che nelle risposte ai due quesiti precedenti abbiamo già ricavato delle espressioni per il valore atteso e la varianza di \tilde{G}_n per un generico valore di n).

Con queste ipotesi possiamo ora ricavare il valore atteso e la varianza del guadagno netto associato alla prima strategia:

$$\begin{aligned} E\left(\tilde{G}_{Strategia\ 1}\right) &= E(\tilde{G}_{18} + \tilde{G}'_{18}) \\ [[\text{proprietà E4}]] &= E(\tilde{G}_{18}) + E(\tilde{G}'_{18}) \\ &[[\text{risposta al quesito b) }]] \\ &= -\frac{1}{37} - \frac{1}{37} = -\frac{2}{37}, \end{aligned}$$

$$\begin{aligned} var\left(\tilde{G}_{Strategia\ 1}\right) &= var(\tilde{G}_{18} + \tilde{G}'_{18}) \\ [[\text{proprietà V8}^*]] &= var(\tilde{G}_{18}) + var(\tilde{G}'_{18}) \\ &[[\text{risposte al quesito c) }]] \\ &= \frac{36^2}{37} \left(\frac{1}{18} - \frac{1}{37}\right) + \frac{36^2}{37} \left(\frac{1}{18} - \frac{1}{37}\right) \\ &= 2 \times \frac{36^2}{37} \left(\frac{1}{18} - \frac{1}{37}\right) \end{aligned}$$

Consideriamo ora invece la seconda strategia. Con questa strategia il guadagno netto è definito come

$$\tilde{G}_{Strategia\ 2} = 2 \times \tilde{G}_{18}.$$

Il valore atteso del guadagno netto associato alla seconda strategia è dunque dato da

$$\begin{aligned} E\left(\tilde{G}_{Strategia\ 2}\right) &= E(2 \times \tilde{G}_{18}) \\ [[\text{proprietà E2}]] &= 2 \times E(\tilde{G}_{18}) \\ &[[\text{risposta al quesito b) }]] \\ &= 2 \times \left(-\frac{1}{37}\right) = -\frac{2}{37}, \end{aligned}$$

e la sua varianza è data da

$$\begin{aligned} var\left(\tilde{G}_{Strategia\ 2}\right) &= var(2 \times \tilde{G}_{18}) \\ [[\text{proprietà V8}]] &= 2^2 \times var(\tilde{G}_{18}) \\ &[[\text{risposte al quesito c) }]] \\ &= 2^2 \times \frac{36^2}{37} \left(\frac{1}{18} - \frac{1}{37}\right) \\ &= 4 \times \frac{36^2}{37} \left(\frac{1}{18} - \frac{1}{37}\right) \end{aligned}$$

Siccome

$$E\left(\tilde{G}_{Strategia\ 1}\right) = E\left(\tilde{G}_{Strategia\ 2}\right) = -\frac{2}{37}$$

ma

$$\begin{aligned} var\left(\tilde{G}_{Strategia\ 1}\right) &= 2 \times \frac{36^2}{37} \left(\frac{1}{18} - \frac{1}{37}\right) \\ &< 4 \times \frac{36^2}{37} \left(\frac{1}{18} - \frac{1}{37}\right) = var\left(\tilde{G}_{Strategia\ 2}\right), \end{aligned}$$

possiamo concludere che secondo il criterio media-varianza di Markowitz la prima strategia (ovvero quella che prevede di fare due giocate) è preferibile.

Tuttavia, vale la pena osservare che

$$\begin{aligned} P(\{\tilde{G}_{Strategia\ 1} > 0\}) &= P(\{\tilde{G}_{18} + \tilde{G}'_{18} > 0\}) \\ &= P(\{\tilde{G}_{18} > 0\} \cap \{\tilde{G}'_{18} > 0\}) \\ &[[\text{indipendenza}]] \\ &= P(\{\tilde{G}_{18} > 0\}) \times P(\{\tilde{G}'_{18} > 0\}) \\ &= \frac{18}{37} \times \frac{18}{37} = \frac{18^2}{37^2} = 0,236 \end{aligned}$$

mentre

$$\begin{aligned} P(\{\tilde{G}_{Strategia\ 2} > 0\}) &= P(\{2\tilde{G}_{18} > 0\}) \\ &= P(\{\tilde{G}_{18} > 0\}) = P(\{\tilde{G}'_{18} > 0\}) \\ &= \frac{18}{37} = 0,486 \end{aligned}$$

Quindi, anche se secondo il criterio media-varianza di Markowitz la prima strategia è preferibile, con la seconda strategia la probabilità di ottenere una vincita netta positiva è molto più elevata.

8.3 La famiglia delle distribuzioni ipergeometriche

Le distribuzioni ipergeometriche possono essere ricavate con riferimento ad un esperimento casuale che consiste in estrazioni con riposizione (oppure in un'unica estrazione in blocco) di n palline da un'urna che contiene N di palline.

Per ricavare le distribuzioni ipergeometriche assumiamo che l'urna contenga r palline di colore rosso e $N - r$ palline di un altro colore, e definiamo \tilde{X} come la variabile casuale che conta il numero di palline rosse tra le n palline estratte dall'urna.

Ovviamente, \tilde{X} può assumere solo valori interi compresi tra 0 e n e, a seconda di quante palline rosse e non rosse sono contenute nell'urna, potrebbe darsi che il supporto della variabile casuale \tilde{X} sia strettamente incluso nell'insieme dei numeri naturali che vanno da 0 a n .

Usando il metodo di assegnazione classico, non è difficile rendersi conto che la funzione di probabilità della variabile casuale \tilde{X} deve essere data da

$$p(x) = \begin{cases} \frac{\binom{r}{x} \binom{N-r}{n-x}}{\binom{N}{n}} & \text{per } x = 0, 1, 2, \dots, n; \\ 0 & \text{altrimenti} \end{cases} \quad (103)$$

dove

- $\binom{r}{x} = 0$ se $x > r$ (se nell'urna ci sono soltanto r palline rosse, si possono estrarre al più r palline rosse)
- e $\binom{N-r}{n-x} = 0$ se $n - x > N - r$, ovvero se $x < n - N + r$ (se nell'urna ci sono soltanto $N - r$ palline di colore diverso dal rosso, allora nelle n estrazioni si devono ottenere almeno $n - (N - r) = n - N + r$ palline rosse).

Per ottenere la funzione di massa di probabilità definita nella formula (103), basta osservare che con le N palline presenti nell'urna si possono formare $\binom{N}{n}$ combinazioni di n palline (casi possibili) e per contare quante di queste combinazioni contengono x palline rosse basta osservare che

- con le r palline rosse presenti nell'urna si possono formare $\binom{r}{x}$ combinazioni di x palline ...

- ... e che ciascuna di queste combinazioni può essere combinata con ciascuna delle $\binom{N-r}{n-x}$ combinazioni di $n-x$ palline che si possono formare con le $N-r$ palline che sono di colore diverso dal rosso.

Il numero di combinazioni di n palline che contengono esattamente x palline rosse (casi favorevoli) è quindi dato dal prodotto

$$\binom{r}{x} \times \binom{N-r}{n-x}.$$

Questo risultato dimostra che la funzione (103) è effettivamente la funzione di massa di probabilità di una variabile casuale che conta il numero di palline rosse che si ottengono in n estrazioni senza reimmissione (oppure in un'unica estrazione in blocco di n palline) da un'urna che contiene r palline rosse e $N-r$ palline di un altro colore. Siccome le probabilità nella (103) devono sommare a 1, possiamo concludere che

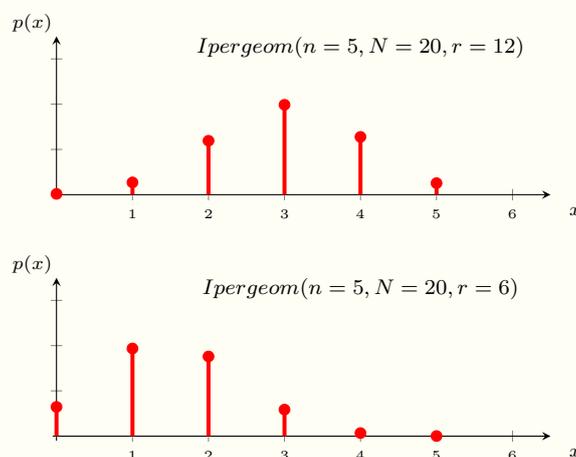
$$\sum_{x=0}^n \binom{r}{x} \binom{N-r}{n-x} = \binom{N}{n}.$$

Questa uguaglianza algebrica è nota come **identità di Vandermonde**.

Le distribuzioni discrete con una funzione di massa di probabilità $p(x)$ come quella definita nella formula (103) sono note come **distribuzioni ipergeometriche** perché le probabilità $p(x)$ possono essere interpretate come termini di una **serie ipergeometrica** (in questa dispensa non forniremo la definizione generale di "serie ipergeometrica"). D'ora in poi, per comodità di notazione, indicheremo che la distribuzione di una variabile casuale \tilde{X} è ipergeometrica con parametri N , r e n scrivendo semplicemente

$$\tilde{X} \sim \text{Ipergeom}(N; r; n).$$

Figura 8.1. I grafici mostrano due funzioni di massa di probabilità ipergeometriche.



Rappresentazione come somma di variabili casuali indicatrici. Variabili casuali che contano il numero di palline rosse che si ottengono in n estrazioni senza riposizione da un'urna, possono essere rappresentate come somme di variabili casuali indicatrici. Infatti, se indichiamo con \tilde{I}_1 la variabile casuale indicatrice che assume il valore 1 se la prima pallina estratta è rossa e che assume il valore 0 altrimenti; con \tilde{I}_2 la variabile casuale indicatrice che assume il valore 1 se la seconda pallina estratta è rossa e che assume il valore 0 altrimenti; ecc., allora non è difficile rendersi conto che la variabile casuale

$$\tilde{X} = \tilde{I}_1 + \tilde{I}_2 + \cdots + \tilde{I}_n \quad (104)$$

rappresenta il numero complessivo di palline rosse che si ottengono nelle n estrazioni. Si noti che questa rappresentazione ha senso anche nel caso di un'unica estrazione in blocco. Nel caso di un'unica estrazione in blocco si può infatti immaginare che dopo aver estratto le n palline, le palline estratte vengano inserite in una nuova urna dalla quale vengono a loro volta estratte una dopo l'altra e senza reimmissione. Questo esperimento aggiuntivo non modifica il numero di palline rosse tra le n palline estratte (e quindi non incide sulla distribuzione della variabile casuale \tilde{X}), ma permette di stabilire un ordine di estrazione fittizio. Nei prossimi paragrafi faremo riferimento alla rappresentazione (104) per dimostrare alcune proprietà delle distribuzioni ipergeometriche.

Valore atteso e varianza. Il valore atteso e la varianza di una distribuzione ipergeometrica sono rispettivamente dati da

$$E(\tilde{X}) = n \times \frac{r}{N} \quad \text{e} \quad \text{var}(\tilde{X}) = n \times \frac{r}{N} \left(1 - \frac{r}{N}\right) \times \frac{N-n}{N-1}. \quad (105)$$

Queste formule possono essere ottenute semplificando direttamente le formule definitorie del valore atteso e della varianza (usando l'identità di Vandermonde), oppure facendo riferimento alla rappresentazione di \tilde{X} come somma di variabili casuali indicatrici. Nella seguente dimostrazione seguiremo la seconda strada.

Dimostrazione (delle formule per il valore atteso e la varianza).

$$\begin{aligned} E(\tilde{X}) &= E(\tilde{I}_1 + \tilde{I}_2 + \cdots + \tilde{I}_n) \\ [[\text{proprietà E4}]] &= E(\tilde{I}_1) + E(\tilde{I}_2) + \cdots + E(\tilde{I}_n) \end{aligned} \quad (106)$$

e siccome le variabili casuali \tilde{I}_i hanno tutte distribuzione bernoulliana con parametro

$$p = \frac{r \times (N-1)(N-2) \cdots (N-n+1)}{N(N-1) \cdots (N-n+1)} = \frac{r}{N}$$

(infatti, tra tutte le $N(N-1) \cdots (N-n+1)$ disposizioni di palline che si possono ottenere attraverso n estrazioni senza reimmissione ce ne sono $r \times (N-1)(N-2) \cdots (N-n+1)$ che contengono una pallina rossa nella i -esima posizione perché ci sono r modi per scegliere

una pallina rossa da mettere nella i -esima posizione e $(N-1)(N-2)\cdots(N-n+1)$ modi per occupare le posizioni rimanenti con le palline rimanenti), possiamo concludere che

$$E(\tilde{I}_i) = \text{[[formula (100)]]} = r/N \quad \text{per ogni } i = 1, 2, \dots, n$$

e che il valore atteso di una distribuzione ipergeometrica è dunque dato da

$$E(\tilde{X}) = \text{[[formula (106)]]} = \frac{r}{N} + \frac{r}{N} + \cdots \text{[[} n \text{ volte]]} \cdots + \frac{r}{N} = n \times \frac{r}{N}.$$

Consideriamo ora invece la varianza. Usando la proprietà V8* della varianza, otteniamo

$$\begin{aligned} \text{var}(\tilde{X}) &= \text{var}(\tilde{I}_1 + \tilde{I}_2 + \cdots + \tilde{I}_n) = \text{[[proprietà V8*]]} = \\ &= \sum_{i=1}^n \sum_{j=1}^n \text{cov}(\tilde{I}_i, \tilde{I}_j) = \text{[[proprietà V2]]} = \sum_{i=1}^n \text{var}(\tilde{I}_i) + 2 \times \sum_{i=1}^n \sum_{i < j} \text{cov}(\tilde{I}_i, \tilde{I}_j). \end{aligned} \quad (107)$$

Per esprimere $\text{var}(\tilde{X})$ come una funzione dei parametri n , N e r , dobbiamo ora trovare delle espressioni per le varianze $\text{var}(\tilde{I}_i)$ e le covarianze $\text{cov}(\tilde{I}_i, \tilde{I}_j)$. Cominciamo dalle varianze $\text{var}(\tilde{I}_i)$. Per ottenere il valore atteso abbiamo già dimostrato che

$$\tilde{I}_i \sim \text{Bernoulli}(p = r/N) \quad \text{per ogni } i = 1, 2, \dots, n,$$

e usando la formula della varianza di distribuzioni bernoulliane vediamo che

$$\text{var}(\tilde{I}_i) = \text{[[formula (101)]]} = p(1-p) = \frac{r}{N} \left(1 - \frac{r}{N}\right). \quad (108)$$

Consideriamo ora invece le covarianze $\text{cov}(\tilde{I}_i, \tilde{I}_j)$. Usando la formula indiretta per la covarianza otteniamo

$$\text{cov}(\tilde{I}_i, \tilde{I}_j) = \text{[[proprietà V5]]} = E(\tilde{I}_i \times \tilde{I}_j) - E(\tilde{I}_i) \times E(\tilde{I}_j).$$

A questo punto osserviamo che anche la variabile casuale $\tilde{I}_{ij} = \tilde{I}_i \times \tilde{I}_j$ è una variabile casuale indicatrice in quanto può assumere soltanto i valori 0 e 1. In particolare, \tilde{I}_{ij} assume il valore 1 se e solo se nell' i -esima e anche nella j -esima estrazione si ottiene una pallina rossa. Per ogni $i \neq j$ la probabilità di questo evento è data da

$$p = \frac{r(r-1) \times (N-2)(N-3)\cdots(N-n+1)}{N(N-1)\cdots(N-n+1)} = \frac{r(r-1)}{N(N-1)}$$

(per ottenere il numero di disposizioni che contengono una pallina rossa nelle posizioni i e j basta osservare che ci sono $r \times (r-1)$ modi per scegliere due palline rosse da mettere nelle posizioni i e j e che le rimanenti $n-2$ posizioni possono poi essere occupate con le $N-2$ palline rimanenti in $(N-2)(N-3)\cdots(N-n+1)$ modi diversi) e quindi possiamo concludere che

$$\tilde{I}_{ij} = \tilde{I}_i \times \tilde{I}_j \sim \text{Bernoulli}\left(p = \frac{r(r-1)}{N(N-1)}\right) \Rightarrow E(\tilde{I}_{ij}) = E(\tilde{I}_i \times \tilde{I}_j) = \frac{r(r-1)}{N(N-1)}.$$

D'altra parte, siccome $E(\tilde{I}_i) = r/N$ per ogni $i = 1, 2, \dots, n$ (vedi sopra dove abbiamo ricavato l'espressione per il valore atteso di \tilde{X}), possiamo concludere che

$$\begin{aligned} \text{cov}(\tilde{I}_i, \tilde{I}_j) &= E(\tilde{I}_i \times \tilde{I}_j) - E(\tilde{I}_i) \times E(\tilde{I}_j) \\ &= \frac{r(r-1)}{N(N-1)} - \frac{r}{N} \times \frac{r}{N} \\ &= -\frac{r}{N} \left(1 - \frac{r}{N}\right) \frac{1}{N-1} \quad \text{per ogni } i \neq j. \end{aligned} \quad (109)$$

Sostituendo la formula (108) delle varianze e la formula (109) delle covarianze nella formula (107), otteniamo la formula di $var(\tilde{X})$ che volevamo dimostrare:

$$\begin{aligned} var(\tilde{X}) &= \sum_{i=1}^n var(\tilde{I}_i) + \sum_{i=1}^n \sum_{j \neq i}^n cov(\tilde{I}_i, \tilde{I}_j) \\ &= \sum_{i=1}^n \left[\frac{r}{N} \left(1 - \frac{r}{N} \right) \right] + \sum_{i=1}^n \sum_{j \neq i}^n \left[-\frac{r}{N} \left(1 - \frac{r}{N} \right) \frac{1}{N-1} \right] \\ &\quad \left[\text{la doppia sommatoria } \sum_{i=1}^n \sum_{j \neq i}^n \text{ è composta da } \dots \right. \\ &\quad \left. \dots n(n-1) \text{ termini che sono tutti uguali tra di loro} \right] \\ &= n \times \left[\frac{r}{N} \left(1 - \frac{r}{N} \right) \right] + n(n-1) \times \left[-\frac{r}{N} \left(1 - \frac{r}{N} \right) \frac{1}{N-1} \right] \\ &= n \times \frac{r}{N} \left(1 - \frac{r}{N} \right) \times \left[1 - \frac{n-1}{N-1} \right] = n \times \frac{r}{N} \left(1 - \frac{r}{N} \right) \times \frac{N-n}{N-1}. \end{aligned}$$

□

Funzione generatrice dei momenti. Siccome il supporto di una distribuzione ipergeometrica è sempre un insieme finito, tutte le distribuzioni ipergeometriche sono dotate di fgm. Tuttavia, sommatoria

$$E(e^{t\tilde{X}}) = \sum_{x=0}^n e^{tx} \frac{\binom{r}{x} \binom{N-r}{n-x}}{\binom{N}{n}}$$

che esprime la fgm ipergeometrica non si lascia semplificare in modo utile per agevolare sviluppi teorici che presenteremo in questa dispensa.

Esercizio 8.3. In un'urna ci sono 20 palline di cui 5 sono di colore rosso. Dall'urna vengono estratte una dopo l'altra e senza reimmissione 6 palline.

- Qual è la probabilità che nessuna delle palline estratte sia rossa?
- Qual è la probabilità che al più due delle palline estratte siano rosse?
- Si calcolino il valore atteso e la varianza del numero di palline rosse estratte.
- Sapendo che nelle prime due estrazioni è stata estratta una pallina rossa, qual è la probabilità che nelle successive quattro estrazioni vengano estratte altre due palline rosse?

Risposte:

- Qual è la probabilità che nessuna delle palline estratte sia rossa?

Sia \tilde{X} la variabile casuale che conta il numero di palline rosse che si ottengono nelle 6 estrazioni. Da quanto abbiamo detto sulle distribuzioni ipergeometriche deduciamo che

$$\tilde{X} \sim \text{Ipergeom.}(N = 20; r = 5; n = 6).$$

La probabilità che nessuna delle 6 palline estratte sia rossa è quindi data da

$$P(\{\tilde{X} = 0\}) = \frac{\binom{5}{0} \binom{15}{6-0}}{\binom{20}{6}} = \frac{1 \times 5005}{38760} = 0,129.$$

b) Qual è la probabilità che al più due delle palline estratte siano rosse?

La probabilità che al più due delle palline estratte siano rosse è data da

$$\begin{aligned} P(\{\tilde{X} \leq 2\}) &= \frac{\binom{5}{0} \binom{15}{6-0}}{\binom{20}{6}} + \frac{\binom{5}{1} \binom{15}{6-1}}{\binom{20}{6}} + \frac{\binom{5}{2} \binom{15}{6-2}}{\binom{20}{6}} \\ &= \frac{1 \times 5005}{38760} + \frac{5 \times 3003}{38760} + \frac{10 \times 1365}{38760} \\ &= 0,129 + 0,387 + 0,352 = 0,868 \end{aligned}$$

c) Si calcolino il valore atteso e la varianza del numero di palline rosse estratte.

Il valore atteso e la varianza di \tilde{X} sono rispettivamente dati da

$$E(\tilde{X}) = 6 \times \frac{5}{20} = 1,5$$

e

$$\text{Var}(\tilde{X}) = 6 \times \frac{5}{20} \times \left(1 - \frac{5}{20}\right) \times \frac{20-6}{20-1} = 0,829$$

d) Sapendo che nelle prime due estrazioni è stata estratta una pallina rossa, qual è la probabilità condizionata che nelle successive quattro estrazioni vengano estratte altre due palline rosse?

Sia \tilde{X}_1 la variabile casuale che conta il numero di palline rosse che si ottengono nelle prime due estrazioni, e \tilde{X}_2 la variabile casuale che conta il numero di palline rosse nelle successive quattro estrazioni.

Per rispondere al quesito dobbiamo calcolare la probabilità condizionata

$$P(\{\tilde{X}_2 = 2\} | \{\tilde{X}_1 = 1\}).$$

Per determinare il valore di questa probabilità condizionata, conviene considerare l'esperimento casuale che consiste nelle sei estrazioni come un esperimento che si svolge in due passi, dove al primo passo vengono estratte (senza

reimmissione) le prime due palline, e dove al secondo passo vengono estratte (senza reimmissione) le rimanenti 4 palline. Ovviamente, con questo tipo di interpretazione, la distribuzione condizionata della variabile \tilde{X}_2 sotto l'ipotesi che si verifichi l'evento $\{\tilde{X}_1 = 1\}$ deve essere

$$Ipergeom.(N = 20 - 2 = 18; r = 5 - 1 = 4; n = 4).$$

Infatti, se nelle prime due estrazioni otteniamo una pallina rossa, allora prima di procedere alle successive $n_2 = 4$ estrazioni nell'urna saranno rimaste $N_2 = 20 - 2 = 18$ palline delle quali $r_2 = 5 - 1 = 4$ di colore rosso. Facendo riferimento alla suddetta distribuzione condizionata, possiamo calcolare direttamente la probabilità condizionata richiesta come

$$P(\{\tilde{X}_2 = 2\}|\{\tilde{X}_1 = 1\}) = \frac{\binom{4}{2} \times \binom{14}{4-2}}{\binom{18}{4}} = \frac{6 \times 91}{3060} = 0,178.$$

Lasciamo come esercizio supplementare verificare che il risultato

$$P(\{\tilde{X}_2 = 2\}|\{\tilde{X}_1 = 1\}) = \frac{\binom{4}{2} \times \binom{14}{4-2}}{\binom{18}{4}} = \frac{6 \times 91}{3060} = 0,178$$

coincide con quello che si ottiene cercando di calcolare la probabilità condizionata richiesta come

$$P(\{\tilde{X}_2 = 2\}|\{\tilde{X}_1 = 1\}) = \frac{P(\{\tilde{X}_2 = 2\} \cap \{\tilde{X}_1 = 1\})}{P(\{\tilde{X}_1 = 1\})}$$

e ragionando direttamente sui casi possibili e sui casi favorevoli.

8.4 La famiglia delle distribuzioni binomiali

Le distribuzioni binomiali sono le distribuzioni delle variabili casuali che nell'ambito di un numero finito n di prove indipendenti eseguite sotto le medesime condizioni (ovvero nell'ambito di un [esperimento binomiale](#)) contano il numero di "successi" che si ottengono.

Ovviamente si tratta di distribuzioni discrete con supporto dato dall'insieme dei numeri naturali che vanno da 0 a n . Come vedremo nella dimostrazione che segue, la funzione di massa di probabilità di una distribuzione binomiale deve essere definita come

$$p(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{se } x = 0, 1, 2, \dots, n; \\ 0 & \text{altrimenti,} \end{cases} \quad (110)$$

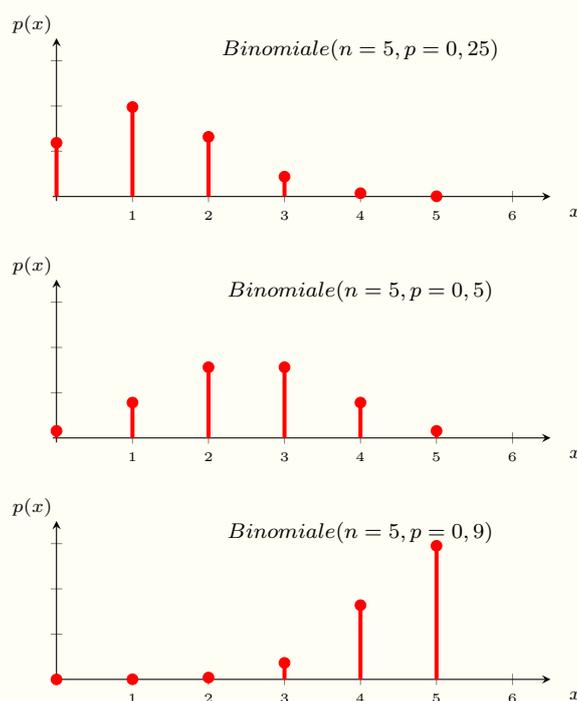
dove $p \in (0, 1)$ è la probabilità di successo comune a tutte le n prove. Il grafico in Figura 8.2 mostra l'andamento di alcune funzioni di massa di probabilità binomiali.

Dimostrazione. Per dimostrare che la funzione (110) sia la funzione di massa di probabilità di una variabile casuale \tilde{X} che conta il numero di successi in n prove indipendenti con probabilità di successo $p \in (0, 1)$, conviene partire dall'osservazione che l'esito finale di una successione di n prove dà luogo ad un evento del tipo $\bigcap_{i=1}^n E_i$ dove $E_i = S_i$ oppure $E_i = \bar{S}_i$ a seconda se l'esito dell' i -esima prova è un successo (S_i) oppure un insuccesso (\bar{S}_i). Siccome stiamo ipotizzando che le n prove siano indipendenti, possiamo concludere che

$$P\left(\bigcap_{i=1}^n E_i\right) = P(E_1) \times P(E_2) \times \cdots \times P(E_n) = p^x \times (1-p)^{n-x}$$

dove x è il numero di successi e $n-x$ è il numero di insuccessi tra gli n eventi E_i coinvolti nell'intersezione. Questo dimostra che la probabilità associata ad una qualunque sequenza di esiti delle n prove dipende solo dal **numero complessivo di successi** che si verificano nelle n prove (ovvero da x), ma non dall'**ordine** in cui questi ultimi si verificano. Siccome ogni sequenza di esiti è incompatibile con tutte quelle diverse da essa, e siccome le posizioni di x successi in una sequenza di n esiti possono essere scelte in $\binom{n}{x}$ modi diversi, la probabilità di osservare x successi in una sequenza di n prove deve essere data da $\binom{n}{x} p^x (1-p)^{n-x}$. \square

Figura 8.2. I grafici mostrano tre funzioni di massa di probabilità binomiali.



Dalla precedente dimostrazione si desume che la probabilità di osservare x successi in una successione di n prove bernoulliane con probabilità di successo p è data dal prodotto tra

- la probabilità $p^x \times (1-p)^{n-x}$ di ottenere x successi e $n-x$ insuccessi in un preciso ordine ...

- e il numero di modi $\binom{n}{x}$ in cui possono verificarsi x successi e $n - x$ insuccessi in una sequenza di n esiti.

Il nome **”binomiale”** delle distribuzioni discrete con funzione di probabilità

$$p(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{se } x = 0, 1, 2, \dots, n; \\ 0 & \text{altrimenti,} \end{cases}, \quad p \in (0, 1),$$

deriva dal fatto che **le probabilità** $p(0), p(1), \dots, p(n)$ **sono i termini dello sviluppo in serie di un binomio**. Com'è noto infatti, un binomio $(a + b)^n$ può sempre essere sviluppato come

$$(a + b)^n = \sum_{x=0}^n \binom{n}{x} a^x b^{n-x}, \quad (111)$$

e ponendo $a = p$ e $b = 1 - p$ si ottiene

$$[p + (1 - p)]^n = \sum_{x=0}^n \binom{n}{x} p^x (1 - p)^{n-x},$$

dove i termini della sommatoria sono le probabilità di una distribuzione binomiale.

D'ora in poi, per comodità di notazione, per indicare che \tilde{X} è una variabile casuale con distribuzione binomiale di parametri n e p , scriveremo semplicemente

$$\tilde{X} \sim \text{Binomiale}(n; p).$$

Tenendo presente il modo in cui abbiamo ottenuto le distribuzioni binomiali non è difficile rendersi conto che nel caso particolare dove $n = 1$ una distribuzione binomiale si riduce ad una distribuzione bernoulliana.

Rappresentazione come somma di variabili casuali indicatrici. Ovviamente, con riferimento ad un [esperimento binomiale](#) che consiste in n prove indipendenti con probabilità di successo p , possiamo sempre definire n variabili casuali indicatrici $\tilde{I}_1, \tilde{I}_2, \dots, \tilde{I}_n$ che descrivono gli esiti delle singole prove: \tilde{I}_i assume il valore 1 se l'esito dell' i -esima prova è un successo, e assume il valore 0 altrimenti. Chiaramente, la somma

$$\tilde{X} = \tilde{I}_1 + \tilde{I}_2 + \dots + \tilde{I}_n \quad (112)$$

restituisce il numero complessivo di successi che si ottengono nelle n prove e, per quanto abbiamo visto in questa sezione,

$$\tilde{X} \sim \text{Binomiale}(n, p).$$

In quanto segue faremo riferimento alla rappresentazione (112) per dimostrare alcune proprietà delle distribuzioni binomiali.

Valore atteso, varianza e fgm. Sfruttando la rappresentazione (112), è facile vedere che ...

- il **valore atteso** di una distribuzione binomiale è dato da

$$\begin{aligned}\mu = E(\tilde{X}) &= [[\text{rappresentazione (112)}]] = E(\tilde{I}_1 + \tilde{I}_2 + \dots + \tilde{I}_n) = \\ &= [[\text{proprietà E4}]] = E(\tilde{I}_1) + E(\tilde{I}_2) + \dots + E(\tilde{I}_n) = p + p + \dots + p = np;\end{aligned}$$

- la **varianza** di una distribuzione binomiale è data da

$$\begin{aligned}\sigma^2 = \text{var}(\tilde{X}) &= [[\text{rappresentazione (112)}]] = \text{var}(\tilde{I}_1 + \tilde{I}_2 + \dots + \tilde{I}_n) = \\ &= [[\text{indipendenza e proprietà V8*}]] = \text{var}(\tilde{I}_1) + \text{var}(\tilde{I}_2) + \dots + \text{var}(\tilde{I}_n) = \\ &= p(1-p) + p(1-p) + \dots + p(1-p) = np(1-p)\end{aligned}$$

- e che la **funzione generatrice dei momenti** di una distribuzione binomiale (che deve esistere perché il supporto di una distribuzione binomiale è sempre finito) è data da

$$\begin{aligned}m_{\tilde{X}}(t) = E\left(e^{t \times \tilde{X}}\right) &= [[\text{rappresentazione (112)}]] = \\ &= E\left(e^{t(\tilde{I}_1 + \tilde{I}_2 + \dots + \tilde{I}_n)}\right) = E\left(e^{t\tilde{I}_1} \times e^{t\tilde{I}_2} \times \dots \times e^{t\tilde{I}_n}\right) = \\ &= [[\text{Teorema 4.10 e proprietà E5}]] = E\left(e^{t\tilde{I}_1}\right) \times E\left(e^{t\tilde{I}_2}\right) \times \dots \times E\left(e^{t\tilde{I}_n}\right) = \\ &= [[\text{formula (102) fgm indicatori}]] = [(1-p) + pe^t] \times \dots \times [(1-p) + pe^t] = \\ &= [(1-p) + pe^t]^n\end{aligned}\tag{113}$$

(questi passaggi logici per ricavare l'espressione della fgm binomiale replicano i ragionamenti della dimostrazione del Teorema 7.3 sulla fgm di somme di variabili casuali indipendenti).

Ovviamente, le suddette formule per il valore atteso, la varianza e la fgm di distribuzioni binomiali possono essere ottenute anche a partire dalle formule definitorie, ovvero semplificando le sommatorie

$$\mu = E(\tilde{X}) = \sum_{x=0}^n x \binom{n}{x} p^x (1-p)^{n-x},$$

$$\sigma^2 = \text{var}(\tilde{X}) = \sum_{x=0}^n (x - \mu)^2 \binom{n}{x} p^x (1-p)^{n-x},$$

e

$$m(t) = E\left(e^{t\tilde{X}}\right) = \sum_{x=0}^n e^{tx} \binom{n}{x} p^x (1-p)^{n-x}.$$

Come esercizio invitiamo il lettore ad effettuare i passaggi algebrici che conducono alle formule che poco sopra abbiamo ricavato sfruttando la rappresentazione (112).

Esercizio 8.4. Si ricavino il momento terzo $\mu_3 = E(\tilde{X}^3)$ e il momento quarto $\mu_4 = E(\tilde{X}^4)$ di una distribuzione binomiale.
 [[Aiuto: si usi la fgm; vedi il Teorema 7.1.]]

Teorema 8.1 (Proprietà riproduttiva delle distribuzioni binomiali). Se

$$\tilde{X}_1 \sim \text{Binomiale}(n_1; p), \quad \tilde{X}_2 \sim \text{Binomiale}(n_2; p), \quad \dots,$$

$$\dots, \quad \tilde{X}_k \sim \text{Binomiale}(n_k; p)$$

(si noti che tutte le distribuzioni binomiali hanno lo stesso parametro p), e se le variabili casuali $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_k$ sono indipendenti, allora

$$\tilde{X} = \tilde{X}_1 + \tilde{X}_2 + \dots + \tilde{X}_k \sim \text{Binomiale}(n = n_1 + n_2 + \dots + n_k; p).$$

Dimostrazione. Le fgm delle k variabili casuali \tilde{X}_i sono date da (vedi la formula (113))

$$m_{\tilde{X}_i}(t) = [pe^t + (1-p)]^{n_i}, \quad t \in \mathbb{R}.$$

Usando il Teorema 7.3 possiamo dunque concludere che la loro somma \tilde{X} sia dotata di fgm e che quest'ultima sia data da

$$\begin{aligned} m_{\tilde{X}}(t) &= m_{\tilde{X}_1}(t) \times m_{\tilde{X}_2}(t) \times \dots \times m_{\tilde{X}_k}(t) \\ &= [pe^t + (1-p)]^{n_1} \times [pe^t + (1-p)]^{n_2} \times \dots \times [pe^t + (1-p)]^{n_k} \\ &= [pe^t + (1-p)]^{n_1 + n_2 + \dots + n_k}. \end{aligned}$$

Siccome l'espressione nell'ultima riga è l'espressione della fgm della distribuzione binomiale con parametri $n = n_1 + n_2 + \dots + n_k$ e p , possiamo invocare il Teorema di unicità della fgm (Teorema 7.4) onde concludere che la distribuzione della variabile casuale \tilde{X} deve essere binomiale con suddetto parametro n e con parametro p uguale al valore del parametro p che è comune alle distribuzioni binomiali delle k variabili casuali \tilde{X}_i che vengono sommate. \square

Relazione tra distribuzioni ipergeometriche e distribuzioni binomiali. Consideriamo un'urna che contiene N palline di cui r sono di colore rosse. Come abbiamo già visto, nel caso di n estrazioni **senza reimmissione** la distribuzione della variabile casuale \tilde{X} che conta il numero di palline rosse tra le n palline estratte è $Ipergeom(n; N; r)$ e il valore atteso e la varianza di \tilde{X} sono dati da (vedi le formule (105))

$$E(\tilde{X}) = n \frac{r}{N} \quad \text{e} \quad \text{var}(\tilde{X}) = n \frac{r}{N} \left(1 - \frac{r}{N}\right) \frac{N-n}{N-1}.$$

Vediamo ora invece che cosa accade se le estrazioni avvengono **con reimmissione**. In questo caso la distribuzione di \tilde{X} è $\text{Binomiale}(n; p = r/N)$ e il valore atteso e la varianza di \tilde{X} sono dunque dati da

$$E(\tilde{X}) = np = n \frac{r}{N} \quad \text{e} \quad \text{var}(\tilde{X}) = np(1-p) = n \frac{r}{N} \left(1 - \frac{r}{N}\right).$$

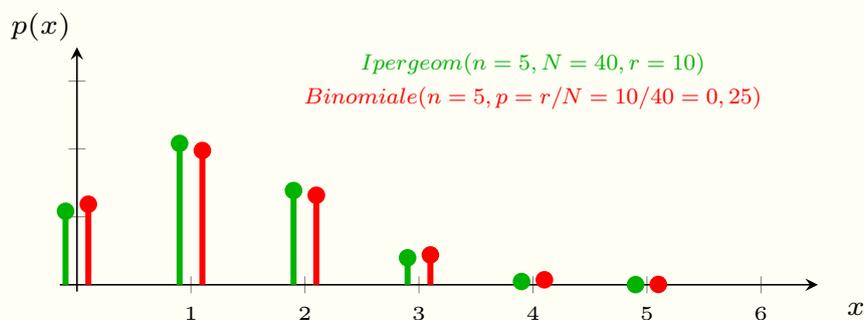
Si noti che il valore atteso $E(\tilde{X})$ è lo stesso in entrambi i casi, ma che nel caso di estrazioni senza reimmissione il valore di $var(\tilde{X})$ è più piccolo a causa della presenza del **fattore di correzione** $(N - n)/(N - 1)$ (ovviamente, nel caso in cui $n = 1$ le due varianze sono uguali perché in tal caso non vi è nessuna differenza tra estrazioni con o senza reimmissione).

Con riferimento ai due schemi di estrazione conviene aggiungere un'altra osservazione. Supponiamo che le estrazioni avvengano con riposizione. Se il numero numero di palline rosse r e il numero $N - r$ di palline di colore diverso sono entrambi molto più grandi del numero di palline n che vengono estratte dall'urna, allora è altamente improbabile che una o più palline vengano estratte più di una volta e pertanto ci aspettiamo che la distribuzione $Binomiale(n; p = r/N)$ e la distribuzione $Ipergeom(n; N; r)$ siano molto simili. Questa congettura è corretta. Infatti, si può dimostrare che

$$\max_{x=1,2,\dots,n} \left| \frac{\binom{n}{x} p^x (1-p)^{n-x}}{\frac{\binom{r}{x} \binom{N-r}{n-x}}{\binom{N}{n}}} - 1 \right| \simeq 0$$

se r e $N - r$ sono entrambi molto più grandi di n (omettiamo la dimostrazione). Il grafico in Figura 8.3 illustra questa approssimazione.

Figura 8.3. Il grafico mostra l'approssimazione binomiale alla distribuzione $Ipergeom(n = 5; N = 40; r = 10)$.



Esercizio 8.5. Sia \tilde{X} una variabile casuale che conta il numero di volte che si ottiene il punteggio sei in $n = 4$ lanci indipendenti di un dado regolare.

- Si calcolino $E(\tilde{X})$ e $var(\tilde{X})$.
- Qual è la probabilità di ottenere 3 volte il punteggio sei?
- Qual è la probabilità di ottenere almeno 3 volte il punteggio sei?
- Qual è la probabilità di ottenere al più 2 volte il punteggio sei?

Risposte:

Siccome l'esperimento casuale in questione consiste in $n = 4$ prove bernoulliane con probabilità di successo $p = 1/6$, assumeremo che $\tilde{X} \sim \text{Binomiale}(n = 4; p = 1/6)$.

a) Si calcolino $E(\tilde{X})$ e $\text{var}(\tilde{X})$.

Usando le formule del valore atteso e della varianza di una distribuzione binomiale otteniamo

$$E(\tilde{X}) = 4 \times \frac{1}{6} = \frac{2}{3}$$

e

$$\text{Var}(\tilde{X}) = 4 \times \frac{1}{6} \times \left(1 - \frac{1}{6}\right) = \frac{5}{9}.$$

b) Qual è la probabilità di ottenere 3 volte il punteggio sei?

$$P(\{\tilde{X} = 3\}) = p(3) = \binom{4}{3} \times \left(\frac{1}{6}\right)^3 \times \left(1 - \frac{1}{6}\right)^{4-3} = 0,0154.$$

c) Qual è la probabilità di ottenere almeno 3 volte il punteggio sei?

$$\begin{aligned} P(\{\tilde{X} \geq 3\}) &= p(3) + p(4) \\ &= \binom{4}{3} \times \left(\frac{1}{6}\right)^3 \times \left(1 - \frac{1}{6}\right)^{4-3} + \binom{4}{4} \times \left(\frac{1}{6}\right)^4 \times \left(1 - \frac{1}{6}\right)^{4-4} \\ &= 0,0154 + 0,0008 = 0,0159. \end{aligned}$$

d) Qual è la probabilità di ottenere al più 2 volte il punteggio sei?

$$\begin{aligned} P(\{\tilde{X} \leq 2\}) &= 1 - P(\{\tilde{X} > 2\}) \\ &= 1 - P(\{\tilde{X} \geq 3\}) \\ & \quad \text{[[risposta del quesito precedente]]} \\ &= 1 - 0,0159 = 0,9841 \end{aligned}$$

Esercizio 8.6. Si risponda ai quesiti dell'Esercizio 8.3 considerando l'ipotesi che le estrazioni siano indipendenti e che avvengano con reimmissione.

Esercizio 8.7. In un famoso gioco bisogna compilare una schedina pronosticando l'esito finale di $n = 13$ partite di calcio scegliendo tra

- vince la squadra che gioca in casa,
- la partita termina in pareggio,

- vince la squadra che gioca in trasferta.

Si consideri uno scommettitore che compila in modo casuale una schedina puntando su un unico esito finale per ciascuna delle 13 partite e sia \tilde{X} la variabile casuale che restituisce il numero di esiti finali indovinati.

- Si calcolino il valore atteso e lo scarto quadratico medio di \tilde{X} .
- Qual è la probabilità di indovinare tutti e 13 gli esiti finali?
- Qual è la probabilità di indovinare almeno 12 esiti finali?
- Qual è la probabilità di indovinare al più 3 esiti finali?
- e*) Qual è la probabilità di indovinare tutti e 13 gli esiti finali con dodici puntate singole e un'unica puntata doppia? Si assuma che tutte le puntate siano casuali.
- f*) Qual è la probabilità di indovinare 12 esiti finali con dodici puntate singole e un'unica puntata doppia? Si assuma che tutte le puntate siano casuali.

Risposte:

- Si calcolino il valore atteso e lo scarto quadratico medio di \tilde{X} .

Per quanto abbiamo visto in questa sezione, possiamo assumere che $\tilde{X} \sim \text{Binomiale}(n = 13; p = 1/3)$ (ricordiamo che \tilde{X} è la variabile casuale che conta il numero di esiti indovinati in $n = 13$ puntate singole).

Il valore atteso e lo scarto quadratico medio di \tilde{X} sono quindi dati da

$$\mu = E(\tilde{X}) = 13 \times \frac{1}{3} = 4.3333$$

e da

$$\sigma = \sqrt{\text{var}(\tilde{X})} = \sqrt{13 \times \frac{1}{3} \left(1 - \frac{1}{3}\right)} = 1.6997.$$

- Qual è la probabilità di indovinare tutti e 13 gli esiti finali?

La probabilità richiesta è data da

$$\begin{aligned} P(\{\tilde{X} = 13\}) &= \binom{13}{13} \left(\frac{1}{3}\right)^{13} \left(1 - \frac{1}{3}\right)^{13-13} \\ &= \frac{1}{3^{13}} = 6 \times 10^{-7}. \end{aligned}$$

c) Qual è la probabilità di indovinare almeno 12 esiti finali?

La probabilità richiesta è data da

$$\begin{aligned} P(\{\tilde{X} \geq 12\}) &= P(\{\tilde{X} = 12\}) + P(\{\tilde{X} = 13\}) \\ &= \binom{13}{12} \left(\frac{1}{3}\right)^{12} \left(1 - \frac{1}{3}\right)^{13-12} + \\ &\quad + \binom{13}{13} \left(\frac{1}{3}\right)^{13} \left(1 - \frac{1}{3}\right)^{13-13} \\ &= 1.63 \times 10^{-5} + 6 \times 10^{-7} \simeq 1.69 \times 10^{-5} \end{aligned}$$

d) Qual è la probabilità di indovinare al più 3 esiti finali?

La probabilità richiesta è data da

$$\begin{aligned} P(\{\tilde{X} \leq 3\}) &= \sum_{x=0}^3 \binom{13}{x} \left(\frac{1}{3}\right)^x \left(1 - \frac{1}{3}\right)^{13-x} \\ &= 0.0051 + 0.0334 + 0.1002 \\ &= 0.1387 \end{aligned}$$

e*) Qual è la probabilità di indovinare tutti e 13 gli esiti finali con dodici puntate singole e un'unica puntata doppia? Si assuma che tutte le puntate siano casuali.

Per rispondere a questa domanda conviene considerare due variabili casuali \tilde{X}_1 e \tilde{X}_2 , dove

- \tilde{X}_1 conta il numero di esiti indovinati nelle $n_1 = 12$ puntate singole
- e dove \tilde{X}_2 conta il numero di esiti indovinati nell'unica puntata doppia.

Per quanto abbiamo visto in questa sezione, possiamo dire che

$$\tilde{X}_1 \sim \text{Binomiale}(n_1 = 12; p = 1/3) \quad \text{e} \quad \tilde{X}_2 \sim \text{Bernoulli}(p = 2/3).$$

Inoltre, siccome le variabili casuali \tilde{X}_1 e \tilde{X}_2 si riferiscono a partite diverse, possiamo anche assumere che \tilde{X}_1 e \tilde{X}_2 siano indipendenti.

L'evento di cui è richiesta la probabilità può essere espresso attraverso le variabili casuali \tilde{X}_1 e \tilde{X}_2 come

$$A = \{\tilde{X}_1 = 12\} \cap \{\tilde{X}_2 = 1\}.$$

Dall'ipotesi di indipendenza discende quindi che

$$\begin{aligned}
 P(A) &= P(\{\tilde{X}_1 = 12\} \cap \{\tilde{X}_2 = 1\}) \\
 [[\text{indipendenza}]] &= P(\{\tilde{X}_1 = 12\}) \times P(\{\tilde{X}_2 = 1\}) \\
 &= \binom{12}{12} \left(\frac{1}{3}\right)^{12} \left(1 - \frac{1}{3}\right)^{12-12} \times \frac{2}{3} \\
 &= \left(\frac{1}{3}\right)^{12} \times \frac{2}{3} = 1.2 \times 10^{-6}.
 \end{aligned}$$

f*) Qual è la probabilità di indovinare 12 esiti finali con dodici puntate singole e un'unica puntata doppia? Si assuma che tutte le puntate siano casuali.

Per rispondere questa domanda conviene fare ancora riferimento alle variabili casuali \tilde{X}_1 e \tilde{X}_2 che abbiamo introdotto nella risposta al quesito precedente. Usando queste ultime possiamo esprimere l'evento di cui è richiesta la probabilità come

$$\begin{aligned}
 B &= [\{\tilde{X}_1 = 12\} \cap \{\tilde{X}_2 = 0\}] \cup \\
 &\quad \cup [\{\tilde{X}_1 = 11\} \cap \{\tilde{X}_2 = 1\}].
 \end{aligned}$$

Siccome i due eventi tra parentesi quadre sono incompatibili, la probabilità richiesta può essere calcolata come

$$\begin{aligned}
 P(B) &= P([\{\tilde{X}_1 = 12\} \cap \{\tilde{X}_2 = 0\}] \cup [\{\tilde{X}_1 = 11\} \cap \{\tilde{X}_2 = 1\}]) \\
 &[[\text{i due eventi tra parentesi quadre sono incompatibili}]] \\
 &= P(\{\tilde{X}_1 = 12\} \cap \{\tilde{X}_2 = 0\}) + P(\{\tilde{X}_1 = 11\} \cap \{\tilde{X}_2 = 1\}) \\
 &[[\tilde{X}_1 \text{ e } \tilde{X}_2 \text{ sono indipendenti}]] \\
 &= P(\{\tilde{X}_1 = 12\}) \times P(\{\tilde{X}_2 = 0\}) + P(\{\tilde{X}_1 = 11\}) \times P(\{\tilde{X}_2 = 1\}) \\
 &= \binom{12}{12} \left(\frac{1}{3}\right)^{12} \left(1 - \frac{1}{3}\right)^{12-12} \times \left(1 - \frac{2}{3}\right) + \\
 &\quad + \binom{12}{11} \left(\frac{1}{3}\right)^{11} \left(1 - \frac{1}{3}\right)^{12-11} \times \frac{2}{3} \\
 &= 6 \times 10^{-7} + 3.01 \times 10^{-5} \simeq 3 \times 10^{-5}
 \end{aligned}$$

8.5 La famiglia delle distribuzioni trinomiali

Le distribuzioni trinomiali sorgono nel contesto di successioni (finite o infinite) di prove indipendenti che vengono eseguite sotto le medesime condizioni quando l'esito di ciascuna prova può essere di **tre** tipi diversi. In quanto segue chiameremo questo tipo di esperimento casuale **"esperimento trinomiale"** e identificheremo i tre esiti che possono

essere osservati al termine di ciascuna prova con le lettere A , B e C .

Con riferimento a n prove di un esperimento trinomiale possiamo definire **tre variabili casuali** che contano quante volte si è verificato ciascuno dei tre esiti possibili. In quanto segue indicheremo dunque con \tilde{X}_A , \tilde{X}_B e \tilde{X}_C le variabili casuali che restituiscono il numero di volte che si è verificato ciascuno dei tre esiti A , B e C , rispettivamente.

Ovviamente, siccome il numero complessivo di prove è uguale a n , la somma delle realizzazioni delle variabili casuali \tilde{X}_A , \tilde{X}_B e \tilde{X}_C deve sempre essere uguale a n :

$$\tilde{X}_A + \tilde{X}_B + \tilde{X}_C = n.$$

Le probabilità di tutti gli eventi legati alle tre variabili casuali \tilde{X}_A , \tilde{X}_B e \tilde{X}_C sono quindi univocamente determinate dalla distribuzione congiunta di soltanto due di esse, e la distribuzione congiunta di due variabili casuali di questo tipo è appunto una **distribuzione trinomiale**. Come vedremo nella dimostrazione che segue, la funzione di massa di probabilità associata ad una distribuzione trinomiale è data da

$$p(x_A, x_B) = \begin{cases} \frac{n!}{x_A! x_B! (n-x_A-x_B)!} p_A^{x_A} p_B^{x_B} (1-p_A-p_B)^{n-x_A-x_B} \\ \text{per } x_A, x_B = 0, 1, 2, \dots, n \text{ tali che } x_A + x_B \leq n, \\ 0 \quad \text{altrimenti,} \end{cases} \quad (114)$$

dove p_A e p_B sono, rispettivamente, la probabilità che in una singola prova si verifichi l'esito A e la probabilità che in una singola prova si verifichi l'esito B . Siccome in ciascuna delle n prove di riferimento deve verificarsi uno (e soltanto uno) dei tre esiti A , B oppure C , e siccome nell'ambito di un esperimento trinomiale ciascuno di questi tre esiti deve potersi verificare, i valori di p_A e p_B devono soddisfare le seguenti disuguaglianze:

$$p_A > 0, \quad p_B > 0, \quad p_A + p_B < 1.$$

Si noti che il fattore $1 - p_A - p_B$ che compare nella formula (114) è dunque la probabilità che in una singola prova si verifichi l'esito C .

Dimostrazione. In questa dimostrazione ricaveremo la formula della funzione di massa di probabilità trinomiale, ovvero la formula (114). Il ragionamento sarà perfettamente analogo a quello che abbiamo già visto quando abbiamo ricavato la formula della funzione di massa di probabilità binomiale.

Per cominciare, consideriamo dunque una sequenza di esiti delle n prove che può essere rappresentata attraverso un evento del tipo $\bigcap_{i=1}^n E_i$ dove ciascun evento E_i è definito come A_i , B_i oppure C_i a seconda se l'esito dell' i -esima prova è del tipo A , B oppure C . Siccome per ipotesi le n prove sono indipendenti, possiamo concludere che

$$P\left(\bigcap_{i=1}^n E_i\right) = P(E_1) \times P(E_2) \times \dots \times P(E_n) = p_A^{x_A} \times p_B^{x_B} \times (1-p_A-p_B)^{n-x_A-x_B}$$

dove x_A è il numero complessivo di eventi A_i , x_B è il numero complessivo di eventi B_i e $n - x_A - x_B$ è il numero complessivo di eventi C_i tra gli n eventi E_i coinvolti nell'intersezione $\bigcap_{i=1}^n E_i$. Questo risultato dimostra che **la probabilità associata a ciascuna sequenza di esiti dipende solo dal numero di volte che si verifica ciascuno dei tre esiti nelle n prove, ma non dall'ordine in cui i tre esiti si verificano**. Siccome ogni sequenza di esiti è incompatibile con tutte le altre sequenze diverse da lei, e siccome le posizioni di x_A esiti del tipo A e di x_B esiti del tipo B in una sequenza di n esiti possono essere scelte in

$$\binom{n}{x_A} \times \binom{n-x_A}{x_B} = \frac{n!}{x_A!(n-x_A)!} \times \frac{(n-x_A)!}{x_B!(n-x_A-x_B)!} = \frac{n!}{x_A!x_B!(n-x_A-x_B)!}$$

modi diversi, la probabilità di osservare x_A esiti del tipo A e di x_B esiti del tipo B deve essere data da

$$\frac{n!}{x_A!x_B!(n-x_A-x_B)!} p_A^{x_A} \times p_B^{x_B} \times (1-p_A-p_B)^{n-x_A-x_B}.$$

□

Come si desume dalla precedente dimostrazione, la probabilità di osservare x_A esiti del tipo A e x_B esiti del tipo B nell'ambito delle n prove di un esperimento trinomiale è data dal prodotto tra

- la probabilità $p_A^{x_A} \times p_B^{x_B} \times (1-p_A-p_B)^{n-x_A-x_B}$ di ottenere x_A esiti del tipo A , x_B esiti del tipo B e $n - x_A - x_B$ esiti del tipo C **in un preciso ordine** (qualunque esso sia)
- e il numero di modi

$$\frac{n!}{x_A!x_B!(n-x_A-x_B)!}$$

in cui possono verificarsi x_A esiti del tipo A , x_B esiti del tipo B e $n - x_A - x_B$ esiti del tipo C in una sequenza di n esiti.

Il nome "trinomiale" delle distribuzioni bivariate con funzione di massa di probabilità congiunta

$$p(x_A, x_B) = \begin{cases} \frac{n!}{x_A!x_B!(n-x_A-x_B)!} p_A^{x_A} p_B^{x_B} (1-p_A-p_B)^{n-x_A-x_B} & \text{per } x_A, x_B = 0, 1, 2, \dots, n \text{ tali che } x_A + x_B \leq n, \\ 0 & \text{altrimenti,} \end{cases}$$

deriva dal fatto che le probabilità congiunte $p(x_A, x_B)$ sono i termini dello sviluppo in serie di un "trinomio". Infatti, un trinomio $(a + b + c)^n$ può sempre essere sviluppato come

$$(a + b + c)^n = \sum_{x_A=0}^n \sum_{x_B=0}^{n-x_A} \frac{n!}{x_A!x_B!(n-x_A-x_B)!} a^{x_A} b^{x_B} c^{n-x_A-x_B}.$$

e ponendo $a = p_A$, $b = p_B$ e $c = 1 - p_A - p_B$ si ottengono le probabilità trinomiali.

D'ora in poi, per comodità di notazione, indicheremo la distribuzione trinomiale con parametri n , p_A e p_B semplicemente con

$$Trinomiale(n; p_A; p_B),$$

e quando vorremo indicare che la distribuzione congiunta di due variabili casuali \tilde{X}_A e \tilde{X}_B è $Trinomiale(n; p_A; p_B)$ scriveremo semplicemente

$$(\tilde{X}_A, \tilde{X}_B) \sim Trinomiale(n; p_A; p_B).$$

Rappresentazione come somme di variabili casuali indicatrici. Ovviamente, anche \tilde{X}_A e \tilde{X}_B possono essere rappresentate come somme di variabili casuali indicatrici indipendenti. Infatti, se indichiamo con

$$\tilde{I}_{A,1}, \quad \tilde{I}_{A,2}, \quad \dots, \quad \tilde{I}_{A,n}$$

le variabili casuali indicatrici che per ciascuna delle n prove indicano se l'esito A si verifica o meno ($\tilde{I}_{A,i}$ assume il valore 1 se nell' i -esima prova si verifica l'esito A , e assume il valore 0 altrimenti), allora il numero complessivo di volte che si verifica l'esito A deve essere dato da

$$\tilde{X}_A = \tilde{I}_{A,1} + \tilde{I}_{A,2} + \dots + \tilde{I}_{A,n}.$$

Definendo in modo analogo le variabili casuali indicatrici

$$\tilde{I}_{B,1}, \quad \tilde{I}_{B,2}, \quad \dots, \quad \tilde{I}_{B,n},$$

vediamo che anche la variabile casuale \tilde{X}_B può essere rappresentata come

$$\tilde{X}_B = \tilde{I}_{B,1} + \tilde{I}_{B,2} + \dots + \tilde{I}_{B,n}.$$

Si noti che soltanto le variabili casuali indicatrici che si riferiscono a prove diverse sono indipendenti (ovvero quelle con pedici diversi), ma che nessuna delle coppie $(\tilde{I}_{A,i}, \tilde{I}_{B,i})$ con lo stesso pedice i è una coppia di variabili casuali indipendenti!!!

Distribuzioni marginali. Considerando gli esiti del tipo A come successi e entrambi gli altri due esiti come insuccessi, vediamo che la variabile casuale \tilde{X}_A può essere interpretata come numero complessivo di successi nelle n prove, e che la distribuzione marginale di \tilde{X}_A è dunque binomiale con parametri n e $p = p_A$. Analogamente, la distribuzione marginale di \tilde{X}_B è binomiale con parametri n e $p = p_B$.

Ovviamente, le distribuzioni marginali delle variabili casuali \tilde{X}_A e \tilde{X}_B possono essere ricavate anche attraverso un approccio algebrico a partire dalla funzione di probabilità congiunta (basta applicare le formule (40) e (41)). Per illustrare il procedimento mostriamo i passaggi algebrici per ottenere la funzione di massa di probabilità marginale di \tilde{X}_A :

per $x_A = 0, 1, 2, \dots, n$ si ottiene

$$\begin{aligned}
 p_A(x_A) &= P(\{\tilde{X}_A = x_A\}) = \sum_{x_B \in \mathcal{S}_B} p(x_A, x_B) \\
 &= \sum_{x_B=1}^{n-x_A} \frac{n!}{x_A! x_B! (n-x_A-x_B)!} p_A^{x_A} p_B^{x_B} (1-p_A-p_B)^{n-x_A-x_B} \\
 &= \frac{n!}{x_A! (n-x_A)!} \times p_A^{x_A} \times \sum_{x_B=1}^{n-x_A} \frac{(n-x_A)!}{x_B! (n-x_A-x_B)!} p_B^{x_B} (1-p_A-p_B)^{n-x_A-x_B} \\
 &= \binom{n}{x_A} \times p_A^{x_A} \times \sum_{x_B=1}^{n-x_A} \binom{n-x_A}{x_B} p_B^{x_B} (1-p_A-p_B)^{n-x_A-x_B} \\
 &[[\text{formula (111) dello sviluppo di un binomio}]] \\
 &= \binom{n}{x_A} p_A^{x_A} \times [p_B + (1-p_A-p_B)]^{n-x_A} \\
 &= \binom{n}{x_A} p_A^{x_A} (1-p_A)^{n-x_A}
 \end{aligned}$$

e l'espressione nell'ultima riga è la formula della funzione di massa di probabilità binomiale con parametri n e $p = p_A$, così come volevamo dimostrare.

Distribuzioni condizionate. Per ricavare le distribuzioni condizionate di \tilde{X}_A basta applicare la definizione di probabilità condizionata. Se la probabilità condizionata

$$P(\{\tilde{X}_A = x_A\} | \{\tilde{X}_B = x_B\})$$

è ben definita (ovvero se $x_B = 0, 1, 2, \dots, n$), si ottiene

$$\begin{aligned}
 p(x_A | x_B) &= P(\{\tilde{X}_A = x_A\} | \{\tilde{X}_B = x_B\}) \\
 &= \begin{cases} \binom{n-x_B}{x_A} \left(\frac{p_A}{1-p_B}\right)^{x_A} \left(1 - \frac{p_A}{1-p_B}\right)^{n-x_B-x_A} & \text{per } x_A = 0, 1, \dots, n-x_B \\ 0 & \text{altrimenti.} \end{cases} \quad (115)
 \end{aligned}$$

Quindi, non solo la distribuzione **marginale** di \tilde{X}_A è binomiale, ma anche tutte le distribuzioni **condizionate** di \tilde{X}_A sono binomiali. In particolare, la distribuzione condizionata di \tilde{X}_A dato $\tilde{X}_B = x_B$ è *Binomiale*($n = n - x_B; p = p_A / (1 - p_B)$). Scambiando i ruoli delle variabili casuali \tilde{X}_A e \tilde{X}_B si vede che la distribuzione condizionata di \tilde{X}_B dato $\tilde{X}_A = x_A$ è invece *Binomiale*($n = n - x_A; p = p_B / (1 - p_A)$).

Dimostrazione (Dimostrazione della formula (115)). Con $x_B = 0, 1, 2, \dots, n$ e $x_A = 0, 1, \dots, n - x_B$ si ottiene

$$\begin{aligned}
 p(x_A | x_B) &= P(\{\tilde{X}_A = x_A\} | \{\tilde{X}_B = x_B\}) = \frac{P(\{\tilde{X}_A = x_A\} \cap \{\tilde{X}_B = x_B\})}{P(\{\tilde{X}_B = x_B\})} = \frac{p(x_A, x_B)}{p(x_B)} \\
 &= \frac{\frac{n!}{x_A! x_B! (n-x_A-x_B)!} p_A^{x_A} p_B^{x_B} (1-p_A-p_B)^{n-x_A-x_B}}{\frac{n!}{x_B! (n-x_B)!} p_B^{x_B} (1-p_B)^{n-x_B}} = \frac{(n-x_B)!}{x_A! (n-x_A-x_B)!} \frac{p_A^{x_A} (1-p_A-p_B)^{n-x_A-x_B}}{(1-p_B)^{n-x_B}} \\
 &= \binom{n-x_B}{x_A} \times \frac{p_A^{x_A} \times (1-p_A-p_B)^{n-x_A-x_B}}{(1-p_B)^{x_A} \times (1-p_B)^{n-x_B-x_A}} = \binom{n-x_B}{x_A} \times \left(\frac{p_A}{1-p_B}\right)^{x_A} \times \left(1 - \frac{p_A}{1-p_B}\right)^{n-x_B-x_A}.
 \end{aligned}$$

□

Covarianza e coefficiente di correlazione lineare. Come dimostreremo tra breve,

$$\text{cov}(\tilde{X}_A, \tilde{X}_B) = -np_A p_B \quad (116)$$

e

$$r(\tilde{X}_A, \tilde{X}_B) = -\sqrt{\frac{p_A p_B}{(1-p_A)(1-p_B)}}. \quad (117)$$

Si noti che i valori di $\text{cov}(\tilde{X}_A, \tilde{X}_B)$ e $r(\tilde{X}_A, \tilde{X}_B)$ sono sempre negativi!!! A pensarci bene questa non è una sorpresa: siccome il numero complessivo di prove n è costante, all'aumentare del numero di volte che si verifica l'esito A il numero di volte che si verifica l'esito B deve diminuire, e viceversa.

Dimostrazione. Per dimostrare la formula (116) della covarianza, osserviamo innanzitutto che (si ricordi le rappresentazioni di \tilde{X}_A e \tilde{X}_B come somme di variabili casuali indicatrici)

$$\text{cov}(\tilde{X}_A, \tilde{X}_B) = \text{cov}\left(\sum_{i=1}^n \tilde{I}_{A,i}, \sum_{i=1}^n \tilde{I}_{B,i}\right) = \text{[[proprietà V4*]]} = \sum_{i=1}^n \sum_{j=1}^n \text{cov}(\tilde{I}_{A,i}, \tilde{I}_{B,j}). \quad (118)$$

Tenendo presente che le variabili casuali indicatrici che si riferiscono a prove diverse sono indipendenti, vediamo che (proprietà V1)

$$\text{cov}(\tilde{I}_{A,i}, \tilde{I}_{B,j}) = 0 \quad \text{se } i \neq j.$$

Sostituendo questo risultato nella (118) otteniamo

$$\text{cov}(\tilde{X}_A, \tilde{X}_B) = \sum_{i=1}^n \text{cov}(\tilde{I}_{A,i}, \tilde{I}_{B,i}). \quad (119)$$

Per esprimere $\text{cov}(\tilde{X}_A, \tilde{X}_B)$ come funzione dei tre parametri n , p_A e p_B , osserviamo ora che

- $E(\tilde{I}_{A,i} \times \tilde{I}_{B,i}) = 0$ perché la variabile casuale $\tilde{I}_{A,i} \times \tilde{I}_{B,i}$ può assumere soltanto il valore 0 (se nell' i -esima prova si verifica l'esito A , allora la variabile casuale indicatrice $\tilde{I}_{B,i}$ è nulla; altrimenti è nulla la variabile casuale indicatrice $\tilde{I}_{A,i}$),
- e che $E(\tilde{I}_{A,i}) = p_A$ e $E(\tilde{I}_{B,i}) = p_B$.

Usando la formula indiretta per il calcolo della covarianza vediamo quindi che

$$\text{cov}(\tilde{I}_{A,i}, \tilde{I}_{B,i}) = E(\tilde{I}_{A,i} \times \tilde{I}_{B,i}) - E(\tilde{I}_{A,i}) \times E(\tilde{I}_{B,i}) = 0 - p_A \times p_B,$$

e sostituendo questo risultato nella (119) otteniamo finalmente la formula

$$\text{cov}(\tilde{X}_A, \tilde{X}_B) = \sum_{i=1}^n \text{cov}(\tilde{I}_{A,i}, \tilde{I}_{B,i}) = \sum_{i=1}^n (-p_A \times p_B) = -np_A p_B,$$

che volevamo dimostrare.

Per dimostrare anche la formula (117) del coefficiente di correlazione lineare, ricordiamo innanzitutto che le distribuzioni marginali delle variabili casuali \tilde{X}_A e \tilde{X}_B sono entrambe binomiali e che dunque

$$\text{var}(\tilde{X}_A) = np_A(1 - p_A) \quad \text{e} \quad \text{var}(\tilde{X}_B) = np_B(1 - p_B).$$

Sostituendo nella formula che definisce il coefficiente di correlazione lineare otteniamo quindi

$$\begin{aligned} r(\tilde{X}_A, \tilde{X}_B) &= \frac{\text{cov}(\tilde{X}_A, \tilde{X}_B)}{\sqrt{\text{var}(\tilde{X}_A) \times \text{var}(\tilde{X}_B)}} = \frac{-np_{APB}}{\sqrt{np_A(1 - p_A) \times np_B(1 - p_B)}} \\ &= \dots = -\sqrt{\frac{p_{APB}}{(1 - p_A)(1 - p_B)}}. \end{aligned}$$

□

Esercizio 8.8. Un'urna contiene 5 palline rosse, 3 palline bianche e 2 palline nere. Dall'urna vengono estratte, con reimmissione, 5 palline.

- Qual è la probabilità di ottenere tre palline rosse?
- Qual è la probabilità di ottenere tre palline rosse e una pallina bianca?
- Sapendo che tre delle palline estratte sono rosse, qual è la probabilità che una sia bianca?
- Tizio vince 2 Euro per ogni pallina rossa estratta e perde 3 Euro per ogni pallina bianca estratta. Si determinino il valore atteso e la varianza della vincita di Tizio.

Risposte:

- Qual è la probabilità di ottenere tre palline rosse?

Indichiamo con \tilde{X}_R la variabile casuale che conta il numero di palline rosse estratte. Siccome le estrazioni avvengono con reimmissione e siccome non c'è motivo di ritenere che gli esiti delle estrazioni si influenzino a vicenda, assumeremo che la distribuzione di \tilde{X}_R sia binomiale con parametri $n = 5$ e $p = p_R = 5/10$. Usando la funzione di massa di probabilità binomiale vediamo che la probabilità richiesta è quindi data da

$$\begin{aligned} P(\{\tilde{X}_R = 3\}) &= \binom{5}{3} \left(\frac{5}{10}\right)^3 \left(1 - \frac{5}{10}\right)^{5-3} \\ &= 0,313. \end{aligned}$$

b) Qual è la probabilità di ottenere tre palline rosse e una pallina bianca?

Indichiamo con \tilde{X}_B la variabile casuale che conta il numero di palline bianche estratte. Siccome le estrazioni avvengono con reimmissione e siccome non c'è motivo di ritenere che gli esiti delle estrazioni si influenzino a vicenda, assumeremo che la distribuzione congiunta delle variabili casuali \tilde{X}_R e \tilde{X}_B sia trinomiale con parametri $n = 5$, $p_R = 5/10$ e $p_B = 3/10$. Usando la funzione di massa di probabilità trinomiale vediamo che la probabilità richiesta è data da

$$\begin{aligned} P\left(\{\tilde{X}_R = 3\} \cap \{\tilde{X}_B = 1\}\right) &= \\ &= \frac{5!}{3!1!1!} \times \left(\frac{5}{10}\right)^3 \times \left(\frac{3}{10}\right)^1 \times \left(\frac{2}{10}\right)^1 \\ &= 0,15. \end{aligned}$$

c) Sapendo che tre delle palline estratte sono rosse, qual è la probabilità che una sia bianca?

Per rispondere al quesito dobbiamo calcolare la probabilità condizionata

$$P\left(\{\tilde{X}_B = 1\} \mid \{\tilde{X}_R = 3\}\right).$$

Nelle risposte ai quesiti a) e b) abbiamo già visto che

$$P(\{\tilde{X}_R = 3\}) = 0,313$$

e che

$$P\left(\{\tilde{X}_B = 1\} \cap \{\tilde{X}_R = 3\}\right) = 0,15.$$

La probabilità condizionata richiesta è dunque data da

$$P\left(\{\tilde{X}_B = 1\} \mid \{\tilde{X}_R = 3\}\right) = \frac{0,15}{0,313} = 0,479.$$

Ovviamente si ottiene esattamente lo stesso risultato facendo riferimento alla funzione di massa di probabilità condizionata $p(x_B \mid x_R = 3)$. Si ricordi infatti che le distribuzioni condizionate di una distribuzione trinomiale sono binomiali. In particolare, la distribuzione condizionata di \tilde{X}_B dato $\tilde{X}_R = 3$ è binomiale con parametri n e p rispettivamente dati da $n - x_R = 5 - 3 = 2$ e

$$\frac{p_B}{1 - p_R} = \frac{\frac{2}{10}}{1 - \frac{5}{10}} = \frac{2}{5}.$$

Usando la corrispondente funzione di massa di probabilità otteniamo

$$\begin{aligned} p(x_B = 1 \mid x_R = 3) &= \binom{2}{1} \times \left(\frac{2}{5}\right)^1 \times \left(1 - \frac{2}{5}\right)^{2-1} \\ &= 0,48 \quad (\simeq 0,479) \end{aligned}$$

esattamente come prima.

- d) Tizio vince 2 Euro per ogni pallina rossa estratta e perde 3 Euro per ogni pallina bianca estratta. Si determinino il valore atteso e la varianza della vincita di Tizio.

Sia \tilde{Y} la variabile casuale che rappresenta la vincita di Tizio. Dal modo in cui si determina la vincita desumiamo che

$$\tilde{Y} = 2 \times \tilde{X}_R - 3 \times \tilde{X}_B.$$

Siccome i valori attesi di \tilde{X}_R e \tilde{X}_B esistono entrambi (si ricordi che le distribuzioni di queste due variabili casuali sono entrambe binomiali), possiamo applicare la proprietà E4* onde concludere che anche il valore atteso di \tilde{Y} esiste e che quest'ultimo è dato da

$$\begin{aligned} E(\tilde{Y}) &= E(2 \times \tilde{X}_R - 3 \times \tilde{X}_B) \\ &= E(2 \times \tilde{X}_R) - E(3 \times \tilde{X}_B) \\ &= 2 \times E(\tilde{X}_R) - 3 \times E(\tilde{X}_B) \\ &= 2 \times np_R - 3 \times np_B \\ &= 2 \times \left(5 \times \frac{5}{10}\right) - 3 \times \left(5 \times \frac{3}{10}\right) \\ &= 2 \times 2,5 - 3 \times 1,5 = 0,5. \end{aligned}$$

Consideriamo ora invece la varianza della vincita \tilde{Y} . Siccome anche le varianze di \tilde{X}_R e \tilde{X}_B esistono, possiamo applicare la proprietà V8* onde concludere che

$$\begin{aligned} var(\tilde{Y}) &= var(2 \times \tilde{X}_R - 3 \times \tilde{X}_B) \\ [[\text{proprietà V8*}]] &= 2^2 \times var(\tilde{X}_R) + (-3)^2 \times var(\tilde{X}_B) + \\ &\quad + 2 \times 2 \times (-3) \times cov(\tilde{X}_R, \tilde{X}_B) \\ &= 2^2 \times np_R(1 - p_R) + (-3)^2 \times np_B(1 - p_B) + \\ &\quad + 2 \times 2 \times (-3) \times (-np_R p_B) \\ &= 2^2 \times \left[5 \times \frac{5}{10} \times \left(1 - \frac{5}{10}\right)\right] + \\ &\quad + (-3)^2 \times \left[5 \times \frac{3}{10} \times \left(1 - \frac{3}{10}\right)\right] + \\ &\quad + 2 \times 2 \times (-3) \times \left(-5 \times \frac{5}{10} \times \frac{3}{10}\right) \\ &= 5 + 9,45 + 9 = 23,45. \end{aligned} \tag{120}$$

8.6 La famiglia delle distribuzioni geometriche (o di Pascal)

Una distribuzione geometrica (o di Pascal) è la distribuzione di una variabile casuale che nell'ambito di una successione di prove indipendenti che vengono eseguite sotto le medesime condizioni (ovvero nell'ambito di un [processo di Bernoulli](#)) conta il numero di prove necessarie per ottenere il primo successo.

Ovviamente si tratta di una distribuzione discreta il cui supporto è dato dall'insieme dei numeri naturali positivi (per ottenere un successo bisogna eseguire almeno una prova e non esiste un tetto massimo al numero di prove necessarie per ottenere un successo).

Non è difficile dimostrare che la funzione di massa di probabilità di una distribuzione geometrica deve essere data da

$$p(x) = \begin{cases} p(1-p)^{x-1} & \text{per } x = 1, 2, 3, \dots; \\ 0 & \text{altrimenti,} \end{cases}$$

dove $p \in (0, 1)$ è la probabilità di successo comune a tutte le prove.

Dimostrazione. Supponiamo che il primo successo si verifichi nella prova numero x (un numero naturale positivo). Si noti che questo evento si verifica se e solo se nelle prime $x - 1$ prove si verificano solo insuccessi e nella prova numero x si verifica un successo. L'evento in questione può dunque essere indicato come

$$\bar{S}_1 \cap \bar{S}_2 \cap \dots \cap \bar{S}_{x-1} \cap S_x,$$

dove S_i è l'evento secondo il quale l'esito dell' i -esima prova è un successo e \bar{S}_i è l'evento complementare (ovvero l'evento secondo il quale l'esito dell' i -esima prova è un insuccesso). Siccome gli eventi coinvolti nell'intersezione si riferiscono a prove diverse, essi devono essere indipendenti e quindi possiamo concludere che

$$\begin{aligned} P(\bar{S}_1 \cap \bar{S}_2 \cap \dots \cap \bar{S}_{x-1} \cap S_x) &= P(\bar{S}_1) \times P(\bar{S}_2) \times \dots \times P(\bar{S}_{x-1}) \times P(S_x) \\ &= (1-p) \times \dots \text{[[} x-1 \text{ volte]]} \dots \times (1-p) \times p \\ &= (1-p)^{x-1} p \end{aligned}$$

come nell'espressione della funzione di massa di probabilità geometrica. □

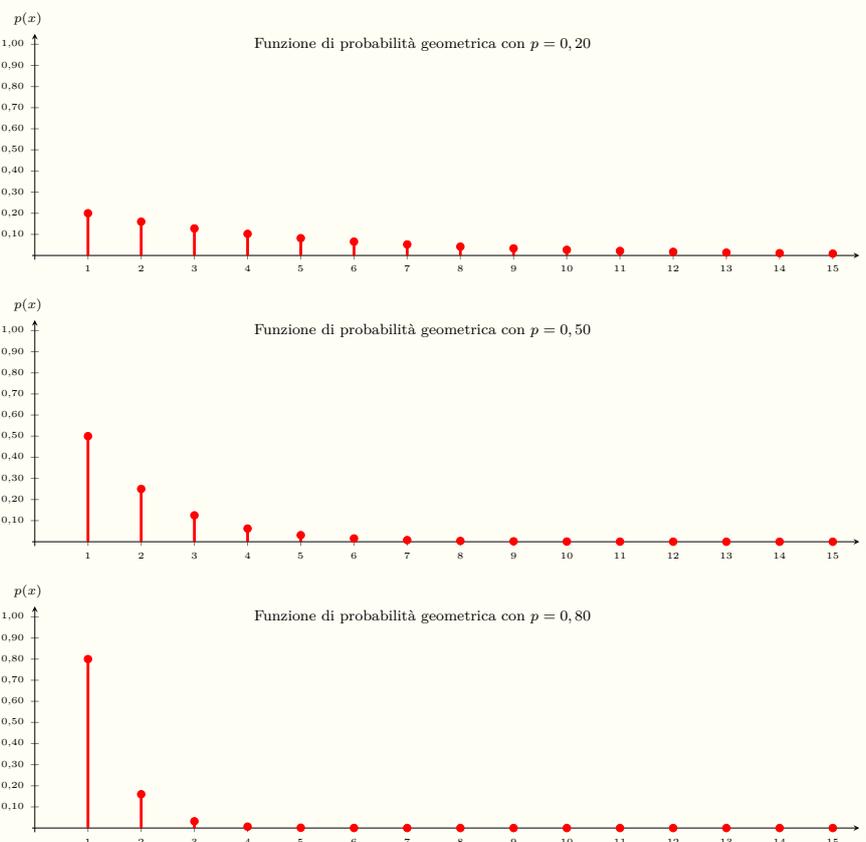
Dalla precedente dimostrazione desumiamo che la probabilità di osservare il primo successo nella prova numero x è data dal prodotto tra

- la probabilità $(1-p)^{x-1}$ di ottenere soltanto insuccessi nelle prime $x - 1$ prove
- e la probabilità p di ottenere un successo nella prova numero x .

D'ora in poi, per comodità di notazione, indicheremo che \tilde{X} è una variabile casuale con distribuzione geometrica di parametro p scrivendo semplicemente

$$\tilde{X} \sim \text{Geometrica}(p).$$

Figura 8.4. I grafici mostrano l'andamento di alcune funzioni di massa di probabilità geometriche.



Serie geometriche. Il nome "geometrica" delle distribuzioni discrete con funzione di massa di probabilità

$$p(x) = \begin{cases} p(1-p)^{x-1} & \text{per } x = 1, 2, 3, \dots; \\ 0 & \text{altrimenti,} \end{cases} \quad 0 < p < 1,$$

deriva dal fatto che le probabilità $p(1)$, $p(2)$, ... sono i termini (traslati) di una serie geometrica, ovvero di una serie del tipo

$$\sum_{t=0}^{\infty} ar^t, \quad a, r \in \mathbb{R}, \quad a \neq 0$$

(se $r = 0$, il valore di $r^0 = 0^0$ deve essere interpretato come un 1). Infatti, ponendo $a = p$, $r = 1 - p$ e $t = x - 1$ si ottengono le probabilità della distribuzione geometrica:

$$ar^t = p(1-p)^{x-1} = p(x), \quad x = 1, 2, \dots$$

Per dimostrare le principali proprietà delle distribuzioni geometriche ci serviranno la formula per calcolare le **somme parziali** di una serie geometrica e la formula per

calcolare il **limite di una serie geometrica** (quando questo limite esiste). Non è difficile verificare che la formula per calcolare le somme parziali è data da

$$\sum_{t=0}^y ar^t = \frac{a(1-r^{y+1})}{1-r}, \quad t = 1, 2, \dots \quad (121)$$

Infatti, per dimostrare questa formula basta osservare che

$$\begin{aligned} \left[\sum_{t=0}^y ar^t \right] \times (1-r) &= [ar^0 + ar^1 + ar^2 + \dots + ar^y] \times (1-r) = \\ &= [ar^0 + ar^1 + ar^2 + \dots + ar^y] \times 1 - [ar^0 + ar^1 + ar^2 + \dots + ar^y] \times r = \\ &= [ar^0 + ar^1 + ar^2 + \dots + ar^y] - [ar^1 + ar^2 + ar^3 + \dots + ar^{y+1}] = \\ &= ar^0 - ar^{y+1} = a(1-r^{y+1}). \end{aligned}$$

Per indagare sul limite di una serie geometrica bisogna invece far tendere y a infinito ad ambo i membri dell'uguaglianza (121). In questo modo vediamo che le somme parziali di una serie geometrica sono

- divergenti se $r \geq 1$
- oscillanti se $r \leq -1$
- convergenti se $-1 < r < 1$.

Il comportamento asintotico delle somme parziali di una serie geometrica dipende quindi dal valore di r che viene chiamato **"ragione"** della serie geometrica. Si noti che la ragione r è il valore costante del rapporto tra un qualunque termine della serie e il termine che lo precede: $ar^{t+1}/(ar^t) = r$. Dalle osservazioni sulle somme parziali deduciamo che il limite di una serie geometrica esiste se e solo se la sua ragione r appartiene all'intervallo $(-1, 1)$ e che in tal caso

$$\sum_{t=0}^{\infty} ar^t = \lim_{y \rightarrow \infty} \sum_{t=0}^y ar^t = \lim_{y \rightarrow \infty} a \times \frac{1-r^{y+1}}{1-r} = \frac{a}{1-r}.$$

Probabilità cumulate. Usando la formula (121) delle somme parziali di una serie geometrica vediamo che nel caso di una variabile casuale \tilde{X} con distribuzione geometrica di parametro p si ottiene

$$\begin{aligned} P(\{\tilde{X} \leq x\}) &= P(\{\tilde{X} = 1\}) + P(\{\tilde{X} = 2\}) + \dots + P(\{\tilde{X} = x\}) \\ &= p(1) + p(2) + \dots + p(x) = \sum_{t=1}^x (1-p)^{t-1} p = \sum_{t=0}^{x-1} (1-p)^t p \\ \text{[[formula (121)]]} &= \frac{p[1 - (1-p)^x]}{1 - (1-p)} = 1 - (1-p)^x. \end{aligned}$$

Vale la pena osservare che questa formula può essere ottenuta anche mediante un ragionamento probabilistico. Infatti, se \tilde{X} è la variabile casuale che conta il numero

di prove che sono necessarie per ottenere il primo successo nell'ambito di un **processo di Bernoulli**, allora possiamo dire che **l'evento $\{\tilde{X} > x\}$ si verifica se e solo se nelle prime x prove si verificano soltanto insuccessi** e virtù delle caratteristiche dei **processi di Bernoulli** (prove indipendenti con probabilità di successo identica in tutte le prove) possiamo immediatamente concludere che

$$P(\{\tilde{X} > x\}) = (1 - p)^x. \quad (122)$$

Usando la legge **L2** vediamo dunque che le probabilità cumulate di una distribuzione geometrica sono date da

$$P(\{\tilde{X} \leq x\}) = [[\text{legge L2}]] = 1 - P(\{\tilde{X} > x\}) = 1 - (1 - p)^x \quad (123)$$

come già visto poco sopra.

La proprietà dell'assenza di memoria. Si considerino i numeri ritardatari nel gioco della roulette (consideriamo una roulette del tipo francese con 37 settori numerati da 0 a 36). Secondo il metodo classico, la probabilità di ottenere lo zero in una singola giocata è data da $p = 1/37$. Siccome gli esiti di giocate diverse dovrebbero essere indipendenti, la variabile casuale \tilde{X} che conta il numero di giocate necessarie per ottenere per la prima volta lo zero dovrebbe avere distribuzione geometrica di parametro $p = 1/37$. Supponiamo ora che lo zero non esca da un certo numero $x - 1$ di giocate e chiediamoci quale sia la probabilità condizionata che esca nella prossima giocata, ovvero nella giocata numero x . Un semplice calcolo mostra che

$$\begin{aligned} P(\{\tilde{X} = x\} | \{\tilde{X} \geq x\}) &= \frac{P(\{\tilde{X} = x\} \cap \{\tilde{X} \geq x\})}{P(\{\tilde{X} \geq x\})} = \\ &= \frac{P(\{\tilde{X} = x\})}{P(\{\tilde{X} \geq x\})} = \frac{(1 - p)^{x-1} p}{(1 - p)^{x-1}} = p, \end{aligned} \quad (124)$$

ovvero che la probabilità condizionata che lo zero esca nella giocata numero x dato che prima non è mai uscito è esattamente uguale alla probabilità $p = 1/37$ che esca in una qualunque giocata!!! Il valore della probabilità condizionata $P(\{\tilde{X} = x\} | \{\tilde{X} \geq x\})$ non dipende dunque dal valore di x , ovvero dal numero di giocate che sono già trascorse senza che uscisse lo zero!!! Questo risultato è un'immediata conseguenza di una proprietà delle distribuzioni geometriche che è nota come **"proprietà dell'assenza di memoria"**. La definizione di questa proprietà è basata sulle cosiddette **"probabilità condizionate di sopravvivenza"** che sono definite come

$$\begin{aligned} P(\{\tilde{X} > x + h\} | \{\tilde{X} > x\}) &= \frac{P(\{\tilde{X} > x + h\} \cap \{\tilde{X} > x\})}{P(\{\tilde{X} > x\})} = \\ &= \frac{P(\{\tilde{X} > x + h\})}{P(\{\tilde{X} > x\})} = \frac{1 - F(x + h)}{1 - F(x)} \end{aligned} \quad (125)$$

con $x, h \geq 0$. Il nome di queste probabilità condizionate deriva dal loro ruolo nell'**analisi di sopravvivenza**, ovvero nella branca della statistica e della teoria della probabilità che si occupa di metodi per descrivere, analizzare e prevedere tempi d'attesa. Il riferimento alla "sopravvivenza" è dovuto al fatto che molti di questi metodi sono stati sviluppati con riferimento a tempi d'attesa che si riferiscono a eventi non graditi quali rotture, guasti, morte di un organismo biologico ecc.. Ovviamente, i metodi dell'analisi di sopravvivenza possono essere applicati a qualsiasi tipo di tempo d'attesa. Per interpretare le **probabilità condizionate di sopravvivenza** conviene quindi fare riferimento al caso dove $F(x)$ è la distribuzione di una variabile casuale \tilde{X} che rappresenta un tempo d'attesa. In questo caso $P(\{\tilde{X} > x + h\}|\{\tilde{X} > x\})$ è la probabilità condizionata che il tempo d'attesa non termini entro le prossime h unità di tempo dato che si è già protratto fino al tempo x . Avendo in mente questo quadro di riferimento possiamo dunque **definire** la proprietà di assenza di memoria dicendo che una distribuzione **discreta** $F(x)$ soddisfa la **proprietà dell'assenza di memoria** se per ogni prefissato orizzonte temporale futuro $h = 0, 1, 2, \dots$ la **probabilità condizionata di sopravvivenza** non dipende dal tempo d'attesa x che è già trascorso. Volendo essere più formali possiamo dire che la distribuzione $F(x)$ di una variabile casuale discreta \tilde{X} soddisfa la proprietà di assenza di memoria se

$$P(\{\tilde{X} > x + h\}|\{\tilde{X} > x\}) = P(\{\tilde{X} > h\}) \quad \text{per ogni } x, h = 0, 1, 2, \dots, \quad (126)$$

ovvero se (vedi la catena di uguaglianze (125))

$$\frac{1 - F(x + h)}{1 - F(x)} = 1 - F(h) \quad \text{per ogni } x, h = 0, 1, 2, \dots \quad (127)$$

Non è difficile dimostrare che **tutte le distribuzioni geometriche soddisfano la proprietà dell'assenza di memoria**. Infatti, se $\tilde{X} \sim Geometrica(p)$ e x e h sono due numeri naturali scelti in modo arbitrario, si ottiene

$$\begin{aligned} P(\{\tilde{X} > x + h\}|\{\tilde{X} > x\}) &= \frac{P(\{\tilde{X} > x + h\} \cap \{\tilde{X} > x\})}{P(\{\tilde{X} > x\})} = \frac{P(\{\tilde{X} > x + h\})}{P(\{\tilde{X} > x\})} = \\ &= \text{[[formula (122)]]} = \frac{(1 - p)^{(x+h)}}{(1 - p)^x} = (1 - p)^h = \\ &= \text{[[formula (122)]]} = P(\{\tilde{X} > h\}) \end{aligned}$$

così come previsto dalla condizione (126) che definisce la proprietà di assenza di memoria.

D'altra parte, si può anche dimostrare che **tra tutte le distribuzioni che concentrano tutta la probabilità sui numeri naturali, le distribuzioni geometriche sono le uniche che soddisfano la proprietà dell'assenza di memoria:**

Dimostrazione. Per dimostrare che tra tutte le distribuzioni che concentrano tutta la probabilità sui numeri naturali le distribuzioni geometriche sono le uniche che soddisfano la proprietà dell'assenza di memoria, assumeremo che la condizione (126) sia soddisfatta (si ricordi che questa condizione

definisce la proprietà di assenza di memoria) e dimostreremo che in questo caso la distribuzione di \tilde{X} deve necessariamente essere una distribuzione geometrica.

Cominciamo la dimostrazione osservando che la condizione (126) può essere soddisfatta solo se $P(\{\tilde{X} > x\}) > 0$ per ogni $x \in \mathbb{R}$. Infatti, se per qualche $x \in \mathbb{R}$ si avesse $P(\{\tilde{X} > x\}) = 0$, allora si avrebbe $P(\{\tilde{X} > y\}) = 0$ per ogni $y \geq x$ (legge L4) e la probabilità condizionata che si trova al primo membro dell'equazione (126) non sarebbe definita per infiniti numeri naturali x .

Consideriamo ora tra le equazioni (126) soltanto quelle con $h = 1$:

$$P(\{\tilde{X} > 1\}) = P(\{\tilde{X} > x + 1\} | \{\tilde{X} > x\}) \quad \text{per ogni } x = 0, 1, 2, \dots$$

Come si evince dalla catena di uguaglianze (125), queste equazioni possono essere espresse come

$$P(\{\tilde{X} > 1\}) = \frac{P(\{\tilde{X} > x + 1\})}{P(\{\tilde{X} > x\})} \quad \text{per ogni } x = 0, 1, 2, \dots$$

Moltiplicando le prime h di queste equazioni vediamo che

$$\prod_{x=0}^{h-1} P(\{\tilde{X} > 1\}) = \prod_{x=0}^{h-1} \frac{P(\{\tilde{X} > x + 1\})}{P(\{\tilde{X} > x\})} \quad \text{per } h = 1, 2, \dots \quad (128)$$

Scrivendo ora per esteso il prodotto che compare al secondo membro di quest'ultima equazione

$$\begin{aligned} \prod_{x=0}^{h-1} \frac{P(\{\tilde{X} > x + 1\})}{P(\{\tilde{X} > x\})} &= \frac{P(\{\tilde{X} > 0 + 1\})}{P(\{\tilde{X} > 0\})} \times \frac{P(\{\tilde{X} > 1 + 1\})}{P(\{\tilde{X} > 1\})} \times \dots \times \frac{P(\{\tilde{X} > h - 1 + 1\})}{P(\{\tilde{X} > h - 1\})} \\ &= \frac{P(\{\tilde{X} > 1\})}{P(\{\tilde{X} > 0\})} \times \frac{P(\{\tilde{X} > 2\})}{P(\{\tilde{X} > 1\})} \times \dots \times \frac{P(\{\tilde{X} > h\})}{P(\{\tilde{X} > h - 1\})} \end{aligned}$$

vediamo che il numeratore di ciascun fattore è uguale al denominatore del fattore successivo e quindi possiamo concludere che

$$\prod_{x=0}^{h-1} \frac{P(\{\tilde{X} > x + 1\})}{P(\{\tilde{X} > x\})} = \frac{P(\{\tilde{X} > h\})}{P(\{\tilde{X} > 0\})}. \quad (129)$$

D'altra parte, tutti i fattori della produttoria al primo membro dell'equazione (128) sono uguali a $P(\{\tilde{X} > 1\})$, e siccome il numero di fattori è $(h - 1) - 0 + 1 = h$, possiamo scrivere

$$\prod_{x=0}^{h-1} P(\{\tilde{X} > 1\}) = P(\{\tilde{X} > 1\})^h. \quad (130)$$

Si noti ora che secondo l'equazione (128), le espressioni che si trovano al secondo membro delle uguaglianze (129) e (130) devono essere uguali:

$$\frac{P(\{\tilde{X} > h\})}{P(\{\tilde{X} > 0\})} = P(\{\tilde{X} > 1\})^h \quad \text{per } h = 1, 2, \dots \quad (131)$$

Ora, per determinare il valore di $P(\{\tilde{X} > 0\})$ consideriamo una qualunque delle equazioni (126) con $h = 0$:

$$\begin{aligned} P(\{\tilde{X} > 0\}) &= P(\{\tilde{X} > x + 0\} | \{\tilde{X} > x\}) \quad \Rightarrow \\ \Rightarrow \quad [[\text{uguaglianze (125)}]] \quad &\Rightarrow \quad P(\{\tilde{X} > 0\}) = \frac{P(\{\tilde{X} > x + 0\})}{P(\{\tilde{X} > x\})} = 1. \end{aligned}$$

Sostituendo $P(\{\tilde{X} > 0\}) = 1$ nelle equazioni (131), e tenendo presente che $P(\{\tilde{X} > 1\})^{h=0} = 1$, vediamo dunque che

$$P(\{\tilde{X} > h\}) = P(\{\tilde{X} > 1\})^h \quad \text{per } h = 0, 1, 2, \dots$$

Ponendo $p = P(\{\tilde{X} \leq 1\})$ possiamo riscrivere la precedente formula come

$$P(\{\tilde{X} > h\}) = [[\text{legge L3}]] = (1-p)^h \quad \text{per } h = 0, 1, 2, \dots \quad (132)$$

A questo punto la dimostrazione è quasi completa. Infatti, la formula (132) è identica alla formula (122) che esprime le probabilità retrocumulate di una distribuzione geometrica. Tuttavia, per completare la dimostrazione dobbiamo ancora dimostrare che il valore di p , che in questa dimostrazione abbiamo definito come $p = P(\{\tilde{X} \leq 1\})$, sia strettamente compreso tra 0 e 1. Chiaramente, in virtù della legge L4, il valore di $p = P(\{\tilde{X} \leq 1\})$ deve appartenere all'intervallo $[0, 1]$. Per dimostrare che il valore di $p = P(\{\tilde{X} \leq 1\})$ non può essere uguale a 1, ricordiamo che all'inizio di questa dimostrazione avevamo osservato che

$$P(\{\tilde{X} > x\}) > 0 \quad \text{per ogni } x \in \mathbb{R},$$

e da questa condizione deduciamo che

$$p = P(\{\tilde{X} \leq 1\}) = 1 - P(\{\tilde{X} > 1\}) < 1.$$

Rimane dunque soltanto più da dimostrare che il valore di $p = P(\{\tilde{X} \leq 1\})$ non può essere nullo. A tal fine conviene fare un ragionamento per assurdo. Se il valore di $p = P(\{\tilde{X} \leq 1\})$ fosse nullo, si avrebbe

$$\begin{aligned} \lim_{h \rightarrow \infty} F(h) &= \lim_{h \rightarrow \infty} P(\{\tilde{X} \leq h\}) = \lim_{h \rightarrow \infty} [1 - P(\{\tilde{X} > h\})] \\ &= \lim_{h \rightarrow \infty} [1 - (1-p)^h] = \lim_{h \rightarrow \infty} [1 - (1-0)^h] = 0 \end{aligned}$$

e questo risultato violerebbe la proprietà F3 che tutte le funzioni di ripartizione devono soddisfare. Il valore di $p = P(\{\tilde{X} \leq 1\})$ che compare nelle equazioni (132) deve quindi essere strettamente compreso tra 0 e 1, e le equazioni (132) ci dicono dunque che le probabilità retrocumulate della distribuzione di \tilde{X} sono quelle di una distribuzione geometrica. Ovviamente, questo significa che la distribuzione di \tilde{X} deve essere geometrica e questa constatazione completa la dimostrazione. \square

Funzione generatrice dei momenti, valore atteso e varianza. Siccome il supporto delle distribuzioni geometriche è illimitato, non è detto che queste distribuzioni siano dotate di fgm. Tuttavia, un semplice calcolo mostra che

$$\begin{aligned} \sum_{x=1}^{\infty} e^{tx} (1-p)^{x-1} p &= p e^t \sum_{x=1}^{\infty} e^{t(x-1)} (1-p)^{x-1} = \\ &= [[\text{serie geometrica}]] = p e^t \sum_{x=1}^{\infty} [e^t (1-p)]^{x-1} \\ &= p e^t \times \frac{1}{1 - e^t (1-p)} = \frac{p}{e^{-t} - (1-p)} \end{aligned}$$

per ogni t tale che $e^t(1-p) \in (-1, 1)$ (si noti che $e^t(1-p)$ è la ragione della serie geometrica). Siccome $e^t(1-p) > 0$ e

$$e^t(1-p) < 1 \quad \Leftrightarrow \quad t < -\ln(1-p),$$

e siccome $-\ln(1-p) > 0$ (si ricordi che il logaritmo naturale di un numero compreso tra 0 e 1 è negativo), possiamo concludere che tutte le distribuzioni geometriche sono dotate di fgm e che quest'ultima è data da

$$m(t) = \frac{p}{e^{-t} - (1-p)} \quad \text{per } t < -\ln(1-p).$$

Calcolando le derivate

$$m'(t) = \frac{pe^{-t}}{[e^{-t} - (1-p)]^2}$$

e

$$m''(t) = -\frac{pe^{-t}}{[e^{-t} - (1-p)]^2} + 2 \times \frac{pe^{-2t}}{[e^{-t} - (1-p)]^3}$$

e ponendo $t = 0$, otteniamo (Teorema 7.1)

$$\mu = \mu_1 = E(\tilde{X}) = \frac{1}{p} \quad \text{e} \quad \mu_2 = E(\tilde{X}^2) = -\frac{1}{p} + \frac{2}{p^2}.$$

Il valore atteso e la varianza di una distribuzione geometrica sono quindi dati da

$$\mu = E(\tilde{X}) = \frac{1}{p}$$

e

$$\sigma^2 = \text{var}(\tilde{X}) = E(\tilde{X}^2) - [E(\tilde{X})]^2 = -\frac{1}{p} + \frac{2}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}.$$

Il ruolo delle distribuzioni geometriche nei processi di Bernoulli. Consideriamo ancora un **processi di Bernoulli** che consiste in una successione infinita di prove. Come abbiamo visto all'inizio di questa sezione, nell'ambito di questo tipo di esperimento casuale, la variabile casuale \tilde{X}_1 che conta il numero di prove necessarie per ottenere un successo ha distribuzione geometrica di parametro p uguale alla probabilità di successo comune alle singole prove. Tuttavia, non abbiamo ancora detto nulla sulle distribuzioni delle variabili casuali $\tilde{X}_2, \tilde{X}_3, \dots$ che contano, rispettivamente, il numero di prove successive al primo successo fino ad arrivare al secondo successo (la variabile casuale \tilde{X}_2), il numero di prove successive al secondo successo fino ad arrivare al terzo successo (la variabile casuale \tilde{X}_3), ecc.. Per completare la descrizione di un processo di Bernoulli dobbiamo quindi ancora chiarire ...

- quale sia la distribuzione marginale delle variabili casuali $\tilde{X}_2, \tilde{X}_3, \dots$
- e quale sia la distribuzione congiunta di una collezione finita di variabili casuali \tilde{X}_i .

Il prossimo teorema contiene le risposte a queste domande.

Teorema 8.2 (Tempi d'attesa nei processi di Bernoulli). Nell'ambito di un **processo di Bernoulli** che consiste in una successione infinita di prove, le variabili casuali $\tilde{X}_1, \tilde{X}_2, \dots$ che contano le prove tra successi consecutivi sono i.i.d. con distribuzione geometrica di parametro p uguale alla probabilità di successo delle singole prove.

Viceversa, se $\tilde{X}_1, \tilde{X}_2, \dots$ sono delle variabili casuali che contano le prove tra successi consecutivi nell'ambito di una successione infinita di prove (di cui a priori non sappiamo se è **processo di Bernoulli** o meno), e se le variabili casuali $\tilde{X}_1, \tilde{X}_2, \dots$ sono

i.i.d. con distribuzione geometrica di parametro p , allora la successione infinita di prove deve essere necessariamente un **processo di Bernoulli** e la probabilità di successo delle singole prove deve essere uguale a p .

Esercizio 8.9 (Distribuzione geometrica). Si consideri una successione di lanci di un dado regolare a sei facce.

- a) Si calcoli la probabilità di ottenere per la prima volta il punteggio sei al terzo lancio.
- b) Si calcoli la probabilità di dover effettuare almeno tre lanci per ottenere il punteggio sei.
- c) Si calcolino il valore atteso e la varianza della variabile casuale \tilde{X} che descrive il numero di lanci necessari per ottenere per la prima volta il punteggio sei.
- d) Sapendo che non si è mai ottenuto il punteggio sei nei primi 100 lanci, qual è la probabilità condizionata di ottenere il punteggio sei nel lancio numero 101?
- e) Sapendo che non si è mai ottenuto il punteggio sei nei primi 100 lanci, qual è la probabilità condizionata di non ottenere il punteggio sei nemmeno nei cinque lanci successivi?
- f) Qual è la probabilità di ottenere due volte il punteggio sei nei primi cinque lanci e che il primo sei successivo al quinto lancio si verifichi al decimo lancio?

Risposte:

- a) Si calcoli la probabilità di ottenere per la prima volta il punteggio sei al terzo lancio.

Per rispondere al quesito considereremo la variabile casuale \tilde{X} che conta il numero di lanci necessari per ottenere per la prima volta il punteggio sei. Secondo quanto abbiamo visto in questa sezione, la distribuzione di questa variabile casuale dovrebbe essere geometrica con parametro $p = 1/6$. Basandoci su questa ipotesi, possiamo concludere che la probabilità di ottenere per la prima volta il punteggio sei a terzo lancio sia data da

$$P(\{\tilde{X} = 3\}) = \left(1 - \frac{1}{6}\right)^{3-1} \times \left(\frac{1}{6}\right) = 0,116.$$

- b) Si calcoli la probabilità di dover effettuare almeno tre lanci per ottenere il punteggio sei.

Per rispondere al quesito possiamo considerare ancora la variabile casuale \tilde{X} che conta il numero di lanci necessari per ottenere per la prima volta il punteggio sei. Infatti, l'evento di cui è richiesta la probabilità può essere espresso come $\{\tilde{X} \geq 3\}$ e la probabilità di questo evento è data da

$$\begin{aligned} P(\{\tilde{X} \geq 3\}) &= 1 - P(\{\tilde{X} < 3\}) = 1 - p(1) - p(2) = \\ &= 1 - \left(1 - \frac{1}{6}\right)^{1-1} \times \left(\frac{1}{6}\right) - \left(1 - \frac{1}{6}\right)^{2-1} \times \left(\frac{1}{6}\right) \\ &= 1 - 0,167 - 0,139 = 0,694. \end{aligned}$$

In alternativa, possiamo anche rispondere osservando che l'evento $\{\tilde{X} \geq 3\}$ si verifica se e solo se in entrambi i primi due lanci si ottiene un punteggio diverso da sei. Siccome gli eventi che si riferiscono a lanci diversi sono indipendenti, possiamo concludere che la probabilità richiesta sia data da

$$P(\bar{S}_1 \cap \bar{S}_2) = P(\bar{S}_1) \times P(\bar{S}_2) = \left(1 - \frac{1}{6}\right) \times \left(1 - \frac{1}{6}\right) = \left(1 - \frac{1}{6}\right)^2 = 0,694$$

come già visto in precedenza.

- c) Si calcolino il valore atteso e la varianza della variabile casuale \tilde{X} che descrive il numero di lanci necessari per ottenere per la prima volta il punteggio sei.

Usando le formule per calcolare il valore e la varianza di una distribuzione geometrica otteniamo

$$E(\tilde{X}) = \frac{1}{1/6} = 6$$

e

$$Var(\tilde{X}) = \frac{1 - \frac{1}{6}}{\left(\frac{1}{6}\right)^2} = 30.$$

- d) Sapendo che non si è mai ottenuto il punteggio sei nei primi 100 lanci, qual è la probabilità condizionata di ottenere il punteggio sei nel lancio numero 101?

Dalla proprietà di assenza di memoria discende che (vedi la formula (124))

$$P(\{\tilde{X} = 101\} | \{\tilde{X} \geq 101\}) = p = \frac{1}{6}.$$

- e) Sapendo che non si è mai ottenuto il punteggio sei nei primi 100 lanci, qual è la probabilità condizionata di non ottenere il punteggio sei nemmeno nei cinque lanci successivi?

Dalla proprietà di assenza di memoria (vedi le condizioni (126)) discende che

$$\begin{aligned} P(\{\tilde{X} > 100 + 5\} | \{\tilde{X} > 100\}) &= P(\{\tilde{X} > 5\}) \\ &= \left(1 - \frac{1}{6}\right)^5 = 0,402. \end{aligned}$$

- f) Qual è la probabilità di ottenere due volte il punteggio sei nei primi cinque lanci e che il primo sei successivo al quinto lancio si verifichi al decimo lancio?

Per rispondere al quesito conviene considerare due variabili casuali:

- la variabile casuale \tilde{X} che conta il numero di volte che si ottiene sei nei primi cinque lanci
- e la variabile casuale \tilde{Y} che, a partire dal quinto lancio, conta il numero di lanci necessari per ottenere un sei.

Siccome le variabili casuali \tilde{X} e \tilde{Y} si riferiscono ai lanci diversi, possiamo assumere che siano indipendenti; per quanto abbiamo visto sui processi di Bernoulli, possiamo inoltre assumere che

$$\tilde{X} \sim \text{Binomiale}(n = 5; p = 1/6) \quad \text{e} \quad \tilde{Y} \sim \text{Geometrica}(p = 1/6).$$

Basandoci su queste ipotesi, possiamo ora calcolare la probabilità richiesta come

$$\begin{aligned} P(\{\tilde{X} = 2\} \cap \{\tilde{Y} = 10 - 5\}) &= \\ &= P(\{\tilde{X} = 2\} \cap \{\tilde{Y} = 5\}) \\ &= P(\{\tilde{X} = 2\}) \times P(\{\tilde{Y} = 5\}) \\ &= \binom{5}{2} \left(\frac{1}{6}\right)^2 \left(1 - \frac{1}{6}\right)^{5-2} \times \left(1 - \frac{1}{6}\right)^{5-1} \frac{1}{6} \\ &= 0,161 \times 0,080 = 0,013. \end{aligned}$$

Esercizio 8.10. Secondo un ingegnere, a fronte di regolari interventi di manutenzione la durata (espressa in anni) di un macchinario può essere descritta attraverso una variabile casuale con distribuzione geometrica di valore atteso $\mu = 10$ anni.

- a) Qual è la probabilità che il macchinario si guasti durante i primi cinque anni di utilizzo?
- b) Sapendo che il macchinario non si è guastato nei primi dieci anni, qual è la probabilità condizionata che non si guasti nemmeno nei cinque anni successivi?

- c) Qual è la durata massima entro la quale il 95% dei macchinari del tipo considerato si guasta? (ipotizzando regolari interventi di manutenzione)

Risposte:

- a) Qual è la probabilità che il macchinario si guasti durante i primi cinque anni di utilizzo?

Indichiamo con \tilde{X} la variabile casuale che descrive la durata del macchinario. Secondo l'ipotesi nella consegna si ha

$$\tilde{X} \sim \text{Geometrica}(p) \quad \text{con} \quad E(\tilde{X}) = 10.$$

Siccome il valore atteso di una distribuzione geometrica è il reciproco del valore del parametro p , possiamo concludere che

$$E(\tilde{X}) = \frac{1}{p} = 10 \quad \Rightarrow \quad p = \frac{1}{10} = 0,1 \quad \Rightarrow$$

$$\Rightarrow \quad P(\{\tilde{X} \leq 5\}) = [[\text{formula (123)}]] = 1 - (1 - 0,1)^5 = 0,410.$$

- b) Sapendo che il macchinario non si è guastato nei primi dieci anni, qual è la probabilità condizionata che non si guasti nemmeno nei cinque anni successivi?

Per rispondere dobbiamo calcolare la probabilità condizionata

$$P(\{\tilde{X} > 10 + 5\} | \{\tilde{X} > 10\}).$$

Siccome per ipotesi la distribuzione di \tilde{X} è geometrica, e siccome le distribuzioni geometriche soddisfano la proprietà di assenza di memoria, possiamo calcolare il valore di $P(\{\tilde{X} > 10 + 5\} | \{\tilde{X} > 10\})$ come (vedi le condizioni (126) che definiscono la proprietà dell'assenza di memoria)

$$\begin{aligned} P(\{\tilde{X} > 10 + 5\} | \{\tilde{X} > 10\}) &= [[\text{assenza di memoria}]] = P(\{\tilde{X} > 5\}) = \\ &= 1 - P(\tilde{X} \leq 5) = [[\text{risposta quesito a)}]] = 1 - 0,410 = 0,590. \end{aligned}$$

- c) Qual è la durata massima entro la quale il 95% dei macchinari del tipo considerato si guasta? (ipotizzando regolari interventi di manutenzione)

Per rispondere al quesito calcoleremo il 95-esimo percentile della distribuzione di \tilde{X} , ovvero il più piccolo valore di x dove le probabilità cumulate della distribuzione geometrica di \tilde{X} superano per la prima volta 0,95 (quindi ci aspettiamo che al più il 5% dei macchinari abbia una durata superiore a x).

Ricordando che le probabilità cumulate di una distribuzione geometrica di parametro p sono data da

$$P(\{\tilde{X} \leq x\}) = \text{[[formula (123)]]} = 1 - (1 - p)^x,$$

possiamo concludere che il più piccolo valore di x dove le probabilità cumulate superano 0,95 sia dato dal più piccolo numero intero x tale che

$$1 - (1 - p)^x \geq 0,95 \quad \Rightarrow \quad x \geq \frac{\ln(1 - 0,95)}{\ln(1 - p)} = \frac{\ln(1 - 0,95)}{\ln(1 - 0,1)} = 28,43$$

A fronte di regolari interventi di manutenzione la durata massima del 95% dei macchinari è all'incirca di 28 anni.

8.7 La famiglia delle distribuzioni binomiali negative

Le distribuzioni binomiali negative sono le distribuzioni di variabili casuali che nell'ambito di una successione di prove indipendenti eseguite sotto le medesime condizioni (ovvero nell'ambito di un [processo bernoulliano](#)) contano il numero di prove necessarie per ottenere un prefissato numero $r \geq 1$ di successi.

Ovviamente si tratta di distribuzioni discrete con supporto dato dall'insieme dei numeri naturali a partire da r (il numero di successi che si vuole raggiungere).

Non è difficile dimostrare che la [funzione di massa di probabilità](#) di una distribuzione binomiale negativa deve essere data da

$$p(x) = \begin{cases} \binom{x-1}{r-1} p^r (1-p)^{x-r} & \text{per } x = r, r+1, r+2, \dots; \\ 0 & \text{altrimenti,} \end{cases} \quad (133)$$

dove

- $p \in (0, 1)$ è la probabilità di successo comune a tutte le prove
- e $r = 1, 2, \dots$ è il prefissato numero di successi che si vuole raggiungere.

Si noti che con $r = 1$ la funzione di massa di probabilità binomiale negativa si riduce ad una funzione di massa di probabilità geometrica (infatti, in questo caso si tratterebbe della distribuzione di una variabile casuale che nell'ambito di un processo bernoulliano conta il numero di prove necessarie per ottenere il primo successo).

Dimostrazione. In questa dimostrazione dimostreremo che la funzione di massa di probabilità di una variabile casuale \tilde{X} che nell'ambito di un [processo di Bernoulli](#) conta il numero di prove necessarie per ottenere r successi deve essere definita come nella formula

(133). A tal fine è sufficiente osservare che il successo numero r si verifica nella prova numero x (assumiamo che $x \geq r$) se e solo se nelle prime $x - 1$ prove si verificano $r - 1$ successi e nella prova numero x si verifica l' r -esimo successo. Come abbiamo visto quando abbiamo ricavato la distribuzione binomiale (positiva), la probabilità di ottenere $r - 1$ successi in $x - 1$ prove è data da

$$\binom{x-1}{r-1} p^{r-1} (1-p)^{(x-1)-(r-1)} = \binom{x-1}{r-1} p^{r-1} (1-p)^{x-r}$$

e moltiplicando questa probabilità per la probabilità p di ottenere un successo nella prova numero x otteniamo il valore di $p(x) = P(\{\tilde{X} = x\})$ che coincide con $\binom{x-1}{r-1} p^r (1-p)^{x-r}$ come volevamo dimostrare (ricordiamo che le prove sono indipendenti). \square

Dalla precedente dimostrazione desumiamo che la probabilità di osservare l' r -esimo successo nella prova numero x è data dal prodotto tra

- la probabilità

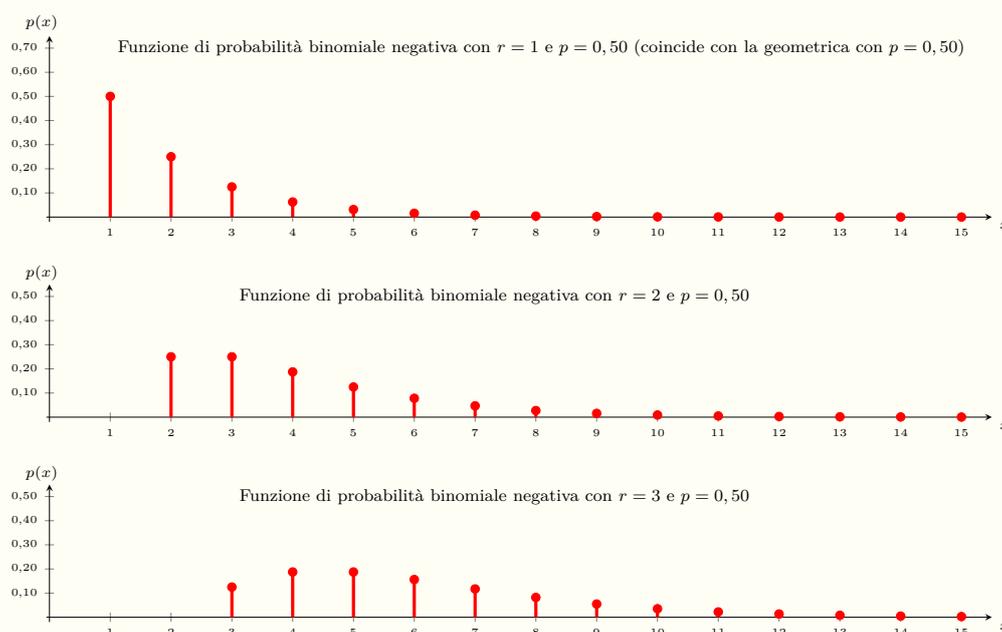
$$\binom{x-1}{r-1} p^{r-1} (1-p)^{(x-1)-(r-1)} = \binom{x-1}{r-1} p^{r-1} (1-p)^{x-r}$$

di ottenere $r - 1$ successi nelle prime $x - 1$ prove

- e la probabilità p di ottenere un successo nella prova numero x .

I grafici in Figura 8.5 mostrano alcune funzioni di massa di probabilità binomiali negative.

Figura 8.5. Funzioni di massa di probabilità binomiali negative.



Il nome ”**binomiale negativa**” delle distribuzioni discrete con funzione di massa di probabilità definita come nella formula (133) deriva dal fatto che le probabilità $p(r)$, $p(r+1)$, \dots sono i termini (traslati) dello sviluppo in serie di un binomio con esponente negativo:

$$(a+b)^{-r} = \sum_{y=0}^{\infty} \binom{y+r-1}{r-1} (-a)^y b^{-(y+r)} \quad \text{per } |a/b| < 1 \text{ e } r = 1, 2, \dots \quad (134)$$

(per $|a/b| \geq 1$ la serie al secondo membro è divergente oppure oscillante). Infatti, ponendo $a = -(1-p)/p$, $b = 1/p$ e $y = x - r$ possiamo facilmente verificare che

$$\binom{y+r-1}{r-1} (-a)^y b^{-(y+r)} = \binom{x-1}{r-1} p^r (1-p)^{x-r} = p(x) \quad \text{per } x = r, r+1, \dots$$

Quando d’ora in poi vorremo indicheremo che la distribuzione di una data variabile casuale \tilde{X} è binomiale negativa con parametri r e p , scriveremo semplicemente

$$\tilde{X} \sim \text{Bin.neg.}(r; p).$$

Probabilità cumulate. Le probabilità cumulate di una distribuzione binomiale negativa possono essere calcolate attraverso la formula

$$P(\{\tilde{X} \leq x\}) = \sum_{t=r}^x \binom{t-1}{r-1} p^r (1-p)^{t-r} = 1 - \sum_{y=0}^{r-1} \binom{x}{y} p^y (1-p)^{x-y}, \quad (135)$$

per $x = r, r+1, r+2, \dots$

Si noti, tuttavia, che l’utilizzo della formula sul lato destro dell’ultima uguaglianza è vantaggioso solo se $x > 2r - 1$. Infatti, per qualunque valore di $x = r, r+1, r+2, \dots$ la sommatoria che compare sul lato destro dell’ultima uguaglianza è sempre composta da r termini, mentre la sommatoria sul lato sinistro è composta da $x - r + 1$ termini. Quindi la sommatoria sul lato destro è composta da meno termini se e solo se

$$r < x - r + 1 \quad \Rightarrow \quad x > 2r - 1.$$

Dimostrazione (della formula (135)). La formula (135) può essere facilmente dimostrata facendo riferimento ad un [processo di Bernoulli](#). Infatti, in un [processo di Bernoulli](#) (e anche una qualunque successione di prove che possono avere soltanto due esiti) il successo numero r si verifica dopo la prova numero x se e solo se nelle prime x prove si sono verificati al più $r - 1$ successi. Se indichiamo con \tilde{X} la variabile casuale che conta il numero di prove necessarie per ottenere r successi, e con \tilde{Y}_x la variabile casuale che conta il numero di successi nelle prime x prove, possiamo dunque scrivere

$$\{\tilde{X} > x\} = \{\tilde{Y}_x \leq r - 1\}$$

Siccome sappiamo che nell’ambito di un processo di Bernoulli si deve avere

$$\tilde{X} \sim \text{Bin.neg.}(r; p) \quad \text{e} \quad \tilde{Y}_x \sim \text{Binomiale}(x; p),$$

possiamo concludere che

$$\begin{aligned} P(\{\tilde{X} > x\}) &= P(\{\tilde{Y}_x \leq r-1\}) = \\ &= \text{[[funzione di probabilità binomiale]]} = \sum_{y=0}^{r-1} \binom{x}{y} p^y (1-p)^{x-y} \end{aligned}$$

e applicando la legge L3 otteniamo quindi la formula (135). \square

Probabilità condizionate di sopravvivenza. Siccome le distribuzioni binomiali negative concentrano tutta la probabilità su sottoinsiemi dei numeri naturali, queste distribuzioni (o delle loro traslazioni) possono essere considerate come distribuzioni per variabili casuali \tilde{X} che rappresentano tempi d'attesa discreti. Come abbiamo già osservato, le distribuzioni binomiali negative con $r = 1$ sono distribuzioni geometriche e queste ultime soddisfano la [proprietà dell'assenza di memoria](#). Per variabili casuali \tilde{X} che hanno distribuzione binomiale negativa con $r = 1$ deve quindi essere soddisfatta la condizione (126) che [definisce](#) la [proprietà di assenza di memoria](#). Ricordiamo che la condizione in questione impone che

$$P(\{\tilde{X} > x+h\} | \{\tilde{X} > x\}) = P(\{\tilde{X} > h\}) \quad \text{per ogni } x, h = 0, 1, 2, \dots$$

Con riferimento a variabili casuali \tilde{X} che invece hanno distribuzione binomiale negativa di parametro $r > 1$, si può dimostrare che per qualsiasi prefissato valore di $h = 0, 1, 2, \dots$ la probabilità marginale

$$P(\{\tilde{X} > r-1+h\}) = \sum_{y=r-1+h+1}^{\infty} \binom{y-1}{r-1} p^r (1-p)^{y-r}$$

è sempre strettamente maggiore di tutte le [probabilità condizionate di sopravvivenza](#)

$$P(\{\tilde{X} > x+h\} | \{\tilde{X} > x\}) = \frac{\sum_{y=x+h+1}^{\infty} \binom{y-1}{r-1} p^r (1-p)^{y-r}}{\sum_{y=x+1}^{\infty} \binom{y-1}{r-1} p^r (1-p)^{y-r}}$$

con lo stesso valore di h e con $x = r, r+1, r+2, \dots$, e che queste ultime (a parità di h) decrescono all'aumentare di $x = r, r+1, r+2, \dots$. Questo significa che se un tempo d'attesa \tilde{X} ha distribuzione binomiale negativa con parametro $r > 1$, allora a partire dall'istante temporale $x = r$ le [probabilità condizionate di sopravvivenza](#) diminuiscono all'aumentare del tempo d'attesa x che è già trascorso.

Sempre con riferimento alle [probabilità condizionate di sopravvivenza](#), si può inoltre dimostrare che per tempi d'attesa \tilde{X} con distribuzione binomiale negativa di parametro $r > 1$ si ha

$$\lim_{x \rightarrow \infty} P(\{\tilde{X} > x+h\} | \{\tilde{X} > x\}) = (1-p)^h \quad \text{per ogni } h = 0, 1, 2, \dots$$

Da questo risultato si evince che le [probabilità condizionate di sopravvivenza](#) tendono ad un limite strettamente positivo quando il tempo d'attesa x che è già trascorso tende a infinito.

Rappresentazione come somma di variabili casuali geometriche indipendenti. Come abbiamo visto all'inizio di questo paragrafo, le variabili casuali che nell'ambito di un [processo di Bernoulli](#) contano le prove necessarie per ottenere un prefissato numero r di successi hanno distribuzione binomiale negativa. Chiaramente, una variabile casuale \tilde{Y} del suddetto tipo può essere rappresentata come somma delle r variabili casuali $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_r$ che contano, rispettivamente, il numero di prove fino ad arrivare al primo successo (la variabile casuale \tilde{X}_1), il numero di prove successive al primo successo fino ad arrivare al secondo successo (la variabile casuale \tilde{X}_2) e così via:

$$\tilde{Y} = \tilde{X}_1 + \tilde{X}_2 + \dots + \tilde{X}_r. \quad (136)$$

Nel Teorema 8.2 abbiamo visto che nell'ambito di un [processo di Bernoulli](#) le variabili casuali $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_r$ sono i.i.d. e che hanno distribuzione geometrica con parametro p uguale alla probabilità di successo comune a tutte le prove. Sfruttando le nostre conoscenze teoriche sulle distribuzioni geometriche possiamo quindi, attraverso la rappresentazione (136), dimostrare alcune proprietà che riguardano le distribuzioni binomiali negative. Nei prossimi paragrafi sfrutteremo questa possibilità.

Proprietà riproduttiva. Consideriamo ancora un [processo di Bernoulli](#) e la successione infinita di variabili casuali $\tilde{X}_1, \tilde{X}_2, \dots$ che contano le prove tra successi consecutivi. Come abbiamo osservato poc'anzi, queste variabili casuali sono i.i.d., hanno distribuzione geometrica e sommando le prime r_1 di esse si ottiene

$$\tilde{Y}_1 = \tilde{X}_1 + \tilde{X}_2 + \dots + \tilde{X}_{r_1} \sim \text{Bin.neg}(r_1; p).$$

Allo stesso modo, sommando le successive r_2 variabili casuali \tilde{X}_i si ottiene

$$\tilde{Y}_2 = \tilde{X}_{r_1+1} + \tilde{X}_{r_1+2} + \dots + \tilde{X}_{r_1+r_2} \sim \text{Bin.neg}(r_2; p).$$

Sommando ora le variabili casuali \tilde{Y}_1 e \tilde{Y}_2 otteniamo

$$\tilde{Y} = \tilde{Y}_1 + \tilde{Y}_2 = \tilde{X}_1 + \tilde{X}_2 + \dots + \tilde{X}_{r_1+r_2}$$

e quindi possiamo concludere che

$$\tilde{Y} = \tilde{Y}_1 + \tilde{Y}_2 \sim \text{Bin.neg}(r = r_1 + r_2; p).$$

Siccome le variabili casuali \tilde{Y}_1 e \tilde{Y}_2 sono indipendenti (si noti infatti che \tilde{Y}_1 e \tilde{Y}_2 sono somme di variabili casuali i.i.d. diverse e che quindi sono soddisfatte le condizioni del Teorema 4.10), questo ragionamento dimostra che sommando due variabili casuali indipendenti che hanno entrambe distribuzione binomiale negativa con lo stesso parametro p , si ottiene ancora una variabile casuale con distribuzione binomiale negativa dello stesso parametro p ma con parametro r uguale alla somma dei parametri r delle distribuzioni binomiali negative delle due variabili casuali che vengono sommate. Iterando questo

ragionamento si può dimostrare il seguente teorema che potrebbe, in alternativa, essere dimostrato attraverso la fgm che ricaveremo tra breve (per dimostrare la proprietà riproduttiva delle distribuzioni binomiali negative attraverso la fgm basta seguire la farsa della dimostrazione del Teorema 8.1 sulla proprietà riproduttiva delle distribuzioni binomiali).

Teorema 8.3 (Proprietà riproduttiva della famiglia delle distribuzioni binomiali negative). Se

$$\tilde{Y}_i \sim \text{Bin.neg}(r = r_i; p), \quad i = 1, 2, \dots, k,$$

sono k variabili casuali indipendenti (si noti che il valore del parametro p è uguale per tutte le k variabili casuali \tilde{Y}_i), allora

$$\tilde{Y} = \tilde{Y}_1 + \tilde{Y}_2 + \dots + \tilde{Y}_k \quad \sim \quad \text{Bin.neg}(r = r_1 + r_2 + \dots + r_k; p).$$

Funzione generatrice dei momenti. Siccome il supporto di una distribuzione binomiale negativa è sempre un insieme illimitato, a priori non possiamo essere sicuri che tutte le distribuzioni binomiali negative siano dotate di fgm. Tuttavia, non è difficile dimostrare che ciò è vero. A tal fine possiamo procedere in due modi:

- per via algebrica dimostrando che per ogni $r = 1, 2, \dots$, ogni $p \in (0, 1)$ e ogni $t < -\ln(1 - p)$ (si noti che $-\ln(1 - p) > 0$ se $p \in (0, 1)$) la serie

$$\sum_{x=r}^{\infty} e^{tx} p(x) = \sum_{x=r}^{\infty} e^{tx} \binom{x-1}{r-1} p^r (1-p)^{x-r}$$

è assolutamente convergente;

- oppure facendo riferimento alla rappresentazione di una variabile casuale binomiale negativa come somma di variabili casuali geometriche indipendenti e invocando il Teorema 7.3. Seguendo questa strada vediamo immediatamente che la fgm binomiale negativa è data da

$$m(t) = \prod_{i=1}^r \frac{p}{e^{-t} - (1-p)} = \frac{p^r}{[e^{-t} - (1-p)]^r} \quad \text{per } t < -\ln(1-p)$$

(si noti che tutti i fattori della produttoria sono uguali all'espressione della fgm geometrica).

Dimostrazione (Dimostrazione algebrica della formula della fgm binomiale negativa). Per ottenere la fgm per via algebrica osserviamo che

$$\begin{aligned} \sum_{x=r}^{\infty} e^{tx} p(x) &= \sum_{x=r}^{\infty} e^{tx} \binom{x-1}{r-1} p^r (1-p)^{x-r} \\ \text{[[cambio variabile: } y = x - r \text{]]} &= \sum_{y=0}^{\infty} e^{t(y+r)} \binom{y+r-1}{r-1} p^r (1-p)^y \\ &= p^r e^{tr} \sum_{y=0}^{\infty} \binom{y+r-1}{r-1} [e^t(1-p)]^y \end{aligned}$$

e che la serie nell'ultima riga è una serie binomiale negativa con $a = -e^t(1-p)$, $b = 1$ e $n = r$ (si confronti con la formula (134)). Ricordando quanto abbiamo detto sulle serie binomiali negative, possiamo concludere che la serie nell'ultima riga della precedente catena di uguaglianze è convergente se e solo se

$$\left| \frac{a}{b} \right| = e^t(1-p) < 1 \quad \Leftrightarrow \quad t < -\ln(1-p)$$

e che in tal caso il limite della serie è dato da

$$\sum_{y=0}^{\infty} \binom{y+r-1}{r-1} (-a)^y b^{-(y+r)} = (a+b)^{-r} = [-e^t(1-p) + 1]^{-r}.$$

La fgm binomiale negativa è quindi definita come

$$m(t) = \frac{p^r e^{tr}}{[1 - e^t(1-p)]^r} = \frac{p^r}{[e^{-t} - (1-p)]^r}, \quad \text{per } t < -\ln(1-p)$$

così come avevamo già dimostrato facendo riferimento alla rappresentazione di una variabile casuale binomiale negativa come somma di variabili casuali geometriche indipendenti. \square

Valore atteso e varianza. Rappresentando una variabile casuale binomiale negativa come somma di variabili casuali geometriche indipendenti, si vede facilmente che

$$\begin{aligned} E(\tilde{X}) &= E(\tilde{X}_1 + \tilde{X}_2 + \cdots + \tilde{X}_r) \\ \text{[[proprietà E4]]} &= E(\tilde{X}_1) + E(\tilde{X}_2) + \cdots + E(\tilde{X}_r) \\ &= \frac{1}{p} + \frac{1}{p} + \cdots \text{ [[} r \text{ volte]]} + \frac{1}{p} = \frac{r}{p} \end{aligned}$$

e che

$$\begin{aligned} \text{var}(\tilde{X}) &= \text{var}(\tilde{X}_1 + \tilde{X}_2 + \cdots + \tilde{X}_r) \\ \text{[[proprietà V8*]]} &= \text{var}(\tilde{X}_1) + \text{var}(\tilde{X}_2) + \cdots + \text{var}(\tilde{X}_r) \\ &= \frac{1-p}{p^2} + \frac{1-p}{p^2} + \cdots \text{ [[} r \text{ volte]]} + \frac{1-p}{p^2} = r \times \frac{1-p}{p^2}. \end{aligned}$$

Il valore atteso e la varianza di una distribuzione binomiale negativa sono dunque dati da

$$\mu = E(\tilde{X}) = \frac{r}{p} \quad \text{e da } \sigma^2 = \text{var}(\tilde{X}) = r \times \frac{1-p}{p^2}.$$

Ovviamente, questi risultati possono essere ottenuti anche attraverso le derivate della fgm (invitiamo il lettore a verificare).

Esercizio 8.11. Si consideri una successione di lanci indipendenti di un dado regolare e sia \tilde{X} la variabile casuale che conta il numero di lanci necessari per ottenere 3 volte il punteggio sei.

- Qual è la probabilità che \tilde{X} sia pari a 10?
- Si calcolino il valore atteso e la varianza di \tilde{X} .
- Sapendo che nei primi quattro lanci si è ottenuto 2 volte il punteggio sei, qual è la probabilità condizionata che siano necessari ancora più di due lanci per ottenere per la terza volta il punteggio sei?

Risposte:

- Qual è la probabilità che \tilde{X} sia pari a 10?

Dalla definizione della variabile casuale \tilde{X} desumiamo che

$$\tilde{X} \sim \text{Bin.neg}(r = 3; p = 1/6).$$

La probabilità richiesta è quindi data da

$$\begin{aligned} P(\{\tilde{X} = 10\}) &= \binom{10-1}{3-1} \left(1 - \frac{1}{6}\right)^{10-3} \left(\frac{1}{6}\right)^3 \\ &= 0,047. \end{aligned}$$

- Si calcolino il valore atteso e la varianza di \tilde{X} .

Usando le espressioni per il valore atteso e la varianza di distribuzioni binomiali negative otteniamo

$$E(\tilde{X}) = \frac{3}{1/6} = 18$$

e

$$\text{Var}(\tilde{X}) = 3 \times \frac{1 - \frac{1}{6}}{\left(\frac{1}{6}\right)^2} = 90.$$

- Sapendo che nei primi quattro lanci si è ottenuto 2 volte il punteggio sei, qual è la probabilità condizionata che siano necessari ancora più di due lanci per ottenere per la terza volta il punteggio sei?

Siccome gli eventi che si riferiscono a lanci diversi sono indipendenti, possiamo concludere che la probabilità richiesta coincide con la probabilità di dover effettuare più di due lanci per ottenere il punteggio sei, ovvero con la probabilità di non ottenere il punteggio 6 in due lanci di un dado. La probabilità condizionata richiesta è dunque data da

$$\left(1 - \frac{1}{6}\right)^2 = 0,694.$$

Esercizio 8.12. Un macchinario contiene tre componenti che entrano in funzione uno dopo l'altro quando quello precedente si guasta. Il macchinario cessa di funzionare quando l'ultimo dei tre componenti si guasta. La durata di ciascuno dei tre componenti è descritta da una variabile casuale geometrica con valore atteso $\mu = 4$ ore.

- Qual è la probabilità che la durata di un singolo componente sia maggiore di 5 ore?
- Sapendo che uno dei tre componenti è già in funzione da 6 ore, qual è la probabilità condizionata che tra 5 ore lo stesso componente sia ancora in funzione?
- Qual è la probabilità che tra 5 ore il macchinario sia ancora in funzione?
- Sapendo che il macchinario è già in funzione da 6 ore, qual è la probabilità condizionata che tra 5 ore sia ancora in funzione?
- Si calcoli la varianza della durata di un singolo componente.
- Si calcolino valore atteso e varianza della durata del macchinario.
- Si calcoli il decimo percentile della durata del macchinario.

Risposte:

- Qual è la probabilità che la durata di un singolo componente sia maggiore di 5 ore?

Siano \tilde{X}_1, \tilde{X}_2 e \tilde{X}_3 le variabili casuali che descrivono le durate dei tre componenti. Siccome per ipotesi

$$\tilde{X}_i \sim \text{Geometrica}(p=?) \quad \text{e} \quad E(\tilde{X}_i) = \mu = 1/p = 4 \text{ ore,}$$

possiamo concludere che $p = 1/4 = 0,25$ e che

$$\begin{aligned} P(\{\tilde{X}_i > 5\}) &= [\text{probabilità di 5 "insuccessi" consecutivi}] \\ &= (1 - 0,25)^5 = 0,237. \end{aligned}$$

- Sapendo che uno dei tre componenti è già in funzione da 6 ore, qual è la probabilità condizionata che tra 5 ore lo stesso componente sia ancora in funzione?

Dalla [proprietà di assenza di memoria](#) della distribuzione geometrica discende che

$$P(\{\tilde{X}_i > 6 + 5\} | \{\tilde{X}_i > 6\}) = P(\{\tilde{X}_i > 5\}) = 0,237$$

(lo stesso risultato del quesito precedente).

c) Qual è la probabilità che tra 5 ore il macchinario sia ancora in funzione?

Sia \tilde{X} la variabile casuale che rappresenta la durata del macchinario. Per ipotesi $\tilde{X} = \tilde{X}_1 + \tilde{X}_2 + \tilde{X}_3$.

Siccome sembra ragionevole che le durate dei tre componenti non si influenzino a vicenda, possiamo assumere che le variabili casuali \tilde{X}_1 , \tilde{X}_2 e \tilde{X}_3 siano indipendenti, e siccome

$$\tilde{X}_i \sim Geometrica(p = 1/\mu = 1/4 = 0,25), \quad i = 1, 2, 3,$$

(vedi la risposta al quesito a)), possiamo concludere che

$$\tilde{X} = \tilde{X}_1 + \tilde{X}_2 + \tilde{X}_3 \sim Bin.neg(r = 3; p = 0,25).$$

La probabilità richiesta è dunque data da

$$\begin{aligned} P(\{\tilde{X} > 5\}) &= 1 - P(\{\tilde{X} \leq 5\}) \\ &= 1 - \sum_{x=3}^5 P(\{\tilde{X} = x\}) \\ &= 1 - \binom{3-1}{3-1} 0,25^3 (1-0,25)^{3-3} + \\ &\quad - \binom{4-1}{3-1} 0,25^3 (1-0,25)^{4-3} \\ &\quad - \binom{5-1}{3-1} 0,25^3 (1-0,25)^{5-3} \\ &= 1 - 0,016 - 0,035 - 0,053 = 0,896. \end{aligned}$$

d) Sapendo che il macchinario è già in funzione da 6 ore, qual è la probabilità condizionata che tra 5 ore sia ancora in funzione?

Per rispondere dobbiamo calcolare la probabilità condizionata

$$P(\{\tilde{X} > 6 + 5\} | \{\tilde{X} > 6\}).$$

Siccome le distribuzioni binomiali negative con $r > 1$ non soddisfano la **proprietà di assenza di memoria**, non possiamo concludere che la probabilità condizionata richiesta sia data da

$$P(\{\tilde{X} > 5\}) = 0,896 \quad [[\text{risposta del quesito c) }]] \quad (137)$$

ma dobbiamo calcolarla *ex novo*:

$$\begin{aligned} P(\{\tilde{X} > 6 + 5\} | \{\tilde{X} > 6\}) &= \frac{P(\{\tilde{X} > 6 + 5\} \cap \{\tilde{X} > 6\})}{P(\{\tilde{X} > 6\})} \\ &= \frac{P(\{\tilde{X} > 6 + 5\})}{P(\{\tilde{X} > 6\})}. \end{aligned}$$

La probabilità al denominatore può essere calcolata come

$$\begin{aligned}
 P(\{\tilde{X} > 6\}) &= P(\{\tilde{X} > 5\}) - P(\{\tilde{X} = 6\}) \\
 [[\text{Risultato (137)}]] &= 0,896 - \binom{6-1}{3-1} 0,25^3 (1-0,25)^{6-3} \\
 &= 0,896 - 0,066 = 0,830.
 \end{aligned}$$

e ragionando in modo analogo otteniamo anche la probabilità al numeratore:

$$\begin{aligned}
 P(\{\tilde{X} > 6 + 5\}) &= P(\{\tilde{X} > 11\}) = P(\{\tilde{X} > 6\}) - \sum_{x=7}^{11} P(\{\tilde{X} = x\}) \\
 [[\text{Risultato precedente}]] &= 0,830 - \sum_{x=7}^{11} \binom{x-1}{3-1} 0,25^3 (1-0,25)^{x-3} \\
 &= 0,830 - 0,104 - 0,104 - 0,100 - 0,094 - 0,086 \\
 &= 0,342.
 \end{aligned}$$

La probabilità condizionata richiesta è dunque data da

$$P(\{\tilde{X} > 6 + 5\} | \{\tilde{X} > 6\}) = \frac{P(\{\tilde{X} > 6 + 5\})}{P(\{\tilde{X} > 6\})} = \frac{0,342}{0,830} = 0,412.$$

Questo risultato è la probabilità che dopo 6 ore di funzionamento il macchinario rimanga in funzione per ulteriori 5 ore oppure più a lungo.

Come ci aspettavamo in virtù del fatto che per $x \geq r$ le [probabilità condizionate di sopravvivenza](#) di una distribuzione binomiale negativa con $r > 1$ sono strettamente decrescenti, il valore di

$$P(\{\tilde{X} > 6 + 5\} | \{\tilde{X} > 6\}) = 0,412$$

è strettamente minore del valore di

$$P(\{\tilde{X} > 5\}) = 0,896 \quad [[\text{risposta del quesito c)}]]$$

(si noti che $P(\{\tilde{X} > 5\})$ è la probabilità che a partire dal momento della sua prima accensione il macchinario rimanga in funzione per 5 ore oppure più a lungo).

- e) Si calcoli la varianza della durata di un singolo componente.

Siccome

$$\tilde{X}_i \sim \text{Geometrica}(p = 1/\mu = 1/4 = 0,25), \quad i = 1, 2, 3,$$

(vedi la risposta al quesito a)), possiamo concludere che

$$\text{var}(\tilde{X}_i) = \frac{1-p}{p^2} = \frac{1-0,25}{0,25^2} = 12.$$

f) Si calcolino valore atteso e varianza della durata del macchinario.

Siccome

$$\tilde{X} = \tilde{X}_1 + \tilde{X}_2 + \tilde{X}_3 \sim \text{Bin.neg}(r = 3; p = 0,25),$$

possiamo concludere che

$$E(\tilde{X}) = \frac{r}{p} = \frac{3}{0,25} = 12$$

e che

$$\text{var}(\tilde{X}) = \frac{r(1-p)}{p^2} = \frac{3(1-0,25)}{0,25^2} = 36.$$

g) Si calcoli il decimo percentile della durata del macchinario.

Per rispondere al quesito dobbiamo cumulare le probabilità della distribuzione $\text{Bin.neg}(r = 3; p = 0,25)$ (ovvero della distribuzione della variabile casuale che rappresenta la durata del macchinario) fino al raggiungimento del valore 0,10.

Usando i risultati della risposta al quesito c) otteniamo

$$p(3) = P(\{\tilde{X} = 3\}) = 0,016 \Rightarrow F(3) = \sum_{x=3}^3 p(x) = 0,016 < 0,10,$$

$$p(4) = P(\{\tilde{X} = 4\}) = 0,035 \Rightarrow F(4) = \sum_{x=3}^4 p(x) = 0,051 < 0,10,$$

$$p(5) = P(\{\tilde{X} = 5\}) = 0,053 \Rightarrow F(5) = \sum_{x=3}^5 p(x) = 0,104 \geq 0,10.$$

Il più piccolo valore di x dove le probabilità cumulate superano 0,10 è $x = x_{0,10} = 5$ e questo valore è quindi il decimo percentile della distribuzione della durata del macchinario. Il valore di $x_{0,10} = 5$ ci dice che con probabilità almeno pari a 0,10 la durata del macchinario non sarà superiore a 5 ore.

8.8 La famiglia delle distribuzioni di Poisson

La famiglia delle distribuzioni di Poisson prende il suo nome dal matematico, fisico, astronomo e statistico francese Siméon-Denis Poisson (1781 - 1840). Si tratta di una famiglia di distribuzioni discrete che hanno per supporto l'insieme dei numeri naturali (compreso lo zero). Di solito le distribuzioni di Poisson vengono ricavate come

- approssimazioni a distribuzioni binomiali con parametro n elevato e parametro p prossimo a zero
- oppure come distribuzioni di variabili casuali che contano "successi" nell'ambito di un **processo di Poisson**.

Tra breve descriveremo entrambe queste possibilità.

La funzione di probabilità di una distribuzione di Poisson è data da

$$p(x) = \begin{cases} \frac{\lambda^x e^{-\lambda}}{x!} & \text{per } x = 0, 1, 2, \dots \\ 0 & \text{altrimenti} \end{cases}$$

dove λ è un parametro che deve essere un numero reale positivo. Come vedremo tra breve, il parametro λ che identifica una distribuzione di Poisson esprime il valore atteso e la varianza della distribuzione.

Notiamo immediatamente che le probabilità $p(x)$ di una distribuzione di Poisson sono i termini (opportunitamente normalizzati) della serie di potenze che definisce la funzione esponenziale:

$$e^\lambda = \sum_{x=0}^{\infty} \frac{\lambda^x}{x!}, \quad \lambda \in \mathbb{R} \quad \Rightarrow \quad p(x) = \frac{\lambda^x}{x! e^\lambda} = \frac{\lambda^x e^{-\lambda}}{x!} \quad \text{per } x = 0, 1, 2, \dots$$

D'ora in poi indicheremo le distribuzioni di Poisson con $Poisson(\lambda)$, e per indicare che la distribuzione di una variabile casuale \tilde{X} è di Poisson con parametro λ , scriveremo semplicemente

$$\tilde{X} \sim Poisson(\lambda).$$

Le distribuzioni di Poisson come approssimazioni di distribuzioni binomiali.

Come abbiamo già accennato, le distribuzioni di Poisson possono essere ricavate come approssimazioni a distribuzioni binomiali con parametro n elevato e parametro p prossimo a zero. Infatti, si può dimostrare che (omettiamo la dimostrazione)

$$\frac{e^{-\lambda} \lambda^x}{x!} = \lim_{n \rightarrow \infty} \binom{n}{x} (\lambda/n)^x \times (1 - \lambda/n)^{n-x} \quad \text{per ogni } x = 0, 1, 2, \dots \quad (138)$$

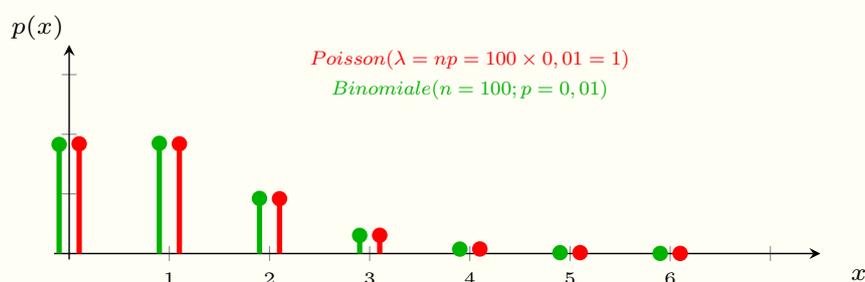
Siccome una distribuzione binomiale con n elevato e $p = \lambda/n$ prossimo a zero è la distribuzione di una variabile casuale che conta il numero di successi nell'ambito di un

esperimento binomiale dove la probabilità di successo comune a tutte le prove è prossima a zero, le distribuzioni di Poisson vengono spesso chiamate **"leggi degli eventi rari"**.

Per illustrare l'approssimazione di Poisson alle distribuzioni binomiali con n elevato e p prossimo a zero, possiamo considerare la distribuzione *Binomiale*($n = 100; p = 0,01$). Dalla formula (138) si evince che la funzione di massa di probabilità di questa distribuzione binomiale dovrebbe essere prossima alla funzione di massa di probabilità della distribuzione di Poisson con parametro $\lambda = np = 100 \times 0,01 = 1$. La tabella e il grafico in Figura 8.6 confermano che l'approssimazione è molto precisa.

Figura 8.6. La tabella e il grafico sottostanti illustrano l'approssimazione di Poisson a una distribuzione binomiale con n elevato e p prossimo a zero.

x	<i>Binomiale</i> ($n = 100; p = 0.01$) $p(x) = \binom{100}{x} 0.01^x 0.99^{100-x}$	<i>Poisson</i> ($\lambda = np = 1$) $p(x) = \frac{1^x e^{-1}}{x!}$
0	0,3660	0,3679
1	0,3697	0,3679
2	0,1849	0,1839
3	0,0610	0,0613
4	0,0149	0,0153
5	0,0029	0,0031
6	0,0005	0,0005



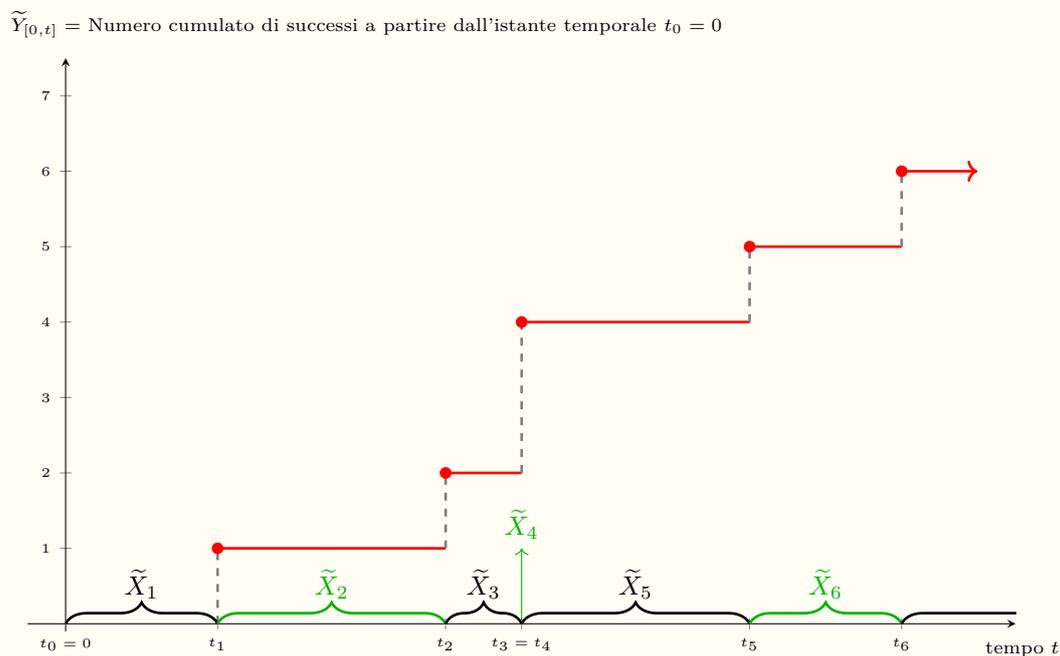
Processi di Poisson. Come abbiamo già accennato, le distribuzioni di Poisson possono essere ricavate facendo riferimento ad un particolare tipo di **processi puntuali** che vengono chiamati **processi di Poisson**. Nella Sezione 8.1 dove abbiamo introdotto il concetto di distribuzione notevole, abbiamo già detto che un processo di Poisson è un esperimento casuale ideale che produce successi in modo molto simile a come li produce un **processo di Bernoulli** dove le prove si svolgono a distanza di tempo molto ridotta e dove la probabilità di successo comune a tutte le prove è prossima a zero. In questo paragrafo definiremo in modo più preciso che cosa si intende per un processo di Poisson, ma prima di formulare una definizione precisa dobbiamo spendere qualche parola sui processi puntuali in generale.

In primo luogo ricordiamo che, per definizione, un **processo puntuale** è un esperimento casuale che consiste nella rilevazione degli istanti temporali in cui si verificano determinati eventi che chiamiamo "successi". Esempi di processi puntuali sono esperimenti casuali che consistono nella rilevazione degli istanti temporali in cui

- arrivano telefonate ad un centralino,
- in cui passano veicoli da un casello di un'autostrada,
- in cui si verificano determinati eventi in un mercato borsistico,
- in cui si verificano determinati eventi cosmici, ...

Come si desume dal grafico in Figura 8.7 ...

Figura 8.7. Il grafico mostra come la realizzazione di un processo puntuale può essere descritta attraverso le variabili casuali $\tilde{Y}_{[0,t]}$ ($t \geq 0$) che contano i successi all'aumentare del tempo t oppure, in modo equivalente, attraverso la successione delle variabili casuali $\tilde{X}_1, \tilde{X}_2, \dots$ che rappresentano i tempi d'attesa tra successi consecutivi.



... la realizzazione di un processo puntuale può essere completamente descritta attraverso le realizzazioni di uno dei seguenti tipi di variabili casuali:

- attraverso le realizzazioni delle variabili casuali $\tilde{Y}_{[0,t]}$, $t \geq 0$, che chiameremo **variabili casuali di conteggio** perché contano i successi nell'intervallo temporale $[0, t]$ al quale si riferiscono...

- ... oppure (in modo equivalente) attraverso le realizzazioni delle variabili casuali $\tilde{X}_1, \tilde{X}_2, \dots$ che rappresentano i **tempi d'attesa tra successi consecutivi** (nel caso di successi contemporanei il tempo d'attesa è nullo).

Infatti, come si può facilmente verificare con l'ausilio del grafico in Figura 8.7, dalle realizzazioni delle variabili casuali di conteggio si possono ricavare le realizzazioni delle variabili casuali che rappresentano i tempi d'attesa perché

$$\tilde{S}_n = \sum_{i=1}^n \tilde{X}_i = \inf \left\{ t > 0 : \tilde{Y}_{[0,t]} \geq n \right\};$$

e dalle realizzazioni delle variabili casuali che rappresentano i tempi d'attesa si possono ricavare le realizzazioni delle variabili casuali di conteggio perché

$$\tilde{Y}_{[0,t]} = \sup \left\{ n \in \mathbb{N} : \tilde{S}_n = \sum_{i=1}^n \tilde{X}_i \leq t \right\}.$$

Oltre alle variabili casuali di conteggio $\tilde{Y}_{[0,t]}$, con riferimento ad un processo puntuale si possono definire anche altre variabili casuali che contano successi, come per esempio

- le variabili casuali

$$\tilde{Y}_{(a,b]} = \tilde{Y}_{[0,b]} - \tilde{Y}_{[0,a]} \quad (139)$$

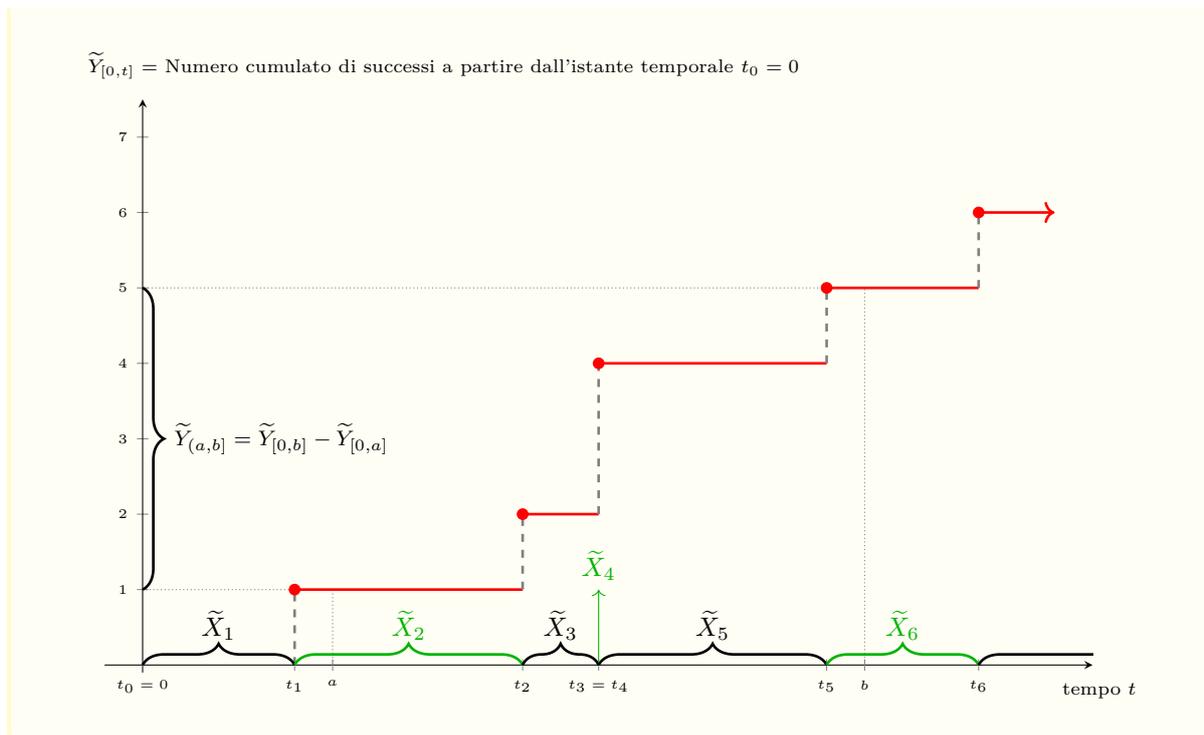
che, come si vede nella Figura 8.8, contano il numero di successi nell'intervallo di tempo $(a, b]$ al quale si riferiscono;

- oppure le variabili casuali

$$\tilde{Y}_{[b,b]} = \lim_{a \rightarrow b} \tilde{Y}_{(a,b]} \quad (140)$$

che invece restituiscono il numero complessivo di successi in un singolo istante temporale $t = b$.

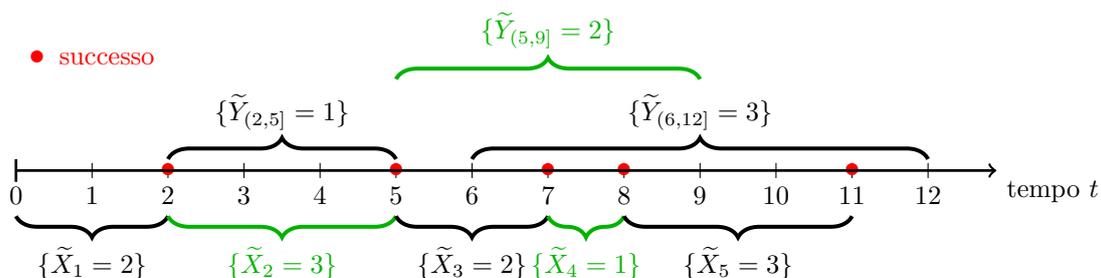
Figura 8.8. Il grafico mostra come sono definite le variabili casuali $\tilde{Y}_{(a,b]}$ che contano i successi in intervalli di tempo semiaperti e limitati.



A seconda se gli istanti temporali dei successi vengono rilevati su un asse temporale discreto o continuo, un processo puntuale viene chiamato **”a tempo discreto”** oppure **”a tempo continuo”**. Nei processi puntuali a tempo discreto si assume spesso che ...

- ... i successi possano verificarsi solo in corrispondenza di multipli interi positivi di una determinata unità di tempo u (l'unità di tempo potrebbe essere, per esempio, un secondo, un minuto, un'ora, un giorno ecc.) ...
- ... e che in un dato istante temporale possa verificarsi al più un singolo successo (non possono dunque verificarsi due o più successi contemporaneamente).

Sotto queste condizioni, un processo puntuale a tempo discreto può essere interpretato come una successione finita o infinita di prove che si svolgono a intervalli di tempo regolari e la realizzazione di un tale processo può essere rappresentata attraverso una successione di punti su un asse temporale discreto come illustrato nel grafico sottostante:



Chiaramente, un **processo di Bernoulli** dove le prove vengono eseguite a intervalli di tempo regolari è un particolare tipo di processo puntuale a tempo discreto.

Per approcciarci alla definizione di **processo di Poisson** conviene ora elencare alcune **proprietà delle variabili casuali di conteggio nell'ambito di un processo di Bernoulli** dove

- a) **la prima prova viene eseguita all'istante temporale $t = u > 0$,**
- b) **e dove le prove successive vengono eseguite una alla volta ogni u unità di tempo.**

Si noti che nell'ambito di un **processo di Bernoulli** con queste caratteristiche, le variabili casuali di conteggio devono soddisfare le seguenti condizioni:

B0) [[All'istante temporale $t = 0$ non possono verificarsi successi.]]

$$P\left(\{\tilde{Y}_{[0,0]} = 0\}\right) = 1.$$

B1) [[E' impossibile che si verifichino due o più successi contemporaneamente.]]

$$P\left(\{\tilde{Y}_{[b,b]} \geq 2\}\right) = 0 \quad \text{per ogni } b \geq 0.$$

B2) [[In qualsiasi intervallo temporale limitato può verificarsi al più un numero finito di successi.]]

$$P\left(\{\tilde{Y}_{(a,b]} < \infty\}\right) = 1 \quad \text{per ogni intervallo limitato } (a, b].$$

B3) [[All'aumentare del tempo t il numero di successi tende sicuramente a infinito.]]

$$P\left(\left\{\lim_{t \rightarrow \infty} \tilde{Y}_{[0,t]} = \infty\right\}\right) = 1.$$

B4) Le variabili casuali $\tilde{Y}_{(a_1, b_1]}$, $\tilde{Y}_{(a_2, b_2]}$, \dots , $\tilde{Y}_{(a_k, b_k]}$ che si riferiscono a intervalli temporali disgiunti sono indipendenti.

B5) **Se $h > 0$ è un multiplo intero di u** , allora la distribuzione di $\tilde{Y}_{(t, t+h]}$ non cambia al variare del punto iniziale t dell'intervallo temporale $(t, t+h]$ di riferimento.

Dimostrazione. In questa dimostrazione dimostreremo che nell'ambito di un **processo di Bernoulli** dove

- a) la prima prova viene eseguita all'istante temporale $t = u > 0$
- b) e dove le prove successive vengono eseguite una alla volta ogni u unità di tempo

le variabili casuali di conteggio $\tilde{Y}_{[0,t]}$ soddisfano le condizioni **B0 - B5**.

- Le proprietà **B0 - B2** sono immediate conseguenze delle condizioni a) e b).

- La dimostrazione della proprietà **B3** è un po' più lunga. In primo luogo osserviamo che l'evento $\left\{ \lim_{t \rightarrow \infty} \tilde{Y}_{[0,t]} = \infty \right\}$ si verifica se e solo se per ogni $r \in \mathbb{N}$ esiste un $s \in \mathbb{N}$ tale che la realizzazione di $\tilde{Y}_{[0,su]}$ sia maggiore di r . Questo fatto può essere espresso come

$$\left\{ \lim_{t \rightarrow \infty} \tilde{Y}_{[0,t]} = \infty \right\} = \bigcap_{r=0}^{\infty} \left(\bigcup_{s=0}^{\infty} \{ \tilde{Y}_{[0,su]} > r \} \right).$$

A questo punto osserviamo che per ogni prefissato valore di r , la successione di eventi

$$A_{s,r} = \{ \tilde{Y}_{[0,su]} > r \}, \quad s = 0, 1, 2, \dots$$

è **non decrescente** (si noti che s è il numero di prove che vengono eseguite durante l'intervallo temporale $[0, su]$) e che pertanto

$$P \left(\bigcup_{s=0}^{\infty} A_{s,r} \right) = [[\text{legge L10}]] = \lim_{s \rightarrow \infty} P(A_{s,r}). \quad (141)$$

D'altra parte, la successione di eventi

$$B_r = \bigcup_{s=0}^{\infty} A_{s,r} = \bigcup_{s=0}^{\infty} \{ \tilde{Y}_{[0,su]} > r \}, \quad r = 1, 2, \dots,$$

è **non crescente** (se r aumenta e s rimane costante l'evento $\{ \tilde{Y}_{[0,su]} > r \}$ diventa più "piccolo"), e per la legge **L11** si deve quindi avere

$$P \left(\bigcap_{r=0}^{\infty} B_r \right) = [[\text{legge L11}]] = \lim_{r \rightarrow \infty} P(B_r). \quad (142)$$

Combinando le formule (141) e (142) si ottiene dunque

$$\begin{aligned} P \left(\left\{ \lim_{t \rightarrow \infty} \tilde{Y}_{[0,t]} = \infty \right\} \right) &= P \left(\bigcap_{r=0}^{\infty} \bigcup_{s=0}^{\infty} \{ \tilde{Y}_{[0,su]} > r \} \right) = [[\text{definizione } B_r]] = P \left(\bigcap_{r=0}^{\infty} B_r \right) = \\ &= [[\text{formula (142)}]] = \lim_{r \rightarrow \infty} P(B_r) = [[\text{definizione } B_r]] = \lim_{r \rightarrow \infty} P \left(\bigcup_{s=0}^{\infty} A_{s,r} \right) = \\ &= [[\text{formula (141)}]] = \lim_{r \rightarrow \infty} \lim_{s \rightarrow \infty} P(A_{s,r}). \end{aligned}$$

Per completare la dimostrazione della proprietà **B3**, ovvero per dimostrare che $P \left(\left\{ \lim_{t \rightarrow \infty} \tilde{Y}_{[0,t]} = \infty \right\} \right) = 1$, è quindi sufficiente dimostrare che

$$\lim_{s \rightarrow \infty} P(A_{s,r}) = 1 \quad \text{per ogni prefissato valore di } r = 1, 2, \dots,$$

ovvero che

$$\lim_{s \rightarrow \infty} P(\bar{A}_{s,r}) = \lim_{s \rightarrow \infty} P \left(\{ \tilde{Y}_{[0,su]} \leq r \} \right) = 0 \quad \text{per ogni prefissato valore di } r = 1, 2, \dots \quad (143)$$

A tal fine osserviamo che

$$\tilde{Y}_{[0,su]} \sim \text{Binomiale}(n = s; p)$$

(p è la probabilità di successo comune a tutte le prove) e che dunque

$$P \left(\{ \tilde{Y}_{[0,su]} \leq r \} \right) = \sum_{y=0}^r \binom{s}{y} p^y (1-p)^{s-y} \quad \text{per } r = 0, 1, 2, \dots, s \quad \text{e per } s = 1, 2, \dots$$

Se manteniamo fisso il valore di r e lasciamo tendere s a più infinito vediamo che il secondo membro della precedente uguaglianza tende a zero (invitiamo il lettore a verificare questa affermazione) e questo fatto dimostra che l'affermazione nella (143) è vera. Con quest'ultima osservazione abbiamo dunque completato la dimostrazione della proprietà **B3**.

- Per dimostrare la proprietà **B4** osserviamo che ciascuna delle k variabili casuali $\tilde{Y}_{(a_i, b_i]}$ può essere rappresentata come la somma delle variabili casuali indicatrici che si riferiscono alle prove che vengono effettuate durante l'intervallo temporale $(a_i, b_i]$ che è indicato nel pedice. Siccome stiamo ipotizzando che gli intervalli temporali $(a_i, b_i]$ siano disgiunti, ciascuna variabile casuale $\tilde{Y}_{(a_i, b_i]}$ è la somma di un diverso insieme di variabili casuali indicatrici, e siccome nell'ambito di un **processo di Bernoulli** le variabili casuali indicatrici che si riferiscono a prove diverse sono indipendenti, possiamo applicare il Teorema 4.10 onde concludere che anche le variabili casuali $\tilde{Y}_{(a_i, b_i]}$ sono indipendenti.
- Per dimostrare la proprietà **B5** basta osservare che, in virtù delle suddette condizioni a) e b), durante un intervallo temporale la cui durata è un multiplo intero h di u vengono eseguite esattamente $n = h$ prove (qualunque sia l'istante temporale iniziale t dell'intervallo) e siccome per ipotesi le prove sono bernoulliane con probabilità di successo p , la distribuzione della variabile casuale $\tilde{Y}_{(t, t+h]}$ deve essere binomiale con parametro $n = h$ e parametro p uguale alla probabilità di successo comune a tutte le prove. Come previsto dalla proprietà **B5**, la distribuzione di $\tilde{Y}_{(t, t+h]}$ rimane dunque invariata al variare dell'istante temporale iniziale $t \geq 0$.

□

A questo punto possiamo finalmente **definire** che cosa si intende per un **processo di Poisson**: un processo di Poisson è un **processo puntuale a tempo continuo** dove ...

- le variabili casuali di conteggio soddisfano le condizioni **B0 - B4** che devono essere soddisfatte anche nel caso di un **processo di Bernoulli** dove la prima prova viene eseguita all'istante temporale $t = u$ e dove le prove successive vengono eseguite una alla volta ogni u unità di tempo, ...
- e che al posto della condizione **B5** soddisfa la condizione

B5*) Per ogni $h > 0$ la distribuzione di $\tilde{Y}_{(t, t+h]}$ non cambia al variare del punto iniziale t dell'intervallo temporale $(t, t + h]$ di riferimento.

Ovviamente, un **processo di Poisson** è soltanto un processo puntuale ideale, ovvero un esperimento casuale ideale che, come abbiamo appena visto, è descritto da un modello probabilistico che ha delle caratteristiche ben precise: un modello probabilistico per un **processo di Poisson** è un qualunque modello probabilistico (Ω, \mathcal{A}, P) che può "ospitare" delle variabili casuali di conteggio in modo tale che le condizioni **B0 - B4** e la condizione **B5*** siano soddisfatte. Tuttavia, per essere sicuri di aver definito qualcosa di sensato dovremmo ora dimostrare l'esistenza di un tale modello probabilistico. Ricordiamo che la questione dell'esistenza è di fondamentale importanza perché se non esistesse un modello probabilistico per un processo di Poisson, allora attraverso gli assiomi di Kolmogorov, le condizioni **B0 - B4** e la condizione **B5*** si potrebbe dimostrare qualsiasi affermazione!!! Per ora daremo per scontata l'esistenza di uno o più modelli probabilistici per un **processo di Poisson**, ma più avanti, quando studieremo le distribuzioni dei tempi d'attesa di un **processo di Poisson**, vedremo che l'esistenza di un modello probabilistico per un **processo**

di Poisson può essere dedotta dal Teorema di esistenza per successioni infinite di variabili casuali indipendenti (ovvero dal Teorema 4.13).

Sempre con riferimento al modo in cui abbiamo **definito** il **processo di Poisson**, sembrerebbe ovvio che **un processo di Poisson generi successi in modo molto simile ad un processo di Bernoulli**. Per indagare su questa ipotesi conviene confrontare le distribuzioni congiunte delle variabili casuali $Y_{(a,b]}$ che si ottengono nei due tipi di processo puntuale. Siccome le variabili casuali $Y_{(a,b]}$ che si riferiscono a intervalli di tempo disgiunti sono indipendenti in entrambi i processi (condizione B4), le distribuzioni congiunte delle variabili casuali $Y_{(a,b]}$ sono univocamente determinate dalle loro distribuzioni marginali (Teorema 4.8) e pertanto basta confrontare le distribuzioni marginali. Nel caso di un **processo di Bernoulli** sappiamo che le distribuzioni marginali delle variabili casuali $Y_{(a,b]}$ sono tutte binomiali (o eventualmente degeneri se durante l'intervallo $(a, b]$ non vengono eseguite prove). Il prossimo teorema (che non dimostreremo) chiarisce invece come sono definite le distribuzioni marginali delle variabili casuali $Y_{(a,b]}$ nell'ambito di un processo di Poisson.

Teorema 8.4 (Primo teorema sui processi di Poisson). Nell'ambito di un **processo di Poisson** le variabili casuali che contano i successi in intervalli di tempo limitati hanno distribuzione di Poisson con parametro λ che è proporzionale alla lunghezza dell'intervallo di tempo di riferimento.

Formalmente: se (Ω, \mathcal{A}, P) è un modello probabilistico per un **processo di Poisson** (ovvero un modello probabilistico che soddisfa le condizioni B0 - B4 e B5*), allora esiste una costante $\theta > 0$ tale che

$$\tilde{Y}_{(t,t+h]} \sim \text{Poisson}(\lambda = \theta \times h)$$

per ogni intervallo limitato $(t, t + h]$ con $t \geq 0$ (ricordiamo che $t = 0$ è l'istante temporale in cui si comincia a rilevare i successi).

Siccome il parametro λ di una distribuzione di Poisson rappresenta il suo valore atteso (lo dimostreremo tra breve), la costante θ nella conclusione del precedente teorema rappresenta il numero atteso di successi in un intervallo di tempo di durata $h = 1$ unità di tempo e per questo motivo la costante θ viene chiamata **tasso di accadimento**.

Conoscendo le distribuzioni marginali delle variabili casuali $\tilde{Y}_{(a,b]}$ nell'ambito di un **processo di Bernoulli** e in un **processo di Poisson**, possiamo ora procedere ad un confronto. Per fare un confronto tra processi che dovrebbero essere "simili", consideriamo

- un **processo di Poisson** con tasso di accadimento θ
- e un **processo di Bernoulli** dove ...

- ... le prove vengono effettuate ogni u unità di tempo con u che è trascurabile rispetto al reciproco del tasso di accadimento θ del processo di Poisson, diciamo in modo tale che

$$\frac{u}{1/\theta} = \theta u \simeq 0,0001$$

(come vedremo più avanti, il reciproco del tasso di accadimento θ di un processo di Poisson è il tempo d'attesa medio tra due successi consecutivi);

- ... e dove la probabilità di successo comune a tutte le prove è data da

$$p = \theta u \simeq 0,0001.$$

Per evitare complicazioni tecniche confronteremo soltanto le distribuzioni marginali delle variabili casuali del tipo $\tilde{Y}_{(a,b]} = \tilde{Y}_{(t,t+h]}$ dove h è un multiplo intero di u , ovvero delle variabili casuali

$$\tilde{Y}_{(t,t+nu]} \quad \text{con } t \geq 0 \text{ e con } n = 1, 2, \dots$$

In primo luogo osserviamo che

- nel [processo di Bernoulli](#) si ha

$$\tilde{Y}_{(t,t+h]} = \tilde{Y}_{(t,t+nu]} \sim \text{Binomiale}(n = h/u; p = \theta u)$$

perché in $h = nu$ unità di tempo vengono eseguite n prove;

- e che nel [processo di Poisson](#) si ha

$$\tilde{Y}_{(t,t+h]} = \tilde{Y}_{(t,t+nu]} \sim \text{Poisson}(\lambda = \theta h = \theta nu = np)$$

così come previsto dalla conclusione del Teorema 8.4.

Per dimostrare che le distribuzioni delle variabili casuali $\tilde{Y}_{(t,t+nu]}$ non cambiano molto passando da un processo all'altro conviene considerare separatamente le variabili casuali $\tilde{Y}_{(t,t+nu]}$ con n piccolo, n intermedio, e n elevato:

- n è piccolo in modo tale che $\lambda = np \leq 0,01$: in questo caso sia la distribuzione $\text{Binomiale}(n = h/u; p = \theta u)$ che la distribuzione $\text{Poisson}(\lambda = \theta h = n\theta u = np)$ concentrano quasi tutta la probabilità nel punto $x = 0$. Infatti, se np è prossimo a zero (diciamo non superiore a 0,01) si ha

$$\begin{aligned} p_{\text{binomiale}}(0) &= \binom{n}{0} p^0 (1-p)^{n-0} = (1-p)^n = e^{n \ln(1-p)} \simeq \\ &\simeq \llbracket \text{siccome } \ln(1-p) \simeq -p \text{ per } p \simeq 0 \rrbracket \simeq e^{-np} \simeq 1 \end{aligned}$$

e

$$p_{\text{poisson}}(0) = e^{-np} (np)^0 / 0! = e^{-np} \simeq 1;$$

- il valore di n è intermedio in modo tale che np sia compreso tra 0,01 e 5: siccome abbiamo scelto u in modo tale $p = \theta u \simeq 0,0001$, anche in questo caso la distribuzione *Binomiale*($n = h/u; p = \theta u$) e la distribuzione di Poisson devono essere molto simili l'una all'altra. Si ricordi infatti che le distribuzioni binomiali con n elevato e p prossimo a zero possono essere approssimate attraverso delle distribuzioni di Poisson (vedi la formula (138)).
- il valore di n è molto elevato in modo tale che np sia maggiore di 5: in questo caso, in virtù del teorema del limite centrale che vedremo più avanti, la distribuzione *Binomiale*($n = h/u; p = \theta u$) e la distribuzione *Poisson*($\lambda = \theta h = n\theta u = np$) devono essere molto simili l'una all'altra perché sono entrambe prossime ad una stessa distribuzione normale (più avanti introdurremo anche questa famiglia di distribuzioni).

Il prossimo teorema riassume le precedenti considerazioni:

Teorema 8.5 (Relazione tra processi di Poisson e processi bernoulliani). Le distribuzioni congiunte delle variabili casuali di conteggio (ovvero delle variabili casuali $\tilde{Y}_{(a,b]}$) nell'ambito di un **processo di Poisson** con tasso di accadimento θ sono molto simili a quelle che si hanno nell'ambito di un **processo di Bernoulli** con

- prove che vengono eseguite ogni u unità di tempo, dove u è trascurabile rispetto al reciproco del tasso di accadimento θ ,
- e con probabilità di successo riferita alle singole prove che è pari a $p = \theta u$.

Nel prossimo esempio faremo alcune valutazioni numeriche sugli errori di approssimazione che sorgono quando si cerca di "simulare" un **processo di Poisson** attraverso un **processo di Bernoulli**.

Esercizio 8.13 (Confronto tra un processo di Poisson e un processo bernoulliano che lo approssima). Si consideri un processo di Poisson nel quale mediamente si verificano $\theta = 3$ successi al minuto (θ è quindi il tasso di accadimento e il tempo viene misurato in minuti).

- Qual è la probabilità che durante un intervallo temporale di 2 minuti si verifichino 4 successi?
- Qual è la probabilità che durante un intervallo temporale di 30 secondi si verifichi 1 successo?
- Qual è la probabilità che durante un intervallo temporale di 2 minuti si verifichino 4 successi e che nei 30 secondi immediatamente successivi si verifichi

1 successo?

- d) Si risponda alle precedenti domande sotto l'ipotesi che i successi vengano generati da un processo di Bernoulli dove le prove vengono eseguite ogni 5 secondi ($u = 5$ secondi è $1/12$ di un minuto ed è abbastanza piccolo rispetto a $1/\theta = 1/3$ di un minuto) e dove la probabilità di successo comune a tutte le prove è data da $p = \theta \times u = 3 \times 1/12 = 1/4$.

Risposte:

- a) Qual è la probabilità che durante un intervallo temporale di 2 minuti si verifichino 4 successi?

Come abbiamo visto, nell'ambito di un processo di Poisson con tasso di accadimento $\theta = 3$ successi al minuto si ha

$$\tilde{Y}_{(t,t+2]} \sim \text{Poisson}(\lambda = \theta \times 2 = 3 \times 2 = 6).$$

Pertanto la probabilità richiesta è data da

$$P(\{\tilde{Y}_{(t,t+2]} = 4\}) = \frac{e^{-6}6^4}{4!} = 0,134.$$

- b) Qual è la probabilità che durante un intervallo temporale di 30 secondi si verifichi 1 successo?

Ragionando come nella risposta precedente vediamo che (si tenga presente che il tasso di accadimento $\theta = 3$ è il numero di successi **al minuto** e pertanto il tempo viene misurato in minuti e non in secondi: 30 secondi sono **1/2** minuto)

$$\tilde{Y}_{(t,t+1/2]} \sim \text{Poisson}(\lambda = \theta \times 1/2 = 3 \times (1/2) = 1,5)$$

e che la probabilità richiesta è data da

$$P(\{\tilde{Y}_{(t,t+1/2]} = 1\}) = \frac{e^{-1,5}1,5^1}{1!} = 0,335.$$

- c) Qual è la probabilità che durante un intervallo temporale di 2 minuti si verifichino 4 successi e che nei 30 secondi immediatamente successivi si verifichi 1 successo?

Siccome nei processi di Poisson le variabili casuali che contano successi in intervalli di tempo disgiunti sono indipendenti, possiamo concludere che

$$\begin{aligned} P(\{\tilde{Y}_{(t,t+2]} = 4\} \cap \{\tilde{Y}_{(t+2,t+2+1/2]} = 1\}) &= \\ &= P(\{\tilde{Y}_{(t,t+2]} = 4\}) \times P(\{\tilde{Y}_{(t+2,t+2+1/2]} = 1\}). \end{aligned} \quad (144)$$

Ora, nelle risposta ai quesiti a) e b) abbiamo visto che

$$P(\{\tilde{Y}_{(t,t+2]} = 4\}) = 0,134$$

e che

$$P(\{\tilde{Y}_{(t,t+1/2]} = 1\}) = 0,335.$$

Siccome nei processi di Poisson le distribuzioni delle variabili casuali $\tilde{Y}_{(t,t+h]}$ non cambiano al variare del punto iniziale t dell'intervallo temporale $(t, t+h]$ di riferimento, possiamo concludere che

$$P(\{\tilde{Y}_{(t+2,t+2+1/2]} = 1\}) = P(\{\tilde{Y}_{(t,t+1/2]} = 1\}) = 0,335.$$

Sostituendo nella (144) vediamo dunque che la probabilità richiesta è data da

$$\begin{aligned} & P(\{\tilde{Y}_{(t,t+2]} = 4\} \cap \{\tilde{Y}_{(t+2,t+2+1/2]} = 1\}) = \\ & = P(\{\tilde{Y}_{(t,t+2]} = 4\}) \times P(\{\tilde{Y}_{(t+2,t+2+1/2]} = 1\}) = \\ & = 0,134 \times 0,335 = 0,045. \end{aligned}$$

- d) Si risponda alle precedenti domande sotto l'ipotesi che i successi vengano generati da un processo di Bernoulli dove le prove vengono eseguite ogni 5 secondi ($u = 1/12$ di un minuto) e dove la probabilità di successo comune a tutte le prove è data da $p = \theta \times u = 3 \times 1/12 = 1/4$.

Nell'ambito di un processo Bernoulli dove viene eseguita una prova ogni 5 secondi ($u = 1/12$ di un minuto) vengono eseguite $n = 2/u = 2/(1/12) = 24$ prove nell'arco di due minuti e pertanto si ha

$$\tilde{Y}_{(t,t+2]} \sim \text{Binomiale}(n = 24; p = 1/4 = 0,25).$$

La probabilità che il processo di Bernoulli generi 4 successi nell'arco di 2 minuti è quindi data da

$$P(\{\tilde{Y}_{(t,t+2]} = 4\}) = \binom{24}{4} 0,25^4 (1 - 0,25)^{24-4} = 0,132$$

ed è molto prossima al valore 0,134 che avevamo ottenuto per il processo di Poisson.

Se invece consideriamo un intervallo temporale di 30 secondi ($1/2$ minuto), allora nel processo di Bernoulli vengono eseguite $n = 30/5 = 6$ prove durante un tale intervallo e pertanto si ha

$$\tilde{Y}_{(t,t+1/2]} \sim \text{Binomiale}(n = 6; p = 1/4 = 0,25).$$

La probabilità che il processo di Bernoulli generi 1 successo nell'arco di 30 secondo è quindi data da

$$P(\{\tilde{Y}_{(t,t+1/2]} = 1\}) = \binom{6}{1} 0,25^1 (1 - 0,25)^{6-1} = 0,356.$$

Anche questo valore non è lontano dal valore 0,335 che abbiamo ottenuto per il processo di Poisson.

Infine, le proprietà [B4](#) e [B5](#) dei processi Bernoulli implicano che anche per questo tipo di processo puntuale la probabilità

$$P(\{\tilde{Y}_{(t,t+2]} = 4\} \cap \{\tilde{Y}_{(t+2,t+2+1/2]} = 1\})$$

può essere calcolata come

$$P(\{\tilde{Y}_{(t,t+2]} = 4\}) \times P(\{\tilde{Y}_{(t,t+1/2]} = 1\}) = 0,132 \times 0,356 = 0,047.$$

Anche in questo caso il risultato è molto simile a quello che si riferisce al processo di Poisson (ricordiamo che per il processo di Poisson la suddetta probabilità è pari a 0,045 - vedi la risposta al quesito c)).

Come abbiamo visto, i [processi di Poisson](#) sono degli esperimenti casuali ideali che generano successi in modo molto simile a come vengono generati da determinati [processi di Bernoulli](#) (che sono anch'essi degli esperimenti casuali ideali).

Con riferimento ciò che osserviamo nel mondo reale, possiamo quindi considerare come [processi di Poisson](#) tutti quei processi puntuali che generano successi che si distribuiscono in modo casuale nel tempo senza effetti di agglomerazione oppure di repulsione. Spesso possono essere considerati come [processi di Poisson](#) esperimenti casuali che consistono nella rilevazione degli istanti temporali in cui . . .

- arrivano chiamate ad un centralino, richieste di connessione ad un *server*, ecc.
- si verificano guasti meccanici su un macchinario,
- si verificano tragedie quali incidenti aerei, morti sul lavoro, esplosioni incidentali, ecc.,
- si verificano eventi lieti quali scoperte archeologiche o scientifiche, scoperte di giacimenti minerari, ecc.,
- si verificano transazioni su un determinato titolo quotato in borsa???

Ovviamente, il riferimento alla dimensione del tempo per la rilevazione dei successi in un processo puntuale è puramente arbitrario. Infatti, nelle applicazioni si è spesso anche

interessati a processi puntuali che si svolgono nello spazio o anche nello spazio-tempo. La teoria sui [processi di Poisson](#) può quindi essere applicata anche ad alcuni esperimenti casuali che consistono nella rilevazione di punti su un asse spaziale come per esempio

- la rilevazione dei punti lungo una strada dove l'asfalto è rovinato;
- la rilevazione dei punti lungo un cavo dove ci sono delle imperfezioni;
- ecc.

Esercizio 8.14. Si assuma che per un dato titolo azionario gli istanti temporali in cui avvengono le transazioni siano generati da un processo di Poisson e che la probabilità di non osservare transazioni durante un intervallo temporale della durata di $h = 1$ minuto sia pari a $0,05$.

- Qual è il numero medio di transazioni nell'arco di un minuto?
- Qual è la probabilità di non osservare transazioni durante un intervallo temporale di 3 minuti?
- Qual è la probabilità di osservare almeno 2 transazioni durante un intervallo temporale di 30 secondi?
- Qual è la probabilità che durante il prossimo minuto ci sia una transazione e che nei due minuti successivi non ci siano transazioni?

Risposte:

- Qual è il numero medio di transazioni nell'arco di un minuto?

Secondo l'ipotesi nella consegna, gli istanti temporali nei quali si verificano le transazioni sarebbero generati da un processo di Poisson sotto il quale la probabilità di non osservare successi nell'arco di 1 minuto è pari a $0,05$:

$$P(\{\tilde{Y}_{(t,t+1]} = 0\}) = 0,05.$$

Siccome il numero medio di transazioni nell'arco di 1 minuto è il tasso di accadimento θ del processo di Poisson, possiamo concludere che

$$\begin{aligned} P(\{\tilde{Y}_{(t,t+1]} = 0\}) &= \frac{e^{-\theta \times 1} \theta^0}{0!} = 0,05 \quad \Rightarrow \\ \Rightarrow e^{-\theta} &= 0,05 \quad \Rightarrow \theta = -\ln 0,05 = 2,9957. \end{aligned}$$

- b) Qual è la probabilità di non osservare transazioni durante un intervallo temporale di 3 minuti?

Siccome $\theta = 2,9957$, possiamo concludere che

$$\tilde{Y}_{(t,t+3]} \sim \text{Poisson}(\lambda = 3 \times 2,9957 = 8,9871).$$

La probabilità di non osservare transazioni durante un intervallo temporale di 3 minuti è quindi data da

$$P(\{\tilde{Y}_{(t,t+3]} = 0\}) = \frac{e^{-8,9871} \times 8,9871^0}{0!} = 0,0001.$$

- c) Qual è la probabilità di osservare almeno 2 transazioni durante un intervallo temporale di 30 secondi?

Siccome 30 secondi corrispondono a 1/2 minuto, possiamo concludere che

$$\tilde{Y}_{(t,t+0,5]} \sim \text{Poisson}(\lambda = 0,5 \times 2,9957 = 1,4979).$$

La probabilità richiesta è quindi data da

$$\begin{aligned} P(\{\tilde{Y}_{(t,t+0,5]} \geq 2\}) &= 1 - P(\{\tilde{Y}_{(t,t+0,5]} < 2\}) \\ &= 1 - \frac{e^{-1,4979} \times 1,4979^0}{0!} - \frac{e^{-1,4979} \times 1,4979^1}{1!} \\ &= 1 - 0,2236 - 0,3349 = 0,4415. \end{aligned}$$

- d) Qual è la probabilità che durante il prossimo minuto ci sia una transazione e che nei due minuti successivi non ci siano transazioni?

Indicando con t l'istante temporale attuale, possiamo esprimere l'evento di cui è richiesta la probabilità come

$$\{\tilde{Y}_{(t,t+1]} = 1\} \cap \{\tilde{Y}_{(t+1,t+3]} = 0\}$$

Siccome i due eventi dell'intersezione si riferiscono a intervalli temporali disgiunti, essi devono essere indipendenti (questa è una caratteristica dei processi di Poisson) e la probabilità richiesta può dunque essere calcolata come

$$\begin{aligned} P(\{\tilde{Y}_{(t,t+1]} = 1\} \cap \{\tilde{Y}_{(t+1,t+3]} = 0\}) &= \\ &= P(\{\tilde{Y}_{(t,t+1]} = 1\}) \times P(\{\tilde{Y}_{(t+1,t+3]} = 0\}). \end{aligned}$$

Ricordando che

$$P(\{\tilde{Y}_{(t,t+1]} = 1\}) = [[\text{risposta quesito d}]] = 0,1498$$

e tenendo presente che

$$\begin{aligned}
 & P(\{\tilde{Y}_{(t+1,t+3]} = 0\}) = \\
 & \text{[[La distribuzione di } \tilde{Y}_{(t,t+h]} \text{ non dipende da } t \text{]]} \\
 & = P(\{Y_{(t,t+(3-2)]} = 0\}) = P(\{Y_{(t,t+2]} = 0\}) = \\
 & \text{[[} \tilde{Y}_{(t,t+2]} \sim \text{Poisson}(\lambda = \theta \times 2 = 2,9957 \times 2) \text{]]} \\
 & = \frac{e^{-\theta \times 2} \times (\theta \times 2)^0}{0!} = \frac{e^{-(2 \times 2,9957)} \times (2 \times 2,9957)^0}{0!} = \\
 & = e^{-(2 \times 2,9957)} = 0,0025,
 \end{aligned}$$

vediamo che la probabilità richiesta è data da

$$\begin{aligned}
 & P(\{\tilde{Y}_{(t,t+1]} = 1\} \cap \{\tilde{Y}_{(t+1,t+3]} = 0\}) = \\
 & = P(\{\tilde{Y}_{(t,t+1]} = 1\}) \times P(\{\tilde{Y}_{(t+1,t+3]} = 0\}) \\
 & = 0,1498 \times 0,0025 = 0,0004.
 \end{aligned}$$

Funzione generatrice dei momenti. Come vedremo tra breve, tutte le distribuzioni di Poisson sono dotate di fgm e la fgm di una distribuzione di Poisson ha un'espressione molto semplice. Infatti, se $p(x)$ è la funzione di probabilità di Poisson, si ottiene

$$\begin{aligned}
 \sum_{x \in S} e^{tx} \times p(x) &= \sum_{x=0}^{\infty} e^{tx} \frac{e^{-\lambda} \lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^t)^x}{x!} = \\
 &= \text{[[definizione di funzione esponenziale]]} = e^{-\lambda} \times e^{\lambda e^t} = e^{\lambda(e^t - 1)}
 \end{aligned}$$

e questo risultato dimostra che tutte le distribuzioni di Poisson sono dotate di fgm e che la fgm di una distribuzione di Poisson è data da

$$m(x) = e^{\lambda(e^t - 1)}, \quad t \in \mathbb{R}. \quad (145)$$

Derivando la fgm $m(x) = e^{\lambda(e^t - 1)}$ vediamo che

$$m'(t) = \frac{dm(t)}{dt} = e^{\lambda(e^t - 1)} \lambda e^t \quad \Rightarrow \quad \mu = E(\tilde{X}) = m'(0) = \lambda,$$

e che

$$\begin{aligned}
 m''(t) &= \frac{d^2 m(t)}{dt^2} = e^{\lambda(e^t - 1)} (\lambda e^t)^2 + e^{\lambda(e^t - 1)} \lambda e^t \quad \Rightarrow \\
 \Rightarrow \quad \mu_2 = E(\tilde{X}^2) = m''(0) &= \lambda^2 + \lambda \quad \Rightarrow \quad \sigma^2 = \text{var}(\tilde{X}) = \mu_2 - \mu^2 = \lambda.
 \end{aligned}$$

Il valore atteso e la varianza di una distribuzione di Poisson sono dunque entrambi uguali a λ !

Proprietà riproduttiva. Anche la famiglia delle distribuzioni di Poisson gode della proprietà riproduttiva:

Teorema 8.6 (Proprietà riproduttiva della famiglia delle distribuzioni di Poisson).

Se

$$\tilde{X}_i \sim \text{Poisson}(\lambda = \lambda_i), \quad i = 1, 2, \dots, k,$$

sono k variabili casuali **indipendenti**, allora

$$\tilde{X} = \tilde{X}_1 + \tilde{X}_2 + \dots + \tilde{X}_k \sim \text{Poisson}(\lambda = \lambda_1 + \lambda_2 + \dots + \lambda_k).$$

Una somma di variabili casuali di Poisson indipendenti è ancora una variabile casuale di Poisson. Il parametro λ della distribuzione di Poisson della somma è uguale alla somma dei parametri λ_i delle distribuzioni di Poisson delle variabili casuali che vengono sommate.

Anche il teorema sulla proprietà riproduttiva della famiglia delle distribuzioni di Poisson può essere dimostrato seguendo la falsariga della dimostrazione del Teorema 8.1 sulla proprietà riproduttiva delle distribuzioni binomiali (invitiamo il lettore a verificare i passaggi logici).

Vale la pena osservare che, in alternativa, la proprietà riproduttiva della famiglia delle distribuzioni di Poisson può essere anche dimostrata facendo riferimento alla teoria dei **processi di Poisson**. Infatti, se

$$(b_0, b_1], \quad (b_1, b_2], \quad (b_2, b_3], \quad \dots, \quad (b_{k-1}, b_k]$$

sono k intervalli di tempo adiacenti e

$$\tilde{Y}_{(b_0, b_1]}, \quad \tilde{Y}_{(b_1, b_2]}, \quad \tilde{Y}_{(b_2, b_3]}, \quad \dots, \quad \tilde{Y}_{(b_{k-1}, b_k]}$$

sono le variabili casuali che contano i successi all'interno di questi intervalli, allora

$$\tilde{Y}_{(b_0, b_k]} = \tilde{Y}_{(b_0, b_1]} + \tilde{Y}_{(b_1, b_2]} + \dots + \tilde{Y}_{(b_{k-1}, b_k]}$$

è la variabile casuale che restituisce il numero complessivo di successi nell'intervallo $(b_0, b_k]$ (che è l'unione degli intervalli $(b_{i-1}, b_i]$) e secondo il Teorema 8.4 si deve avere

$$\tilde{Y}_{(b_{i-1}, b_i]} \sim \text{Poisson}(\lambda_i = \theta(b_i - b_{i-1})) \quad \text{per } i = 1, 2, \dots, k,$$

e anche

$$\tilde{Y}_{(b_0, b_k]} \sim \text{Poisson}(\lambda = \theta(b_k - b_0)).$$

Siccome la somma dei parametri $\lambda_i = \theta(b_i - b_{i-1})$ è uguale a $\lambda = \theta(b_k - b_0)$, e siccome i valori di θ e quelli degli estremi b_i degli intervalli possono essere scelti in modo arbitrario (a patto che $b_{i-1} < b_i$ per ogni $i = 1, 2, \dots, k$), questo ragionamento fornisce una dimostrazione alternativa della proprietà riproduttiva delle distribuzioni di Poisson.

Probabilità condizionate di sopravvivenza. Siccome le distribuzioni di Poisson hanno per supporto l'insieme dei numeri naturali e quindi si prestano come distribuzioni

per variabili casuali che descrivono tempi d'attesa (su un asse di tempo discreto), conviene spendere qualche parola sull'andamento delle [probabilità condizionate di sopravvivenza](#)

$$P(\{\tilde{X} > x + h\} | \{\tilde{X} > x\}) = \frac{\sum_{t=x+h+1}^{\infty} \frac{\lambda^t e^{-\lambda}}{t!}}{\sum_{t=x+1}^{\infty} \frac{\lambda^t e^{-\lambda}}{t!}}, \quad x, h = 0, 1, 2, \dots$$

di variabili casuali \tilde{X} che hanno distribuzione di Poisson. Si può dimostrare che anche per le distribuzioni di Poisson, così come per le distribuzioni binomiali negative con $r > 1$, le probabilità condizionate di sopravvivenza con prefissati valori di $h = 1, 2, \dots$ sono tutte strettamente minori della corrispondente probabilità marginale

$$P(\{\tilde{X} > h\})$$

e che decrescono all'aumentare del tempo d'attesa x che è già trascorso. Tuttavia, diversamente da ciò che accade con le distribuzioni binomiali negative, nel caso delle distribuzioni di Poisson le probabilità condizionate di sopravvivenza tendono a 0 se il tempo d'attesa x che è già trascorso tende a infinito.

Per non appesantire troppo la trattazione non dimostreremo nessuna delle precedenti affermazioni sulle [probabilità condizionate di sopravvivenza](#) delle distribuzioni di Poisson, ma ci limitiamo a presentare un esercizio che illustra l'impiego delle distribuzioni di Poisson nell'analisi di sopravvivenza. La consegna dell'esercizio sarà identica a quella dell'Esercizio [8.12](#) tranne per l'ipotesi che le durate delle componenti del macchinario abbiano una distribuzione di Poisson piuttosto che una distribuzione geometrica.

Esercizio 8.15 (Distribuzione di Poisson nell'analisi di durate). Un macchinario contiene tre componenti che entrano in funzione uno dopo l'altro quando quello precedente si guasta. Il macchinario cessa di funzionare quando l'ultimo dei tre componenti si guasta. La durata di ciascuno dei tre componenti è descritta da una variabile casuale con distribuzione di Poisson di valore atteso $\mu = 4$ ore.

- Qual è la probabilità che la durata di un componente sia maggiore di 5 ore?
- Sapendo che un componente è già in funzione da 6 ore, qual è la probabilità condizionata che tra 5 ore lo stesso componente sia ancora in funzione?
- Qual è la probabilità che tra 5 ore il macchinario sia ancora in funzione?
- Sapendo che il macchinario è già in funzione da 6 ore, qual è la probabilità condizionata che tra 5 ore sia ancora in funzione?
- Si calcoli la varianza della durata di un singolo componente.
- Si calcolino valore atteso e varianza della durata del macchinario.

- g) Si calcoli il decimo percentile della durata del macchinario a partire dal momento della prima accensione.

Risposte:

- a) Qual è la probabilità che la durata di un componente sia maggiore di 5 ore? Siano \tilde{X}_1 , \tilde{X}_2 e \tilde{X}_3 le variabili casuali che rappresentano le durate dei tre componenti. Siccome per ipotesi

$$\tilde{X}_i \sim \text{Poisson}(\lambda = ?) \quad \text{e} \quad E(\tilde{X}_i) = \lambda = 4 \text{ ore,}$$

possiamo concludere che

$$\begin{aligned} P(\{\tilde{X}_i > 5\}) &= 1 - P(\{\tilde{X}_i \leq 5\}) \\ &= 1 - \left(\frac{e^{-4}4^0}{0!} + \frac{e^{-4}4^1}{1!} + \frac{e^{-4}4^2}{2!} + \frac{e^{-4}4^3}{3!} + \frac{e^{-4}4^4}{4!} + \frac{e^{-4}4^5}{5!} \right) \\ &= 1 - (0,018 + 0,073 + 0,147 + 0,195 + 0,195 + 0,156) \\ &= 1 - 0,784 = 0,216. \end{aligned}$$

Per fare un confronto con il caso dove la durata \tilde{X}_i di un componente ha distribuzione geometrica con lo stesso valore atteso (ovvero $\mu = E(\tilde{X}_i) = 4$ ore), ricordiamo che nell'Esercizio 8.12 abbiamo ottenuto

$$P(\{\tilde{X}_i > 5\}) = 0,237.$$

- b) Sapendo che un componente è già in funzione da 6 ore, qual è la probabilità condizionata che tra 5 ore lo stesso componente sia ancora in funzione?

Per rispondere al quesito dobbiamo calcolare il valore di

$$\begin{aligned} P(\{\tilde{X}_i > 6 + 5\} | \{\tilde{X}_i > 6\}) &= \frac{P(\{\tilde{X}_i > 6 + 5\} \cap \{\tilde{X}_i > 6\})}{P(\{\tilde{X}_i > 6\})} \\ &= \frac{P(\{\tilde{X}_i > 6 + 5\})}{P(\{\tilde{X}_i > 6\})} = \frac{P(\{\tilde{X}_i > 11\})}{P(\{\tilde{X}_i > 6\})} \end{aligned}$$

Sfruttando il risultato del quesito precedente troviamo che

$$\begin{aligned} P(\{\tilde{X}_i > 6\}) &= P(\{\tilde{X}_i > 5\}) - P(\{\tilde{X}_i = 6\}) \\ &= 0,216 - \frac{e^{-4}4^6}{6!} = 0,216 - 0,104 = 0,112 \end{aligned}$$

e che

$$\begin{aligned} P(\{\tilde{X}_i > 11\}) &= P(\{\tilde{X}_i > 6\}) - \sum_{x=7}^{11} P(\{\tilde{X}_i = x\}) \\ &= 0,112 - \left(\frac{e^{-4}4^7}{7!} + \frac{e^{-4}4^8}{8!} + \frac{e^{-4}4^9}{9!} + \frac{e^{-4}4^{10}}{10!} + \frac{e^{-4}4^{11}}{11!} \right) \\ &= 0,112 - (0,060 + 0,030 + 0,013 + 0,005 + 0,002) = 0,002. \end{aligned}$$

La probabilità condizionata richiesta è dunque data da

$$P(\{\tilde{X}_i > 6 + 5\} | \{\tilde{X}_i > 6\}) = \frac{P(\{\tilde{X}_i > 11\})}{P(\{\tilde{X}_i > 6\})} = \frac{0,002}{0,112} = 0,018.$$

Se un componente è già in funzione da 6 ore, la probabilità condizionata che rimanga in funzione per ulteriori 5 ore oppure più a lungo è quindi solo pari 0,018.

Come ci aspettavamo in virtù del fatto che le probabilità condizionate di sopravvivenza delle distribuzioni di Poisson decrescono all'aumentare del tempo d'attesa x che è già trascorso, la probabilità condizionata di sopravvivenza $P(\{\tilde{X}_i > 6 + 5\} | \{\tilde{X}_i > 6\}) = 0,018$ è strettamente minore della probabilità marginale $P(\{\tilde{X}_i > 5\}) = 0,216$, ovvero della probabilità che un componente nuovo rimanga in funzione per più di cinque ore.

Per fare un confronto con il caso dove la durata \tilde{X}_i di un componente ha distribuzione geometrica con lo stesso valore atteso (ovvero $\mu = E(\tilde{X}_i) = 4$ ore), ricordiamo che nell'Esercizio 8.12 abbiamo ottenuto

$$P(\{\tilde{X}_i > 6+5\} | \{\tilde{X}_i > 6\}) = [[\text{assenza di memoria}]] = P(\{\tilde{X}_i > 5\}) = 0,237.$$

c) Qual è la probabilità che tra 5 ore il macchinario sia ancora in funzione?

Sia \tilde{X} la variabile casuale che descrive la durata del macchinario. Per ipotesi $\tilde{X} = \tilde{X}_1 + \tilde{X}_2 + \tilde{X}_3$.

Siccome sembra ragionevole che le durate dei tre componenti non si influenzino a vicenda, possiamo assumere che le variabili casuali \tilde{X}_1 , \tilde{X}_2 e \tilde{X}_3 siano indipendenti. Tenendo presente che

$$\tilde{X}_i \sim \text{Poisson}(\lambda_i = 4), \quad i = 1, 2, 3,$$

possiamo dunque concludere che (proprietà riproduttiva della famiglia delle distribuzioni di Poisson)

$$\tilde{X} = \tilde{X}_1 + \tilde{X}_2 + \tilde{X}_3 \sim \text{Poisson}(\lambda = 4 + 4 + 4 = 12).$$

La probabilità richiesta è dunque data da

$$\begin{aligned}
 P(\{\tilde{X} > 5\}) &= 1 - P(\{\tilde{X} \leq 5\}) = 1 - \sum_{x=3}^5 P(\{\tilde{X} = x\}) \\
 &= 1 - \left(\frac{e^{-12}12^0}{0!} + \frac{e^{-12}12^1}{1!} + \frac{e^{-12}12^2}{2!} + \frac{e^{-12}12^3}{3!} + \frac{e^{-12}12^4}{4!} + \right. \\
 &\quad \left. + \frac{e^{-12}12^5}{5!} \right) \\
 &= 1 - (0,000 + 0,000 + 0,000 + 0,002 + 0,005 + 0,013) = 0,980.
 \end{aligned}$$

Per fare un confronto con il caso dove la durata \tilde{X}_i di un componente ha distribuzione geometrica con lo stesso valore atteso (ovvero $\mu = E(\tilde{X}_i) = 4$ ore), ricordiamo che nell'Esercizio 8.12 abbiamo ottenuto

$$P(\{\tilde{X} > 5\}) = 0,896.$$

- d) Sapendo che il macchinario è già in funzione da 6 ore, qual è la probabilità condizionata che tra 5 ore sia ancora in funzione?

Per rispondere dobbiamo calcolare il valore di

$$\begin{aligned}
 P(\{\tilde{X} > 6 + 5\} | \{\tilde{X} > 6\}) &= \frac{P(\{\tilde{X} > 6 + 5\} \cap \{\tilde{X} > 6\})}{P(\{\tilde{X} > 6\})} \\
 &= \frac{P(\{\tilde{X} > 6 + 5\})}{P(\{\tilde{X} > 6\})}.
 \end{aligned}$$

La probabilità al denominatore può essere calcolata come

$$\begin{aligned}
 P(\{\tilde{X} > 6\}) &= P(\{\tilde{X} > 5\}) - P(\{\tilde{X} = 6\}) \\
 \text{[[Risposta quesito c)]]} &= 0,980 - \frac{e^{-12}12^6}{6!} \\
 &= 0,980 - 0,025 = 0,955
 \end{aligned}$$

e ragionando allo stesso modo vediamo che la probabilità al numeratore è data da

$$\begin{aligned}
 P(\{\tilde{X} > 6 + 5\}) &= P(\{\tilde{X} > 11\}) = P(\{\tilde{X} > 6\}) - \sum_{x=7}^{11} P(\{\tilde{X} = x\}) \\
 \text{[[Risultato precedente]]} & \\
 &= 0,955 - \sum_{x=7}^{11} \frac{e^{-12}12^x}{x!} \\
 &= 0,955 - 0,044 - 0,066 - 0,087 - 0,105 - 0,114 = 0,539.
 \end{aligned}$$

La probabilità condizionata richiesta è dunque data da

$$P(\{\tilde{X} > 6 + 5\}|\{\tilde{X} > 6\}) = \frac{P(\{\tilde{X} > 6 + 5\})}{P(\{\tilde{X} > 6\})} = \frac{0,539}{0,955} = 0,564.$$

Questo risultato ci dice qual è la probabilità che dopo 6 ore di funzionamento il macchinario rimanga in funzione per ulteriori 5 ore oppure più a lungo.

Si noti che il valore di

$$P(\{\tilde{X} > 6 + 5\}|\{\tilde{X} > 6\}) = 0,564$$

è molto più piccolo del valore di

$$P(\{\tilde{X} > 5\}) = 0,980 \quad [[\text{risposta del quesito c) }]]$$

che invece rappresenta la probabilità che a partire dal momento della sua prima accensione il macchinario rimanga in funzione 5 ore oppure più a lungo.

Per fare un confronto con il caso dove la durata \tilde{X}_i di un componente ha distribuzione geometrica con lo stesso valore atteso (ovvero $\mu = E(\tilde{X}_i) = 4$ ore), ricordiamo che nell'Esercizio 8.12 abbiamo ottenuto

$$P(\{\tilde{X} > 6 + 5\}|\{\tilde{X} > 6\}) = 0,412.$$

- e) Si calcoli la varianza della durata di un singolo componente.

Siccome per ipotesi

$$\tilde{X}_i \sim \text{Poisson}(\lambda = \lambda_i = 4), \quad i = 1, 2, 3,$$

possiamo concludere che

$$\text{var}(\tilde{X}_i) = E(\tilde{X}_i) = \lambda_i = 4.$$

Per fare un confronto con il caso dove la durata \tilde{X}_i di un componente ha distribuzione geometrica con lo stesso valore atteso (ovvero $\mu = E(\tilde{X}_i) = 4$ ore), ricordiamo che nell'Esercizio 8.12 abbiamo ottenuto

$$E(\tilde{X}_i) = 4 < 12 = \text{var}(\tilde{X}_i)$$

- f) Si calcolino valore atteso e varianza della durata del macchinario.

Siccome

$$\tilde{X} = \tilde{X}_1 + \tilde{X}_2 + \tilde{X}_3 \quad \sim \quad \text{Poisson}(\lambda = 12),$$

possiamo concludere che

$$E(\tilde{X}) = \text{var}(\tilde{X}) = \lambda = 12.$$

Per fare un confronto con il caso dove la durata \tilde{X}_i di un componente ha distribuzione geometrica con lo stesso valore atteso (ovvero $\mu = E(\tilde{X}_i) = 4$ ore), ricordiamo che nell'Esercizio 8.12 abbiamo ottenuto

$$E(\tilde{X}) = 12 < 36 = \text{var}(\tilde{X}).$$

- g) Si calcoli il decimo percentile della durata del macchinario a partire dal momento della prima accensione.

Per rispondere al quesito dobbiamo cumulare le probabilità della distribuzione *Poisson* ($\lambda = 12$) (ovvero della distribuzione della variabile casuale che descrive la durata del macchinario) fino a raggiungere o superare il valore 0,10:

$$\begin{aligned} p(0) &= P(\{\tilde{X} = 0\}) = \frac{e^{-12}12^0}{0!} = 0,000 \Rightarrow F(0) = \sum_{x=0}^0 p(x) = 0,000 < 0,10, \\ p(1) &= P(\{\tilde{X} = 1\}) = \frac{e^{-12}12^1}{1!} = 0,000 \Rightarrow F(1) = \sum_{x=0}^1 p(x) = 0,000 < 0,10, \\ p(2) &= P(\{\tilde{X} = 2\}) = \frac{e^{-12}12^2}{2!} = 0,000 \Rightarrow F(2) = \sum_{x=0}^2 p(x) = 0,000 < 0,10, \\ p(3) &= P(\{\tilde{X} = 3\}) = \frac{e^{-12}12^3}{3!} = 0,002 \Rightarrow F(3) = \sum_{x=0}^3 p(x) = 0,002 < 0,10, \\ p(4) &= P(\{\tilde{X} = 4\}) = \frac{e^{-12}12^4}{4!} = 0,005 \Rightarrow F(4) = \sum_{x=0}^4 p(x) = 0,007 < 0,10, \\ p(5) &= P(\{\tilde{X} = 5\}) = \frac{e^{-12}12^5}{5!} = 0,013 \Rightarrow F(5) = \sum_{x=0}^5 p(x) = 0,020 < 0,10, \\ p(6) &= P(\{\tilde{X} = 6\}) = \frac{e^{-12}12^6}{6!} = 0,025 \Rightarrow F(6) = \sum_{x=0}^6 p(x) = 0,045 < 0,10, \\ p(7) &= P(\{\tilde{X} = 7\}) = \frac{e^{-12}12^7}{7!} = 0,044 \Rightarrow F(7) = \sum_{x=0}^7 p(x) = 0,089 < 0,10, \\ p(8) &= P(\{\tilde{X} = 8\}) = \frac{e^{-12}12^8}{8!} = 0,066 \Rightarrow F(8) = \sum_{x=0}^8 p(x) = 0,155 \geq 0,10 \end{aligned}$$

Il più piccolo valore di x dove la funzione di ripartizione $F(x)$ supera 0,10 è $x = x_{0,10} = 8$ e questo valore è quindi il decimo percentile della distribuzione

della durata del macchinario. Il valore di $x_{0,10} = 8$ ci dice che con probabilità almeno pari a 0,10 la durata del macchinario non sarà superiore a 8 ore.

Per fare un confronto con il caso dove la durata \tilde{X}_i di un componente ha distribuzione geometrica con lo stesso valore atteso (ovvero $\mu = E(\tilde{X}_i) = 4$ ore), ricordiamo che nell'Esercizio 8.12 abbiamo ottenuto $x_{0,10} = 5$ ore.

8.9 La famiglia delle distribuzioni esponenziali

La famiglia delle distribuzioni esponenziali è la famiglia di tutte le distribuzioni assolutamente continue che hanno una funzione di densità del tipo

$$f(x) = \begin{cases} \theta e^{-\theta x} & \text{per } x > 0 \\ 0 & \text{altrimenti} \end{cases} \quad (146)$$

con $\theta > 0$. Nell'Esercizio 4.7 abbiamo già dimostrato che tutte le funzioni (146) con $\theta > 0$ soddisfano le tre proprietà f0 - f2 che le qualificano come delle funzioni di densità.

Come si desume dalla definizione delle funzioni di densità (146), le distribuzioni esponenziali concentrano tutta la probabilità sull'insieme dei numeri reali positivi e pertanto si prestano come distribuzioni per variabili casuali che descrivono durate e tempi d'attesa. Infatti, come vedremo tra breve, le distribuzioni esponenziali hanno un ruolo centrale nello studio di durate e tempi d'attesa perché

- sono le uniche distribuzioni continue che soddisfano la **versione continua** della [proprietà di assenza di memoria](#)
- e sono le distribuzioni dei tempi d'attesa tra successi consecutivi generati da [processi di Poisson](#).

D'ora in poi, per comodità di notazione, indicheremo la distribuzione esponenziale di parametro $\theta > 0$ con *Esponenziale*(θ), e per indicare che la distribuzione di una variabile casuale \tilde{X} è esponenziale con parametro θ scriveremo semplicemente

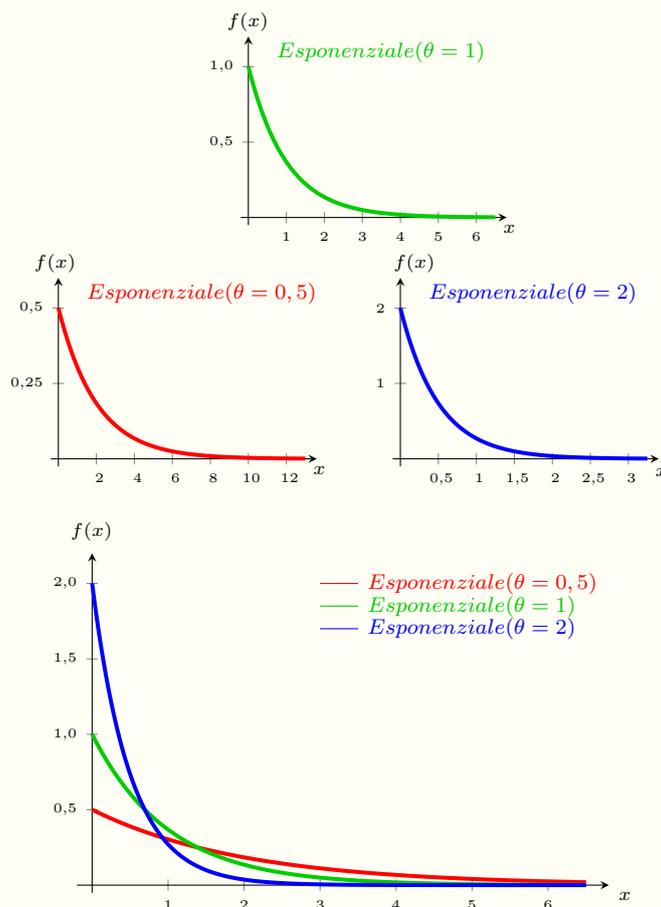
$$\tilde{X} \sim \text{Esponenziale}(\theta).$$

Il ruolo del parametro θ . Come si desume dalla formula (146), le funzioni densità esponenziali hanno un andamento molto semplice e il parametro θ che le identifica è solo un **parametro di scala**, ovvero un parametro che incide soltanto sulla scala dei valori riportati sugli assi cartesiani, ma non sulla "forma" della funzione di densità. Infatti, per disegnare la funzione di densità esponenziale corrispondente ad un qualunque valore positivo del parametro θ , basta disegnare la funzione di densità esponenziale con $\theta = 1$ e

- dividere per θ i valori riportati sull'asse delle ascisse
- e moltiplicare per θ i valori riportati sull'asse delle ordinate.

I grafici in Figura 8.9 mostrano alcune funzioni di densità esponenziali.

Figura 8.9. I grafici mostrano alcune distribuzioni esponenziali. Si noti che graficamente appaiono tutte identiche, l'unica cosa che cambia sono le scale di valori sui due assi cartesiani.



Come dimostreremo tra breve, il modo in cui il parametro θ agisce sulle scale di valori riportate sugli assi cartesiani implica che

$$\tilde{X} \sim \text{Esponenziale}(\theta) \quad \Rightarrow \quad \tilde{Y} = b\tilde{X} \sim \text{Esponenziale}(\theta/b) \quad \text{per ogni } b > 0. \quad (147)$$

In virtù di questo fatto il **parametro di scala** θ viene spesso qualificato come "inverso".

Dimostrazione (θ è un **parametro di scala** inverso). Se $\tilde{X} \sim \text{Esponenziale}(\theta)$ e $\tilde{Y} = b\tilde{X}$, allora la funzione di ripartizione della variabile casuale \tilde{Y} deve essere data da

$$\begin{aligned} F(y) &= P(\{\tilde{Y} \leq y\}) = P(\{b\tilde{X} \leq y\}) = P(\{\tilde{X} \leq y/b\}) \\ &= \begin{cases} \int_0^{y/b} \theta e^{-\theta t} dt & \text{per } y \geq 0 \\ 0 & \text{per } y < 0. \end{cases} \end{aligned}$$

Eseguendo il cambio di variabile $t = x/b \Rightarrow dt = du/b$, si vede che per $y \geq 0$ la funzione di ripartizione di \tilde{Y} può essere espressa come

$$F(y) = \int_0^{y/b} \theta e^{-\theta t} dt = \int_0^y (\theta/b) e^{-(\theta/b)u} du = \int_{-\infty}^y f(u) du,$$

dove

$$f(t) = \begin{cases} (\theta/b)e^{-(\theta/b)u} & \text{per } u > 0 \\ 0 & \text{altrimenti.} \end{cases}$$

Questo ragionamento dimostra quindi che $\tilde{Y} = b\tilde{X} \sim \text{Esponenziale}(\theta/b)$ come volevamo dimostrare. \square

Funzione di ripartizione e quantili. Per ricavare la funzione di ripartizione di una distribuzione esponenziale basta integrare la funzione di densità (146). Infatti, non è difficile verificare che

$$F(x) = \int_{-\infty}^x f(t) dt = \begin{cases} \int_0^x \theta e^{-\theta t} dt = 1 - e^{-\theta x} & \text{per } x \geq 0 \\ 0 & \text{per } x < 0. \end{cases} \quad (148)$$

Invertendo la funzione di ripartizione si trova anche una formula per calcolare i quantili x_p di una distribuzione esponenziale: infatti, per qualsiasi $p \in (0, 1)$ si ha

$$F(x) = p \Leftrightarrow x = x_p = -\frac{1}{\theta} \ln(1 - p).$$

Momenti e funzione generatrice dei momenti. Non è difficile verificare che tutte le distribuzioni esponenziali sono dotate di fgm e che per una distribuzione esponenziale esiste quindi sempre tutta l'infinita successione dei momenti.

Infatti, l'integrale

$$\begin{aligned} \int_{-\infty}^{\infty} e^{tx} f(x) dx &= \int_0^{\infty} e^{tx} \theta e^{-\theta x} dx \\ &= \int_0^{\infty} \theta e^{-(\theta-t)x} dx = \begin{cases} \left[\frac{\theta e^{-(\theta-t)x}}{-(\theta-t)} \right]_0^{\infty} = \frac{\theta}{\theta-t} & \text{se } t < \theta \\ \infty & \text{se } t \geq \theta \end{cases} \end{aligned}$$

converge se e solo se $t < \theta$, e siccome le distribuzioni esponenziali sono definite soltanto per $\theta > 0$, possiamo concludere che la fgm di una distribuzione esponenziale sia data da

$$m(t) = \frac{\theta}{\theta - t} \quad \text{per } t < \theta.$$

Derivando la fgm vediamo dunque che

$$\begin{aligned} m(t) &= \frac{\theta}{\theta - t} \quad \text{per } t < \theta \Rightarrow \\ \Rightarrow m'(t) &= \frac{dm(t)}{dt} = \frac{\theta}{(\theta - t)^2} \Rightarrow \mu = E(\tilde{X}) = m'(0) = \frac{1}{\theta} \\ \Rightarrow m''(t) &= \frac{d^2 m(t)}{dt^2} = \frac{2\theta}{(\theta - t)^3} \Rightarrow \mu_2 = E(\tilde{X}^2) = m''(0) = \frac{2}{\theta^2} \Rightarrow \\ \Rightarrow \sigma^2 &= \text{var}(\tilde{X}) = \mu_2 - \mu^2 = \frac{1}{\theta^2}. \end{aligned}$$

Il valore atteso e la varianza di una distribuzione esponenziale sono quindi dati da $\mu = E(\tilde{X}) = 1/\theta$ e da $\sigma^2 = \text{var}(\tilde{X}) = 1/\theta^2$.

Ovviamente, le formule per i momenti (e dunque anche quelle per il valore atteso e la varianza) delle distribuzioni esponenziali possono essere ricavate anche in modo diretto semplificando l'integrale

$$\mu_k = E(\tilde{X}^k) = \int_{-\infty}^{\infty} x^k f(x) dx = \int_0^{\infty} x^k \theta e^{-\theta x} dx$$

(basta usare il metodo dell'**integrazione per parti**), ma questo modo di procedere è sicuramente più complicato di quello basato sulla fgm.

Assenza di memoria. In precedenza abbiamo accennato al fatto che le distribuzioni esponenziali siano le uniche distribuzioni che soddisfano la **versione continua della proprietà di assenza di memoria**. La **versione continua della proprietà di assenza di memoria** prevede che

$$P(\{\tilde{X} > x + h\} | \{\tilde{X} > x\}) = P(\tilde{X} > h) \quad \text{per ogni } x, h \geq 0, \quad (149)$$

mentre la versione discreta della **proprietà di assenza di memoria** prevede che la suddetta uguaglianza sia soddisfatta soltanto se x e h sono dei numeri naturali (vedi la condizione (126)). Chiaramente, anche nel caso continuo la **proprietà di assenza di memoria** è una proprietà che riguarda soltanto la distribuzione $F(x)$ di una variabile casuale, e volendo essere pignoli dovrebbe quindi essere definita come

$$\frac{1 - F(x + h)}{1 - F(x)} = 1 - F(h) \quad \text{per ogni } x, h \geq 0. \quad (150)$$

Per verificare che tutte le distribuzioni esponenziali soddisfano la versione continua della proprietà di assenza di memoria è quindi sufficiente osservare che per qualsiasi distribuzione esponenziale $F(x)$ si ottiene

$$\frac{1 - F(x + h)}{1 - F(x)} = \frac{1 - [1 - e^{-\theta(x+h)}]}{1 - [1 - e^{-\theta x}]} = e^{-\theta h} = 1 - F(h) \quad \text{per ogni } x, h \geq 0.$$

D'altra parte, come abbiamo accennato poco sopra, si potrebbe anche dimostrare che **le distribuzioni esponenziali sono le uniche distribuzioni che soddisfano la versione continua della proprietà di assenza di memoria**, ma per non appesantire troppo la trattazione non dimostreremo questa affermazione.

Il ruolo delle distribuzioni esponenziali nei processi di Poisson. Come abbiamo visto nel Teorema 8.4, in un **processo di Poisson** con tasso di accadimento $\theta > 0$ si ha

$$\tilde{Y}_{(0,h]} \sim \text{Poisson}(\lambda = \theta h).$$

La probabilità che durante l'intervallo temporale $(0, h]$ non si verificano successi è quindi data da

$$P(\{\tilde{Y}_{(0,h]} = 0\}) = \frac{e^{-\theta h}(\theta h)^0}{0!} = e^{-\theta h}.$$

Tuttavia, l'evento $\{\tilde{Y}_{(0,h]} = 0\}$ può essere anche definito attraverso la variabile casuale \tilde{X}_1 che rappresenta il tempo d'attesa fino al primo successo. Infatti, non è difficile rendersi conto che per ogni $h > 0$ si ha

$$\{\tilde{Y}_{(0,h]} = 0\} = \{\tilde{X}_1 > h\}.$$

Ne consegue che nell'ambito di un processo di Poisson con tasso di accadimento θ si ha

$$P(\{\tilde{X}_1 > h\}) = P(\{\tilde{Y}_{(0,h]} = 0\}) = e^{-\theta h} \quad \text{per ogni } h > 0,$$

ovvero

$$P(\{\tilde{X}_1 \leq h\}) = 1 - P(\{\tilde{X}_1 > h\}) = 1 - e^{-\theta h} \quad \text{per ogni } h > 0.$$

Siccome $\lim_{h \rightarrow 0^+} 1 - e^{-\theta h} = 1 - e^{\theta \times 0} = 0$, possiamo dunque concludere che

$$P(\{\tilde{X}_1 \leq h\}) = \begin{cases} 1 - e^{-\theta h} & \text{per } h \geq 0 \\ 0 & \text{per } h < 0, \end{cases}$$

ovvero che $\tilde{X}_1 \sim \text{Esponenziale}(\theta)$ (si ricordino le proprietà **F0 - F3** che tutte le funzioni di ripartizione devono soddisfare). In realtà si può dimostrare molto di più. Infatti, con riferimento ai tempi di attesa dei **processi di Poisson** vale il seguente teorema:

Teorema 8.7 (Secondo teorema sui processi di Poisson). Nell'ambito di un **processo di Poisson** le variabili casuali $\tilde{X}_1, \tilde{X}_2, \dots$ che descrivono i tempi d'attesa tra successi consecutivi sono i.i.d. con distribuzione esponenziale di parametro θ uguale al tasso di accadimento del processo.

Se nell'ambito di un processo puntuale (di cui a priori non sappiamo se è un processo di Poisson) i tempi d'attesa sono rappresentati da variabili casuali i.i.d. con distribuzione esponenziale di parametro θ , allora il processo puntuale di riferimento deve necessariamente essere un **processo di Poisson**.

Si ricordi che fino ad ora non abbiamo ancora dimostrato l'esistenza di un **processo di Poisson**, ovvero non abbiamo ancora dimostrato l'esistenza di un modello probabilistico (Ω, \mathcal{A}, P) (e di annesse variabili casuali) che soddisfa le condizioni **B0 - B4** e la condizione **B5***. Facendo riferimento al Teorema 8.7 possiamo finalmente colmare questa lacuna. Infatti, come abbiamo già visto, le realizzazioni dei tempi d'attesa di un processo puntuale descrivono univocamente la realizzazione del processo puntuale (vedi il grafico in Figura 8.7) e un modello probabilistico per variabili casuali che rappresentano i tempi d'attesa

di un processo puntuale può quindi essere considerato come un modello probabilistico per il processo puntuale stesso. In virtù del Teorema 8.7 possiamo quindi affermare che un modello probabilistico per variabili casuali i.i.d. con distribuzione esponenziale di parametro θ può essere considerato come un modello probabilistico per un processo di Poisson con tasso di accadimento θ . Siccome l'esistenza di un tale modello probabilistico è assicurata dal Teorema 4.13, possiamo concludere che qualsiasi valore di $\theta > 0$ esiste un modello probabilistico per un **processo di Poisson** con tasso di accadimento uguale a θ . Ricordiamo che questo risultato teorico è di fondamentale importanza: senza di esso ci rimarrebbe il dubbio che attraverso le condizioni B0 - B4, la condizione B5* e gli assiomi di Kolmogorov si possa dimostrare qualsiasi affermazione.

Esercizio 8.16. Si assuma che la distribuzione temporale degli arrivi ad uno sportello sia un processo di Poisson e che nell'arco di 10 minuti arrivino mediamente 2 clienti.

- a) Si calcoli il novantesimo percentile del tempo d'attesa tra due clienti consecutivi.
- b) Qual è la probabilità che il tempo d'attesa tra due clienti consecutivi superi 15 minuti?
- c) Sapendo che da un'ora non si è presentato nessun cliente, qual è la probabilità che anche nei prossimi 15 minuti non si presenti nessun cliente?

Risposte:

- a) Si calcoli il novantesimo percentile del tempo d'attesa tra due clienti consecutivi.

Per ipotesi, il processo puntuale che genera gli arrivi è un processo di Poisson con tasso di accadimento

$$\theta = 2/10 = 1/5 = 0,20 \text{ clienti al minuto.}$$

Pertanto (Teorema 8.7), i tempi d'attesa (espressi in minuti) tra clienti consecutivi sono descritti da variabili casuali indipendenti che hanno tutte distribuzione esponenziale di parametro $\theta = 0,20$.

Il novantesimo percentile della distribuzione esponenziale comune a tutti i tempi d'attesa è dunque dato da

$$x_{0,9} = -\frac{1}{0,2} \ln(1 - 0,9) = 11,513 \text{ minuti.}$$

- b) Qual è la probabilità che il tempo d'attesa tra due clienti consecutivi superi 15 minuti?

Siccome i tempi d'attesa (espressi in minuti) hanno tutti distribuzione esponenziale di parametro $\theta = 0,20$, la probabilità che il tempo d'attesa tra due clienti consecutivi superi 15 minuti è data da

$$P(\{\tilde{X}_i > 15\}) = 1 - F_X(15) = e^{-0,2 \times 15} = 0,050.$$

- c) Sapendo che da un'ora non si è presentato nessun cliente, qual è la probabilità che anche nei prossimi 15 minuti non si presenti nessun cliente?

Dalla proprietà di assenza di memoria delle distribuzioni esponenziali discende che

$$P(\{X_i > 60 + 15\} | \{X_i > 60\}) = P(\{X_i > 15\}) = 0,050$$

come nella risposta al quesito b).

Esercizio 8.17. Si supponga che i tempi d'attesa tra accessi consecutivi ad un server siano descritti da variabili casuali esponenziali indipendenti e che mediamente trascorrono 2 minuti tra due accessi consecutivi.

- a) Qual è la probabilità di dover attendere per più di 3 minuti prima del quinto accesso?
- b) Si determinino il valore atteso e la varianza del tempo d'attesa prima del quinto accesso.

Risposte:

- a) Qual è la probabilità di dover attendere per più di 3 minuti prima del quinto accesso?

Siccome per ipotesi i tempi d'attesa sono descritti da variabili casuali i.i.d. con distribuzione esponenziale di valore atteso $1/\theta = 2$ minuti, possiamo concludere che gli istanti temporali in cui si verificano gli accessi siano generati da un processo di Poisson con tasso di accadimento $\theta = 1/2 = 0,5$ accessi al minuto. Ne consegue che

$$\tilde{Y}_{(t,t+3]} \sim \text{Poisson}(\lambda = \theta \times 3 = 0,5 \times 3 = 1,5),$$

e la probabilità richiesta è dunque data da

$$\begin{aligned} P(\{\tilde{Y}_{(t,t+3]} < 5\}) &= \sum_{y=0}^4 \frac{1,5^y \times e^{-1,5}}{y!} = \\ &= 0,223 + 0,335 + 0,251 + 0,126 + 0,047 = 0,982. \end{aligned}$$

- b) Si determinino il valore atteso e la varianza del tempo d'attesa prima del quinto accesso.

Il valore atteso del tempo d'attesa prima del quinto accesso è dato da

$$\begin{aligned} E(\tilde{X}_1 + \tilde{X}_2 + \dots + \tilde{X}_5) &= [[\text{proprietà E4}]] = \\ &= E(\tilde{X}_1) + E(\tilde{X}_2) + \dots + E(\tilde{X}_5) \\ &= \frac{1}{\theta} + \frac{1}{\theta} + \dots + \frac{1}{\theta} \quad [[5 \text{ volte}]] = 5 \times \frac{1}{\theta} = 5 \times 2 = 10, \end{aligned}$$

e la varianza del tempo d'attesa prima del quinto accesso è invece data da

$$\begin{aligned} Var(\tilde{X}_1 + \tilde{X}_2 + \dots + \tilde{X}_5) &= \\ &[[\text{proprietà V8* \& indipendenza}]] \\ &= Var(\tilde{X}_1) + Var(\tilde{X}_2) + \dots + Var(\tilde{X}_5) \\ &= \frac{1}{\theta^2} + \frac{1}{\theta^2} + \dots + \frac{1}{\theta^2} \quad [[5 \text{ volte}]] \\ &= 5 \times \frac{1}{\theta^2} = 5 \times 2^2 = 20. \end{aligned}$$

8.10 La famiglia delle distribuzioni gamma

La famiglia delle distribuzioni gamma è la famiglia di tutte le distribuzioni assolutamente continue che hanno una funzione di densità del tipo

$$f(x) = \begin{cases} \frac{\theta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\theta x} & \text{per } x > 0 \\ 0 & \text{altrimenti} \end{cases} \quad (151)$$

dove α e θ sono due parametri reali positivi e dove

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy$$

è la **funzione matematica gamma**.

Per verificare che le funzioni (151) siano effettivamente delle funzioni di densità, basta osservare che

- queste funzioni sono definite per ogni $x \in \mathbb{R}$ (così come previsto dalla proprietà f0);
- $\Gamma(\alpha) > 0$ per ogni $\alpha > 0$ (tra breve dimostreremo questa proprietà della funzione matematica gamma) e che nessuna delle suddette funzioni $f(x)$ può quindi assumere valori negativi (così come previsto dalla proprietà f1);

•

$$\begin{aligned} \int_{-\infty}^{\infty} f(x)dx &= \int_0^{\infty} \frac{\theta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\theta x} dx = [[y = \theta x \Rightarrow dy = \theta dx]] = \\ &= \int_0^{\infty} \frac{\theta^\alpha}{\Gamma(\alpha)} \left(\frac{y}{\theta}\right)^{\alpha-1} e^{-y} \frac{1}{\theta} dy = \frac{1}{\Gamma(\alpha)} \int_0^{\infty} y^{\alpha-1} e^{-y} dy \end{aligned}$$

e che l'espressione sul lato destro dell'ultima uguaglianza è uguale a 1 perché l'integrale è uguale a $\Gamma(\alpha)$ (così come previsto dalla proprietà f2).

Come si desume dalla definizione delle funzioni di densità gamma, le distribuzioni gamma concentrano tutta la probabilità sull'insieme dei numeri reali positivi e pertanto si prestano come distribuzioni per variabili casuali che descrivono durate e tempi d'attesa.

D'ora in poi, per comodità di notazione, indicheremo le distribuzioni gamma con $\text{Gamma}(\alpha; \theta)$ e scriveremo $\tilde{X} \sim \text{Gamma}(\alpha; \theta)$ per indicare che \tilde{X} è una variabile casuale con distribuzione $\text{Gamma}(\alpha; \theta)$.

Principali proprietà della funzione matematica $\Gamma(\alpha)$. Per dimostrare le principali proprietà delle distribuzioni gamma, ci serviranno due proprietà fondamentali della funzione matematica gamma:

$$\Gamma(\alpha) = \int_0^{\infty} y^{\alpha-1} e^{-y} dy, \quad \alpha > 0.$$

f0) Per ogni $\alpha \in (0, 1)$ il corrispondente valore di $\Gamma(\alpha)$ è un numero reale positivo.

Dimostrazione. Ovviamente $\Gamma(\alpha) > 0$ per ogni $\alpha > 0$ perché $y^{\alpha-1} e^{-y} > 0$ per ogni $y > 0$ e ogni $\alpha \in \mathbb{R}$. Per dimostrare la proprietà f0 è quindi sufficiente dimostrare che $\Gamma(\alpha) < \infty$ per ogni $\alpha \in (0, 1)$. A tal fine basta osservare che per ogni $\alpha \in (0, 1)$ si ha

$$\begin{aligned} \Gamma(\alpha) &= \int_0^{\infty} y^{\alpha-1} e^{-y} dy = \int_0^1 y^{\alpha-1} e^{-y} dy + \int_1^{\infty} y^{\alpha-1} e^{-y} dy \leq \int_0^1 y^{\alpha-1} dy + \int_1^{\infty} e^{-y} dy \\ &= \left[\frac{y^\alpha}{\alpha} \right]_0^1 + [-e^{-y}]_1^{\infty} = \frac{1}{\alpha} - 0 + 0 - (-1) = \frac{1}{\alpha} + 1 < \infty. \end{aligned}$$

□

f1) In corrispondenza di $\alpha = 1$ si ha $\Gamma(1) = 1$.

Dimostrazione. La dimostrazione è immediata:

$$\Gamma(1) = \int_0^{\infty} y^{1-1} e^{-y} dy = \int_0^{\infty} y^0 e^{-y} dy = \int_0^{\infty} e^{-y} dy = [-e^{-y}]_0^{\infty} = 0 - (-1) = 1.$$

□

$\Gamma 2)$ Per ogni $\alpha > 1$ si ha

$$\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1),$$

e se $\alpha > 0$ è un numero intero otteniamo dunque

$$\Gamma(\alpha) = (\alpha - 1)! \quad [[\text{il fattoriale di } \alpha - 1]].$$

Dimostrazione. Integrando per parti otteniamo

$$\Gamma(\alpha) = \int_0^{\infty} y^{\alpha-1} e^{-y} dy = \left[y^{\alpha-1} \times \frac{e^{-y}}{(-1)} \right]_0^{\infty} - \int_0^{\infty} (\alpha-1)y^{\alpha-2} \times \frac{e^{-y}}{(-1)} dy,$$

e siccome $\alpha > 1$ implica

$$\left[y^{\alpha-1} \times \frac{e^{-y}}{(-1)} \right]_0^{\infty} = 0,$$

possiamo concludere che con $\alpha > 1$ si ottiene

$$\Gamma(\alpha) = (\alpha - 1) \int_0^{\infty} y^{\alpha-2} e^{-y} dy = (\alpha - 1)\Gamma(\alpha - 1).$$

come volevasi dimostrare. □

Usando le proprietà $\Gamma 0$ - $\Gamma 2$ si vede che per ogni $\alpha > 0$ il corrispondente valore di $\Gamma(\alpha)$ è un numero reale positivo.

Il ruolo dei parametri α e θ . Come si desume dal modo in cui abbiamo definito le funzioni di densità gamma,

$$f(x) = \begin{cases} \frac{\theta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\theta x} & \text{per } x > 0 \\ 0 & \text{altrimenti} \end{cases}, \quad \alpha, \theta > 0,$$

il parametro θ è un **parametro di scala** inverso che non incide sulla forma delle funzioni di densità gamma. Infatti, esprimendo le funzioni di densità gamma come

$$f(x) = \begin{cases} \frac{\theta}{\Gamma(\alpha)} (\theta x)^{\alpha-1} e^{-\theta x} & \text{per } x > 0 \\ 0 & \text{altrimenti} \end{cases}, \quad \alpha, \theta > 0,$$

vediamo che per ottenere il grafico di un funzione di densità gamma con $\theta \neq 1$ è sufficiente disegnare il grafico della funzione di densità gamma con lo stesso valore del parametro α e con parametro θ uguale a 1 e poi

- dividere per θ i valori riportati sull'asse delle ascisse

- e moltiplicare per θ i valori riportati sull'asse delle ordinate.

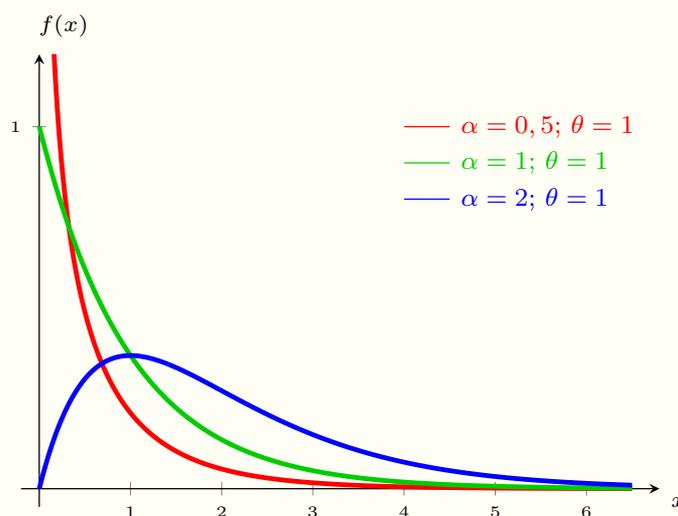
Secondo la parametrizzazione che usiamo in questa dispensa, anche per le distribuzioni gamma il parametro θ è dunque un **parametro di scala** inverso nel senso che

$$\tilde{X} \sim \text{Gamma}(\alpha; \theta) \quad \Rightarrow \quad \tilde{Y} = b\tilde{X} \sim \text{Gamma}(\alpha; \theta/b) \quad \text{per ogni } b > 0 .$$

Questa implicazione può essere facilmente dimostrata seguendo la falsariga della dimostrazione dell'analogia implicazione (147) che si riferisce alle distribuzioni esponenziali.

Siccome θ è un parametro di scala inverso, possiamo analizzare la "forma" di una funzione di densità gamma considerando soltanto il caso in cui $\theta = 1$. Ovviamente, la forma di una funzione di densità gamma può a questo punto dipendere soltanto più dal valore del parametro α . Il grafico in Figura 8.10 mostra l'andamento di alcune funzioni di densità gamma con diversi valori del parametro α e con θ fisso e uguale a 1.

Figura 8.10. Il grafico sottostante mostra l'andamento di alcune funzioni di densità gamma con diversi valori del parametro α e con $\theta = 1$.



Come si vede in Figura 8.10, a seconda del valore di α il grafico di una funzione di densità gamma

$$f(x) = \begin{cases} \frac{\theta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\theta x} & \text{per } x > 0 \\ 0 & \text{altrimenti} \end{cases}, \quad \alpha, \theta > 0,$$

può avere un andamento molto diverso. Infatti, non è difficile verificare che

- $\alpha < 1 \Rightarrow \lim_{x \rightarrow 0^+} f(x) = +\infty$;
- le funzioni di densità gamma con $\alpha = 1$ sono funzioni di densità esponenziali (ricordiamo che $\Gamma(1) = 1$) e che in questo caso $\lim_{x \rightarrow 0^+} f(x) = \theta$;

- $\alpha > 1 \Rightarrow \lim_{x \rightarrow 0^+} f(x) = 0$.

Per descrivere in modo più dettagliato il ruolo del parametro α conviene indagare anche sull'esistenza e sulla posizione di eventuali punti di massimo di una funzione di densità gamma:

$$f(x) = \begin{cases} \frac{\theta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\theta x} & \text{per } x > 0 \\ 0 & \text{altrimenti} \end{cases}, \quad \alpha, \theta > 0.$$

A tal fine osserviamo che per $x > 0$

$$f'(x) = \frac{df(x)}{dx} = \frac{\theta^\alpha}{\Gamma(\alpha)} x^{\alpha-2} e^{-\theta x} [(\alpha - 1) - \theta x]$$

e da questa espressione deduciamo che per $x > 0$

- le funzioni di densità gamma con $\alpha \leq 1$ sono strettamente decrescenti;
- le funzioni di densità gamma con $\alpha > 1$ hanno un unico punto di massimo (ovvero un'unica **"moda"**) che si trova in corrispondenza di

$$f'(x) = 0 \quad \Leftrightarrow \quad x = \frac{\alpha - 1}{\theta}.$$

Le distribuzioni chi-quadrato. Come abbiamo visto poc'anzi, la famiglia delle distribuzioni gamma contiene tutte le distribuzioni esponenziali (infatti, ponendo $\alpha = 1$ si possono ottenere tutte le distribuzioni esponenziali) e anche molte altre distribuzioni. In particolare, la famiglia delle distribuzioni gamma contiene anche tutte le cosiddette **distribuzioni chi-quadrato**. Queste ultime sono infatti le distribuzioni che si ottengono ponendo $\theta = 1/2$ e $\alpha = k/2$ con $k = 1, 2, \dots$. Il valore di $k = 1, 2, \dots$ che identifica una distribuzione chi-quadrato viene chiamato **"numero di gradi di libertà"**. Le distribuzioni chi-quadrato hanno un ruolo molto importante nell'inferenza statistica sul quale faremo solo qualche cenno marginale perché esula dagli argomenti trattati in questa dispensa.

Funzione di ripartizione. In generale, la funzione di ripartizione di una distribuzione gamma, ovvero la funzione

$$F(x) = \int_{-\infty}^x f(t) dt = \begin{cases} \int_0^x \frac{\theta^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\theta t} dt & \text{per } x > 0, \\ 0 & \text{per } x \leq 0, \end{cases}$$

non è una **funzione elementare**²⁸ e i suoi valori non possono dunque essere calcolati attraverso una tradizionale calcolatrice tascabile. Tuttavia, **se il parametro α è un**

²⁸In questa dispensa non daremo una definizione rigorosa del concetto di **"funzione elementare"**. Per "funzione elementare" intenderemo (in modo un po' vago) una funzione i cui valori possono essere calcolati attraverso una tradizionale calcolatrice tascabile.

numero intero, la funzione di ripartizione gamma può essere espressa come

$$F(x) = \begin{cases} 1 - \sum_{y=0}^{\alpha-1} \frac{(\theta x)^y e^{-\theta x}}{y!} & \text{per } x > 0, \\ 0 & \text{per } x \leq 0. \end{cases} \quad (152)$$

Questa formula può essere dimostrata attraverso il principio di induzione oppure, come vedremo più avanti, facendo riferimento alla teoria dei [processi di Poisson](#).

Dimostrazione (Dimostrazione della formula (152) attraverso il principio di induzione). Per dimostrare la formula (152) che, ricordiamo, è valida soltanto nel caso in cui α è un numero **intero positivo**, dimostreremo che (per ogni $\alpha = 1, 2, \dots$)

$$1 - F(x) = \int_x^{\infty} \frac{\theta^\alpha}{\Gamma(\alpha)} \times s^{\alpha-1} e^{-\theta s} ds = \sum_{y=0}^{\alpha-1} \frac{(\theta x)^y e^{-\theta x}}{y!} \quad \text{per ogni } x > 0. \quad (153)$$

Procederemo per induzione sul valore di α . Per $\alpha = 1$ la funzione di densità gamma si riduce a quella esponenziale e in questo caso abbiamo già visto che

$$F(x) = \begin{cases} 0 & \text{per } x \leq 0 \\ 1 - e^{-\theta x} & \text{per } x > 0 \end{cases}$$

è la funzione di ripartizione. Per $\alpha = 1$ la formula (153) restituisce dunque il risultato corretto.

Assumiamo ora che la formula (153) sia valida anche per un generico numero naturale positivo α (**ipotesi di induzione**). Per compiere il **passo induttivo** e completare la dimostrazione, dobbiamo soltanto più dimostrare che in questo caso la formula (153) deve essere valida anche per $\alpha + 1$. I seguenti passaggi algebrici confermano che questo è vero:

$$\begin{aligned} 1 - F(x) &= \int_x^{\infty} f(s) ds = \int_x^{\infty} \frac{\theta^{\alpha+1}}{\Gamma(\alpha+1)} \times s^{(\alpha+1)-1} e^{-\theta s} ds = \\ &= \left[\frac{\theta^{\alpha+1}}{\Gamma(\alpha+1)} \times s^\alpha \frac{e^{-\theta s}}{(-\theta)} \right]_x^{\infty} - \int_x^{\infty} \frac{\theta^{\alpha+1}}{\Gamma(\alpha+1)} \times \alpha s^{\alpha-1} \frac{e^{-\theta s}}{(-\theta)} ds = \\ &= \frac{(\theta x)^\alpha e^{-\theta x}}{\alpha!} + \int_x^{\infty} \frac{\theta^\alpha}{\Gamma(\alpha)} \times s^{\alpha-1} e^{-\theta s} ds = [[\text{ipotesi di induzione}]] = \\ &= \frac{(\theta x)^\alpha e^{-\theta x}}{\alpha!} + \sum_{y=0}^{\alpha-1} \frac{(\theta x)^y e^{-\theta x}}{y!} = \sum_{y=0}^{\alpha} \frac{(\theta x)^y e^{-\theta x}}{y!}. \end{aligned}$$

□

Funzione generatrice dei momenti, valore atteso e varianza. Anche le distribuzioni gamma sono tutte dotate di fgm e la fgm delle distribuzioni gamma è definita come

$$m(t) = \left(\frac{\theta}{\theta - t} \right)^\alpha \quad \text{per } t < \theta. \quad (154)$$

Dimostrazione (Fgm delle distribuzioni gamma). In primo luogo dimostreremo che tutte le distribuzioni gamma sono dotate di fgm. A tal fine osserviamo che per funzioni di densità gamma si ha

$$\int_{-\infty}^{\infty} e^{tx} f(x) dx = \int_0^{\infty} e^{tx} \frac{\theta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\theta x} dx = \frac{\theta^\alpha}{\Gamma(\alpha)} \int_0^{\infty} x^{\alpha-1} e^{-(\theta-t)x} dx \quad (155)$$

che l'integrale alla destra dell'ultima uguaglianza è divergente se $t \geq \theta$. Infatti,

- con $\theta > t$ l'integrale è divergente perché la funzione integranda tende a infinito per $x \rightarrow \infty$,
- e con $\theta = t$ l'integrale è divergente perché con $\alpha > 0$ si ottiene

$$\int_0^{\infty} x^{\alpha-1} e^{-(\theta-t)x} dx = \int_0^{\infty} x^{\alpha-1} e^{0 \times x} dx = \int_0^{\infty} x^{\alpha-1} dx = \left[\frac{y^\alpha}{\alpha} \right]_0^{\infty} = \infty.$$

D'altra parte, per $t < \theta$ l'integrale sul lato destro dell'ultima uguaglianza nella (155) è convergente perché

$$\begin{aligned} \int_0^{\infty} x^{\alpha-1} e^{-(\theta-t)x} dx &= \text{[[cambio variabile: } y = (\theta-t)x \Rightarrow dy = (\theta-t)dx \text{]]} = \\ &= \int_0^{\infty} \left(\frac{y}{\theta-t} \right)^{\alpha-1} e^{-y} \frac{1}{\theta-t} dy = \frac{1}{(\theta-t)^\alpha} \int_0^{\infty} y^{\alpha-1} e^{-y} dy = \frac{1}{(\theta-t)^\alpha} \times \Gamma(\alpha). \end{aligned} \quad (156)$$

Siccome il parametro θ di una distribuzione gamma è un numero reale positivo, possiamo dunque concludere che tutte le distribuzioni gamma sono dotate di fgm e che per ogni $t < \theta$ quest'ultima è definita come

$$m(t) = \int_0^{\infty} \frac{\theta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-(\theta-t)x} dx = \text{[[formule (155) e (156)]]} = \frac{\theta^\alpha}{\Gamma(\alpha)} \times \frac{1}{(\theta-t)^\alpha} \times \Gamma(\alpha) = \left(\frac{\theta}{\theta-t} \right)^\alpha.$$

□

Come abbiamo appena visto, la fgm di una distribuzione gamma è definita come

$$m(t) = \left(\frac{\theta}{\theta-t} \right)^\alpha = \theta^\alpha (\theta-t)^{-\alpha} \quad \text{per } t < \theta.$$

Derivando questa funzione otteniamo delle espressioni per il valore atteso e la varianza di una distribuzione gamma (Teorema 7.1):

•

$$\begin{aligned} m'(t) &= \frac{dm(t)}{dt} = \theta^\alpha (-\alpha) (\theta-t)^{-\alpha-1} (-1) = \alpha \theta^\alpha (\theta-t)^{-\alpha-1} \Rightarrow \\ &\Rightarrow \mu = E(\tilde{X}) = m'(0) = \alpha/\theta; \end{aligned} \quad (157)$$

•

$$\begin{aligned} m''(t) &= \frac{d^2 m(t)}{dt^2} = \alpha \theta^\alpha (-\alpha-1) (\theta-t)^{-\alpha-2} (-1) = \alpha(\alpha+1) \theta^\alpha (\theta-t)^{-\alpha-2} \Rightarrow \\ &\Rightarrow \mu_2 = E(\tilde{X}^2) = m''(0) = \alpha(\alpha+1)/\theta^2 \Rightarrow \\ &\Rightarrow \sigma^2 = \text{var}(\tilde{X}) = \mu_2 - \mu^2 = \alpha/\theta^2. \end{aligned} \quad (158)$$

Proprietà riproduttiva. Seguendo la falsariga della dimostrazione della proprietà riproduttiva delle distribuzioni binomiali (vedi il Teorema 8.1), si dimostra facilmente che anche le distribuzioni gamma soddisfano questa proprietà.

Teorema 8.8 (Proprietà riproduttiva delle distribuzioni gamma). Se

$$\begin{aligned} \tilde{X}_1 &\sim \text{Gamma}(\alpha_1; \theta), & \tilde{X}_2 &\sim \text{Gamma}(\alpha_2; \theta), & \dots, \\ & & \dots, & & \tilde{X}_k &\sim \text{Gamma}(\alpha_k; \theta) \end{aligned}$$

(si noti che tutte le distribuzioni gamma hanno lo stesso parametro θ), e se e variabili casuali $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_k$ sono indipendenti, allora

$$\tilde{X} = \tilde{X}_1 + \tilde{X}_2 + \dots + \tilde{X}_k \sim \text{Gamma}(\alpha = \alpha_1 + \alpha_2 + \dots + \alpha_k; \theta).$$

In particolare, se

$$\begin{aligned} \tilde{X}_1 &\sim \text{Esponenziale}(\theta), & \tilde{X}_2 &\sim \text{Esponenziale}(\theta), & \dots, \\ & & \dots, & & \tilde{X}_k &\sim \text{Esponenziale}(\theta), \end{aligned}$$

e se e variabili casuali $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_k$ sono indipendenti, allora

$$\tilde{X} = \tilde{X}_1 + \tilde{X}_2 + \dots + \tilde{X}_k \sim \text{Gamma}(\alpha = k; \theta).$$

Si noti che la seconda parte del precedente teorema segue dal fatto che tutte le distribuzioni esponenziali sono distribuzioni gamma con parametro $\alpha = 1$.

Combinando la seconda parte del Teorema 8.8 con il fatto che nei [processi di Poisson](#) i tempi d'attesa sono descritti da variabili casuali i.i.d. che hanno distribuzione esponenziale (vedi il Teorema 8.7), vediamo che nell'ambito di un [processo di Poisson](#) le somme dei tempi d'attesa hanno distribuzione gamma, così come nei [processi di Bernoulli](#) hanno distribuzione binomiale negativa. Sfruttando questo fatto possiamo ottenere una dimostrazione alternativa della formula (152) che esprime le funzioni di ripartizione gamma con parametro α intero:

Dimostrazione (Dimostrazione alternativa della formula (152)). Consideriamo un [processi di Poisson](#) con tasso di accadimento θ e indichiamo con \tilde{X} la variabile casuale che rappresenta il tempo d'attesa fino al successo numero α (stiamo assumendo che α sia un numero intero positivo), e con $\tilde{Y}_{(0,x]}$ la variabile casuale che conta i successi che si sono verificati durante l'intervallo temporale $(0, x]$. Siccome

$$\{\tilde{X} \leq x\} = \{\tilde{Y}_{(0,x]} \geq \alpha\},$$

(si noti che i due eventi sarebbero comunque equivalenti anche se il processo puntuale di riferimento non fosse un [processo di Poisson](#)), le probabilità di questi due eventi devono essere identiche. Siccome nell'ambito di un [processo di Poisson](#) i tempi d'attesa tra successi consecutivi sono i.i.d. con distribuzione esponenziale di parametro θ uguale al tasso di accadimento del processo (si veda il Teorema 8.7), e siccome il tempo d'attesa fino al successo numero α è uguale alla somma dei primi α tempi d'attesa, possiamo fare riferimento alla seconda parte del Teorema 8.8 onde concludere che

$$\tilde{X} \sim \text{Gamma}(\alpha; \theta)$$

(ricordiamo che α è il numero di successi e che θ è il tasso di accadimento del [processo di Poisson](#)). D'altra parte, secondo il Teorema 8.4, nell'ambito di un [processo di Poisson](#) con tasso di accadimento θ si deve avere

$$\tilde{Y}_{(0,x]} \sim \text{Poisson}(\lambda = \theta x).$$

Quindi possiamo concludere che

$$P(\{\tilde{X} \leq x\}) = P(\{\tilde{Y}_{(0,x]} \geq \alpha\}) = 1 - \sum_{y=0}^{\alpha-1} \frac{\lambda^y e^{-\theta x}}{y!}$$

e questa espressione coincide con quella nella formula (152) che volevamo dimostrare. \square

Facendo riferimento alla seconda parte del Teorema 8.8 possiamo anche dimostrare le espressioni per il valore atteso e la varianza di una distribuzione gamma **sotto l'ipotesi che il parametro α sia un numero intero**. Infatti, se α è un numero intero e se

$$\begin{aligned} \tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_\alpha &\sim \text{i.i.d. Esponenziale}(\theta) \Rightarrow \\ &\Rightarrow \text{[[seconda parte del Teorema 8.8]]} \Rightarrow \\ &\Rightarrow \tilde{X} = \tilde{X}_1 + \tilde{X}_2 + \dots + \tilde{X}_\alpha \sim \text{Gamma}(\alpha; \theta), \end{aligned}$$

allora possiamo scrivere

$$\begin{aligned} \mu = E(\tilde{X}) &= E(\tilde{X}_1 + \tilde{X}_2 + \dots + \tilde{X}_\alpha) \\ \text{[[proprietà E4]]} &= E(\tilde{X}_1) + E(\tilde{X}_2) + \dots + E(\tilde{X}_\alpha) \\ &\text{[[valore atteso delle distribuzioni esponenziali]]} \\ &= \frac{1}{\theta} + \frac{1}{\theta} + \dots \text{[[} \alpha \text{ volte]]} + \frac{1}{\theta} = \frac{\alpha}{\theta} \end{aligned}$$

e

$$\begin{aligned} \sigma^2 = \text{var}(\tilde{X}) &= \text{var}(\tilde{X}_1 + \tilde{X}_2 + \dots + \tilde{X}_\alpha) \\ &\text{[[proprietà V8*]]} \\ &= \text{var}(\tilde{X}_1) + \text{var}(\tilde{X}_2) + \dots + \text{var}(\tilde{X}_\alpha) \\ &\text{[[varianza delle distribuzioni esponenziali]]} \\ &= \frac{1}{\theta^2} + \frac{1}{\theta^2} + \dots \text{[[} \alpha \text{ volte]]} + \frac{1}{\theta^2} = \frac{\alpha}{\theta^2}. \end{aligned}$$

Si noti che queste due formule sono identiche alle formule (157) e (158) che abbiamo ricavato a partire dalla fgm.

Probabilità condizionate di sopravvivenza. Chiaramente, visto il loro ruolo nell'ambito dei [processi di Poisson](#), anche le distribuzioni gamma si prestano come distribuzioni per tempi d'attesa e quindi conviene fare qualche breve cenno sull'andamento delle [probabilità condizionate di sopravvivenza](#) di variabili casuali \tilde{X} che hanno distribuzioni

di gamma. Si noti che le [probabilità condizionate di sopravvivenza](#) delle distribuzioni gamma sono definite come

$$P(\{\tilde{X} > x + h\}|\{\tilde{X} > x\}) = \frac{\int_{x+h}^{\infty} \frac{\theta}{\Gamma(\alpha)} (\theta t)^{\alpha-1} e^{-\theta t} dt}{\int_x^{\infty} \frac{\theta}{\Gamma(\alpha)} (\theta t)^{\alpha-1} e^{-\theta t} dt}, \quad \text{per } x, h \geq 0,$$

e che con $x = 0$ si ha

$$P(\{\tilde{X} > x + h\}|\{\tilde{X} > x\}) = P(\{\tilde{X} > h\}) = \int_h^{\infty} \frac{\theta}{\Gamma(\alpha)} (\theta t)^{\alpha-1} e^{-\theta t} dt, \quad \text{per } h \geq 0.$$

Con le distribuzioni gamma si possono ottenere tre tipi di andamento (per non appesantire troppo la trattazione non dimostreremo le affermazioni sull'andamento delle [probabilità condizionate di sopravvivenza](#) delle distribuzioni gamma):

- Andamento decrescente. Se $\alpha < 1$, la [probabilità condizionata di sopravvivenza](#), a parità di orizzonte temporale futuro h , è strettamente decrescente all'aumentare del tempo d'attesa x che è già trascorso.
- Andamento costante. Se $\alpha = 1$ (si ricordi che le distribuzioni gamma con $\alpha = 1$ sono distribuzioni esponenziali), la [probabilità condizionata di sopravvivenza](#), a parità di orizzonte temporale futuro h , rimane costante all'aumentare del tempo d'attesa x che è già trascorso.
- Andamento crescente. Se $\alpha > 1$, la [probabilità condizionata di sopravvivenza](#), a parità di orizzonte temporale futuro h , è strettamente crescente all'aumentare del tempo d'attesa x che è già trascorso (si noti che nell'ambito di un [processo di Poisson](#) le distribuzioni gamma con α intero e maggiore di 1 hanno un ruolo analogo a quello delle distribuzioni binomiali negative con $r > 1$ nell'ambito di un [processo di Bernoulli](#)).

Inoltre, usando la regola di de L'Hôpital si può facilmente dimostrare che

$$\lim_{x \rightarrow \infty} P(\{\tilde{X} > x + h\}|\{\tilde{X} > x\}) = \lim_{x \rightarrow \infty} \frac{\int_{x+h}^{\infty} \frac{\theta}{\Gamma(\alpha)} (\theta t)^{\alpha-1} e^{-\theta t} dt}{\int_x^{\infty} \frac{\theta}{\Gamma(\alpha)} (\theta t)^{\alpha-1} e^{-\theta t} dt} = e^{-\theta h}.$$

Questo risultato ci dice che, per qualsiasi prefissato valore dell'orizzonte temporale futuro h , se il tempo d'attesa x che è già trascorso tende a infinito, la [probabilità condizionata di sopravvivenza](#) tende ad un limite che è strettamente positivo.

Per illustrare un'applicazione delle distribuzioni gamma mostreremo ora come si può rispondere quesiti dell'Esercizio 8.16 facendo riferimento alle distribuzioni gamma.

Esercizio 8.18. Si supponga che i tempi d'attesa tra accessi consecutivi ad un server siano descritti da variabili casuali esponenziali indipendenti e che mediamente

trascorrano 2 minuti tra due accessi consecutivi.

- a) Qual è la probabilità di dover attendere per più di 3 minuti prima del quinto accesso?
- b) Si determinino il valore atteso e la varianza del tempo d'attesa prima del quinto accesso.

Risposte:

- a) Qual è la probabilità di dover attendere per più di 3 minuti prima del quinto accesso?

Siano $\tilde{X}_1, \tilde{X}_2, \dots$ le variabili casuali che rappresentano i tempi d'attesa tra gli accessi consecutivi al server e sia $\tilde{W}_5 = \tilde{X}_1 + \tilde{X}_2 + \dots + \tilde{X}_5$ la variabile casuale che rappresenta il tempo d'attesa fino al quinto accesso. Siccome stiamo assumendo che le variabili casuali \tilde{X}_i siano i.i.d. con distribuzione esponenziale di parametro

$$E(\tilde{X}_i) = \frac{1}{\theta} = 2 \quad \Rightarrow \quad \theta = 1/2,$$

possiamo concludere che (Teorema 8.8))

$$\tilde{W}_5 \sim \text{Gamma}(\alpha = 5; \theta = 1/2).$$

Siccome il valore del parametro α è un numero naturale, possiamo applicare la formula (152) onde calcolare la probabilità richiesta:

$$\begin{aligned} P(\{\tilde{W}_5 > 3\}) &= \sum_{y=0}^{5-1} \frac{(0,5 \times 3)^y \times e^{-(0,5 \times 3)}}{y!} \\ &= 0,223 + 0,335 + 0,251 + 0,126 + 0,047 = 0,982. \end{aligned}$$

Si noti che questo risultato coincide con il valore di $P(\{\tilde{Y}_{(t,t+3]} < 5\})$ che abbiamo ottenuto nell'Esempio ??.

- b) Si determinino il valore atteso e la varianza del tempo d'attesa prima del quinto accesso.

Usando le formule per il valore atteso e la varianza della distribuzione gamma otteniamo direttamente

$$E(\tilde{W}_5) = \frac{\alpha}{\theta} = \frac{5}{1/2} = 10$$

e

$$\text{Var}(\widetilde{W}_5) = \frac{\alpha}{\theta^2} = \frac{5}{(1/2)^2} = 20.$$

Si noti che nella risposta al quesito b) dell'Esempio ?? abbiamo ottenuto lo stesso risultato senza ricorrere alle formule per il valore atteso e la varianza delle distribuzioni gamma, ma usando soltanto le formule per il valore atteso e la varianza delle distribuzioni esponenziali e le proprietà generali del valore atteso e della varianza.

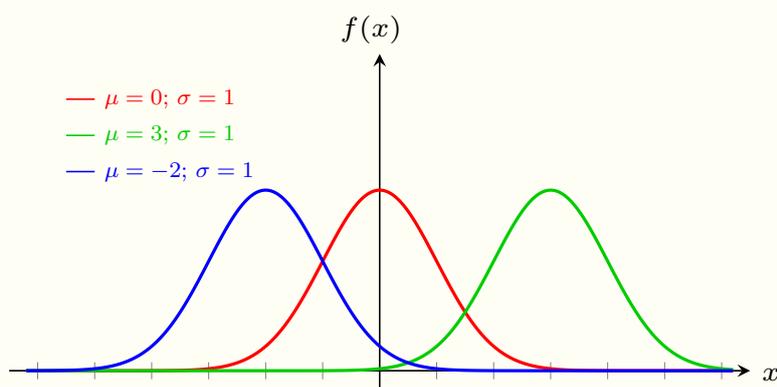
8.11 La famiglia delle distribuzioni normali

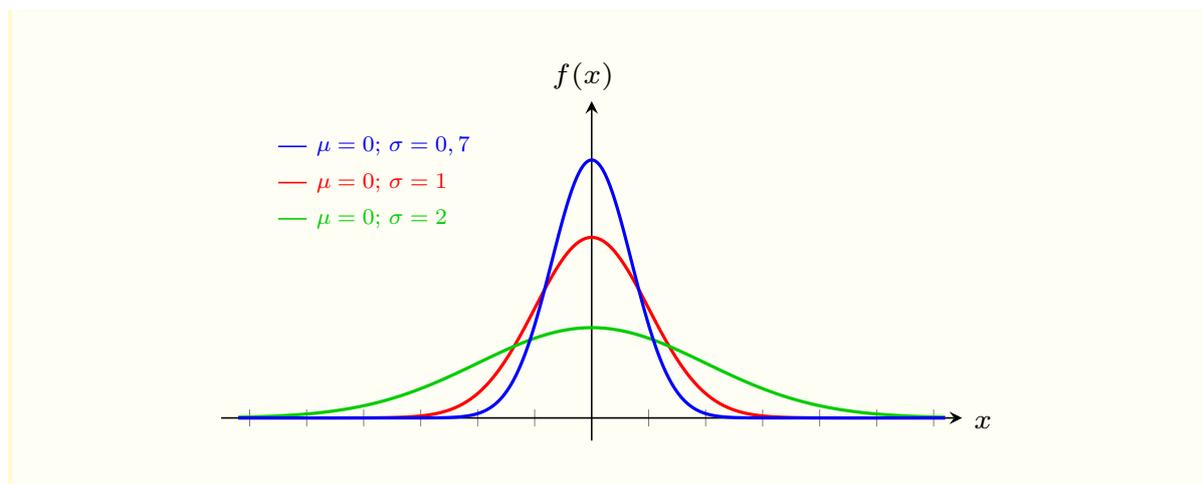
La famiglia delle distribuzioni normali è la famiglia di tutte le distribuzioni assolutamente continue che hanno una funzione di densità del tipo

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad x \in \mathbb{R},$$

dove il parametro μ può essere un numero reale qualunque, e dove il parametro σ deve essere un numero reale strettamente positivo. Come vedremo tra breve, i parametri μ e σ rappresentano, rispettivamente, il valore atteso e lo scarto quadratico medio di una distribuzione normale. I grafici in Figura 8.11 mostrano alcune funzioni di densità normali.

Figura 8.11. I grafici mostrano alcune funzioni di densità normali.





In questa dispensa, per comodità di notazione, scriveremo $Normale(\mu; \sigma)$ per indicare la distribuzione normale con valore atteso μ e scarto quadratico medio σ , e come con tutte le altre distribuzioni scriveremo

$$\tilde{X} \sim Normale(\mu; \sigma)$$

per indicare che la distribuzione di una data variabile casuale \tilde{X} è $Normale(\mu; \sigma)$. A proposito della notazione $Normale(\mu; \sigma)$, conviene mettere in guardia sul fatto che alcuni autori identificano le distribuzioni normali attraverso la varianza σ^2 piuttosto che attraverso lo scarto quadratico medio σ .

Le distribuzioni normali hanno un ruolo centrale in tante applicazioni della teoria della probabilità. Infatti, le osservazioni di molte variabili che descrivono fenomeni reali tendono a distribuirsi secondo la legge normale (in altre parole: gli istogrammi che si ottengono a partire dalle osservazioni di molte variabili sono spesso molto simili ad una funzione di densità normale) e per descrivere i processi che generano queste osservazioni si fa dunque spesso riferimento a distribuzioni normali.

D'altra parte, anche quando non si dispone di dati empirici, spesso si può comunque giustificare l'impiego di distribuzioni normali facendo riferimento al cosiddetto **"teorema del limite centrale"** (d'ora in poi TLC). Infatti, secondo questo teorema la distribuzione della somma di un gran numero di variabili casuali che hanno tutte più o meno la stessa varianza e le cui covarianze sono in un certo senso trascurabili può essere approssimata attraverso una distribuzione normale.

Tra breve riporteremo l'enunciato di quella che può essere considerata la versione base del TLC (ne esistono tante altre adattate a diversi problemi specifici) e vedremo come attraverso la versione base del TLC si possono ottenere delle approssimazioni normali per molte distribuzioni notevoli.

Il nome "normale" delle distribuzioni che analizzeremo in questa sezione sembra essere dovuto al fatto che Johann Carl Friedrich Gauss (1777 - 1855) dedusse queste distribuzioni

a partire dal sistema di "equazioni normali" che bisogna risolvere quando si applica il metodo dei minimi quadrati. In virtù di questo fatto, le distribuzioni normali vengono spesso chiamate "**gaussiane**", anche se un autore noto e influente come Karl Pearson (1857 - 1936) preferisce chiamarle "distribuzioni di Gauss-Laplace" per dare il giusto riconoscimento ad alcuni importanti contributi teorici di Pierre-Simon Laplace (1749-1827).²⁹ Infatti, a quanto pare Laplace fu il primo a dimostrare l'uguaglianza

$$\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}$$

per determinare la costante di normalizzazione delle distribuzioni normali (usando questa uguaglianza si può facilmente dimostrare che tutte le funzioni di densità normali soddisfano la proprietà f2), e inoltre sembra che Laplace fu anche il primo a dimostrare il teorema del limite centrale di cui abbiamo già parlato poc'anzi.

Grafici delle funzioni di densità normali. Come si desume dalla loro definizione

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad x \in \mathbb{R},$$

le funzioni di densità normali sono simmetriche rispetto al punto $x = \mu$ e decrescono e tendono a zero man mano che ci si allontana dal punto $x = \mu$. Aumentando o diminuendo il valore del parametro μ la funzione di densità normale non cambia "forma", ma si sposta, a seconda del caso, verso destra oppure verso sinistra (vedi i grafici in Figura 8.11).

Da queste osservazioni possiamo dedurre alcune importanti proprietà della famiglia delle distribuzioni normali. Innanzitutto, deduciamo che μ è la "moda" (ovvero il punto di massimo) di una funzione di densità normale. E siccome la mediana e il valore atteso (se esiste) di una distribuzione assolutamente continua che ha una funzione di densità simmetrica coincidono con il centro di simmetria della funzione di densità (lasciamo al lettore il compito di dimostrare questa affermazione), possiamo inoltre concludere che il parametro μ di una distribuzione normale rappresenta anche la sua mediana e il suo valore atteso (di cui tra breve dimostreremo anche l'esistenza). Infine, siccome il parametro μ non incide sulla "forma" della funzione di densità normale, ma determina soltanto la posizione del suo centro di simmetria, possiamo anche concludere

$$\begin{aligned} \tilde{X} &\sim \text{Normale}(\mu = \mu_{\tilde{X}}; \sigma = \sigma_{\tilde{X}}) \quad \Rightarrow \\ \Rightarrow \tilde{Y} = a + \tilde{X} &\sim \text{Normale}(\mu = a + \mu_{\tilde{X}}; \sigma = \sigma_{\tilde{X}}) \quad \text{per ogni } a \in \mathbb{R}, \end{aligned} \quad (159)$$

ovvero che μ è un cosiddetto **parametro di posizione**.

²⁹Vedi Pearson K. (1905), "Das Fehlergesetz und Seine Verallgemeinerungen Durch Fechner und Pearson." *A Rejoinder, Biometrika, Vol 4, No. 1/2, pp. 169-212.*

Dimostrazione (Dimostrazione dell'implicazione (159)). Per dimostrare l'implicazione (159) basta osservare che

$$\begin{aligned} P(\{\tilde{Y} \leq y\}) &= P(\{a + \tilde{X} \leq y\}) = P(\{\tilde{X} \leq y - a\}) = \int_{-\infty}^{y-a} \frac{1}{\sigma_{\tilde{X}}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu_{\tilde{X}}}{\sigma_{\tilde{X}}}\right)^2} dx = \\ &= [[\text{cambio di variabile: } t = x + a \Rightarrow dt = dx]] = \\ &= \int_{-\infty}^y \frac{1}{\sigma_{\tilde{X}}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-(a+\mu_{\tilde{X}})}{\sigma_{\tilde{X}}}\right)^2} dt \end{aligned}$$

e che l'ultimo integrale è la funzione di ripartizione di una distribuzione normale con $\mu = a + \mu_{\tilde{X}}$ e con $\sigma = \sigma_{\tilde{X}}$. \square

Avendo descritto il ruolo del parametro μ , rimane da interpretare quello del parametro σ . Come si vede ispezionando l'espressione della funzione di densità

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad x \in \mathbb{R},$$

il parametro σ è un parametro di scala, nel senso che il grafico di una funzione di densità normale con $\sigma \neq 1$ può essere ottenuto a partire dal grafico della funzione di densità normale con lo stesso valore del parametro μ e con $\sigma = 1$, modificando soltanto le scale di valori riportate sui due assi cartesiani:

- i valori sull'asse delle ascisse devono essere moltiplicati per il valore di σ
- mentre i valori sull'asse delle ordinate devono essere divisi per il valore di σ .

Si noti che l'effetto del valore del parametro σ sulle scale di valori riportate sugli assi cartesiani è opposto a quello del parametro θ delle distribuzioni esponenziali e gamma. In virtù di questo fatto, il parametro σ reagisce in modo diverso a trasformazioni di scala di variabili casuali \tilde{X} che hanno distribuzione normale. Seguendo la falsariga della dimostrazione dell'implicazione (147) sulle trasformazioni di scala di variabili casuali con distribuzioni esponenziali, si può infatti dimostrare che

$$\begin{aligned} \tilde{X} &\sim \text{Normale}(\mu = \mu_{\tilde{X}}; \sigma = \sigma_{\tilde{X}}) \Rightarrow \\ \Rightarrow \tilde{Y} = b\tilde{X} &\sim \text{Normale}(\mu = b\mu_{\tilde{X}}; \sigma = b\sigma_{\tilde{X}}) \quad \text{per ogni } b > 0, \end{aligned}$$

e modificando leggermente la dimostrazione (lasciamo al lettore il compito di individuare e apportare le modifiche necessarie), si può anche dimostrare che

$$\begin{aligned} \tilde{X} &\sim \text{Normale}(\mu = \mu_{\tilde{X}}; \sigma = \sigma_{\tilde{X}}) \Rightarrow \\ \Rightarrow \tilde{Y} = b\tilde{X} &\sim \text{Normale}(\mu = b\mu_{\tilde{X}}; \sigma = |b|\sigma_{\tilde{X}}) \quad \text{per ogni } b \in \mathbb{R}. \end{aligned} \tag{160}$$

Combinando le implicazioni (159) e (160) vediamo dunque che **la famiglia delle distribuzioni normali è chiusa rispetto a trasformazioni affini invertibili**:

Teorema 8.9 (Trasformazioni affini di una singola variabile casuale con distribuzione normale). Se

$$\tilde{X} \sim \text{Normale}(\mu = \mu_{\tilde{X}}; \sigma = \sigma_{\tilde{X}}),$$

allora

$$\tilde{Y} = a + b\tilde{X} \sim \text{Normale}(\mu = a + b\mu_{\tilde{X}}; \sigma = |b|\sigma_{\tilde{X}})$$

per ogni $a \in \mathbb{R}$ e ogni $b \neq 0$.

Dimostrazione. La dimostrazione del Teorema 8.9 è molto semplice: per ogni $a \in \mathbb{R}$ e ogni $b \neq 0$ si ha

$$\begin{aligned} \tilde{X} \sim \text{Normale}(\mu = \mu_{\tilde{X}}; \sigma = \sigma_{\tilde{X}}) &\Rightarrow \\ &\Rightarrow \text{[[implicazione (160)]]} \Rightarrow \\ \Rightarrow \tilde{Y}' = b\tilde{X} \sim \text{Normale}(\mu = b\mu_{\tilde{X}}; \sigma = |b|\sigma_{\tilde{X}}) &\Rightarrow \\ &\Rightarrow \text{[[implicazione (159)]]} \Rightarrow \\ \Rightarrow \tilde{Y} = a + \tilde{Y}' = a + b\tilde{X} \sim \text{Normale}(\mu = a + b\mu_{\tilde{X}}; \sigma = |b|\sigma_{\tilde{X}}). \end{aligned}$$

□

In alternativa, il Teorema 8.9 può essere facilmente dimostrato attraverso l'espressione della fgm che ricaveremo tra breve (per ottenere una dimostrazione basata sulla fgm conviene invocare i Teoremi 7.2 e 7.4).

Funzione di ripartizione e quantili. Anche la funzione di ripartizione normale

$$F(x) = \int_{-\infty}^x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt, \quad x \in \mathbb{R},$$

non è una funzione elementare e i suoi valori non possono dunque essere calcolati con l'ausilio di una tradizionale calcolatrice scientifica (intendiamo una calcolatrice che non sia dotata della funzione "erf(x)" e nemmeno della funzione gamma incompleta).

Tuttavia, attraverso il metodo della **"standardizzazione"** (che non è altro che un'applicazione del Teorema 8.9 sulle trasformazioni affini), tutte le funzioni di ripartizione normali possono essere ricondotte alla cosiddetta **funzione di ripartizione normale standard**, ovvero alla funzione di ripartizione della distribuzione normale con $\mu = 0$ e $\sigma = 1$. Per "standarizzare" una variabile casuale \tilde{X} con distribuzione normale, basta applicarle la trasformazione affine

$$\tilde{X} \mapsto \tilde{Z} = \frac{\tilde{X} - \mu_{\tilde{X}}}{\sigma_{\tilde{X}}} = \frac{\mu_{\tilde{X}}}{\sigma_{\tilde{X}}} - \frac{1}{\sigma_{\tilde{X}}}\tilde{X}. \quad (161)$$

Usando il Teorema 8.9 sulle trasformazioni affini vediamo che

$$\tilde{Z} \sim \text{Normale}\left(\mu = \frac{\mu_{\tilde{X}}}{\sigma_{\tilde{X}}} - \frac{\mu_{\tilde{X}}}{\sigma_{\tilde{X}}} = 0; \sigma = \left|\frac{1}{\sigma_{\tilde{X}}}\right| \sigma_{\tilde{X}} = 1\right), \quad (162)$$

ovvero che la distribuzione di \tilde{Z} è normale con $\mu = 0$ e $\sigma = 1$. Attraverso l'operazione di "standardizzazione", ovvero attraverso la trasformazione affine (161), qualunque variabile casuale con distribuzione normale può essere dunque trasformata in una variabile casuale con distribuzione normale standard. Questo fatto può essere sfruttato per determinare i valori che assume una funzione di ripartizione normale. Infatti, se \tilde{X} è una variabile casuale con distribuzione normale di parametri $\mu = \mu_{\tilde{X}} \in \mathbb{R}$ e $\sigma = \sigma_{\tilde{X}} > 0$ qualunque, possiamo scrivere

$$\begin{aligned} F(x) = P(\{\tilde{X} \leq x\}) &= [[\text{standardizzazione}]] = P\left(\left\{\frac{\tilde{X} - \mu_{\tilde{X}}}{\sigma_{\tilde{X}}} \leq \frac{x - \mu_{\tilde{X}}}{\sigma_{\tilde{X}}}\right\}\right) = \\ &= [[(161)]] = P\left(\left\{\tilde{Z} \leq \frac{x - \mu_{\tilde{X}}}{\sigma_{\tilde{X}}}\right\}\right) = \\ &= [[(162)]] = \int_{-\infty}^{\frac{x - \mu_{\tilde{X}}}{\sigma_{\tilde{X}}}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}s^2} ds \end{aligned}$$

e questo ragionamento dimostra che una funzione di ripartizione normale con parametri $\mu \in \mathbb{R}$ e $\sigma > 0$ qualunque può essere espressa come

$$F(x) = \int_{-\infty}^x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt = \int_{-\infty}^{\frac{x-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}s^2} ds = \Phi\left(\frac{x-\mu}{\sigma}\right) \quad (163)$$

dove

$$\Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}s^2} ds, \quad z \in \mathbb{R},$$

è la funzione di ripartizione normale standard, ovvero la funzione di ripartizione della distribuzione normale con $\mu = 0$ e $\sigma = 1$.

Per calcolare i valori di una funzione di ripartizione normale qualunque basta quindi conoscere i valori che assume la funzione di ripartizione normale standard $\Phi(z)$. I valori che assume questa funzione in corrispondenza di una fitta griglia di valori di $z \in \mathbb{R}$ possono essere facilmente ricavati da un'apposita tabella che è nota come "tavola della distribuzione normale standard". Prima di esaminare la tavola della distribuzione normale standard conviene visualizzare che cosa rappresentano i valori di

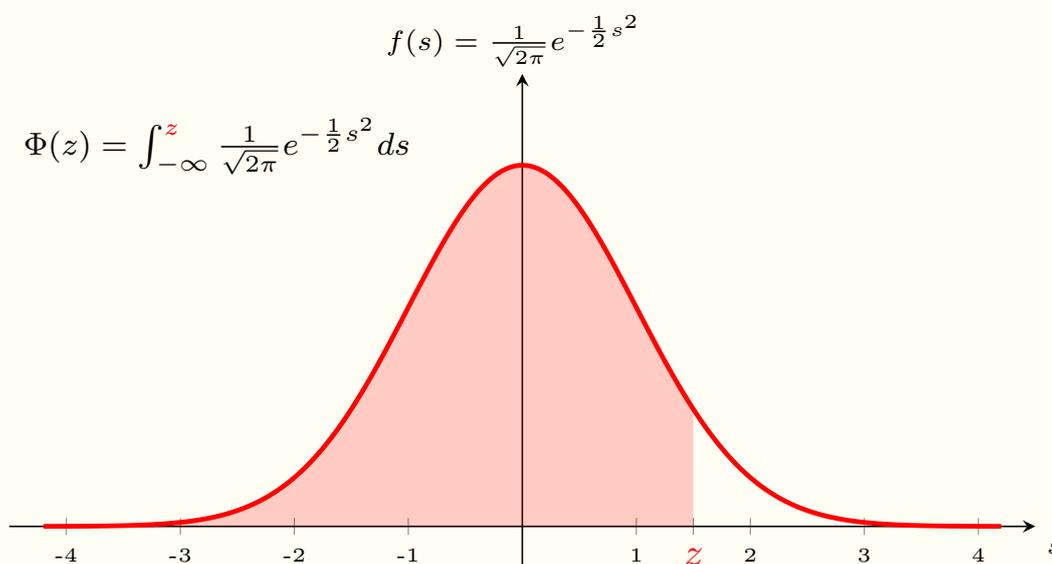
$$\Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}s^2} ds$$

Come si desume dal grafico in Figura 8.12,

$$\Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}s^2} ds$$

è l'area sottesa alla funzione di densità normale standard (ovvero alla funzione di densità normale con $\mu = 0$ e $\sigma = 1$) alla sinistra del punto z sull'asse delle ascisse.

Figura 8.12. Il grafico mostra come è definita la funzione di ripartizione normale standard.



Come si vede nel grafico in Figura 8.12, la funzione di densità normale standard concentra quasi tutta la probabilità sull'intervallo che parte da $z = -3,5$ e che arriva fino a $z = +3,5$ e per questo motivo la [tavola della distribuzione normale standard](#) riporta solo i valori di

$$\Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}s^2} ds$$

per una fitta griglia di valori di z che sono compresi tra $-3,5$ e $+3,5$. Infatti, guardando il grafico non è difficile rendersi conto che (si ricordi che l'area complessiva sottesa ad una qualunque funzione di densità è sempre uguale a 1)

- $\Phi(z) \simeq 0$ per $z \leq -3,5$
- e che $\Phi(z) \simeq 1$ per $z \geq +3,5$.

Di fatto, la [tavola della distribuzione normale standard](#) riporta solo i valori di $\Phi(z)$ per z compreso tra 0 e $+3,5$, ma non quelli che corrispondono a valori negativi di z . Questa omissione non crea problemi. Infatti, tenendo presente che la funzione di densità normale standard è simmetrica rispetto al punto $z = 0$ e ricordando che l'area complessiva sottesa alla funzione di densità normale standard deve essere pari a 1, non è difficile rendersi conto che (vedi Figura 8.12)

$$\Phi(-z) = 1 - \Phi(z) \quad \text{per ogni } z \in \mathbb{R}$$

e che i valori di $\Phi(z)$ che corrispondono a valori negativi di z possono dunque essere facilmente ricavati da quelli che corrispondono a valori positivi di z .

A questo punto siamo pronti per esaminare la [tavola della distribuzione normale standard](#). Le righe di questa tavola sono intestate al valore di z troncato alla prima cifra decimale e nell'intestazione delle colonne troviamo invece la seconda cifra decimale di z . Per trovare il valore di

$$\Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}s^2} ds$$

che corrisponde a $z = 1,23 = 1,2 + 0,03$, per esempio, dobbiamo quindi cercare la casella della tabella che corrisponde

- alla riga che è intestata al valore 1,2
- e alla colonna che è intestata al valore 0,03.

Cercando suddetta casella nella [tavola della distribuzione normale standard](#) troviamo $\Phi(1,23) = 0,8907$.

Siccome la distribuzione normale standard

$$\Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}s^2} ds$$

è continua e strettamente crescente, possiamo sfruttare la [tavola della distribuzione normale standard](#) anche per [determinare approssimazioni ai valori di \$\Phi\(z\)\$ che corrispondono a valori di \$z\$ con tre o più cifre decimali](#). Volendo un'approssimazione precisa si possono infatti interpolare i valori di $\Phi(z_0)$ e $\Phi(z_1)$ che corrispondono ai due valori z_0 e z_1 che approssimano, rispettivamente per eccesso e per difetto, il valore di z desiderato. Altrimenti, ci si può anche accontentare di un'approssimazione meno precisa cercando semplicemente il valore di $\Phi(z)$ che corrisponde al valore desiderato di z dopo che quest'ultimo è stato arrotondato alla seconda cifra decimale.

Esercizio 8.19 (Uso della tavola della distribuzione normale standard). Sia \tilde{X} una variabile casuale con distribuzione normale di parametri $\mu = 10$ e $\sigma = 3$. Si determinino i valori di

$$P(\{\tilde{X} \leq 12\}), \quad P(\{\tilde{X} \leq 9\}), \quad P(\{\tilde{X} > 12\}), \quad P(\{\tilde{X} > 9\})$$

$$P(\{9 < \tilde{X} \leq 12\}), \quad P(\{9 \leq \tilde{X} < 12\}), \quad P(\{9 \leq \tilde{X} \leq 12\}).$$

Risposta:

Per determinare le probabilità richieste basta ricordare che qualunque funzione di ripartizione normale può essere ricondotta alla funzione di ripartizione normale standard attraverso la formula (163):

$$F(x) = \int_{-\infty}^x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt = \int_{-\infty}^{\frac{x-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}s^2} ds = \Phi\left(\frac{x-\mu}{\sigma}\right).$$

Usando la [tavola della distribuzione normale standard](#) possiamo quindi determinare i valori di tutte le probabilità richieste:

•

$$P(\{\tilde{X} \leq 12\}) = \Phi\left(\frac{12-10}{3}\right) = \Phi(2/3) \simeq \Phi(0,67),$$

e siccome $\Phi(0,67) = 0,74857$ possiamo concludere che

$$P(\{\tilde{X} \leq 12\}) = 0,74857;$$

•

$$P(\{\tilde{X} \leq 9\}) = \Phi\left(\frac{9-10}{3}\right) = \Phi(-1/3) \simeq \Phi(-0,33),$$

e usando la relazione $\Phi(-z) = 1 - \Phi(z)$ vediamo che

$$\Phi(-0,33) = 1 - \Phi(0,33) = 1 - 0,6293 = 0,3707.$$

La probabilità richiesta è dunque data da

$$P(\{X \leq 9\}) = 1 - 0,6293 = 0,3707.$$

- Usando i valori delle probabilità che abbiamo appena trovato otteniamo anche i valori di $P(\{\tilde{X} > 12\})$ e di $P(\{\tilde{X} > 9\})$:

$$P(\{\tilde{X} > 12\}) = 1 - P(\{\tilde{X} \leq 12\}) = 1 - 0,74857 = 0,25143,$$

$$P(\{\tilde{X} > 9\}) = 1 - P(\{\tilde{X} \leq 9\}) = 1 - 0,3707 = 0,6293;$$

- Usando ancora i valori di $P(\{\tilde{X} \leq 12\}) = 0,74857$ e $P(\{\tilde{X} \leq 9\}) = 0,3707$ otteniamo

$$\begin{aligned} P(\{9 < \tilde{X} \leq 12\}) &= P(\{\tilde{X} \leq 12\}) - P(\{\tilde{X} \leq 9\}) \\ &= 0,74857 - 0,3707 = 0,37787. \end{aligned}$$

Siccome le distribuzioni normali sono assolutamente continue e dunque anche continue nel senso meno restrittivo del termine, possiamo concludere che anche i valori delle probabilità $P(\{9 \leq \tilde{X} < 12\})$ e $P(\{9 \leq \tilde{X} \leq 12\})$ coincidono con il valore di $P(\{9 < \tilde{X} \leq 12\})$ che abbiamo appena calcolato.

Anche i **quantili** di una distribuzione normale con parametri μ e σ qualsiasi possono essere ricavati con l'ausilio della [tavola della distribuzione normale standard](#). Per spiegare la procedura conviene ricordare ancora una volta che $\Phi(z)$ è una funzione continua e

strettamente crescente di $z \in \mathbb{R}$ e che anche la funzione di ripartizione

$$F(x) = \int_{-\infty}^x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt = \int_{-\infty}^{\frac{x-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}s^2} ds = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

è quindi una funzione continua e strettamente crescente di $x \in \mathbb{R}$. Per ogni $p \in (0, 1)$ deve quindi esistere un unico valore x_p tale che

$$F(x_p) = p, \quad \text{ovvero tale che} \quad \Phi\left(\frac{x_p - \mu}{\sigma}\right) = p,$$

e il valore x_p in questione è quindi il quantile di ordine p della distribuzione $Normale(\mu; \sigma)$.

Per ricavare il valore di x_p basta osservare che

$$\Phi\left(\frac{x_p - \mu}{\sigma}\right) = p \quad \Leftrightarrow \quad \frac{x_p - \mu}{\sigma} = z_p,$$

dove z_p è il valore di z che lascia il $p \times 100\%$ dell'area complessiva sottesa alla funzione di densità normale standard alla sua sinistra. Il valore di z_p può essere ricavato dalla tavola della distribuzione normale standard. Risolvendo l'equazione

$$\frac{x_p - \mu}{\sigma} = z_p$$

per x_p si ottiene dunque il quantile di ordine p della distribuzione $Normale(\mu; \sigma)$:

$$x_p = \mu + z_p\sigma \quad \text{per } p \in (0, 1).$$

Esercizio 8.20 (Quantili di distribuzioni normali). Si determinino (i) la mediana, (ii) il noventesimo percentile, (iii) il decimo percentile e (iv) il quantile di ordine $p = 0,02$ della distribuzione $Normale(\mu = 10; \sigma = 3)$.

Risposte:

Come abbiamo appena visto, i quantili di una distribuzione normale possono essere ottenuti attraverso la formula

$$x_p = \mu + z_p\sigma \quad \text{per } p \in (0, 1).$$

Siccome $z_{0,5} = 0$ (ricordiamo che la funzione di densità normale standard è simmetrica rispetto a $z = 0$), possiamo concludere che la mediana della distribuzione $Normale(\mu = 10; \sigma = 3)$ sia data da

$$x_{0,5} = 10 + z_{0,5} \times 3 = 10 + 0 \times 3 = 10.$$

[[Avremmo potuto anche rispondere direttamente a questa domanda ricordando il fatto che tutte le funzioni di densità normali sono simmetriche rispetto al valore del loro parametro μ .]]

Consideriamo ora invece il novantesimo percentile

$$x_{0,9} = 10 + z_{0,9} \times 3.$$

Dalla [tavola della distribuzione normale standard](#) desumiamo che $z_{0,9} = 1,28$ e quindi possiamo concludere che

$$x_{0,9} = 10 + 1,28 \times 3 = 13,84.$$

Passiamo ora al decimo percentile

$$x_{0,1} = 10 + z_{0,1} \times 3. \quad (164)$$

Notiamo che il valore di $z_{0,1}$ non è direttamente reperibile sulla [tavola della distribuzione normale standard](#) perché quest'ultima riporta soltanto i valori di $\Phi(z)$ che corrispondono a valori positivi di z . Tuttavia, siccome la funzione di densità normale standard è simmetrica rispetto a $z = 0$, possiamo concludere che

$$z_p = -z_{1-p} \quad \text{per ogni } p \in (0, 1).$$

Usando questa relazione vediamo che

$$z_{0,1} = -z_{1-0,1} = -z_{0,9} = -1,28$$

e sostituendo nella formula (164) otteniamo

$$x_{0,1} = 10 - 1,28 \times 3 = 6,16.$$

Anche il valore del quantile di ordine $p = 0,02$ può essere ricavato attraverso il ragionamento che ci ha condotto al valore di $x_{0,1} = 6,16$:

$$x_{0,02} = 10 + z_{0,02} \times 3,$$

e siccome

$$z_{0,02} = -z_{1-0,02} = -z_{0,98} = -2,05,$$

possiamo concludere che

$$x_{0,02} = 10 - 2,05 \times 3 = 3,85.$$

Esercizio 8.21. Si consideri un risparmiatore che ha investito 50 mila euro in un fondo e sia \tilde{X} la variabile casuale che descrive il rendimento aritmetico annuale del

fondo. Si supponga che la distribuzione di \tilde{X} sia normale con parametri

$$\mu_{\tilde{X}} = +0,01 = +1\% \quad \text{e} \quad \sigma_{\tilde{X}} = 0,04 = 4\%.$$

Sia \tilde{Y} la variabile casuale che descrive il valore dell'investimento dopo un anno (in migliaia di euro).

- Si calcoli $P(\{\tilde{Y} > 52\})$.
- Si determini il primo percentile di \tilde{Y} .

Risposte:

- Si calcoli $P(\{\tilde{Y} > 52\})$.

Per rispondere osserviamo innanzitutto che il valore finale dell'investimento è dato da

$$\tilde{Y} = 50(1 + \tilde{X}) = 50 + 50\tilde{X}.$$

Siccome \tilde{Y} è una trasformazione affine del rendimento annuale \tilde{X} , e siccome per ipotesi

$$\tilde{X} \sim \text{Normale}(\mu_{\tilde{X}} = +0,01; \sigma_{\tilde{X}} = 0,04),$$

possiamo concludere che (Teorema 8.9)

$$\tilde{Y} \sim \text{Normale}(\mu_{\tilde{Y}}; \sigma_{\tilde{Y}}),$$

dove

$$\mu_{\tilde{Y}} = 50 + 50 \times 0,01 = 50,5$$

e

$$\sigma_{\tilde{Y}} = |50| \times 0,04 = 2.$$

La probabilità richiesta è quindi data da

$$\begin{aligned} P(\{\tilde{Y} > 52\}) &= 1 - \Phi\left(\frac{52 - 50,5}{2}\right) = 1 - \Phi(0,75) = \\ &= [[\text{tavola } \Phi(\cdot)]] = 1 - 0,7734 = 0,2266. \end{aligned}$$

- Si determini il primo percentile di \tilde{Y} .

Il primo percentile di \tilde{Y} è dato da

$$y_{0,01} = \mu_{\tilde{Y}} + z_{0,01}\sigma_{\tilde{Y}}.$$

Come si desume dalla [tavola della distribuzione normale standard](#),

$$z_{0,01} = -z_{0,99} = -2,33,$$

e usando i valori di $\mu_{\tilde{Y}} = 50,5$ e $\sigma_{\tilde{Y}} = 2$ che abbiamo determinato nella risposta al quesito precedente, vediamo che

$$y_{0,01} = 50,5 + (-2,33) \times 2 = 45,84.$$

Questo valore ci dice che nell'1% dei casi peggiori il valore finale dell'investimento sarà minore di 45840 euro.

Si noti che nel precedente esercizio abbiamo ipotizzato una distribuzione normale per una variabile casuale \tilde{X} che rappresenta il rendimento aritmetico di un fondo anche se per definizione il rendimento aritmetico di un fondo non può essere inferiore al $-100\% = -1$ e qualunque distribuzione normale associa probabilità positiva all'evento $\{\tilde{X} < -1\}$.

Per giustificare l'ipotesi di normalità possiamo osservare che nell'esercizio vengono ipotizzati valori dei parametri $\mu_{\tilde{X}} = 0,01$ e $\sigma_{\tilde{X}} = 0,04$ che sono entrambi molto prossimi a zero e che pertanto la probabilità dell'evento $\{\tilde{X} < -1\}$ è trascurabile:

$$P(\{\tilde{X} < -1\}) = \Phi\left(\frac{-1 - 0,01}{0,04}\right) = \Phi(-25,25) \simeq 0.$$

In alternativa, la forzatura può essere risolta interpretando il rendimento \tilde{X} come **logaritmico** piuttosto che **aritmetico** (si ricordi che le variazioni logaritmiche di grandezze positive possono assumere qualsiasi valore tra $-\infty$ e $+\infty$). Tuttavia, ipotizzando che il rendimento sia di tipo logaritmico non si potrebbe più invocare il Teorema 8.9 per concludere che la distribuzione del valore finale dell'investimento sia normale perché se \tilde{X} rappresentasse il rendimento logaritmico, il valore finale del fondo sarebbe dato da

$$\tilde{Y} = 50 \times e^{\tilde{X}}$$

e non sarebbe più una trasformazione affine di \tilde{X} . Tra breve, quando introdurremo le distribuzioni lognormali, vedremo che in questo caso la distribuzione di \tilde{Y} sarebbe lognormale.

Funzione generatrice dei momenti. Come abbiamo già accennato, tutte le distribuzioni normali sono dotate di valore atteso e varianza. Nella seguente dimostrazione vedremo che tutte le distribuzioni normali sono anche dotate di fgm e che quest'ultima è definita come

$$m(t) = E\left(e^{t\tilde{X}}\right) = e^{t\mu + \frac{1}{2}t^2\sigma^2} \quad \text{per } t \in \mathbb{R}. \quad (165)$$

Dimostrazione. Per dimostrare che tutte le distribuzioni normali sono dotate di fgm, dimostreremo che l'integrale

$$\int_{-\infty}^{\infty} e^{tx} f(x) dx = \int_{-\infty}^{\infty} e^{tx} \times \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \quad (166)$$

è convergente per ogni $t, \mu \in \mathbb{R}$ e per ogni $\sigma > 0$. A tal fine conviene osservare che la funzione integranda può essere espressa come

$$e^{tx} \times \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} = \frac{1}{\sigma\sqrt{2\pi}} e^{tx - \frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

e che l'argomento della funzione esponenziale (ovvero l'esponente del numero di Nepero "e") può essere riscritto come

$$\begin{aligned} & -\frac{1}{2\sigma^2}x^2 + x\left(t + \frac{\mu}{\sigma^2}\right) - \frac{\mu^2}{2\sigma^2} = -\frac{1}{2\sigma^2} [x^2 - 2(t\sigma^2 + \mu)x + 2\mu^2] = \\ & = -\frac{1}{2\sigma^2} [x^2 - 2(t\sigma^2 + \mu)x + 2\mu^2 + (t\sigma^2 + \mu)^2 - (t\sigma^2 + \mu)^2] = \\ & = -\frac{1}{2\sigma^2} [x^2 - 2(t\sigma^2 + \mu)x + (t\sigma^2 + \mu)^2 + 2\mu^2 - (t\sigma^2 + \mu)^2] = \\ & = -\frac{1}{2\sigma^2} \{ [x - (t\sigma^2 + \mu)]^2 + 2\mu^2 - t^2\sigma^4 - 2t\mu\sigma^2 - \mu^2 \} = \\ & = -\frac{1}{2\sigma^2} \{ [x - (t\sigma^2 + \mu)]^2 + \mu^2 - t^2\sigma^4 - 2t\mu\sigma^2 \} = \\ & = -\frac{1}{2\sigma^2} [x - (t\sigma^2 + \mu)]^2 + t\mu + \frac{1}{2}t^2\sigma^2 \end{aligned}$$

La funzione integranda nell'equazione (166) può quindi essere espressa come

$$e^{tx} \times \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}[x-(t\sigma^2+\mu)]^2} \times e^{t\mu + \frac{1}{2}t^2\sigma^2}$$

e l'integrale è quindi dato da

$$\begin{aligned} \int_{-\infty}^{\infty} e^{tx} \times \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx &= \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}[x-(t\sigma^2+\mu)]^2} \times e^{t\mu + \frac{1}{2}t^2\sigma^2} dx = \\ &= e^{t\mu + \frac{1}{2}t^2\sigma^2} \times \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}[x-(t\sigma^2+\mu)]^2} dx = [[\text{proprietà f2}]] = e^{t\mu + \frac{1}{2}t^2\sigma^2}. \end{aligned}$$

Si noti che nell'ultimo passaggio abbiamo invocato la proprietà f2 perché la funzione integranda è una funzione di densità normale il cui integrale, come quello di qualsiasi altra funzione di densità, deve essere uguale a 1. Questo ragionamento dimostra quindi che

$$\int_{-\infty}^{\infty} e^{tx} f(x) dx = e^{t\mu + \frac{1}{2}t^2\sigma^2}$$

per ogni $t, \mu \in \mathbb{R}$ e ogni $\sigma > 0$. Quindi possiamo concludere che tutte le distribuzioni normali sono dotate di fgm e che la fgm di una distribuzione normale è data da

$$m(t) = e^{t\mu + \frac{1}{2}t^2\sigma^2} \quad \text{per } t \in \mathbb{R}.$$

□

Valore atteso e varianza. Derivando la fgm (165) si può dimostrare che il valore atteso e lo scarto quadratico medio di una distribuzione normale coincidono con i due parametri μ e σ :

$$m'(t) = (\mu + t\sigma^2)e^{t\mu + \frac{1}{2}t^2\sigma^2} \quad \Rightarrow \quad E(\tilde{X}) = m'(0) = \mu;$$

$$\begin{aligned} m''(t) &= \sigma^2 e^{t\mu + \frac{1}{2}t^2\sigma^2} + (\mu + t\sigma^2)^2 e^{t\mu + \frac{1}{2}t^2\sigma^2} \quad \Rightarrow \\ \Rightarrow \quad \text{var}(\tilde{X}) &= E(\tilde{X}^2) - [E(\tilde{X})]^2 = m''(0) - [m'(0)]^2 = \sigma^2 + \mu^2 - \mu^2 = \sigma^2. \end{aligned}$$

Proprietà riproduttiva. Usando l'espressione (165) della fgm di distribuzioni normali, e seguendo la falsariga delle dimostrazione del Teorema 8.1 sulla proprietà riproduttiva delle distribuzioni binomiali, si può dimostrare che anche le distribuzioni normali soddisfano la proprietà riproduttiva:

Teorema 8.10 (Proprietà riproduttiva delle distribuzioni normali). Se

$$\begin{aligned} \tilde{X}_1 \sim \text{Normale}(\mu = \mu_1; \sigma = \sigma_1), \quad \tilde{X}_2 \sim \text{Normale}(\mu = \mu_2; \sigma = \sigma_2), \quad \dots \\ \dots, \quad \tilde{X}_k \sim \text{Normale}(\mu = \mu_k; \sigma = \sigma_k) \end{aligned}$$

sono variabili casuali normali **indipendenti**, allora la variabile casuale $\tilde{Y} = \tilde{X}_1 + \tilde{X}_2 + \dots + \tilde{X}_k$ ha distribuzione normale di parametri

$$\mu = \mu_1 + \mu_2 + \dots + \mu_k \quad \text{e} \quad \sigma = \sqrt{\sigma_1^2 + \sigma_2^2 + \dots + \sigma_k^2}.$$

Trasformazioni affini di variabili casuali normali indipendenti. Supponiamo che $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_k$ siano k variabili casuali normali indipendenti e consideriamo una loro trasformazione affine

$$\tilde{Y} = a + b_1\tilde{X}_1 + b_2\tilde{X}_2 + \dots + b_k\tilde{X}_k$$

con coefficienti b_i diversi da zero. Siccome $\tilde{Y}_1 = a + b_1\tilde{X}_1$ è una trasformazione affine della variabile casuale \tilde{X}_1 , possiamo concludere che (Teorema 8.9)

$$\tilde{Y}_1 = a + b_1\tilde{X}_1 \sim \text{Normale}(\mu = a + b_1\mu_{\tilde{X}_1}; \sigma = |b_1|\sigma_{\tilde{X}_1}).$$

Allo stesso modo vediamo che per ciascuna delle variabili casuali $\tilde{Y}_i = b_i\tilde{X}_i$ con $i = 2, 3, \dots, k$ si ha

$$\tilde{Y}_i = b_i\tilde{X}_i \sim \text{Normale}(\mu = b_i\mu_{\tilde{X}_i}; \sigma = |b_i|\sigma_{\tilde{X}_i}).$$

Ora, siccome stiamo ipotizzando che le variabili casuali \tilde{X}_i siano indipendenti, possiamo concludere che anche le variabili casuali \tilde{Y}_i devono essere indipendenti (Teorema 4.10), e applicando il Teorema 8.10 vediamo dunque che la trasformazione affine

$$\tilde{Y} = a + b_1\tilde{X}_1 + b_2\tilde{X}_2 + \dots + b_k\tilde{X}_k$$

deve avere distribuzione normale con parametri

$$\mu = \mu_{\tilde{Y}} = a + b_1\mu_1 + b_2\mu_2 + \dots + b_k\mu_k$$

e

$$\sigma = \sigma_{\tilde{Y}} = \sqrt{b_1^2\sigma_{\tilde{X}_1}^2 + b_2^2\sigma_{\tilde{X}_2}^2 + \dots + b_k^2\sigma_{\tilde{X}_k}^2}.$$

Si noti che questo risultato rimane valido anche se uno o più di uno dei k coefficienti b_i sono nulli, a patto che almeno uno di essi sia diverso da zero (infatti, se uno o più coefficienti b_i sono nulli, basta ignorare i corrispondenti termini nella definizione della trasformazione affine \tilde{Y}). Questo ragionamento dimostra dunque il seguente teorema:

Teorema 8.11 (Trasformazioni affini di variabili casuali normali indipendenti). Se $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_k$ sono k variabili casuali indipendenti e

$$\tilde{X}_i \sim \text{Normale}(\mu = \mu_i; \sigma = \sigma_i) \quad \text{per } i = 1, 2, \dots, k,$$

e se

$$\tilde{Y} = a + b_1\tilde{X}_1 + b_2\tilde{X}_2 + \dots + b_k\tilde{X}_k$$

con almeno una costante b_i diversa da zero, allora la variabile casuale \tilde{Y} ha distribuzione normale di parametri

$$\mu = a + b_1\mu_1 + b_2\mu_2 + \dots + b_k\mu_k$$

e

$$\sigma = \sqrt{b_1^2\sigma_1^2 + b_2^2\sigma_2^2 + \dots + b_k^2\sigma_k^2}.$$

Distribuzioni normali multivariate. Consideriamo ora k variabili casuali i.i.d. con distribuzione normale standard e indichiamo queste variabili casuali con $\tilde{Z}_1, \tilde{Z}_2, \dots, \tilde{Z}_h$. Applicando delle trasformazioni affini alle variabili casuali \tilde{Z}_i possiamo definire k variabili casuali \tilde{Y}_i :

$$\begin{aligned} \tilde{Y}_1 &= a_1 + b_{11}\tilde{Z}_1 + b_{12}\tilde{Z}_2 + \dots + b_{1h}\tilde{Z}_h \\ \tilde{Y}_2 &= a_2 + b_{21}\tilde{Z}_1 + b_{22}\tilde{Z}_2 + \dots + b_{2h}\tilde{Z}_h \\ &\vdots \\ \tilde{Y}_k &= a_k + b_{k1}\tilde{Z}_1 + b_{k2}\tilde{Z}_2 + \dots + b_{kh}\tilde{Z}_h. \end{aligned} \tag{167}$$

Le distribuzioni congiunte di $k = 2$ o $k \geq 2$ variabili casuali \tilde{Y}_i che si possono ottenere in questo modo vengono chiamate **distribuzioni normali multivariate**. In questa dispensa non tratteremo in modo approfondito queste distribuzioni ma ci limitiamo a menzionare una proprietà che di fatto caratterizza questa famiglia di distribuzioni multivariate (le distribuzioni normali multivariate sono quindi le uniche distribuzioni che soddisfano questa proprietà). La proprietà in questione è la seguente: **Tutte le trasformazioni affini di variabili casuali \tilde{Y}_i che hanno distribuzione congiunta normale multivariata hanno distribuzione normale o degenerare!!!** (per distribuzione degenerare intendiamo la distribuzione di una variabile casuale **degenerare**, ovvero una distribuzione che concentra tutta la probabilità in un singolo punto). Per dimostrare questa proprietà delle distribuzioni normali multivariate basta considerare le variabili casuali \tilde{Y}_i nella (167) e osservare che qualsiasi loro trasformazione affine

$$\tilde{Y}' = a' + b'_1\tilde{Y}_1 + b'_2\tilde{Y}_2 + \dots + b'_k\tilde{Y}_k$$

può essere espressa come una trasformazione affine delle variabili casuali \tilde{Z}_i che per ipotesi sono i.i.d. con distribuzione normale standard: secondo il Teorema 8.11 la distribuzione della variabile casuale \tilde{Y}' deve essere normale oppure degenera.

Dimostrare che le distribuzioni normali multivariate sono le uniche distribuzioni che soddisfano la suddetta proprietà, ovvero dimostrare l'affermazione ...

"tutte le trasformazioni affini di k variabili casuali $\tilde{Y}_1, \tilde{Y}_2, \dots, \tilde{Y}_k$ hanno distribuzione normale o degenera solo se la distribuzione congiunta delle k variabili casuali \tilde{Y}_i è una distribuzione normale multivariata"

... richiede invece conoscenze matematiche più avanzate e per questo motivo non dimostreremo questa affermazione.

Esercizio 8.22. Un investitore ha investito $C_1 = 100$ mila euro nel titolo A e $C_2 = 150$ mila euro nel titolo B .

Si supponga che i rendimenti aritmetici annuali dei due titoli siano descritti da variabili casuali \tilde{X}_1 (rendimento del titolo A) e \tilde{X}_2 (rendimento del titolo B) che hanno come valori attesi

$$\mu_1 = E(\tilde{X}_1) = 0,06 \quad \text{e} \quad \mu_2 = E(\tilde{X}_2) = 0,03$$

e come matrice di varianza e covarianza

$$\Sigma = \begin{bmatrix} 0,0009 & -0,0001 \\ -0,0001 & 0,0002 \end{bmatrix}.$$

Si assuma inoltre che la distribuzione congiunta dei rendimenti dei due titoli sia normale multivariata.

- Si calcoli la probabilità che il rendimento del portafoglio dell'investitore sia superiore a 0,05.
- Si calcoli il primo percentile del rendimento del portafoglio dell'investitore.

Risposte:

- Si calcoli la probabilità che il rendimento del portafoglio dell'investitore sia superiore a 0,05.

Nell'Esercizio 6.2 abbiamo già visto che il rendimento atteso del portafoglio è dato da

$$\mu_P = E(\tilde{Y}_P) = 0,042$$

e che la varianza del rendimento del portafoglio è invece data da

$$\sigma_P^2 = \text{var}(\tilde{Y}_P) = 0,000168 \quad \Rightarrow \quad \sigma_P = \sqrt{0,000168} = 0,0130.$$

Siccome \tilde{Y}_P è una trasformazione affine dei rendimenti aritmetici \tilde{X}_1 e \tilde{X}_2 (si ricordi la formula (84)), e siccome tutte le trasformazioni affini di variabili casuali con distribuzione congiunta normale multivariata hanno distribuzione normale oppure degenerare, possiamo concludere che la probabilità richiesta sia data da

$$P(\{\tilde{Y}_P > 0,05\}) = 1 - \Phi\left(\frac{0,05 - 0,042}{0,0130}\right) \simeq 1 - \Phi(0,62) = 1 - 0,7324 = 0,2676.$$

- b) Si calcoli il primo percentile del rendimento del portafoglio dell'investitore.

Il primo percentile di \tilde{Y}_P è dato da

$$\begin{aligned} y_{0,01} &= 0,042 + z_{0,01} \times 0,0130 = [[\text{tavola } \Phi(\cdot)]] = \\ &= 0,042 - 2,33 \times 0,0130 = 0,0117. \end{aligned}$$

Questo risultato ci dice che soltanto nell'1% dei casi peggiori il rendimento del portafoglio sarà minore dell'1,17%.

Esercizio 8.23. Secondo un investitore, i rendimenti logaritmici mensili di un titolo sono descritti da variabili casuali i.i.d. con distribuzione normale di valore atteso $\mu = 0,01$ e scarto quadratico medio $\sigma = 0,004$.

- Si calcoli la probabilità che il rendimento logaritmico annuale sia superiore a 0,15.
- Si calcoli il primo percentile del rendimento del rendimento logaritmico annuale.
- Quanti mesi dura il più breve periodo di mantenimento affinché il primo percentile del corrispondente rendimento logaritmico sia maggiore di 0,01?

Risposte:

- Si calcoli la probabilità che il rendimento logaritmico annuale sia superiore a 0,15.

Riprendendo la notazione della Sezione 6.3.2, indicheremo con $\tilde{X}_1, \tilde{X}_2, \dots$ la successione di variabili casuali che rappresentano i rendimenti logaritmici

mensili e con \tilde{Y}_k il rendimento logaritmico complessivo che si riferisce ad un periodo di mantenimento di k mesi.

Siccome

$$\tilde{Y}_{12} = \text{[[formula (89)]]} = \tilde{X}_1 + \tilde{X}_2 + \dots + \tilde{X}_k$$

è la somma di dodici rendimenti logaritmici mensili (si ricordi la formula (89)), e siccome per ipotesi

$$\tilde{X}_1, \tilde{X}_2, \dots \sim \text{i.i.d. Normale}(\mu = 0, 1; \sigma = 0, 004),$$

possiamo applicare il Teorema 8.10 onde concludere che

$$\tilde{Y}_{12} \sim \text{Normale}(\mu = 12 \times 0, 01 = 0, 12; \sigma = \sqrt{12 \times 0, 004^2} = 0, 0139).$$

Usando la distribuzione di \tilde{Y}_{12} vediamo che la probabilità richiesta è data da

$$P(\{\tilde{Y}_{12} > 0, 15\}) = 1 - \Phi\left(\frac{0, 15 - 0, 12}{0, 0139}\right) \simeq 1 - \Phi(2, 16) = 1 - 0, 9846 = 0, 0154.$$

Si noti che i parametri $\mu = 0, 12$ e $\sigma = 0, 0139$ della distribuzione normale di \tilde{Y}_{12} coincidono con il valore atteso

$$E(\tilde{Y}_{12}) = 0, 12$$

e lo scarto quadratico medio

$$\text{var}(\tilde{Y}_{12}) = 0, 000192 \Rightarrow \sigma = \sqrt{0, 000192} = 0, 0139$$

che avevamo già determinato nella risposta al quesito a) dell'Esercizio 6.4.

- b) Si calcoli il primo percentile del rendimento logaritmico annuale.

Usando la distribuzione normale di \tilde{Y}_{12} (vedi la risposta al quesito precedente), vediamo che il primo percentile del rendimento logaritmico annuale è dato da

$$y_{0,01} = 0, 12 + z_{0,01} \times 0, 0139 = \text{[[tavola } \Phi(\cdot) \text{]]} = 0, 12 - 2, 33 \times 0, 0139 = 0, 0876.$$

Questo risultato ci dice che soltanto nell'1% dei casi peggiori il rendimento annuale sarà minore dell'8, 76%.

- c) Quanti mesi dura il più breve periodo di mantenimento affinché il primo percentile del corrispondente rendimento logaritmico sia maggiore di 0, 01?

Generalizzando il ragionamento della risposta al quesito a) vediamo che

$$\tilde{Y}_k \sim \text{Normale}(\mu = k \times 0, 01 = 0, 12; \sigma = 0, 004 \times \sqrt{k}).$$

Con riferimento ad un generico numero k di mesi, il primo percentile della distribuzione di \tilde{Y}_k è dunque dato da

$$\begin{aligned} y_{0,01}(k) &= k \times 0,01 + z_{0,01} \times 0,004 \times \sqrt{k} = \\ &= [[\text{tavola } \Phi(\cdot)]] = k \times 0,01 - 2,33 \times 0,004 \times \sqrt{k}. \end{aligned}$$

Il periodo di mantenimento minimo affinché il primo percentile del rendimento logaritmico del titolo sia maggiore di 0,01 è dunque dato dal più piccolo valore di k tale che

$$y_{0,01}(k) > 0,01 \quad \Rightarrow \quad k \times 0,01 - 2,33 \times 0,004 \times \sqrt{k} > 0,01.$$

Per determinare questo valore di k procederemo per tentativi:

$$k = 1 \Rightarrow 1 \times 0,01 - 2,33 \times 0,004 \times \sqrt{1} = 0,00068$$

$$k = 2 \Rightarrow 2 \times 0,01 - 2,33 \times 0,004 \times \sqrt{2} = 0,00682$$

$$k = 3 \Rightarrow 3 \times 0,01 - 2,33 \times 0,004 \times \sqrt{3} = 0,01386$$

Da questi risultati deduciamo la durata minima del periodo di mantenimento affinché il primo percentile del corrispondente rendimento logaritmico sia maggiore di 0,01 è di $k = 3$ mesi.

Con riferimento agli ultimi due esercizi conviene osservare che per attività finanziarie "rischiose" come azioni o obbligazioni a lunga scadenza, l'ipotesi di rendimenti (aritmetici o logaritmici) normali o multinormali viene spesso considerata come irrealistica perché dati empirici suggeriscono che le code delle distribuzioni dei rendimenti di attività rischiose siano "più pesanti" delle code di distribuzioni normali.

La relazione con le distribuzioni chi-quadrato. Anche se in questa dispensa non verrà mai invocato, per molte questioni legate all'inferenza statistica il prossimo teorema è di fondamentale importanza.

Teorema 8.12 (Distribuzioni chi-quadrato). Se $\tilde{Z}_1, \tilde{Z}_2, \dots, \tilde{Z}_k$ sono k variabili casuali i.i.d. con distribuzione normale standard e

$$\tilde{Y}_k = \tilde{Z}_1^2 + \tilde{Z}_2^2 + \dots + \tilde{Z}_k^2,$$

allora

$$\tilde{Y}_k \sim \text{Gamma}(\alpha = k/2; \theta = 1/2).$$

[[Ricordiamo che le distribuzioni gamma con parametro $\theta = 1/2$ e parametro $\alpha = k/2$ (con k un numero intero positivo) vengono chiamate **distribuzioni chi-quadrato** e che il valore del parametro $k = 1, 2, \dots$ che le identifica viene chiamato **numero di gradi di libertà**.]]

Dimostrazione. Il teorema può essere facilmente dimostrato attraverso la teoria sulle fgm. Consideriamo inizialmente il caso in cui $k = 1$. La fgm di $\tilde{Y}_1 = \tilde{Z}_1^2$ (se esiste) è data da

$$\begin{aligned} m_{\tilde{Y}_1}(t) &= E\left(e^{t\tilde{Y}_1}\right) = E\left(e^{t\tilde{Z}_1^2}\right) = \int_{-\infty}^{\infty} e^{tz_1^2} \times \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z_1^2} dz_1 = \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z_1^2(\frac{1}{2}-t)} dz_1 = \frac{1}{\sqrt{(1/2)-t}} \int_{-\infty}^{\infty} \frac{\sqrt{(1/2)-t}}{2\pi} e^{-z_1^2(\frac{1}{2}-t)} dz_1 \end{aligned}$$

Si noti ora che l'ultimo integrale è convergente se e solo se $t < 1/2$ perché soltanto in questo caso la funzione integranda è una funzione di densità. Infatti,

- se $t < 1/2$ la funzione integranda è la funzione di densità normale con $\mu = 0$ e $\sigma = (1/2) - t > 0$;
- se $t = 1/2$ la funzione integranda è nulla per ogni $z_1 \in \mathbb{R}$;
- se $t > 1/2$ la funzione integranda è una funzione a valori complessi.

Per $t < 1/2$ l'ultimo integrale è dunque uguale a 1 e quindi possiamo concludere che la variabile casuale $\tilde{Y}_1 = \tilde{Z}_1^2$ è dotata di fgm e che quest'ultima è data da

$$m_{\tilde{Y}_1}(t) = \frac{1}{\sqrt{(1/2)-t}} = \left(\frac{1}{(1/2)-t}\right)^{1/2} \quad \text{per } t < 1/2.$$

Siccome $m_{\tilde{Y}_1}(t)$ è la fgm della distribuzione gamma con parametri $\alpha = \theta = 1/2$, possiamo invocare il Teorema di unicità della fgm (Teorema 7.4) onde concludere che $\tilde{Y}_1 \sim \text{Gamma}(\alpha = 1/2; \theta = 1/2)$.

Il precedente ragionamento dimostra dunque che la conclusione del teorema è vera se $k = 1$. Usando la proprietà riproduttiva delle distribuzioni gamma (vedi il Teorema 8.8) si vede che la conclusione del teorema rimane vera anche se $k > 1$. \square

Il teorema del limite centrale. Come abbiamo già accennato, il teorema del limite centrale (d'ora in poi TLC) nelle sue versioni più generali afferma che sommando un gran numero di variabili casuali con varianze più o meno simili e con covarianze che sono in un certo senso trascurabili, si ottiene una variabile casuale che ha distribuzione approssimativamente normale. Per evitare condizioni sufficienti eccessivamente complicate, forniremo solo quella che può essere considerata la versione base del TLC ma che può comunque essere applicata in una vasta gamma di situazioni. Non forniremo una dimostrazione del TLC perché richiederebbe conoscenze matematiche avanzate.

Teorema 8.13 (Teorema del limite centrale, TLC). Sia $\tilde{X}_1, \tilde{X}_2, \dots$ una successione infinita di variabili casuali indipendenti e identicamente distribuite (i.i.d.) con valore atteso $\mu_{\tilde{X}} = E(\tilde{X}_i)$ e con varianza $\sigma_{\tilde{X}}^2 = \text{var}(\tilde{X}_i)$ che esistono entrambi (si noti che il valore atteso $\mu_{\tilde{X}}$ e la varianza $\sigma_{\tilde{X}}^2$ devono essere uguali per tutte le variabili casuali \tilde{X}_i perché per ipotesi queste ultime hanno tutte la stessa distribuzione), e sia $\tilde{Y}_1, \tilde{Y}_2, \dots$ la successione delle variabili casuali che sono definite come

$$\tilde{Y}_n = \sum_{i=1}^n \tilde{X}_i \quad \text{per } n = 1, 2, \dots$$

Allora,

$$\lim_{n \rightarrow \infty} \sup_{y \in \mathbb{R}} \left| P \left(\left\{ \tilde{Y}_n \leq y \right\} \right) - \Phi \left(\frac{y - n\mu_{\tilde{X}}}{\sigma_{\tilde{X}}\sqrt{n}} \right) \right| = 0$$

e per n sufficientemente elevato la distribuzione della variabile casuale \tilde{Y}_n è quindi molto prossima alla distribuzione normale con $\mu = n\mu_{\tilde{X}}$ e con $\sigma = \sigma_{\tilde{X}}\sqrt{n}$.

Per chiarire l'enunciato del TLC conviene fare alcune osservazioni:

- 1) Rispetto a ciò che già sapevamo, il TLC aggiunge "solo" un'informazione sulla forma delle distribuzioni delle variabili casuali $\tilde{Y}_n = \tilde{X}_1 + \dots + \tilde{X}_n$ con n elevato.

Infatti, già prima di enunciare il TLC sapevamo che per ogni $n = 1, 2, \dots$ si avesse

$$\begin{aligned} E(\tilde{Y}_n) &= E(\tilde{X}_1 + \dots + \tilde{X}_n) = \text{[[proprietà E4*]]} = \\ &= E(\tilde{X}_1) + E(\tilde{X}_2) + \dots + E(\tilde{X}_n) = \mu_{\tilde{X}} + \dots + \mu_{\tilde{X}} = n\mu_{\tilde{X}} \end{aligned}$$

e

$$\begin{aligned} \text{Var}(\tilde{Y}_n) &= \text{Var}(\tilde{X}_1 + \dots + \tilde{X}_n) = \text{[[proprietà V3*]]} = \\ &= \text{Var}(\tilde{X}_1) + \dots + \text{Var}(\tilde{X}_n) = \sigma_{\tilde{X}}^2 + \dots + \sigma_{\tilde{X}}^2 = n\sigma_{\tilde{X}}^2. \end{aligned}$$

se $\tilde{X}_1, \tilde{X}_2, \dots$ è una successione di variabili casuali i.i.d. con valore atteso $\mu_{\tilde{X}} = E(\tilde{X}_i)$ e con varianza $\sigma_{\tilde{X}}^2 = \text{var}(\tilde{X}_i)$ che esistono entrambi (così come ipotizzato nel TLC).

Rispetto a questo fatto che ci era già noto, la conclusione del TLC aggiunge "solo" l'informazione che all'aumentare di n la distribuzione della variabile casuale \tilde{Y}_n tenda ad assumere la forma di una distribuzione normale.

- 2) Dalla conclusione del TLC discende che **per n sufficientemente elevato** si possa usare l'approssimazione

$$P \left(\left\{ \tilde{Y}_n \leq y \right\} \right) \simeq \Phi \left(\frac{y - n\mu_{\tilde{X}}}{\sigma_{\tilde{X}}\sqrt{n}} \right), \quad (168)$$

e che l'errore di approssimazione sia prossimo a zero uniformemente rispetto a $y \in \mathbb{R}$.

Rimane da chiarire che cosa si intende con "n sufficientemente elevato". A quanto deve ammontare il valore di n affinché l'errore di approssimazione nella (168) sia trascurabile?

Per rispondere a questa domanda conviene ricordarsi del Teorema 8.10 sulla proprietà riproduttiva delle distribuzioni normali. Secondo questo teorema, se la distribuzione comune agli addendi \tilde{X}_i è normale, allora deve essere normale anche la

distribuzione di tutte le variabili casuali \tilde{Y}_n già a partire da $n = 1$. **Se la distribuzione comune agli addendi \tilde{X}_i è normale, l'errore di approssimazione nella (168) è dunque nullo per ogni $n = 1, 2, \dots$ e per ogni $y \in \mathbb{R}!!!$** Partendo da questa considerazione si può dimostrare che a parità di n l'errore di approssimazione nella (168) aumenta se la distribuzione comune agli addendi \tilde{X}_i si "allontana" dalla famiglia delle distribuzioni normali. In particolare, si può dimostrare che a parità di n l'errore di approssimazione è tanto più elevato quanto più la distribuzione comune agli addendi \tilde{X}_i è asimmetrica e/o quanto più le sue code sono pesanti. Da queste considerazioni desumiamo che la risposta alla domanda

"A quanto deve ammontare il valore di n affinché l'errore di approssimazione nella (168) sia trascurabile?"

dipende dalla distribuzione degli addendi \tilde{X}_i .

- 3) Siccome molte famiglie di distribuzioni soddisfano la proprietà riproduttiva, secondo il TLC devono esistere approssimazioni normali per determinate distribuzioni che appartengono a tali famiglie:

- **Approssimazioni normali per distribuzioni binomiali.** Dalla proprietà riproduttiva delle distribuzioni binomiali (Teorema 8.1) discende che la somma di n variabili casuali i.i.d. con distribuzione bernoulliana di parametro p abbia distribuzione binomiale di parametri n e p :

$$\begin{aligned} \tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_n &\sim i.i.d. \text{ Bernoulli}(p) \quad \Rightarrow \\ \Rightarrow \tilde{Y}_n = \tilde{X}_1 + \tilde{X}_2 + \dots + \tilde{X}_n &\sim \text{Binomiale}(n; p). \end{aligned}$$

Dal TLC discende quindi che per n sufficientemente elevato una distribuzione $\text{Binomiale}(n; p)$ possa essere approssimata attraverso la distribuzione normale di parametri

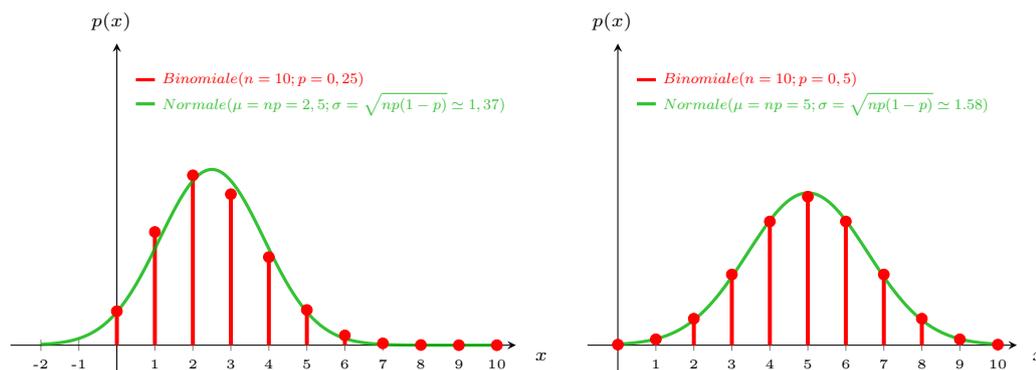
$$\mu = E(\tilde{Y}_n) = np$$

e

$$\sigma^2 = \text{Var}(\tilde{Y}_n) = np(1 - p).$$

Nel caso di una somma di n variabili casuali di Bernoulli si può dimostrare che l'errore di approssimazione nella (168) sia trascurabile se $np(1 - p)$ è maggiore di 5.

Se invece n è elevato e p è molto piccolo in modo tale che $np(1 - p) \leq 5$, allora la distribuzione $\text{Binomiale}(n; p)$ può essere approssimata in modo più preciso attraverso la distribuzione $\text{Poisson}(\lambda = np)$ piuttosto che attraverso la distribuzione $\text{Normale}(\mu = np; \sigma = \sqrt{n \times p \times (1 - p)})$ suggerita dal TLC. I due grafici sottostanti mostrano le approssimazioni normali a due distribuzioni binomiali:



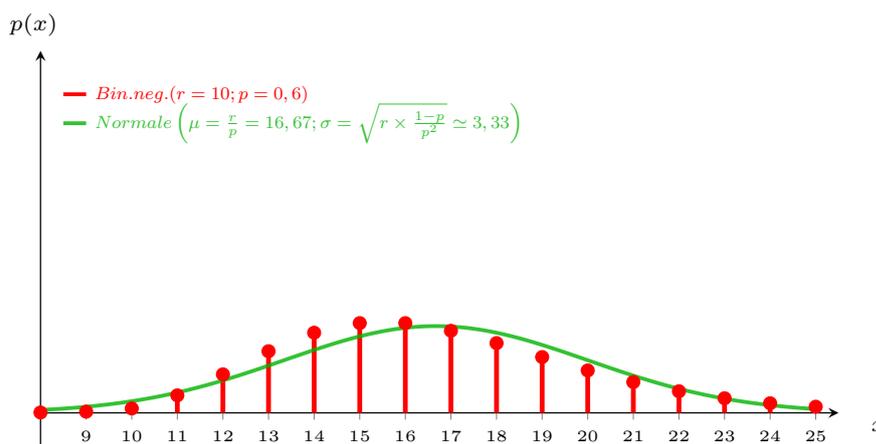
- **Approssimazioni normali per distribuzioni binomiali negative.** Dalla proprietà riproduttiva delle distribuzioni binomiali negative (Teorema 8.3) discende che la somma di r variabili casuali i.i.d. con distribuzione geometrica di parametro p abbia distribuzione binomiale negativa con parametri r e p :

$$\begin{aligned} \tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_n &\sim \text{i.i.d. Geometrica}(p) \Rightarrow \\ \Rightarrow \tilde{Y}_r = \tilde{X}_1 + \tilde{X}_2 + \dots + \tilde{X}_r &\sim \text{Bin.neg.}(r; p). \end{aligned}$$

In virtù del TLC ci aspettiamo quindi che per r sufficientemente elevato la distribuzione $\text{Bin.neg.}(r; p)$ possa essere approssimata attraverso la distribuzione normale di parametri

$$\mu = E(\tilde{Y}_r) = r \times \frac{1}{p} \quad \text{e} \quad \sigma^2 = \text{Var}(\tilde{Y}_r) = r \times \frac{1-p}{p^2}.$$

Il grafico sottostante mostra l'approssimazione normale alla distribuzione binomiale negativa con $r = 10$ e $p = 0,6$.



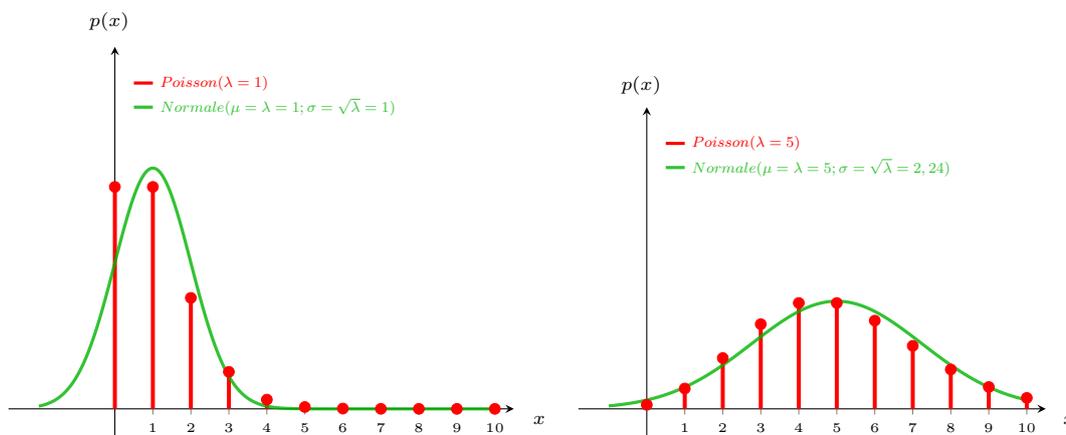
- **Approssimazioni normali per distribuzioni di Poisson.** Dalla proprietà riproduttiva delle distribuzioni di Poisson (Teorema 8.6) discende che

$$\begin{aligned} \tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_n &\sim \text{i.i.d. Poisson}(\lambda = \lambda_{\tilde{X}}) \Rightarrow \\ \Rightarrow \tilde{Y}_n = \tilde{X}_1 + \tilde{X}_2 + \dots + \tilde{X}_n &\sim \text{Poisson}(\lambda = n\lambda_{\tilde{X}}). \end{aligned}$$

In virtù del TLC possiamo dunque concludere che per n sufficientemente elevato la distribuzione $Poisson(\lambda = n\lambda_{\tilde{X}})$ possa essere approssimata attraverso la distribuzione normale di parametri

$$\mu = E(\tilde{Y}_n) = n\lambda_{\tilde{X}} \quad \text{e} \quad \sigma^2 = Var(\tilde{Y}_n) = n\lambda_{\tilde{X}}.$$

L'errore di approssimazione che si commette approssimando la distribuzione $Poisson(\lambda = n\lambda_{\tilde{X}})$ con la distribuzione $Normale(\mu = n\lambda_{\tilde{X}}; \sigma = \sqrt{n\lambda_{\tilde{X}}})$ dipende dal valore di $\lambda = n\lambda_{\tilde{X}}$. Per valori piccoli di $\lambda = n\lambda_{\tilde{X}}$ (ovvero se n non è abbastanza elevato in relazione al valore di $\lambda_{\tilde{X}}$) l'errore di approssimazione è elevato come si vede nel primo dei due grafici sottostanti.



- **Approssimazioni normali per distribuzioni gamma.** Dalla proprietà riproduttiva delle distribuzioni gamma (Teorema 8.8) discende che

$$\begin{aligned} \tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_n &\sim \text{i.i.d. Gamma}(\alpha_{\tilde{X}}, \theta_{\tilde{X}}) \Rightarrow \\ \Rightarrow \tilde{Y}_n = \tilde{X}_1 + \tilde{X}_2 + \dots + \tilde{X}_n &\sim \text{Gamma}(\alpha = n\alpha_{\tilde{X}}; \theta = \theta_{\tilde{X}}). \end{aligned}$$

In virtù del TLC possiamo dunque concludere che per n sufficientemente elevato la distribuzione $Gamma(\alpha = n\alpha_{\tilde{X}}; \theta = \theta_{\tilde{X}})$ possa essere approssimata attraverso la distribuzione normale di parametri

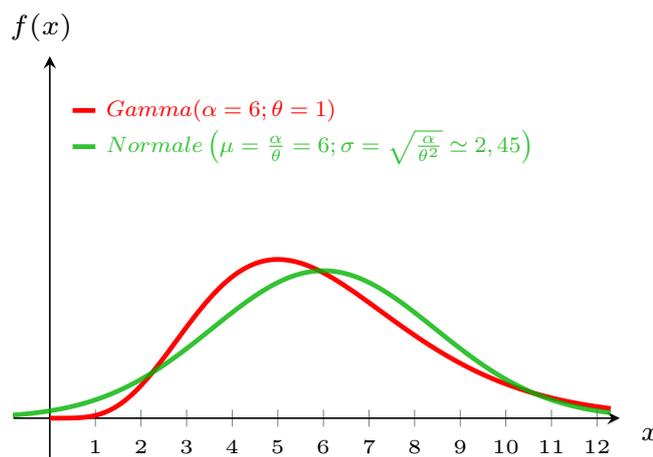
$$\mu = E(Y_n) = n\frac{\alpha_{\tilde{X}}}{\theta_{\tilde{X}}} \quad \text{e} \quad \sigma^2 = Var(Y_n) = n\frac{\alpha_{\tilde{X}}}{\theta_{\tilde{X}}^2}.$$

L'errore di approssimazione che si commette approssimando la distribuzione $Gamma(\alpha = n\alpha_{\tilde{X}}; \theta = \theta_{\tilde{X}})$ attraverso la distribuzione

$$Normale\left(\mu = n\frac{\alpha_{\tilde{X}}}{\theta_{\tilde{X}}}; \sigma = \sqrt{n\alpha_{\tilde{X}}/\theta_{\tilde{X}}^2}\right)$$

dipende dal valore di $\alpha = n\alpha_{\tilde{X}}$. Per valori piccoli di $\alpha = n\alpha_{\tilde{X}}$ (ovvero se n è piccolo in relazione a $\alpha_{\tilde{X}}$) l'errore di approssimazione è elevato.

Il grafico sottostante mostra l'approssimazione normale alla distribuzione $Gamma(\alpha = 6; \theta = 1)$:



- **Approssimazioni normali per distribuzioni chi-quadrato.** Siccome la [distribuzione chi-quadrato](#) con k gradi di libertà è un caso particolare di distribuzione gamma che si ottiene ponendo $\alpha = k/2$ e $\theta = 1/2$ (vedi il Teorema 8.12), anche le [distribuzioni chi-quadrato](#) con un elevato numero di gradi di libertà possono essere approssimate attraverso distribuzioni normali. Siccome il valore atteso e lo scarto quadratico medio di una [distribuzione chi-quadrato](#) sono rispettivamente dati da $\mu = k$ e $\sigma = \sqrt{2k}$ (invitiamo il lettore a verificare), possiamo concludere che questi ultimi siano anche i parametri che identificano un'approssimazione normale alla [distribuzione chi-quadrato](#) con k gradi di libertà.

Esercizio 8.24. Un esperimento casuale consiste in $n = 100$ lanci di un dado regolare.

- a) Qual è la probabilità di ottenere al più 20 volte il punteggio sei?
- b) Qual è la probabilità di ottenere esattamente 20 volte il punteggio sei?
- c) Si determini l'ottavo decile della variabile casuale che conta il numero di volte che si ottiene il punteggio sei nei 100 lanci del dado.

Risposte:

- a) Qual è la probabilità di ottenere al più 20 volte il punteggio sei?

Sia \tilde{Y} la variabile casuale che conta il numero di volte che si ottiene il punteggio sei nei 100 lanci del dado. Per quanto abbiamo detto sulle distribuzioni binomiali, possiamo assumere che

$$\tilde{Y} \sim \text{Binomiale}(n = 100; p = 1/6).$$

Volendo calcolare il valore esatto di

$$P(\{\tilde{Y} \leq 20\}) = \sum_{x=0}^{20} \binom{100}{x} \left(\frac{1}{6}\right)^x \left(1 - \frac{1}{6}\right)^{100-x}$$

ci troviamo di fronte al problema di dover calcolare una somma di 20 termini nei quali compaiono funzioni fattoriali che restituiscono dei numeri molto elevati. Calcolare il valore esatto di $P(\{\tilde{Y} \leq 20\})$ senza l'ausilio di un *computer* richiede dunque molto tempo.

Tuttavia, se disponiamo della [tavola della distribuzione normale standard](#), possiamo sfruttare l'approssimazione normale alla distribuzione binomiale di \tilde{Y} onde ottenere con poca fatica un'approssimazione molto precisa al valore di $P(\{\tilde{Y} \leq 20\})$. Infatti, siccome il valore di

$$np(1-p) = 100 \times \frac{1}{6} \times \left(1 - \frac{1}{6}\right) = 13,889$$

è molto più grande di 5, possiamo concludere che la distribuzione *Binomiale*($n; p$) di \tilde{Y} sia molto prossima alla distribuzione normale di parametri

$$\mu = np = 100 \times \frac{1}{6} = 16,667$$

e

$$\sigma = \sqrt{np(1-p)} = \sqrt{100 \times \frac{1}{6} \times \left(1 - \frac{1}{6}\right)} = 3,727,$$

e per questo motivo possiamo calcolare un'approssimazione al valore di $P(\{\tilde{Y} \leq 20\})$ attraverso la formula

$$\begin{aligned} P(\{\tilde{Y} \leq 20\}) &\simeq \Phi\left(\frac{20 - np}{\sqrt{np(1-p)}}\right) = \Phi\left(\frac{20 - 16,667}{3,727}\right) = \\ &= \Phi(0,89) = [[\text{tavola } \Phi(z)]] = 0,8133. \end{aligned} \quad (169)$$

Per confronto, il valore esatto di $P(\{\tilde{Y} \leq 20\})$ (che può essere ottenuto con l'ausilio di un *computer*) è dato da

$$P(\{\tilde{Y} \leq 20\}) = \sum_{x=0}^{20} \binom{100}{x} \left(\frac{1}{6}\right)^x \left(1 - \frac{1}{6}\right)^{100-x} = 0,8481.$$

Si noti che l'errore di approssimazione che abbiamo commesso attraverso la formula (169) è molto piccolo.

A volte, per tener conto del fatto che le distribuzioni binomiali sono discrete mentre quelle normali sono continue, al posto dell'approssimazione (169) si

calcola l'approssimazione

$$\begin{aligned} P(\{\tilde{Y} \leq 20+0,5\}) &\simeq \Phi\left(\frac{20,5 - np}{\sqrt{np(1-p)}}\right) = \Phi\left(\frac{20,5 - 16,667}{3,727}\right) = \\ &= \Phi(1,03) = [[\text{tavola } \Phi(z)]] = 0,8485 \end{aligned}$$

applicando la cosiddetta **"correzione di continuità"**. Si noti che in questo caso la correzione di continuità ha effettivamente migliorato la precisione dell'approssimazione, ma tra breve vedremo un esempio dove attraverso la correzione di continuità la precisione peggiora. Si noti in ogni caso che l'impatto della correzione di continuità sul valore numerico dell'approssimazione tende a zero all'aumentare di n .

- b) Qual è la probabilità di ottenere esattamente 20 volte il punteggio sei?

Usando l'approssimazione normale alla distribuzione *Binomiale* ($n = 100; p = 1/6$), si può anche ottenere un'approssimazione al valore di $P(\{\tilde{Y} = 20\})$:

$$\begin{aligned} P(\{\tilde{Y} = 20\}) &= P(\{\tilde{Y} \leq 20,5\}) - P(\{\tilde{Y} \leq 19,5\}) = \\ &\simeq \Phi\left(\frac{20,5 - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{19,5 - np}{\sqrt{np(1-p)}}\right) = \\ &= \Phi\left(\frac{20,5 - 16,667}{3,727}\right) - \Phi\left(\frac{19,5 - 16,667}{3,727}\right) = \\ &= \Phi(1,03) - \Phi(0,76) = \\ &= [[\text{tavola } \Phi(z)]] = 0,8485 - 0,7764 = 0,0721 \end{aligned}$$

In questo caso il valore esatto di $P(\{\tilde{Y} = 20\})$ è dato da

$$P(\tilde{Y} = 20) = \binom{100}{20} \left(\frac{1}{6}\right)^{20} \left(1 - \frac{1}{6}\right)^{100-20} = 0,0679,$$

ma per ottenere questo valore con l'ausilio di una calcolatrice tascabile siamo costretti a calcolare il valore di $100!/80!$ che è un numero piuttosto elevato.

- c) Si determini l'ottavo decile della variabile casuale che conta il numero di volte che si ottiene il punteggio sei nei 100 lanci del dado.

Siccome la distribuzione *Binomiale* ($n = 100; p = 1/6$) di \tilde{Y} può essere approssimata attraverso la distribuzione *Normale* ($\mu = np; \sigma = \sqrt{np(1-p)}$), sembra lecito aspettarsi che l'ottavo decile di \tilde{Y} sia prossimo all'ottavo decile della suddetta distribuzione normale, ovvero al valore di

$$\begin{aligned} y_{0,80} &= np + z_{0,80} \times \sqrt{np(1-p)} = [[\text{tavola } \Phi(z)]] = \\ &= 100 \times \frac{1}{6} + 0,84 \times \sqrt{100 \times \frac{1}{6} \left(1 - \frac{1}{6}\right)} = 19,80. \end{aligned}$$

Per valutare la precisione di questa approssimazione, possiamo determinare il valore esatto dell'ottavo decile di \tilde{Y} attraverso il calcolo delle probabilità cumulate. In questo modo si può verificare (eventualmente con l'ausilio di un *computer* oppure di una calcolatrice tascabile) che il valore esatto dell'ottavo decile di \tilde{Y} sia dato da

$$y_{0,80} = 20$$

Si noti che per determinare questo valore con una calcolatrice tascabile di tipo tradizionale bisogna calcolare ben 21 probabilità cumulate.

Esercizio 8.25. Sia \tilde{Y} una variabile casuale con distribuzione di Poisson di parametro $\lambda = 10$.

- Si determini $P(\{\tilde{Y} \leq 16\})$.
- Si determini il primo quartile di \tilde{Y} .

Risposte:

- Si determini $P(\{\tilde{Y} \leq 16\})$.

Anche in questo caso, il valore esatto

$$P(\{\tilde{Y} \leq 16\}) = \sum_{x=0}^{16} \frac{e^{-10} \times 10^x}{x!} = 0,9730$$

può essere ottenuto solo attraverso numerose operazioni di calcolo. Se disponiamo della [tavola della distribuzione normale standard](#), possiamo invece sfruttare l'approssimazione normale alla distribuzione $Poisson(\lambda = 10)$ e risparmiare le suddette operazioni di calcolo. Infatti, usando l'approssimazione normale otteniamo

$$\begin{aligned} P(\{\tilde{Y} \leq 16\}) &\simeq \Phi\left(\frac{16 - \lambda}{\sqrt{\lambda}}\right) = \\ &= \Phi\left(\frac{16 - 10}{\sqrt{10}}\right) = \text{[[tavola } \Phi(z) \text{]]} = \Phi(1,90) = 0,9713 \end{aligned}$$

e volendo considerare anche la "correzione di continuità" otteniamo

$$\begin{aligned} P(\{\tilde{Y} \leq 16\}) &\simeq \Phi\left(\frac{16+0,5 - \lambda}{\sqrt{\lambda}}\right) = \\ &= \Phi\left(\frac{16+0,5 - 10}{\sqrt{10}}\right) = \Phi(2,06) = \text{[[tavola } \Phi(z) \text{]]} = 0,9803. \end{aligned}$$

Come si desume dal confronto con il valore esatto $P(\{\tilde{Y} \leq 16\}) = 0,9730$, in questo caso la "correzione di continuità" ha (lievemente) peggiorato la precisione dell'approssimazione.

b) Si determini il primo quartile di \tilde{Y} .

Per calcolare il valore esatto del primo quartile della distribuzione $Poisson(\lambda = 10)$ bisogna calcolare le probabilità cumulate. Dopo al nona probabilità cumulata si vede che il valore esatto del primo quartile è dato da

$$y_{0,25} = 8.$$

Per evitare il calcolo delle probabilità cumulate possiamo accontentarci dell'approssimazione al valore di $y_{0,25}$ che si ottiene a partire dall'approssimazione normale suggerita dal TLC. L'approssimazione in questione è data da

$$y_{0,25} \simeq \lambda + z_{0,25}\sqrt{\lambda} = [[\text{tavola } \Phi(z)]] = 10 - 0,67 \times \sqrt{10} = 7,88$$

ed è molto prossima al valore esatto. Ovviamente, questa approssimazione può essere calcolata solo se si conosce il valore di $z_{0,25} = 0,67$.

Esercizio 8.26. Sia \tilde{Y} una variabile casuale con distribuzione binomiale negativa di parametri $r = 10$ e $p = 0,75$.

- a) Si calcoli $P(\{\tilde{Y} \leq 16\})$.
 b) Si determini il nono decile di \tilde{Y} .

Risposte:

- a) Si calcoli $P(\{\tilde{Y} \leq 16\})$.

Anche in questo caso, il valore esatto di

$$P(\{\tilde{Y} \leq 16\}) = \sum_{y=10}^{16} \binom{y-1}{10-1} 0,75^y (1-0,75)^{10-y} = 0,9598.$$

può essere ottenuto solo attraverso numerose operazioni di calcolo. Se disponiamo della [tavola della distribuzione normale standard](#), possiamo risparmiarci le suddette operazioni di calcolo sfruttando l'approssimazione normale alla distribuzione $Bin.neg(r = 10; p = 0,75)$:

$$\begin{aligned} P(\{\tilde{Y} \leq 16\}) &\simeq \Phi\left(\frac{16 - \frac{r}{p}}{\sqrt{\frac{r(1-p)}{p^2}}}\right) = \\ &= \Phi\left(\frac{16 - \frac{10}{0,75}}{\sqrt{\frac{10(1-0,75)}{0,75^2}}}\right) = \Phi(1,26) = [[\text{tavola } \Phi(z)]] = 0,8962. \end{aligned}$$

Volendo considerare anche la "correzione di continuità" otteniamo invece

$$P(\{\tilde{Y} \leq 16\}) \simeq \Phi\left(\frac{16+0,5 - \frac{10}{0,75}}{\sqrt{\frac{10(1-0,75)}{0,75^2}}}\right) =$$

$$= \Phi(1,50) = [[\text{tavola } \Phi(z)]] = 0,9332$$

che, in questo caso, fornisce un'approssimazione più precisa.

b) Si determini il nono decile di \tilde{Y} .

Per determinare il nono decile della distribuzione $Bin.neg(r = 10; p = 0,75)$ di \tilde{Y} bisogna calcolare le probabilità cumulate fino ad uguagliare o superare il valore 0,90. Così facendo si trova il valore esatto del nono decile che è dato da

$$y_{0,90} = 16.$$

Sfruttando l'approssimazione normale alla distribuzione $Bin.neg(r = 10; p = 0,75)$, si può calcolare un'approssimazione al valore esatto $y_{0,90} = 16$ senza ricorrere al calcolo delle probabilità cumulate:

$$y_{0,90} \simeq \frac{r}{p} + z_{0,90} \sqrt{\frac{r(1-p)}{p^2}} =$$

$$= [[\text{tavola } \Phi(z)]] = \frac{10}{0,75} + 1,28 \sqrt{\frac{10(1-0,75)}{0,75^2}} = 16,03$$

Si noti, tuttavia, che questa approssimazione può essere calcolata solo se si conosce il valore di $z_{0,90} = 1,28$.

Esercizio 8.27. Sia \tilde{Y} una variabile casuale con distribuzione gamma di parametri $\alpha = 10$ e $\theta = 2$.

- Si calcoli $P(\{\tilde{Y} \leq 9\})$.
- Si determini il terzo quartile di \tilde{Y} .

Risposte:

- Si calcoli $P(\{\tilde{Y} \leq 9\})$.

Come abbiamo visto (vedi la formula (152)), la funzione di ripartizione di una distribuzione gamma con parametro α intero è data da

$$F(y) = P(\{\tilde{Y} \leq y\}) = \begin{cases} 1 - \sum_{x=0}^{\alpha-1} \frac{e^{-\theta y} (\theta y)^x}{x!} & \text{se } y \geq 0 \\ 0 & \text{se } y < 0 \end{cases} \quad (170)$$

e la probabilità richiesta può dunque essere calcolata come

$$P(\{\tilde{Y} \leq 9\}) = 1 - \sum_{x=0}^{10-1} \frac{e^{-2 \times 9} (2 \times 9)^x}{x!} = 0,9846.$$

Per calcolare il valore di $P(\{\tilde{Y} \leq 9\}) = 0,9846$ attraverso la formula (170) dobbiamo però effettuare numerose operazioni di calcolo che possiamo evitare se disponiamo della [tavola della distribuzione normale standard](#). Infatti, siccome il valore del parametro α della distribuzione $\text{Gamma}(\alpha = 10; \theta = 2)$ è abbastanza elevato, questa distribuzione gamma dovrebbe essere abbastanza prossima alla distribuzione normale di parametri

$$\mu = \frac{\alpha}{\theta} = \frac{10}{2} = 5 \quad \text{e} \quad \sigma = \sqrt{\frac{\alpha}{\theta^2}} = \sqrt{\frac{10}{2^2}} = 1,581.$$

Facendo riferimento a questa approssimazione possiamo dunque concludere che

$$P(\{\tilde{Y} \leq 9\}) \simeq \Phi\left(\frac{9 - \alpha/\theta}{\sqrt{\alpha/\theta^2}}\right) = \Phi\left(\frac{9 - 10/2}{\sqrt{10/2^2}}\right) = \Phi(2,53) = 0,9943.$$

Si noti che l'approssimazione

$$F(y) = P(\{\tilde{Y} \leq y\}) \simeq \Phi\left(\frac{y - \alpha/\theta}{\sqrt{\alpha/\theta^2}}\right)$$

può essere sfruttata anche se il valore del parametro α non è un numero intero, ovvero quando il valore di

$$F(y) = P(\{\tilde{Y} \leq y\}) = \int_0^y \frac{\theta^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\theta t} dt$$

non può essere calcolato attraverso la formula (170).

b) Si determini il terzo quartile di \tilde{Y} .

In generale, i quantili di una distribuzione gamma non sono delle funzioni elementari e per calcolarli bisogna quindi ricorrere ad un opportuno algoritmo che, a seconda della precisione desiderata, potrebbe richiedere numerose operazioni di calcolo. In ogni caso, se il valore del parametro α è sufficientemente elevato, possiamo fare riferimento all'approssimazione normale suggerita dal TLC per determinare delle approssimazioni ai quantili di una distribuzione gamma.

Nel caso in questione abbiamo $\alpha = 10$, e l'approssimazione normale dà luogo alla seguente approssimazione per il terzo quartile della distribuzione $\text{Gamma}(\alpha = 10; \theta = 2)$:

$$y_{0,75} \simeq \frac{\alpha}{\theta} + z_{0,75} \sqrt{\frac{\alpha}{\theta^2}} = [[\text{tavola } \Phi(z)]] = \frac{10}{2} + 0,67 \times \sqrt{\frac{10}{2^2}} = 5,30.$$

Si noti che questa approssimazione può essere calcolata solo se si conosce il valore di $z_{0,75} = 0,67$.

Per confronto, il valore "esatto" di $y_{0,75}$, determinato con l'ausilio di un *computer*, è dato da

$$y_{0,75} = 5,9569.$$

8.12 La famiglia delle distribuzioni lognormali

La famiglia delle distribuzioni lognormali è la famiglia di tutte le distribuzioni assolutamente continue che hanno una funzione di densità del tipo

$$f(x) = \begin{cases} \frac{1}{\sqrt{2\pi\delta}} \frac{1}{x} e^{-\frac{1}{2}\left(\frac{\ln x - \gamma}{\delta}\right)^2} & \text{per } x > 0 \\ 0 & \text{per } x \leq 0 \end{cases}$$

dove il parametro γ può essere un numero reale qualunque, e dove il parametro δ è un numero reale positivo. Come si evince dalla definizione delle funzioni di densità, le distribuzioni lognormali concentrano tutta la probabilità sull'insieme dei numeri reali positivi e pertanto si prestano come distribuzioni per variabili casuali che descrivono prezzi, volumi, distanze, pesi, durate ecc.

In questa dispensa scriveremo $\text{Log}N(\gamma; \delta)$ per indicare la distribuzione lognormale con parametri γ e δ , e scriveremo

$$\tilde{X} \sim \text{Log}N(\gamma; \delta)$$

per indicare che \tilde{X} è una variabile casuale con distribuzione $\text{Log}N(\gamma; \delta)$.

La famiglia delle distribuzioni lognormali prende il suo nome dal fatto che

$$\tilde{X} \sim \text{Log}N(\gamma; \delta) \Leftrightarrow \tilde{Y} = \ln \tilde{X} \sim \text{Normale}(\mu = \gamma; \sigma = \delta) \quad (171)$$

(tra breve dimostreremo questa doppia implicazione). Questa relazione tra le distribuzioni normali e quelle lognormali è spesso il motivo per il quale queste ultime compaiono nelle applicazioni della teoria della probabilità. Infatti, in molte applicazioni si è interessati al valore incerto \tilde{X}_t di una qualche variabile non negativa X (come per esempio il prezzo di un titolo finanziario, il peso o le dimensioni di un organismo vivente, ecc.) in un qualche istante temporale futuro t , e le circostanze sotto le quali il valore della variabile

X passa dal suo valore iniziale x_0 (che assumiamo essere noto) al valore finale \tilde{X}_t (incerto) suggeriscono che la **variazione logaritmica**

$$\tilde{Y}_t = \ln \tilde{X}_t - \ln x_0 = \ln \left(\frac{\tilde{X}_t}{x_0} \right)$$

sia descritta da una variabile casuale con distribuzione normale di determinati parametri $\mu = \mu_t$ e $\sigma = \sigma_t$. In virtù di quanto abbiamo già visto sulle trasformazioni affini variabili casuali con distribuzione normale, possiamo dunque dire che

$$\ln \tilde{X}_t = \ln x_0 + \tilde{Y}_t \sim \text{Normale}(\mu = \mu_t + \ln x_0; \sigma = \sigma_t)$$

e dalla relazione (171) si evince quindi che

$$\tilde{X}_t \sim \text{LogN}(\gamma = \mu_t + \ln x_0; \delta = \sigma_t).$$

Per fare un esempio di un'applicazione dove la variazione logaritmica di una variabile X ha distribuzione normale, conviene citare il **modello di Black e Scholes** che è basato sull'ipotesi che l'andamento del prezzo di un'attività finanziaria sia descritto da un cosiddetto **"moto browniano geometrico"**. Questa ipotesi implica che con riferimento a qualsiasi collezione di intervalli di tempo disgiunti, le corrispondenti variazioni logaritmiche del prezzo X dell'attività finanziaria siano descritte da variabili casuali normali indipendenti con parametri μ e σ^2 che sono entrambi proporzionali alle durate degli intervalli di tempo di riferimento (con due costanti di proporzionalità che rimangono costanti nel tempo; si noti che se σ^2 è proporzionale alla durata di un intervallo di tempo, allora σ deve essere proporzionale alla radice della durata). Sulla base di questa e ulteriori ipotesi, Black e Scholes derivarono la celebre formula per il prezzo di opzioni *call* o *put* di tipo europeo che pubblicarono in un articolo scientifico.³⁰

Dimostrazione (Dimostrazione della relazione (171)). Dimostreremo che

$$\tilde{X} \sim \text{LogN}(\gamma; \delta) \Leftrightarrow \tilde{Y} = \ln \tilde{X} \sim \text{Normale}(\mu = \gamma; \sigma = \delta).$$

Cominciamo con l'implicazione "⇒" e assumiamo dunque che $\tilde{X} \sim \text{LogN}(\gamma; \delta)$. In questo caso si ha

$$\begin{aligned} F_{\tilde{Y}}(y) &= P(\{\tilde{Y} \leq y\}) = P(\{\ln \tilde{X} \leq y\}) = P(\{\tilde{X} \leq e^y\}) = \\ &= \int_0^{e^y} \frac{1}{\sqrt{2\pi\delta}} \frac{1}{x} e^{-\frac{1}{2}\left(\frac{\ln x - \gamma}{\delta}\right)^2} dx = \\ &[[\text{cambio di variabile: } t = \ln x \Rightarrow dt = dx/x]] = \\ &= \int_{-\infty}^y \frac{1}{\sqrt{2\pi\delta}} e^{-\frac{1}{2}\left(\frac{t-\gamma}{\delta}\right)^2} dt \end{aligned}$$

e siccome l'espressione nell'ultima riga è la funzione di ripartizione $\text{Normale}(\mu = \gamma; \sigma = \delta)$ possiamo concludere che $\tilde{Y} = \ln \tilde{X} \sim \text{Normale}(\mu = \gamma; \sigma = \delta)$.

³⁰Black, Fischer; Myron Scholes (1973). "The Pricing of Options and Corporate Liabilities". *Journal of Political Economy*. 81 (3): 637–654.

Viceversa, assumiamo ora che $\tilde{Y} = \ln \tilde{X} \sim \text{Normale}(\mu = \gamma; \sigma = \delta)$. In questo caso possiamo scrivere

$$F_{\tilde{X}}(x) = P(\{\tilde{X} \leq x\}) = P(\{e^{\tilde{Y}} \leq x\}) \quad (172)$$

e siccome $e^y > 0$ per ogni $y \in \mathbb{R}$, possiamo concludere che

$$F_{\tilde{X}}(x) = P(\{\tilde{X} \leq x\}) = P(\{e^{\tilde{Y}} \leq x\}) = P(\emptyset) = \llbracket \text{legge L1} \rrbracket = 0 \quad \text{se } x \leq 0.$$

D'altra parte, se $x > 0$, possiamo proseguire nella (172) e scrivere

$$\begin{aligned} F_{\tilde{X}}(x) &= P(\{\tilde{X} \leq x\}) = P(\{e^{\tilde{Y}} \leq x\}) = P(\{\tilde{Y} \leq \ln x\}) = \\ &= \int_{-\infty}^{\ln x} \frac{1}{\sqrt{2\pi\delta}} e^{-\frac{1}{2}\left(\frac{y-\gamma}{\delta}\right)^2} dy = \\ &\llbracket \text{cambio di variabile: } y = \ln t \Rightarrow dy = dt/t \rrbracket = \\ &= \int_0^x \frac{1}{\sqrt{2\pi\delta}} \frac{1}{t} e^{-\frac{1}{2}\left(\frac{\ln t - \gamma}{\delta}\right)^2} dt. \end{aligned}$$

Questo ragionamento dimostra dunque che $F_{\tilde{X}}(x) = P(\{\tilde{X} \leq x\}) = \int_{-\infty}^x f(t)dt$, dove

$$f(t) = \begin{cases} \frac{1}{\sqrt{2\pi\delta}} \frac{1}{t} e^{-\frac{1}{2}\left(\frac{\ln t - \gamma}{\delta}\right)^2} & \text{per } t > 0 \\ 0 & \text{per } t \leq 0 \end{cases},$$

è la funzione di densità lognormale di parametri γ e δ . Quindi possiamo concludere che $\tilde{X} \sim \text{LogN}(\gamma; \delta)$ come volevamo dimostrare. \square

Funzione di ripartizione. Dalla relazione (171), che qui per comodità riscriviamo

$$\tilde{X} \sim \text{LogN}(\gamma; \delta) \quad \Leftrightarrow \quad \tilde{Y} = \ln \tilde{X} \sim \text{Normale}(\mu = \gamma; \sigma = \delta),$$

si evince immediatamente che la funzione di ripartizione lognormale sia data da

$$F(x) = P(\{\tilde{X} \leq x\}) = \begin{cases} P(\{\tilde{Y} \leq \ln x\}) = \Phi\left(\frac{\ln x - \gamma}{\delta}\right) & \text{per } x > 0 \\ 0 & \text{per } x \leq 0 \quad \llbracket [f(x) = 0 \text{ se } x \leq 0] \rrbracket. \end{cases} \quad (173)$$

Partendo da questa formula si ricava facilmente un'espressione per i quantili delle distribuzioni lognormali:

$$F(x_p) = p \quad \Rightarrow \quad \Phi\left(\frac{\ln x_p - \gamma}{\delta}\right) = p \quad \Rightarrow \quad \frac{\ln x_p - \gamma}{\delta} = z_p \quad \Rightarrow \quad x_p = e^{\gamma + \delta z_p} \quad (174)$$

(come al solito, z_p indica il quantile di ordine p della distribuzione normale standard).

Grafico delle funzioni di densità lognormali. Come si desume dalla formula defintoria

$$f(x) = \begin{cases} \frac{1}{\sqrt{2\pi\delta}} \frac{1}{x} e^{-\frac{1}{2}\left(\frac{\ln x - \gamma}{\delta}\right)^2} & \text{per } x > 0, \\ 0 & \text{per } x \leq 0, \end{cases}$$

le funzioni di densità lognormali sono positive se e solo se $x > 0$. Per analizzare il comportamento delle funzioni di densità lognormali in prossimità di $x = 0$, conviene osservare che

$$\frac{1}{x} e^{-\frac{1}{2}\left(\frac{\ln x - \gamma}{\delta}\right)^2} = e^{-\frac{1}{2}\left(\frac{\ln x - \gamma}{\delta}\right)^2 - \ln x} = e^{-\frac{1}{2\delta^2}(\ln x)^2 + \left(\frac{\gamma}{\delta} - 1\right)\ln x - \frac{\gamma^2}{2\delta^2}}$$

e che il limite per x che tende a zero da destra dell'espressione sul lato destro dell'ultima uguaglianza è sempre nullo (per qualsiasi valori dei parametri $\gamma \in \mathbb{R}$ e $\delta > 0$) perché quando x tende a zero da destra, $(\ln x)^2$ tende a $+\infty$ più velocemente di $|\ln x|$. Da queste osservazioni deduciamo che **tutte le funzioni di densità lognormali sono continue nel punto $x = 0$** .

Ora siccome abbiamo appena dimostrato che

$$\lim_{x \rightarrow 0} f(x) = 0$$

e siccome le funzioni di densità lognormali sono continue e dunque si deve anche avere

$$\lim_{x \rightarrow \infty} f(x) = 0$$

(altrimenti sarebbe violata la proprietà f2), tutte le funzioni di densità lognormali devono avere almeno un punto di massimo x_{moda} positivo. Per trovare il valore della moda oppure i valori delle mode (per ora non possiamo ancora escludere che ce ne sia più di una) di una funzione di densità lognormale, calcoliamo la derivata $f'(x) = df(x)/dx$ e cerchiamo i punti x che la rendono nulla:

$$f'(x) = \frac{df(x)}{dx} = -\frac{1}{\sqrt{2\pi}\delta} \frac{1}{x^2} e^{-\frac{1}{2}\left(\frac{\ln x - \gamma}{\delta}\right)^2} \left[1 + \frac{\ln x - \gamma}{\delta^2}\right]$$

$$\Rightarrow f'(x) = 0 \Leftrightarrow 1 + \frac{\ln x - \gamma}{\delta^2} = 0 \Leftrightarrow x = x_{moda} = e^{\gamma - \delta^2}.$$

Da questo ragionamento desumiamo che **tutte le distribuzioni lognormali hanno un'unica moda che si trova in corrispondenza del punto $x_{moda} = e^{\gamma - \delta^2}$** .

A questo punto conviene analizzare più da vicino il ruolo del parametro γ . Infatti, il fatto che i quantili

$$x_p = e^{\gamma + \delta z_p} = e^\gamma \times e^{\delta z_p}$$

siano proporzionali a e^γ , suggerisce che $\gamma' = e^\gamma$ sia un parametro di scala e che $\ln \gamma' = \gamma$ sia dunque un **parametro di scala "logaritmico"**. Per confermare questa intuizione, basta "riparametrizzare" la funzione di densità lognormale

$$f(x) = \begin{cases} \frac{1}{\sqrt{2\pi}\delta} \frac{1}{x} e^{-\frac{1}{2}\left(\frac{\ln x - \gamma}{\delta}\right)^2} & \text{per } x > 0 \\ 0 & \text{per } x \leq 0 \end{cases}$$

ponendo $\gamma' = e^\gamma$ e $\delta' = \delta$:

$$f(x) = \begin{cases} \frac{1}{\sqrt{2\pi}\gamma'\delta'} \frac{\gamma'}{x} e^{-\frac{1}{2}\left(\frac{\ln(x/\gamma')}{\delta'}\right)^2} & \text{per } x > 0 \\ 0 & \text{per } x \leq 0 \end{cases}$$

Dall'espressione della funzione di densità riparametrizzata si evince che per disegnare il grafico di una funzione di densità loglineare con $\gamma' = e^\gamma \neq 1$, basta disegnare il grafico della funzione di densità loglineare con lo stesso valore del parametro $\delta' = \delta$ e con $\gamma' = e^\gamma = 1$ e poi

- **moltiplicare** per il valore di $\gamma' = e^\gamma$ i valori riportati sull'asse delle ascisse
- e **dividere** per il valore di $\gamma' = e^\gamma$ i valori riportati sull'asse delle ordinate.

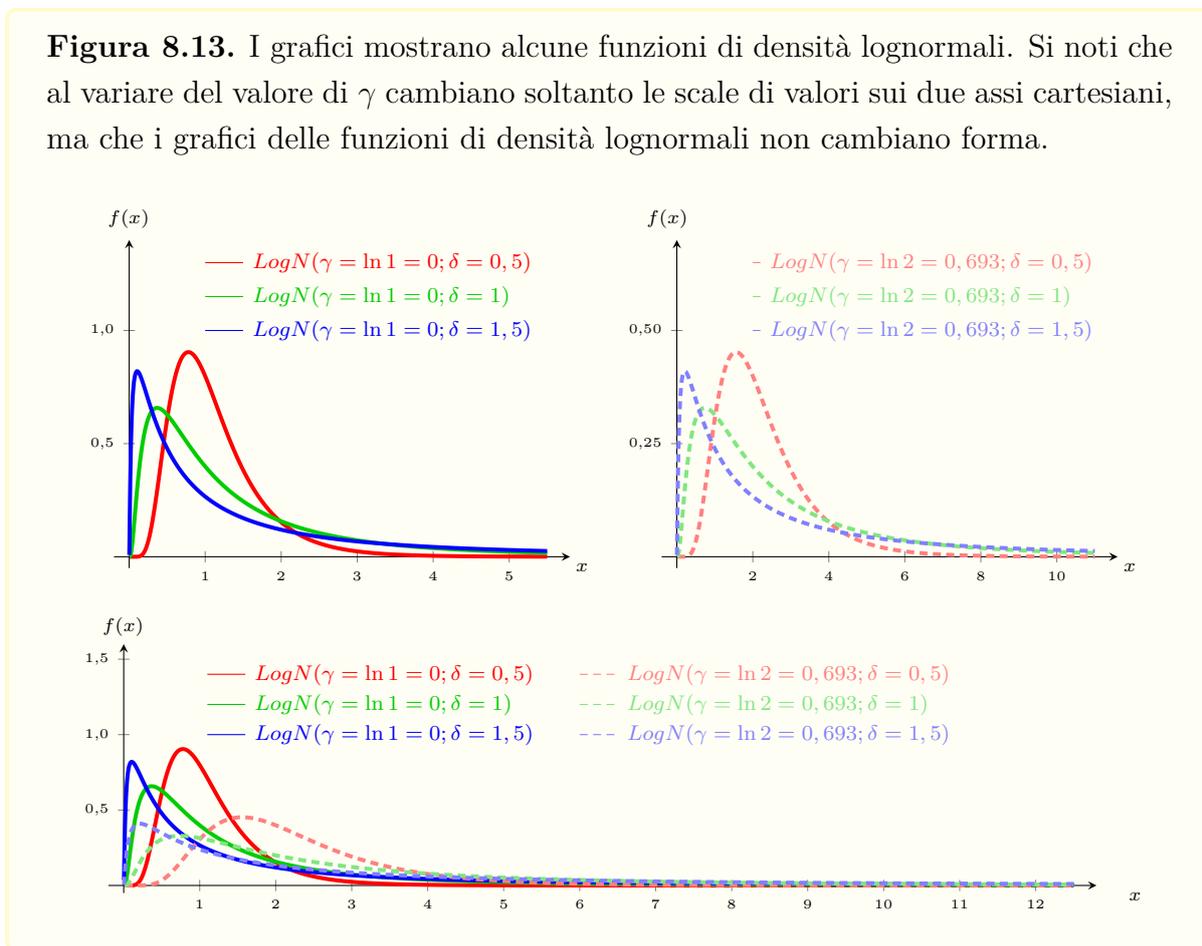
Siccome i valori sull'asse delle ascisse vengono moltiplicati per il valore del parametro di scala $\gamma' = e^\gamma$, quest'ultimo deve essere un **parametro di scala "diretto"** e non un parametro di scala "inverso" come il parametro θ delle distribuzioni esponenziali. Infatti, seguendo la falsariga della dimostrazione dell'implicazione (147) si può facilmente dimostrare che

$$\tilde{X} \sim \text{LogN}(\gamma = \ln \gamma'; \delta) \quad \Rightarrow \quad \tilde{Y} = b\tilde{X} \sim \text{LogN}(\gamma + \ln b = \ln(b\gamma'); \delta) \quad (175)$$

per ogni $b > 0$.

I grafici in Figura 8.13 mostrano alcune funzioni di densità lognormali con diversi valori dei parametri γ e δ .

Figura 8.13. I grafici mostrano alcune funzioni di densità lognormali. Si noti che al variare del valore di γ cambiano soltanto le scale di valori sui due assi cartesiani, ma che i grafici delle funzioni di densità lognormali non cambiano forma.



Approssimazione normale. Dalla relazione (171), che qui per comodità riscriviamo

$$\tilde{X} \sim \text{LogN}(\gamma; \delta) \quad \Leftrightarrow \quad \tilde{Y} = \ln \tilde{X} \sim \text{Normale}(\mu = \gamma; \sigma = \delta),$$

discende che le distribuzioni lognormali **con parametro δ prossimo a zero** possano essere approssimate attraverso opportune distribuzioni normali.

Infatti, se $\tilde{X} \sim \text{LogN}(\gamma; \delta)$ e pertanto $\tilde{Y} = \ln \tilde{X} \sim \text{Normale}(\mu = \gamma; \sigma = \delta)$, allora possiamo scrivere

$$\tilde{X} = e^{\ln \tilde{X}} = e^{\tilde{Y}} = e^{\gamma + (\tilde{Y} - \gamma)} = e^{\gamma} \times e^{\tilde{Y} - \gamma} = e^{\gamma} \times e^{\tilde{Y}'},$$

dove (Teorema 8.9)

$$\tilde{Y}' = \tilde{Y} - \gamma \sim \text{Normale}(0; \delta).$$

Ora, se il valore di $\delta > 0$ è molto prossimo a zero, la distribuzione $\text{Normale}(0, \delta)$ della variabile casuale \tilde{Y}' concentrerà quasi tutta la probabilità in prossimità dello zero e pertanto possiamo concludere che

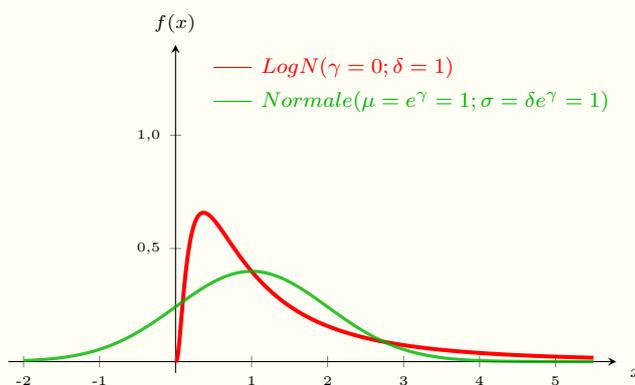
$$P(\{\tilde{X} = e^{\gamma} \times e^{\tilde{Y}'} \simeq e^{\gamma} \times (1 + \tilde{Y}') = e^{\gamma} + e^{\gamma} \tilde{Y}'\}) \simeq 1$$

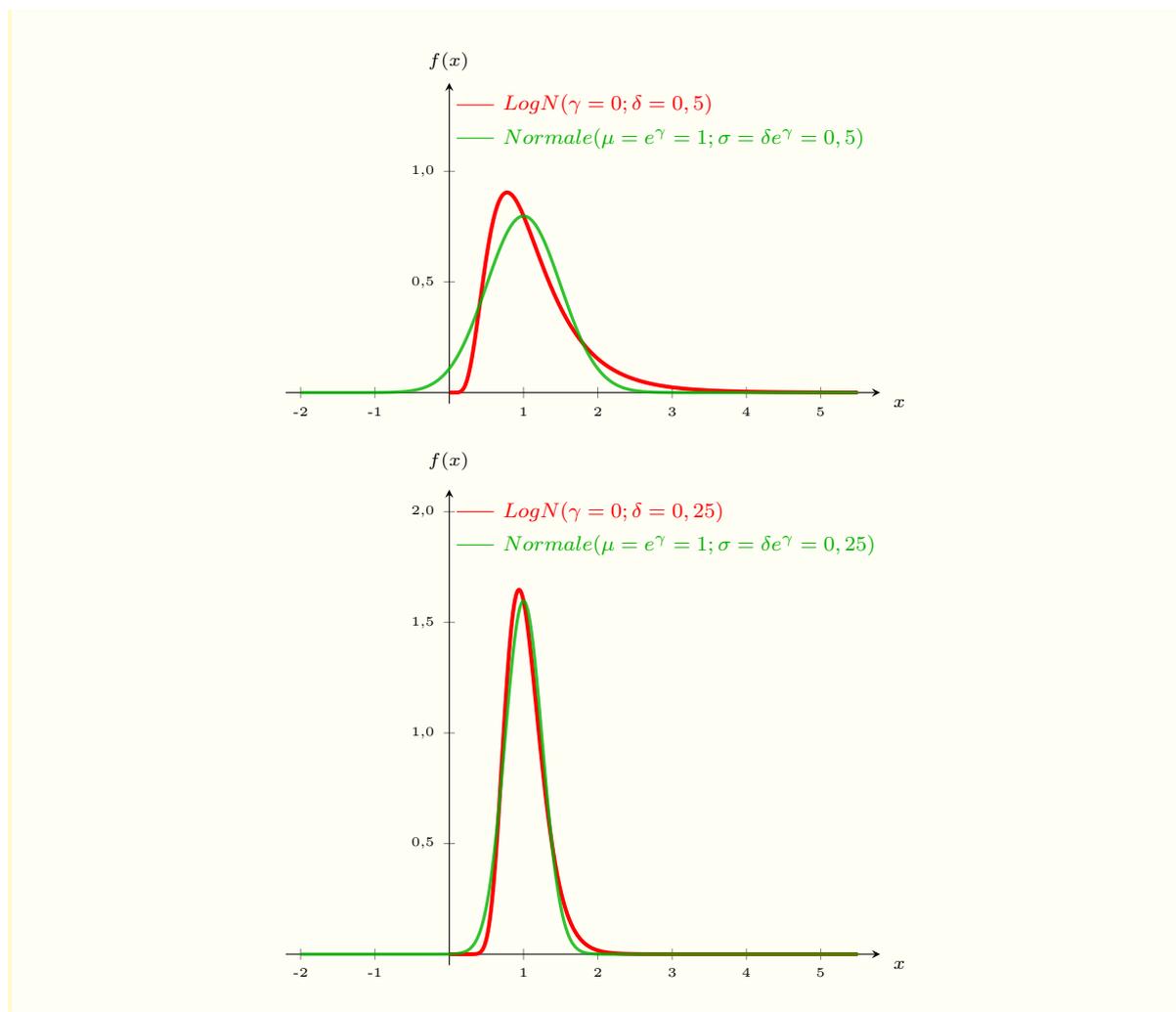
(ricordiamo che $e^y \simeq 1 + y$ se $y \simeq 0$). Siccome

$$\tilde{Y}' \sim \text{Normale}(0; \delta) \Rightarrow [[\text{Teorema 8.9}]] \Rightarrow e^{\gamma} + e^{\gamma} \tilde{Y}' \sim \text{Normale}(e^{\gamma}; \delta e^{\gamma}),$$

questo ragionamento suggerisce che per $\delta > 0$ prossimo a zero la distribuzione $\text{LogN}(\gamma; \delta)$ di \tilde{X} sia molto prossima alla distribuzione $\text{Normale}(e^{\gamma}; \delta e^{\gamma})$ della variabile casuale $e^{\gamma} + e^{\gamma} \tilde{Y}'$!!! Chiaramente, il ragionamento che ci ha condotto a questa conclusione andrebbe formalizzato un po' meglio per poter essere considerato come una dimostrazione, ma per evitare troppi tecnicismi rinunceremo a farlo. In compenso presentiamo alcuni grafici che illustrano l'approssimazione normale alle distribuzioni lognormali con parametro δ prossimo a zero (vedi Figura 8.14).

Figura 8.14. I grafici mostrano come una distribuzione lognormale con parametro δ prossimo a zero può essere approssimata attraverso la distribuzione normale con $\mu = e^{\gamma}$ e $\sigma = e^{\gamma} \delta$.





Momenti. I momenti delle distribuzioni lognormali possono essere facilmente ricavati attraverso la relazione (171), ovvero la relazione

$$\tilde{X} \sim \text{LogN}(\gamma; \delta) \Leftrightarrow \tilde{Y} = \ln \tilde{X} \sim \text{Normale}(\mu = \gamma; \sigma = \delta).$$

Infatti, da questa relazione discende che i momenti di una distribuzione lognormale possono essere calcolati come

$$\mu_k = E(\tilde{X}^k) = E\left(\left[e^{\tilde{Y}}\right]^k\right) = E\left(e^{k\tilde{Y}}\right) = m_{\tilde{Y}}(k), \quad k = 1, 2, \dots$$

dove

$$m_{\tilde{Y}}(t) = E\left(e^{t\tilde{Y}}\right) = \text{[[formula (165)]]} = e^{t\gamma + \frac{1}{2}t^2\delta^2}, \quad t \in \mathbb{R},$$

è la fgm della distribuzione normale di parametri $\mu = \gamma$ e $\sigma = \delta$. Questo ragionamento dimostra che **per tutte le distribuzioni lognormali esiste l'intera successione dei momenti** e che questi ultimi possono essere calcolati come

$$\mu_k = E(\tilde{X}^k) = e^{k\gamma + \frac{1}{2}k^2\delta^2}, \quad k = 1, 2, \dots$$

Usando questa formula vediamo che il valore atteso di una distribuzione lognormale è dato da

$$\mu = E(\tilde{X}) = e^{\gamma + \frac{1}{2}\delta^2} \quad (176)$$

e che la varianza di una distribuzione lognormale è invece data da

$$\begin{aligned} \mu_2 &= E(\tilde{X}^2) = e^{2\gamma + \frac{1}{2}2^2\delta^2} \Rightarrow \\ \Rightarrow \sigma^2 &= \text{var}(\tilde{X}) = \mu_2 - \mu^2 = e^{2\gamma + \delta^2} (e^{\delta^2} - 1). \end{aligned} \quad (177)$$

Per quanto riguarda invece la fgm, non è difficile dimostrare che **nessuna distribuzione lognormale è dotata di fgm** perché per qualsiasi funzione di densità lognormale l'integrale

$$\int_{-\infty}^{\infty} e^{tx} f(x) dx = \int_0^{\infty} e^{tx} \frac{1}{\sqrt{2\pi\delta}} \frac{1}{x} e^{-\frac{1}{2}\left(\frac{\ln x - \gamma}{\delta}\right)^2} dx = \frac{1}{\sqrt{2\pi\delta}} \int_0^{\infty} e^{tx - \ln x - \frac{1}{2}\left(\frac{\ln x - \gamma}{\delta}\right)^2} dx$$

diverge a $+\infty$ per ogni $t > 0$. Per dimostrare questa affermazione basta notare che per $t > 0$ la funzione integranda diverge a $+\infty$ per x che tende a $+\infty$: infatti,

$$\lim_{x \rightarrow \infty} tx - \ln x - \frac{1}{2} \left(\frac{\ln x - \gamma}{\delta} \right)^2 = \infty \quad \text{per ogni } \gamma \in \mathbb{R} \text{ e ogni } t, \delta > 0.$$

In definitiva, abbiamo dunque dimostrato che **per tutte le distribuzioni lognormali esistono i momenti di ogni ordine, ma che nessuna distribuzione lognormale è dotata di fgm.**

Esercizio 8.28. Si consideri un risparmiatore che ha investito 50 mila euro in un fondo e sia \tilde{X} la variabile casuale che descrive il rendimento **logaritmico** annuale del fondo. Si supponga che la distribuzione di \tilde{X} sia normale con parametri

$$\mu_{\tilde{X}} = +0,01 = +1\% \quad \text{e} \quad \sigma_{\tilde{X}} = 0,04 = 4\%.$$

Sia \tilde{Y} la variabile casuale che descrive il valore dell'investimento dopo un anno (in migliaia di euro).

- Si calcoli $P(\{\tilde{Y} > 52\})$.
- Si determini il primo percentile di \tilde{Y} .

Risposte:

Si noti che la consegna di questo esercizio è identica a quella dell'Esercizio 8.21 tranne per il fatto che al posto del rendimento aritmetico in questo esercizio viene considerato il rendimento logaritmico.

a) Si calcoli $P(\{\tilde{Y} > 52\})$.

Per rispondere osserviamo innanzitutto che il valore finale dell'investimento è dato da

$$\tilde{Y} = \text{[[formula (88)]]} = 50 \times e^{\tilde{X}}.$$

Siccome

$$\begin{aligned} \tilde{X} &\sim \text{Normale}(\mu = 0,01; \sigma = 0,04) \Rightarrow \text{[[implicazione (171)]]} \Rightarrow \\ \Rightarrow e^{\tilde{X}} &\sim \text{Lognormale}(\gamma = 0,01; \delta = 0,04) \Rightarrow \text{[[implicazione (175)]]} \Rightarrow \\ \Rightarrow \tilde{Y} &= 50 \times e^{\tilde{X}} \sim \text{LogN}(\gamma = 0,01 + \ln 50; \delta = 0,04), \end{aligned}$$

possiamo concludere che la probabilità richiesta sia data da

$$\begin{aligned} P(\{\tilde{Y} > 52\}) &= \text{[[formula (173)]]} = 1 - \Phi\left(\frac{\ln 52 - 0,01 - \ln 50}{0,04}\right) = \\ &\simeq 1 - \Phi(0,73) = \text{[[tavola } \Phi(\cdot) \text{]]} = 1 - 0,7673 = 0,2327. \end{aligned}$$

Si noti che questo risultato è molto prossimo al valore $P(\{\tilde{Y} > 52\}) = 0,2266$ che abbiamo ottenuto nell'Esercizio 8.21 dove al posto del rendimento logaritmico abbiamo considerato quello aritmetico. Questo fatto non è una sorpresa visto la differenza tra un rendimento logaritmico e il corrispondente rendimento aritmetico è trascurabile se uno dei due rendimenti è prossimo a zero (in tal caso deve dunque essere prossimo a zero anche l'altro) e visto che per entrambi i tipi di rendimento viene ipotizzata la stessa distribuzione normale che concentra quasi tutta la probabilità in prossimità del punto $x = 0$.

b) Si determini il primo percentile di \tilde{Y} .

Nella risposta al quesito precedente abbiamo visto che

$$\tilde{Y} = 50 \times e^{\tilde{X}} \sim \text{LogN}(\gamma = 0,01 + \ln 50; \delta = 0,04).$$

Usando la formula (174) vediamo dunque che il primo percentile di \tilde{Y} è dato da

$$y_{0,01} = e^{\gamma + z_{0,01}\delta} = \text{[[tavola } \Phi(\cdot) \text{]]} = e^{0,01 + \ln 50 - 2,33 \times 0,04} = 46,008.$$

Questo valore ci dice che nell'1% dei casi peggiori il valore finale dell'investimento sarà minore di 46008 euro.

Si noti che anche il valore di $y_{0,01} = 46,008$ è molto prossimo a quello di $y_{0,01} = 45,840$ che abbiamo ottenuto nell'Esercizio 8.21.

Esercizio 8.29. Un investitore ha investito $C_0 = 50$ mila euro in un titolo. Secondo l'investitore, i rendimenti logaritmici mensili del titolo sono descritti da variabili casuali i.i.d. con distribuzione normale di parametri $\mu = 0,01$ e $\sigma = 0,004$.

- Qual è la probabilità che dopo un anno il valore dell'investimento sia maggiore di 57 mila euro?
- Si supponga che l'investitore mantenga l'investimento per un anno. Si determinino il primo e il novantanovesimo percentile del valore finale dell'investimento.
- Si supponga ancora che l'investitore mantenga l'investimento per un anno. Si determinino il valore atteso e la varianza del valore finale dell'investimento.
- Qual è la distribuzione della variabile casuale che rappresenta il valore finale dell'investimento al termine di un periodo di mantenimento di k mesi.
- Qual è la durata minima del periodo di mantenimento affinché il valore atteso del valore finale dell'investimento sia maggiore del novantanovesimo percentile della sua stessa distribuzione?

Risposte:

- Qual è la probabilità che dopo un anno il valore dell'investimento sia maggiore di 57 mila euro?

Indichiamo con \tilde{C}_1 la variabile casuale che rappresenta il valore dell'investimento al termine del primo mese, con \tilde{C}_2 la variabile casuale che rappresenta il valore dell'investimento al termine del secondo mese, ecc. Sulla base di queste variabili casuali possiamo definire i rendimenti logaritmici mensili come

$$\begin{aligned}\tilde{X}_1 &= \ln \tilde{C}_1 - \ln C_0, \\ \tilde{X}_2 &= \ln \tilde{C}_2 - \ln \tilde{C}_1, \\ &\vdots\end{aligned}$$

Sommando i primi $k = 12$ rendimenti logaritmici mensili otteniamo

$$\tilde{Y}_{12} = \tilde{X}_1 + \tilde{X}_2 + \cdots + \tilde{X}_{12} = \ln \tilde{C}_{12} - \ln C_0 = \ln \frac{\tilde{C}_{12}}{C_0}$$

e da questo risultato deduciamo che il valore finale dell'investimento al termine di un periodo di mantenimento di un anno possa essere espresso come

$$\tilde{C}_{12} = C_0 \times e^{\tilde{Y}_{12}} = 50 \times e^{\tilde{Y}_{12}}.$$

Siccome per ipotesi

$$\tilde{X}_1, \tilde{X}_2, \dots \sim \text{i.i.d. Normale}(\mu = \mu_1 = 0,01; \sigma = \sigma_1 = 0,004),$$

possiamo applicare la proprietà riproduttiva delle distribuzioni normali (Teorema 8.10) onde concludere che

$$\begin{aligned} \tilde{Y}_{12} &= \tilde{X}_1 + \tilde{X}_2 + \dots + \tilde{X}_{12} \sim \\ &\sim \text{Normale}(\mu = \mu_{12} = 12 \times 0,01 = 0,12; \sigma = \sigma_{12} = 0,004 \times \sqrt{12}) \end{aligned}$$

(si ricordi che questo risultato lo avevamo già ricavato nella risposta al quesito a) dell'Esercizio 8.23). In virtù della relazione tra le distribuzioni normali e quelle lognormali (vedi la (171)), possiamo dunque concludere che

$$e^{\tilde{Y}_{12}} \sim \text{LogN}(\gamma = 0,12; \delta = 0,004 \times \sqrt{12} = 0,0139),$$

e usando l'implicazione (175) vediamo che

$$\tilde{C}_{12} = 50 \times e^{\tilde{Y}_{12}} \sim \text{LogN}(\gamma = 0,12 + \ln 50 = 4,032; \delta = 0,0139). \quad (178)$$

La probabilità richiesta è dunque data da

$$\begin{aligned} P(\{\tilde{C}_{12} > 57\}) &= \text{[[formula (173)]]} = 1 - \Phi\left(\frac{\ln 57 - 4,032}{0,0139}\right) = \\ &\simeq 1 - \Phi(0,80) = \text{[[tavola } \Phi(\cdot) \text{]]} = 1 - 0,7881 = 0,2119. \end{aligned}$$

- b) Si supponga che l'investitore mantenga l'investimento per un anno. Si determinino il primo e il novantanovesimo percentile del valore finale dell'investimento.

Nella risposta al quesito precedente abbiamo visto che il valore finale dell'investimento al termine di un periodo di mantenimento di un anno è rappresentato da una variabile con distribuzione lognormale di parametri $\gamma = 4,032$ e $\delta = 0,0139$ (vedi la (178)). Il primo e il novantanovesimo percentile di questa distribuzione sono dati da

$$\begin{aligned} c_{0,01} &= \text{[[formula (174)]]} = e^{4,032 + z_{0,01} \times 0,0139} = \\ &= \text{[[tavola } \Phi(\cdot) \text{]]} = e^{4,032 - 2,33 \times 0,0139} = 54,577 \end{aligned}$$

$$\begin{aligned} c_{0,99} &= \text{[[formula (174)]]} = e^{4,032 + z_{0,99} \times 0,0139} = \\ &= \text{[[tavola } \Phi(\cdot) \text{]]} = e^{4,032 + 2,33 \times 0,0139} = 58,229. \end{aligned}$$

Questi valori ci dicono che con probabilità pari al 98% tra un anno il valore dell'investimento sarà compreso tra 54577 euro e 58229 euro.

- c) Si supponga ancora che l'investitore mantenga l'investimento per un anno. Si determinino il valore atteso e la varianza del valore finale dell'investimento.

Nella risposta al quesito a) abbiamo visto che il valore finale dell'investimento al termine di un periodo di mantenimento di un anno è rappresentato da una variabile casuale con distribuzione lognormale di parametri $\gamma = 4,032$ e $\delta = 0,0139$ (vedi la (178)). Il valore atteso e la varianza di questa distribuzione sono dati da

$$E(\tilde{C}_{12}) = \text{[[formula (176)]]} = e^{\gamma + \frac{1}{2}\delta^2} = e^{4,032 + \frac{1}{2} \times 0,0139^2} = 56,379$$

$$\begin{aligned} \text{var}(\tilde{C}_{12}) &= \text{[[formula (177)]]} = e^{2\gamma + \delta^2} (e^{\delta^2} - 1) = \\ &= e^{2 \times 4,032 + 0,0139^2} (e^{0,0139^2} - 1) = 0,614194. \end{aligned}$$

- d) Qual è la distribuzione della variabile casuale che rappresenta il valore finale dell'investimento al termine di un periodo di mantenimento di k mesi.

Come nella risposta al quesito a) indicheremo con \tilde{C}_k è la variabile casuale che rappresenta il valore finale dell'investimento dopo k mesi di mantenimento. Per rispondere al quesito dobbiamo determinare le distribuzioni delle variabili casuali \tilde{C}_k . Ragionando come nella risposta al quesito a) non è difficile rendersi conto che

$$\tilde{C}_k \sim \text{LogN}(\gamma = k \times 0,01 + \ln 50; \delta = 0,004 \times \sqrt{k}),$$

ovvero che le distribuzioni delle variabili casuali \tilde{C}_k sono lognormali con parametri

$$\gamma = \gamma_k = k \times 0,01 + \ln 50$$

e

$$\delta = \delta_k = 0,004 \times \sqrt{k}.$$

- e) Qual è la durata minima del periodo di mantenimento affinché il valore atteso del valore finale dell'investimento sia maggiore del novantanovesimo percentile della sua stessa distribuzione?

Nella risposta al quesito precedente abbiamo visto che il valore finale dell'investimento al termine di un periodo di mantenimento di k mesi è rappresentato da una variabile casuale con distribuzione lognormale di parametri $\gamma = \gamma_k = k \times 0,01 + \ln 50$ e $\delta = \delta_k = 0,004 \times \sqrt{k}$.

Il novantanovesimo percentile di questa distribuzione è dato da

$$\begin{aligned} c_{0,99;k} &= \text{[[formula (174)]]} = e^{\gamma_k + z_{0,99} \times \delta_k} = e^{k \times 0,01 + \ln 50 + z_{0,99} \times 0,004 \times \sqrt{k}} = \\ &= \text{[[tavola } \Phi(\cdot) \text{]]} = 50 \times e^{k \times 0,01 + 2,33 \times 0,004 \times \sqrt{k}}, \end{aligned}$$

mentre il suo valore atteso è dato da

$$\begin{aligned} E(\tilde{C}_k) &= \text{[[formula (176)]]} = e^{\gamma k + \frac{1}{2} \delta_k^2} = \\ &= e^{k \times 0,01 + \ln 50 + \frac{1}{2} \times 0,004^2 \times k} = 50 \times e^{k \times 0,01 + \frac{1}{2} \times 0,004^2 \times k}. \end{aligned}$$

Si noti che all'aumentare della durata k del periodo di mantenimento, il valore atteso $E(\tilde{C}_k)$ aumenta più velocemente del novantanovesimo percentile $c_{0,99;k}$. Il valore di k dove il valore atteso $E(\tilde{C}_k)$ raggiunge il novantanovesimo percentile $c_{0,99;k}$ è la soluzione dell'equazione

$$E(\tilde{C}_k) = c_{0,99;k} \quad \Rightarrow \quad 50 \times e^{k \times 0,01 + \frac{1}{2} \times 0,004^2 \times k} = 50 \times e^{k \times 0,01 + 2,33 \times 0,004 \times \sqrt{k}}.$$

Dopo alcuni passaggi algebrici

$$\begin{aligned} 50 \times e^{k \times 0,01 + \frac{1}{2} \times 0,004^2 \times k} &= 50 \times e^{k \times 0,01 + 2,33 \times 0,004 \times \sqrt{k}} \quad \Rightarrow \\ \Rightarrow \quad k \times 0,01 + \frac{1}{2} \times 0,004^2 \times k &= k \times 0,01 + 2,33 \times 0,004 \times \sqrt{k} \quad \Rightarrow \\ \frac{1}{2} \times 0,004^2 \times k &= 2,33 \times 0,004 \times \sqrt{k} \quad \Rightarrow \\ \sqrt{k} &= \frac{2,33 \times 0,004}{\frac{1}{2} \times 0,004^2} \quad \Rightarrow \quad k = \frac{4 \times 2,33^2}{0,004^2} = 1357225 \text{ mesi.} \end{aligned}$$

vediamo che la durata minima del periodo di mantenimento affinché il valore atteso del valore finale dell'investimento sia maggiore del novantanovesimo percentile della sua stessa distribuzione è di $1357225 + 1$ mesi, ovvero di 113102 anni e 2 mesi.

Ovviamente, non ha senso considerare un periodo di mantenimento così lungo, ma lo scopo di questa domanda è di mettere in luce come sotto le ipotesi fatte nella consegna la distribuzione del valore finale dell'investimento diventa sempre più asimmetrica all'aumentare del periodo di mantenimento

8.13 La famiglia delle distribuzioni di Pareto

La famiglia delle distribuzioni di Pareto è la famiglia di tutte le distribuzioni assolutamente continue che hanno una funzione di densità del tipo

$$f(x) = \begin{cases} \theta x_0^\theta x^{-(\theta+1)} & \text{per } x > x_0 \\ 0 & \text{altrimenti} \end{cases}$$

dove x_0 e θ sono due parametri reali positivi.

Queste distribuzioni portano il nome dell'ingegnere, economista e sociologo italiano Vilfredo Pareto (1848 - 1923) che vedeva in esse dei modelli per distribuzioni di redditi

e ricchezza. Tuttora le distribuzioni di Pareto hanno un ruolo importante in campo economico e finanziario. Infatti, numerosi studi empirici suggeriscono che le code delle distribuzioni di variabili casuali che rappresentano utili o perdite derivanti da operazioni economiche e finanziarie siano approssimabili attraverso distribuzioni di Pareto.

La caratteristica saliente delle distribuzioni di Pareto è l'andamento asintotico della funzione di densità per $x \rightarrow \infty$. Infatti, per $x > x_0$ le funzioni di densità di Pareto sono funzioni potenza con esponente negativo e pertanto tendono a zero molto più lentamente delle funzioni di densità esponenziali e gamma, e anche delle funzioni di densità lognormali. Questo andamento della coda destra delle loro funzioni di densità comporta che le distribuzioni di Pareto associno delle probabilità molto più elevate a valori estremi. Infatti, come vedremo tra breve, si può dimostrare che

$$\tilde{X} \sim \text{Esponenziale}(\theta) \quad \Leftrightarrow \quad \tilde{Y} = e^{\tilde{X}} \sim \text{Pareto}(x_0 = 1; \theta) \quad (179)$$

dove $\text{Pareto}(x_0; \theta)$ indica la distribuzione di Pareto con parametri x_0 e θ .

Grafico delle funzioni di densità di Pareto. Come si desume dalla formula definitoria

$$f(x) = \begin{cases} \theta x_0^\theta x^{-(\theta+1)} & \text{per } x > x_0 \\ 0 & \text{altrimenti,} \end{cases}$$

le funzioni di densità di Pareto sono limitate e decrescenti sull'intervallo (x_0, ∞) , dove sono definite come una funzione potenza di esponente $-(\theta + 1) < -1$. **Il valore del parametro θ regola quindi la velocità con cui una funzione di densità di Pareto tende a 0 quando $x \rightarrow \infty$.**

Analizziamo ora invece il ruolo del parametro x_0 . Come si desume dall'espressione che definisce le funzioni di densità di Pareto, x_0 separa la parte del dominio dove una funzione di densità di Pareto è nulla dalla parte del dominio dove è positiva. Riscrivendo la funzione di densità di Pareto come

$$f(x) = \begin{cases} \frac{\theta}{x_0} (x/x_0)^{-(\theta+1)} & \text{per } x > x_0 \\ 0 & \text{altrimenti,} \end{cases}$$

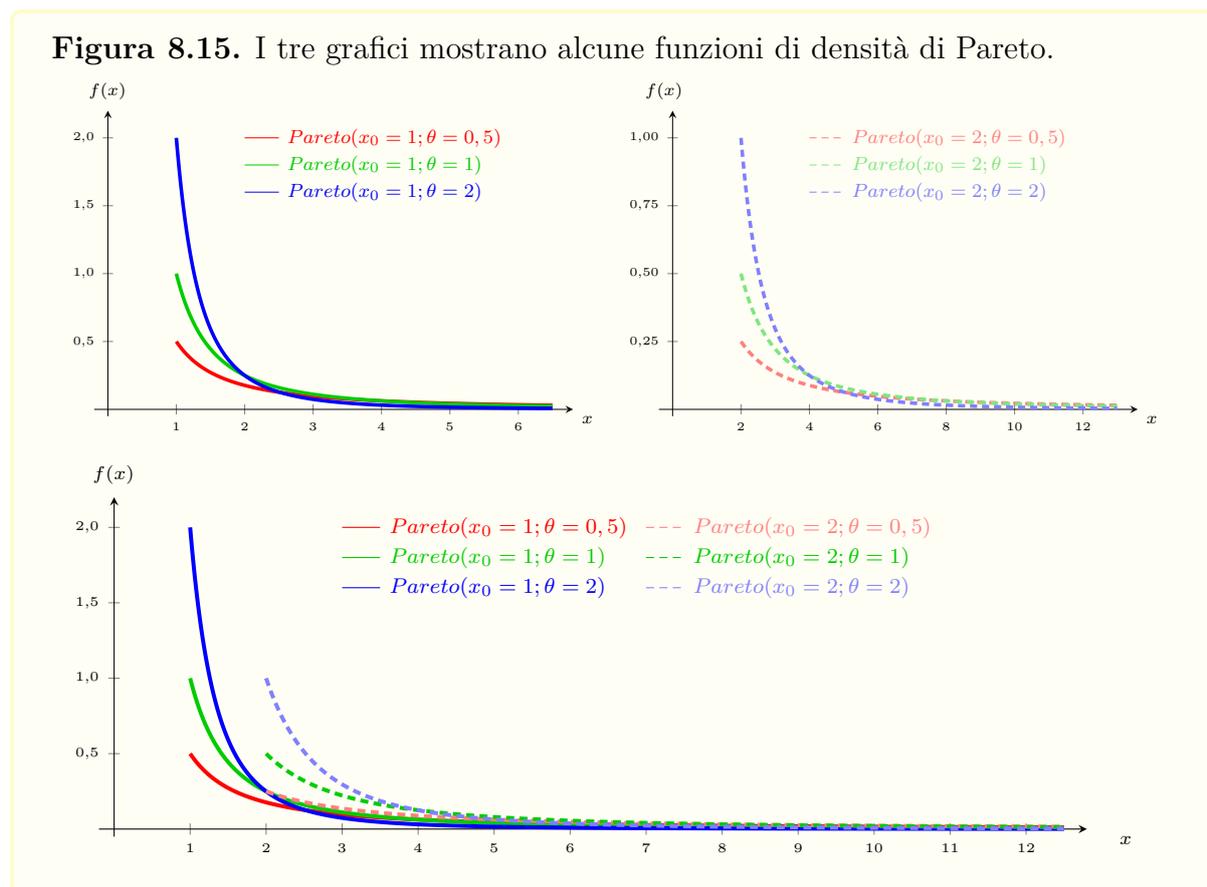
ci rendiamo immediatamente conto che x_0 **è anche un parametro di scala**. Infatti, per disegnare il grafico di una funzione di densità di Pareto con parametro $x_0 \neq 1$ basta disegnare il grafico della funzione di densità di Pareto con lo stesso valore del parametro θ e con $x_0 = 1$ e

- moltiplicare per x_0 i valori riportati sull'asse delle ascisse
- e dividere per x_0 i valori riportati sull'asse delle ordinate.

Siccome i valori sull'asse delle ascisse vengono moltiplicati per il parametro di scala x_0 , quest'ultimo deve essere un **parametro di scala "diretto"** e non un parametro di scala inverso come il parametro θ delle distribuzioni esponenziali. Infatti, seguendo la falsariga della dimostrazione dell'implicazione (147), si può facilmente dimostrare che

$$\tilde{X} \sim \text{Pareto}(x_0, \theta) \Leftrightarrow \tilde{Y} = b\tilde{X} \sim \text{Pareto}(x_0 b; \theta) \text{ per ogni } b > 0.$$

I tre grafici in Figura 8.15 mostrano alcune funzioni di densità di Pareto.



Funzione di ripartizione e quantili. Integrando la funzione di densità

$$f(x) = \begin{cases} \theta x_0^\theta x^{-(\theta+1)} & \text{per } x > x_0 \\ 0 & \text{altrimenti,} \end{cases}$$

vediamo che le funzioni di ripartizione di Pareto sono definite come

$$\begin{aligned} F(x) = P(\{\tilde{X} \leq x\}) &= \int_{-\infty}^x f(t) dt \\ &= \begin{cases} \int_{x_0}^x \theta x_0^\theta t^{-(\theta+1)} dt & \text{per } x \geq x_0 \\ 0 & \text{per } x < x_0. \end{cases} \quad (180) \\ &= \begin{cases} 1 - \left(\frac{x_0}{x}\right)^\theta & \text{per } x \geq x_0 \\ 0 & \text{per } x < x_0. \end{cases} \end{aligned}$$

Risolvendo l'equazione $F(x_p) = p$ possiamo anche facilmente verificare che i quantili di una distribuzione di Pareto sono dati da

$$x_p = x_0(1 - p)^{-1/\theta}, \quad p \in (0, 1). \quad (181)$$

Avendo ricavato l'espressione della funzione di ripartizione di una generica distribuzione di Pareto, possiamo finalmente dimostrare la relazione tra le distribuzioni di Pareto e quelle esponenziali (vedi la doppia implicazione (179)).

Dimostrazione (Dimostrazione della (179)). Dimostreremo prima l'implicazione

$$\tilde{X} \sim \text{Esponenziale}(\theta) \quad \Rightarrow \quad \tilde{Y} = e^{\tilde{X}} \sim \text{Pareto}(x_0 = 1; \theta).$$

Se $\tilde{X} \sim \text{Esponenziale}(\theta)$ e y è un numero reale positivo, allora si deve avere

$$P(\{\tilde{Y} \leq y\}) = P(\{e^{\tilde{X}} \leq y\}) = P(\{\tilde{X} \leq \ln y\}) = \begin{cases} 1 - e^{-\theta \ln y} = 1 - \left(\frac{1}{y}\right)^\theta & \text{se } y \geq 1, \\ 0 & \text{se } y < 1 \end{cases}$$

(ricordiamo che $\ln y \geq 0$ se e solo se $y \geq 1$). Siccome il valore di $P(\{\tilde{Y} \leq y\})$ non può essere negativo (assioma K1) e non può aumentare se y diminuisce (legge L4), possiamo concludere che

$$P(\{\tilde{Y} \leq y\}) = 0$$

anche se $y \leq 0$. In definitiva abbiamo dunque dimostrato che la funzione di ripartizione di \tilde{Y} sia data da

$$F(y) = P(\{\tilde{Y} \leq y\}) = \begin{cases} 1 - \left(\frac{1}{y}\right)^\theta & \text{se } y \geq 1, \\ 0 & \text{se } y < 1. \end{cases}$$

e siccome questa è la funzione di ripartizione della distribuzione $\text{Pareto}(x_0 = 1; \theta)$, possiamo concludere che $\tilde{Y} = e^{\tilde{X}} \sim \text{Pareto}(x_0 = 1; \theta)$ come volevamo dimostrare.

Consideriamo ora invece l'implicazione

$$\tilde{Y} = e^{\tilde{X}} \sim \text{Pareto}(x_0 = 1; \theta) \quad \Rightarrow \quad \tilde{X} \sim \text{Esponenziale}(\theta).$$

Se $\tilde{Y} = e^{\tilde{X}} \sim \text{Pareto}(x_0 = 1; \theta)$, allora si deve avere

$$P(\{\tilde{X} \leq x\}) = P(\{\ln \tilde{Y} \leq x\}) = P(\{\tilde{Y} \leq e^x\}) = \begin{cases} 1 - \left(\frac{1}{e^x}\right)^\theta = 1 - e^{-\theta x} & \text{se } x \geq 0, \\ 0 & \text{se } x < 0 \end{cases}$$

(ricordiamo che $e^x \geq 1 = x_0$ se e solo se $x \geq 0$), e siccome $F(x) = P(\{\tilde{X} \leq x\})$ è la funzione di ripartizione della distribuzione $\text{Esponenziale}(\theta)$, possiamo concludere che $\tilde{X} \sim \text{Esponenziale}(\theta)$ come volevamo dimostrare. \square

Momenti. Per una distribuzione di Pareto esistono solo i momenti $\mu_k = E(X^k)$ di ordine k inferiore al valore del parametro θ . Infatti, se $f(x)$ è una funzione di densità di Pareto, si ottiene

$$\int_{-x_0}^{\infty} x^k f(x) dx = \int_{x_0}^{\infty} x^k \theta x_0^\theta x^{-(\theta+1)} dx = \theta x_0^\theta \int_{x_0}^{\infty} x^{-(\theta-k+1)} dx$$

e l'ultimo integrale converge se e solo se $k < \theta$, e in questo caso il suo valore è dato da

$$\left[\frac{x^{-(\theta-k)}}{-(\theta-k)} \right]_{x_0}^{\infty} = \frac{x_0^{-\theta+k}}{\theta-k}.$$

Questo ragionamento dimostra che **i momenti di ordine $k < \theta$ di una distribuzione di Pareto sono dati da**

$$\mu_k = E(\tilde{X}^k) = \frac{x_0^k \theta}{\theta - k}$$

e che **i momenti di ordine $k \geq \theta$ non esistono.**

Quindi possiamo concludere che

- il valore atteso $\mu = E(\tilde{X})$ di una distribuzione di Pareto esiste se e solo se $\theta > 1$ e che in tal caso si ha

$$\mu = E(\tilde{X}) = \frac{x_0 \theta}{\theta - 1}; \quad (182)$$

- la varianza $\sigma^2 = \text{var}(\tilde{X})$ di una distribuzione di Pareto esiste se e solo se $\theta > 2$ (si ricordi che la varianza di una distribuzione esiste se e solo se esiste anche il momento del secondo ordine) e che in tal caso si ha

$$\sigma^2 = \text{var}(\tilde{X}) = \frac{x_0^2 \theta}{(\theta - 1)^2 (\theta - 2)} \quad (183)$$

(invitiamo il lettore a verificare che la formula della varianza sia corretta).

Inoltre, possiamo anche concludere che **nessuna distribuzione di Pareto è dotata di fgm**. Infatti, secondo il Teorema 7.1 la fgm di una distribuzione esiste solo se esiste anche tutta l'infinita successione dei momenti .

Esercizio 8.30. Sia \tilde{X} la variabile casuale che rappresenta l'ammontare del risarcimento (espresso in migliaia di euro) che una compagnia assicurativa deve pagare nel caso un dato evento si verificasse. Si supponga che la distribuzione di \tilde{X} sia paretiana con parametri $x_0 = 1$ e $\theta = 3$.

- Si calcolino il valore atteso e la varianza di \tilde{X} .
- Si calcoli il novantesimo percentile di \tilde{X} .
- Qual è la probabilità che, nel caso in cui l'evento in questione si verificasse, il risarcimento sia superiore a duemila euro?
- Ipotizzando che l'eventuale risarcimento sia superiore a 10 mila euro, qual è la probabilità condizionata che sia anche superiore a 12 mila euro?
- Si calcolino il valore atteso e la varianza di \tilde{X} .

Siccome il parametro θ della distribuzione paretiana di \tilde{X} è maggiore di 2, il valore atteso e la varianza di \tilde{X} esistono entrambi e sono dati da

$$E(\tilde{X}) = \text{[[formula (182)]]} = \frac{x_0\theta}{\theta - 1} = \frac{1 \times 3}{3 - 1} = \frac{3}{2}$$

e da

$$\text{var}(\tilde{X}) = \text{[[formula (183)]]} = \frac{x_0^2\theta}{(\theta - 1)^2(\theta - 2)} = \frac{1^2 \times 3}{(3 - 1)^2(3 - 2)} = \frac{3}{4}.$$

b) Si calcoli il novantesimo percentile di \tilde{X} .

Il novantesimo percentile di \tilde{X} è dato da

$$x_{0,9} = \text{[[formula (181)]]} = x_0(1 - 0,9)^{-1/\theta} = 1 \times (1 - 0,9)^{-1/3} = 2,154.$$

c) Qual è la probabilità che, nel caso in cui l'evento in questione si verificasse, il risarcimento sia superiore a duemila euro?

La probabilità richiesta è data da

$$P(\{\tilde{X} > 2\}) = \text{[[legge L3 & formula (180)]]} = \left(\frac{x_0}{2}\right)^\theta = \frac{1}{2^3} = 0,125.$$

d) Ipotizzando che l'eventuale risarcimento sia superiore a 10 mila euro, qual è la probabilità condizionata che sia anche superiore a 12 mila euro?

La probabilità condizionata richiesta è data da

$$\begin{aligned} P(\{\tilde{X} > 12\}|\{\tilde{X} > 10\}) &= \frac{P(\{\tilde{X} > 12\} \cap \{\tilde{X} > 10\})}{P(\{\tilde{X} > 10\})} = \frac{P(\{\tilde{X} > 12\})}{P(\{\tilde{X} > 10\})} = \\ &= \text{[[legge L3 & formula (180)]]} = \frac{\left(\frac{x_0}{12}\right)^\theta}{\left(\frac{x_0}{10}\right)^\theta} = \frac{(1/12)^3}{(1/10)^3} = 0,579. \end{aligned}$$

9 Inferenza

L'**inferenza** è la branca della statistica e della teoria della probabilità che predispone metodi per *inferire* qualcosa sul processo che potrebbe aver generato dei dati che osserviamo nel mondo reale.

Nell'inferenza statistica si parte dunque da dati che si osservano nel mondo reale e questi dati vengono considerati come realizzazioni di cosiddette **variabili casuali campionarie**. I metodi che predispone l'inferenza statistica ci dicono quindi come si possono utilizzare le realizzazioni delle variabili casuali campionarie per

- **stimare** parametri che descrivono l'ignota distribuzione congiunta delle variabili casuali campionarie
- e per **verificare ipotesi** che riguardano l'ignota distribuzione congiunta delle variabili casuali campionarie.

In questa dispensa presenteremo soltanto alcuni metodi per affrontare **problemi di stima** e ci limiteremo a considerare soltanto problemi di stima molto specifici, ovvero problemi di stima dove sull'ignota distribuzione congiunta delle variabili casuali campionarie disponiamo di **conoscenze a priori** molto dettagliate. Infatti, assumeremo sempre che le variabili casuali campionarie siano i.i.d. e che la distribuzione comune ad esse appartenga ad una determinata famiglia di distribuzioni \mathcal{F} . Per evitare problemi di stima banali, assumeremo sempre che \mathcal{F} contenga almeno due distribuzioni che presentino dei valori diversi per almeno uno dei parametri che vogliamo stimare (altrimenti, non avremmo un problema di stima). Per distinguere l'ignota distribuzione comune alle variabili casuali campionarie dalle altre distribuzioni che appartengono alla famiglia \mathcal{F} , indicheremo la prima con $F^*(x)$, e scriveremo $F(x)$ per indicare una generica distribuzione che appartiene alla famiglia \mathcal{F} . Considerando questo tipo di contesto, vedremo come a partire dalle realizzazioni delle variabili casuali campionarie si ottengono delle stime per il valore atteso e la varianza di $F^*(x)$ che soddisfano un **criterio di "ottimalità"** che definiremo tra breve. Siccome daremo sempre per scontato che il valore atteso e la varianza di $F^*(x)$ esistano, considereremo soltanto famiglie di distribuzioni \mathcal{F} che contengono soltanto distribuzioni che sono dotate di varianza (si ricordi che in questo caso deve esistere anche il valore atteso).

Chiaramente, i problemi di stima che abbiamo appena descritto sono molto specifici soprattutto per l'ipotesi che le variabili casuali campionarie siano i.i.d.. Questa ipotesi può essere ritenuta valida se le variabili casuali campionarie rappresentano i risultati di misurazioni che vengono effettuate in modo indipendente e sotto le medesime condizioni come, per esempio,

- quando si utilizza la stessa bilancia per misurare più volte il peso di uno stesso oggetto, oppure di oggetti i cui pesi dovrebbero differire soltanto per cause accidentali
- oppure, più in generale, quando si utilizza lo stesso metodo per effettuare un qualunque tipo di misurazione e gli esiti delle misurazioni sono influenzati soltanto da fattori accidentali.

D'altra parte, l'ipotesi di variabili casuali campionarie i.i.d. non può essere ritenuta valida se ciascuna di esse rappresenta l'esito di una misurazione effettuata sotto condizioni diverse e/o se si ritiene che le realizzazioni di alcune di esse influenzino quelle delle altre.

Nonostante la specificità dei problemi di stima che affronteremo in questa dispensa, attraverso di essi riusciremo comunque a presentare molti concetti che sono di fondamentale importanza anche in problemi più complessi.

D'ora in poi useremo la seguente notazione: indicheremo con

$$\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_n,$$

le variabili casuali campionarie e con

$$x_1, x_2, \dots, x_n$$

le loro realizzazioni. Come abbiamo già detto, l'ignota distribuzione comune alle variabili casuali campionarie verrà indicata con $F^*(x)$, e una generica distribuzione che appartiene alla famiglia \mathcal{F} (della quale assumiamo che $F^*(x) \in \mathcal{F}$) verrà invece indicata semplicemente con $F(x)$. Se $F(x) \in \mathcal{F}$, allora scriveremo μ_F per indicare il valore atteso della distribuzione $F(x)$ e σ_F^2 per indicare la sua varianza. Più in generale, scriveremo θ_F per indicare il valore che un determinato parametro θ (come il valore atteso o la varianza) assume in corrispondenza di una data distribuzione $F(x) \in \mathcal{F}$. Si noti che attraverso questo tipo di notazione, l'ignoto valore del parametro θ che si vuole stimare viene indicato con θ_{F^*} . Nei casi particolari dove θ è il valore atteso o la varianza, al posto di θ_{F^*} scriveremo μ_{F^*} e $\sigma_{F^*}^2$, rispettivamente.

9.1 Stima puntuale

Supponiamo dunque di voler stimare l'ignoto valore θ_{F^*} di un parametro θ che descrive l'ignota distribuzione $F^*(x)$ comune alle variabili casuali campionarie.

Avendo osservato le realizzazioni x_1, x_2, \dots, x_n delle variabili casuali campionarie possiamo tentare di stimare l'ignoto valore di θ_{F^*} attraverso una loro funzione

$$t(x_1, x_2, \dots, x_n).$$

Per esempio, se vogliamo stimare il valore atteso $\mu = \mu_{F^*}$ (ovvero se θ è il valore atteso), possiamo calcolare la funzione

$$t(x_1, x_2, \dots, x_n) = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

(ovvero la media delle realizzazioni delle variabili casuali campionarie), ma nulla ci impedisce di considerare una funzione diversa come, per esempio, la mediana degli n valori x_i . Allo stesso modo, se vogliamo stimare la varianza $\sigma^2 = \sigma_{F^*}^2$ (ovvero se θ è la varianza), possiamo calcolare la funzione

$$t(x_1, x_2, \dots, x_n) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

(ovvero la varianza delle realizzazioni delle variabili casuali campionarie), ma anche in questo caso potremmo considerare una funzione diversa.

Il valore $t(x_1, x_2, \dots, x_n)$ che si calcola al fine di stimare l'ignoto valore di un parametro θ viene chiamato **"stima puntuale"**. Si noti che il valore di una stima puntuale $t(x_1, x_2, \dots, x_n)$ può essere considerato come realizzazione di una corrispondente variabile casuale

$$\tilde{T} = t(\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_n),$$

che viene chiamata **"stimatore"**.

Come abbiamo appena visto, per stimare l'ignoto valore di un parametro θ si possono considerare infinite funzioni $t(x_1, x_2, \dots, x_n)$ diverse e quindi si pone il problema di capire da quale di queste infinite funzioni possiamo aspettarci di ottenere una stima puntuale precisa, ovvero una stima puntuale che sia prossima all'ignoto valore θ_{F^*} che vogliamo stimare. Ovviamente, non possiamo valutare la precisione di una stima puntuale $t(x_1, x_2, \dots, x_n)$ calcolando l'errore di stima

$$t(x_1, x_2, \dots, x_n) - \theta_{F^*}$$

perché non conosciamo il valore di θ_{F^*} . Per deciderci a favore di una stima puntuale piuttosto che di un'altra dobbiamo quindi mettere in atto un'altra strategia. Le strategie più usate nelle applicazioni sono tutte basate sul cosiddetto **errore quadratico medio (EQM)** che è definito come

$$EQM_{F^*}(\tilde{T}) = \begin{cases} E_{F^*}(|\tilde{T} - \theta_{F^*}|^2) & \text{se il valore atteso esiste} \\ \infty & \text{altrimenti.} \end{cases}$$

Nella formula che definisce l'EQM, $E_{F^*}(|\tilde{T} - \theta_{F^*}|^2)$ è il valore atteso dell'errore di stima elevato al quadrato **sotto l'ipotesi che la distribuzione comune alle variabili casuali campionarie sia la distribuzione $F(x)$ indicata nei pedici di $E_{F^*}(\cdot)$ e di θ_{F^*}** . Per calcolare l'errore quadratico medio bisogna quindi fare un'ipotesi sull'ignota distribuzione comune alle variabili casuali campionarie, e **ad ogni possibile distribuzione $F(x) \in \mathcal{F}$ corrisponde dunque un valore dell'EQM**. In altre parole, l'EQM di uno stimatore dev'essere considerato come una funzione che ha per dominio l'insieme \mathcal{F} (ricordiamo che \mathcal{F} è un insieme di distribuzioni $F(x)$ del quale assumiamo che $F^*(x) \in \mathcal{F}$).

Per facilitare il calcolo dell'EQM conviene osservare che il valore atteso $E_{F^*}(|\tilde{T} - \theta_{F^*}|^2)$, se esiste, può essere calcolato a partire dal valore atteso $E_{F^*}(\tilde{T})$ e dalla varianza $var_{F^*}(\tilde{T}) = E_{F^*}([\tilde{T} - E_{F^*}(\tilde{T})]^2)$. Infatti, si può dimostrare che il valore atteso $E_{F^*}(|\tilde{T} - \theta_{F^*}|^2)$ esiste se e solo se esiste anche la varianza $var_{F^*}(\tilde{T})$ (si ricordi che in questo caso deve esistere anche il valore atteso $E_{F^*}(\tilde{T})$) e che in quest'ultimo caso l'EQM può essere calcolato come

$$EQM_{F^*}(\tilde{T}) = E_{F^*}(|\tilde{T} - \theta_{F^*}|^2) = var_{F^*}(\tilde{T}) + [E_{F^*}(\tilde{T}) - \theta_{F^*}]^2 \quad (184)$$

L'espressione tra parentesi quadre sul lato destro dell'ultima uguaglianza viene chiamata **distorsione** (in inglese "bias") dello stimatore \tilde{T} . Se per ogni $F(x) \in \mathcal{F}$ la distorsione di uno stimatore \tilde{T} è nulla, ovvero se

$$E_F(\tilde{T}) = \theta_F \quad \text{per ogni } F(x) \in \mathcal{F},$$

allora si dice che \tilde{T} è uno stimatore **corretto**.

Dimostrazione (della scomposizione dell'EQM). Dimosteremo soltanto che l'esistenza di $var_F(\tilde{T})$ implichi l'esistenza del valore atteso $E_F(|\tilde{T} - \theta_F|^2)$ e che quest'ultimo possa essere calcolato secondo la formula (184). Assumiamo dunque che $var_F(\tilde{T}) = E_F(|\tilde{T} - E_F(\tilde{T})|^2)$ esista e che dunque esista anche il valore atteso $E_F(\tilde{T})$. La dimostrazione procede come segue: siccome

$$\begin{aligned} |\tilde{T} - \theta_F|^2 &= |[\tilde{T} - E_F(\tilde{T})] + [E_F(\tilde{T}) - \theta_F]|^2 \\ &= [\tilde{T} - E_F(\tilde{T})]^2 + [E_F(\tilde{T}) - \theta_F]^2 + 2[\tilde{T} - E_F(\tilde{T})][E_F(\tilde{T}) - \theta_F] \\ &= [\tilde{T} - E_F(\tilde{T})]^2 + [E_F(\tilde{T}) - \theta_F]^2 + 2\tilde{T}[E_F(\tilde{T}) - \theta_F] - 2E_F(\tilde{T})[E_F(\tilde{T}) - \theta_F], \end{aligned}$$

possiamo applicare la proprietà E4* onde concludere che il valore atteso $E_F(|\tilde{T} - \theta_F|^2)$ esista e che quest'ultimo possa essere calcolato come

$$\begin{aligned} E_F(|\tilde{T} - \theta_F|^2) &= E_F(|\tilde{T} - E_F(\tilde{T})|^2) + [E_F(\tilde{T}) - \theta_F]^2 + 2E_F(\tilde{T})[E_F(\tilde{T}) - \theta_F] + \\ &\quad - 2E_F(\tilde{T})[E_F(\tilde{T}) - \theta_F] = var_F(\tilde{T}) + [E_F(\tilde{T}) - \theta_F]^2 \end{aligned}$$

così come previsto dalla formula (184). □

Torniamo ora al problema della scelta di una stima puntuale. Come abbiamo accennato poc'anzi, facendo riferimento all'EQM si possono definire vari criteri di scelta, ovvero vari concetti di "ottimalità" per stime puntuali e/o per i corrispondenti stimatori. In questa dispensa presenteremo soltanto uno di questi concetti: quello che di solito viene descritto per primo nei libri di testo introduttivi. Per chiarire le motivazioni che conducono alla definizione del concetto di ottimalità che presenteremo, conviene partire con un passo falso e dare una definizione di "ottimalità" più ambiziosa che a prima vista potrebbe sembrare quella più ovvia, ma che tranne in problemi di stima molto particolari non può essere soddisfatta da nessuna stima puntuale. La definizione di ottimalità più ambiziosa prevede di considerare come "ottimali" soltanto le stime puntuali $t(x_1, x_2, \dots, x_n)$ che provengono da stimatori $\tilde{T} = t(\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_n)$ con **EQM uniformemente minimo**, ovvero da stimatori \tilde{T} con

$$\begin{aligned} EQM_F(\tilde{T}) &\leq EQM_F(\tilde{T}^\dagger) \\ \text{per ogni } F(x) \in \mathcal{F} &\text{ e per ogni altro stimatore } \tilde{T}^\dagger. \end{aligned} \tag{185}$$

Chiaramente, una stima puntuale "ottimale" nel senso che abbiamo appena definito esiste se e solo se esiste uno stimatore \tilde{T} con EQM uniformemente minimo. Tuttavia, come

vedremo tra breve, **uno stimatore con EQM uniformemente minimo esiste soltanto in problemi di stima molto particolari** che in realtà non sono dei veri e propri problemi di stima. Per chiarire la natura dei problemi di stima dove esiste uno stimatore \tilde{T} con EQM uniformemente minimo, assumiamo che \tilde{T} sia un tale stimatore (ovvero uno stimatore che soddisfa la condizione (185)) e consideriamo i seguenti casi:

- Il caso in cui $EQM_F(\tilde{T}) = 0$ per ogni $F(x) \in \mathcal{F}$. In questo caso si ha

$$\begin{aligned} EQM_F(\tilde{T}) &= [[\text{formula (184)}]] = var_F(\tilde{T}) + [E_F(\tilde{T}) - \theta_F]^2 = 0 \quad \Rightarrow \\ &\Rightarrow \quad var_F(\tilde{T}) = 0 \text{ e } E_F(\tilde{T}) = \theta_F \quad \Rightarrow \\ &\Rightarrow [[\text{proprietà E0 e V0}]] \Rightarrow \quad P_F(\{\tilde{T} = \theta_F\}) = 1 \text{ per ogni } F(x) \in \mathcal{F} \end{aligned}$$

e quindi esiste uno stimatore che sempre e comunque restituisce il valore che si vuole stimare. Ovviamente, questo caso si verifica se e solo se l'esatto valore del parametro θ che si vuole stimare può essere dedotto da qualsiasi possibile realizzazione delle variabili casuali campionarie. I problemi di stima dove questo è possibile sono in realtà dei problemi di deduzione logica (può trattarsi di problemi molto banali, ma si possono escogitare anche dei problemi di questo tipo dove è molto difficile trovare uno stimatore \tilde{T} con EQM nullo).

- Il caso in cui $EQM_F(\tilde{T}) > 0$ per almeno una distribuzione $F(x) \in \mathcal{F}$. In questo caso possiamo indicare con $F(x) = F^\dagger(x)$ una distribuzione per la quale

$$EQM_{F^\dagger}(\tilde{T}) > 0$$

e costruire uno stimatore $\tilde{T}^\dagger = t^\dagger(\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_n)$ con

$$EQM_{F^\dagger}(\tilde{T}^\dagger) = 0 < EQM_{F^\dagger}(\tilde{T}).$$

Così facendo dimostriamo che \tilde{T} non può essere uno stimatore con EQM uniformemente minimo, ovvero che \tilde{T} non può essere uno stimatore che soddisfa la condizione di ottimalità (185).

Per costruire lo stimatore \tilde{T}^\dagger che spodesta lo stimatore \tilde{T} basta porre

$$t^\dagger(x_1, x_2, \dots, x_n) = \theta_{F^\dagger} \quad \text{per ogni } (x_1, x_2, \dots, x_n) \in \mathbb{R}^n.$$

In questo modo si ottiene

$$\begin{aligned} &P_F(\{\tilde{T}^\dagger = \theta_{F^\dagger}\}) = 1 \quad \Rightarrow \\ \Rightarrow E_F(\tilde{T}^\dagger) &= [[\text{proprietà E0}]] = \theta_{F^\dagger} \quad \text{e} \quad var_F(\tilde{T}^\dagger) = [[\text{proprietà V0}]] = 0 \quad \Rightarrow \\ \Rightarrow EQM_F(\tilde{T}^\dagger) &= [[\text{formula (184)}]] = 0 + [\theta_{F^\dagger} - \theta_F]^2 \Rightarrow \quad (186) \\ &\Rightarrow EQM_{F^\dagger}(\tilde{T}^\dagger) = 0, \end{aligned}$$

e siccome stiamo assumendo che $EQM_{F^\dagger}(\tilde{T}) > 0$, possiamo concludere che lo stimatore \tilde{T} non soddisfa la condizione di ottimalità (185).

Come si desume precedente ragionamento, uno stimatore \tilde{T} con EQM uniformemente minimo esiste soltanto in problemi di stima dove esiste uno stimatore che non sbaglia mai. In tutti gli altri problemi di stima, una volta scelto uno stimatore \tilde{T} , si può sempre trovare un altro stimatore \tilde{T}^\dagger con

$$EQM_{F^\dagger}(\tilde{T}^\dagger) = 0 < EQM_{F^\dagger}(\tilde{T})$$

per qualche $F^\dagger(x) \in \mathcal{F}$ e nessuno stimatore \tilde{T} può quindi soddisfare la condizione di ottimalità (185). Tuttavia, gli stimatori \tilde{T}^\dagger che abbiamo considerato per ottenere $EQM_{F^\dagger}(\tilde{T}^\dagger) = 0$ sono variabili casuali degeneri che ignorano completamente le variabili casuali campionarie e l'unico valore θ_{F^\dagger} che si ottiene da uno stimatore di questo tipo può essere considerato come un tentativo di "indovinare" l'ignoto valore di θ_{F^*} senza considerare le realizzazioni delle variabili casuali campionarie. Ovviamente, nelle applicazioni non si dovrebbero mai usare stimatori di questo tipo (ovvero non bisognerebbe mai ignorare le realizzazioni delle variabili casuali campionarie e tentare soltanto di indovinare l'ignoto valore di un parametro): infatti, come si desume dalla formula (186) che qui per comodità riscriviamo

$$EQM_F(\tilde{T}^\dagger) = [\theta_{F^\dagger} - \theta_F]^2,$$

l'EQM di uno stimatore degenero è molto elevato se il valore di θ_F è molto lontano dall'unica stima puntuale θ_{F^\dagger} che si ottiene dallo stimatore degenero.

Dalle considerazioni fin qui svolte si desume che gli stimatori degeneri sono sostanzialmente inutili e che creano soltanto problemi perché impediscono l'esistenza di stimatori con EQM uniformemente minimo. Gli stimatori degeneri, e molti altri stimatori che sono "quasi" degeneri, non dovrebbero quindi essere tenuti in considerazione nella definizione di un criterio di ottimalità. Per eliminare questi stimatori, restringeremo dunque l'attenzione alla classe degli stimatori **corretti**. Per dimostrare che questa restrizione elimina efficacemente tutti gli stimatori degeneri basta osservare che se uno stimatore \tilde{T}^\dagger assume sempre solo un determinato valore $\theta^\dagger \in \mathbb{R}$, allora si deve avere

$$E_F(\tilde{T}^\dagger) = [[\text{proprietà E0}]] = \theta^\dagger \quad \text{per ogni } F(x) \in \mathcal{F},$$

e siccome stiamo assumendo che \mathcal{F} contenga due distribuzioni $F_1(x)$ e $F_2(x)$ con $\theta_{F_1} \neq \theta_{F_2}$, uno stimatore degenero non può essere **corretto** perché almeno uno dei due valori θ_{F_1} e/o θ_{F_2} deve essere diverso da θ^\dagger .

Il criterio di ottimalità che considereremo in questa dispensa è quindi il seguente: una stima puntuale $t(x_1, x_2, \dots, x_n)$ è "ottimale" se proviene da uno stimatore \tilde{T} che è **corretto** e che soddisfa la condizione

$$EQM_F(\tilde{T}) \leq EQM_F(\tilde{T}^\dagger) \tag{187}$$

per ogni $F(x) \in \mathcal{F}$ e per ogni altro stimatore \tilde{T}^\dagger che è uno stimatore corretto.

Ora, siccome per gli stimatori corretti l'EQM si riduce alla varianza dello stimatore (vedi la formula (184)), il suddetto criterio di ottimalità può essere riformulato dicendo che

una stima puntuale $t(x_1, x_2, \dots, x_n)$ è "ottimale" se proviene da uno stimatore \tilde{T} che è **corretto** e che soddisfa la condizione

$$\text{var}_F(\tilde{T}) \leq \text{var}_F(\tilde{T}^\dagger) \quad (188)$$

per ogni $F(x) \in \mathcal{F}$ e per ogni altro stimatore \tilde{T}^\dagger che è uno stimatore corretto.

Uno stimatore \tilde{T} che soddisfa questa condizione di ottimalità viene chiamato **stimatore corretto a varianza uniformemente minima** o **stimatore UMVUE**, dove UMVUE è l'acronimo dell'espressione inglese "*Uniformly Minimum Variance Unbiased Estimator*".

Si noti che la proprietà di essere uno stimatore UMVUE dipende crucialmente dalla famiglia \mathcal{F} . Infatti, se \tilde{T} è uno stimatore UMVUE per un determinato parametro θ con riferimento ad una determinata famiglia \mathcal{F} , e se \mathcal{F}' è una famiglia di distribuzioni più ampia che include la famiglia \mathcal{F} , allora può succedere che \tilde{T} non sia più UMVUE con riferimento a \mathcal{F}' per due motivi:

- \mathcal{F}' contiene qualche distribuzione $F(x)$ (che non è contenuta in \mathcal{F}) tale che

$$E_F(\tilde{T}) \neq \theta_F$$

e questo significa che \tilde{T} perde la proprietà della correttezza e quindi non può essere uno stimatore UMVUE rispetto alla famiglia \mathcal{F}' ;

- \tilde{T} non perde la proprietà della correttezza ma \mathcal{F}' contiene qualche distribuzione $F(x)$ (che non è contenuta in \mathcal{F}) tale che

$$\text{var}_F(\tilde{T}^\dagger) \leq \text{var}_F(\tilde{T})$$

per un altro stimatore corretto \tilde{T}^\dagger .

D'altra parte, se \mathcal{F}'' è una famiglia di distribuzioni più ristretta che è inclusa nella famiglia \mathcal{F} , allora può succedere che uno stimatore \tilde{T} che è UMVUE rispetto a \mathcal{F} non sia più UMVUE con riferimento a \mathcal{F}'' perché qualche stimatore \tilde{T}^\dagger che non è corretto rispetto a \mathcal{F} è invece corretto rispetto a \mathcal{F}'' (si noti che restringendo la famiglia di distribuzioni di riferimento l'insieme degli stimatori corretti rimane immutato oppure diventa più ampio) e perché

$$\text{var}_F(\tilde{T}^\dagger) \leq \text{var}_F(\tilde{T}) \quad \text{per qualche } F(x) \in \mathcal{F}''.$$

Consideriamo ora invece il **problema dell'esistenza di stimatori UMVUE**. Con riferimento alla definizione di stimatore UMVUE ci si potrebbe infatti chiedere se esistono dei problemi di stima non banali dove si possono trovare degli stimatori UMVUE. La risposta a questa domanda è affermativa e tra breve descriveremo alcuni problemi di stima dove esiste addirittura soltanto un unico stimatore UMVUE. A prima vista potrebbe sembrare che in questi problemi di stima si dovrebbe tenere in considerazione soltanto

la stima puntuale che proviene dall'unico stimatore UMVUE. Non è così. Infatti, per definire il concetto di stimatore UMVUE abbiamo ristretto l'attenzione agli stimatori **corretti**, ma al di fuori di questa classe di stimatori possono esistere degli stimatori che sono "quasi corretti" e che per ogni $F \in \mathcal{F}$ hanno EQM strettamente minore di quello dello stimatore UMVUE. Tra breve vedremo anche un esempio dove questo accade (senza fornire una dimostrazione).

9.1.1 La media campionaria

In questa sezione introdurremo uno stimatore che per l'inferenza statistica è di fondamentale importanza. Lo stimatore in questione è la **media campionaria**, ovvero la media aritmetica delle variabili casuali campionarie:

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n \tilde{X}_i$$

(si noti che la "n" al pedice di \bar{X}_n indica il numero di variabili casuali campionarie di cui assumiamo di poter osservare le realizzazioni). Per indicare la realizzazione della media campionaria scriveremo

$$\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i.$$

Qui di seguito riportiamo un elenco delle principali proprietà della media campionaria. Come si evince da questo elenco, in alcuni problemi di stima molto importanti dove il valore atteso μ delle variabili casuali campionarie è ignoto, la media campionaria è l'unico stimatore UMVUE per l'ignoto valore di μ .

Proprietà della media campionaria:

M1) (Correttezza). Se il valore atteso $\mu_F = E_F(\tilde{X}_i)$ esiste, allora deve esistere anche il valore atteso $E_F(\bar{X}_n)$ e

$$E_F(\bar{X}_n) = \mu_F.$$

Con riferimento a qualsiasi famiglia di distribuzioni \mathcal{F} che contiene soltanto distribuzioni $F(x)$ che sono dotate di valore atteso $\mu = \mu_F = E_F(\tilde{X}_i)$, la media campionaria \bar{X}_n è dunque uno stimatore corretto per μ :

$$E_F(\bar{X}_n) = \mu_F \quad \text{per ogni } F(x) \in \mathcal{F}.$$

Dimostrazione. La dimostrazione della proprietà M1 è una semplice applicazione della proprietà E4* del valore atteso: siccome \bar{X}_n è una trasformazione lineare delle variabili casuali campionarie \tilde{X}_i e

siccome stiamo ipotizzando che $F(x)$ sia una distribuzione per la quale $\mu_F = E_F(\tilde{X}_i)$ esiste, possiamo concludere che $E_F(\bar{X}_n)$ esiste e che

$$\begin{aligned} E_F(\bar{X}_n) &= E_F\left(\frac{1}{n} \sum_{i=1}^n \tilde{X}_i\right) = \text{[[proprietà E4*]]} = \\ &= \frac{1}{n} \left[E_F(\tilde{X}_1) + E_F(\tilde{X}_2) + \cdots + E_F(\tilde{X}_n) \right] = \frac{1}{n} \times n\mu_F = \mu_F. \end{aligned}$$

□

M2) Se la varianza $\sigma_F^2 = \text{var}(\tilde{X}_i)$ comune alle variabili casuali campionarie esiste, allora esiste anche la varianza $\text{var}_F(\bar{X}_n)$ e

$$\text{var}_F(\bar{X}_n) = \frac{\sigma_F^2}{n}.$$

Dimostrazione. La dimostrazione della proprietà M2 è una semplice applicazione della proprietà V8* della varianza: siccome \bar{X}_n è una trasformazione lineare delle variabili casuali campionarie e siccome stiamo ipotizzando che $F(x)$ sia una distribuzione per la quale $\sigma_F^2 = \text{var}_F(\tilde{X}_i)$ esiste, possiamo concludere che $\text{var}_F(\bar{X}_n)$ esiste e che

$$\begin{aligned} \text{var}_F(\bar{X}_n) &= \text{var}_F\left(\frac{1}{n} \sum_{i=1}^n \tilde{X}_i\right) = \text{var}_F\left(\sum_{i=1}^n \frac{\tilde{X}_i}{n}\right) = \text{[[proprietà V8* & indipendenza]]} = \\ &= \sum_{i=1}^n \frac{1}{n^2} \text{var}_F(\tilde{X}_i) = \sum_{i=1}^n \frac{1}{n^2} \sigma_F^2 = \frac{1}{n^2} \times n\sigma_F^2 = \frac{\sigma_F^2}{n}. \end{aligned}$$

□

M3) (Consistenza). Se il valore atteso $\mu_F = E_F(\tilde{X}_i)$ esiste, si ha

$$\lim_{n \rightarrow \infty} P_F(\{|\bar{X}_n - \mu_F| > \epsilon\}) = 0 \quad \text{per ogni } \epsilon > 0.$$

Con riferimento a qualsiasi famiglia di distribuzioni che contiene soltanto distribuzioni $F(x)$ che sono dotate di valore atteso $\mu = \mu_F = E_F(\tilde{X}_i)$, la **successione delle medie campionarie** è dunque una **successione di stimatori consistente** per μ , ovvero una successione di stimatori tale che

$$\begin{aligned} \lim_{n \rightarrow \infty} P_F(\{|\bar{X}_n - \mu_F| > \epsilon\}) &= 0 \\ \text{per ogni } \epsilon > 0 \text{ e per ogni } F(x) \in \mathcal{F}. \end{aligned}$$

La proprietà di consistenza della media campionaria è anche nota come **legge debole dei grandi numeri**.^a Questa proprietà può essere espressa in modo meno formale dicendo che all'aumentare della numerosità campionaria n la distribuzione della media campionaria concentra sempre più probabilità vicino all'ignoto valore atteso che vogliamo stimare.

^aOltre alla proprietà di consistenza qui enunciata, la media campionaria soddisfa anche una proprietà più forte che è nota come legge "forte" dei grandi numeri.

Dimostrazione. Per semplicità dimostreremo la proprietà della "consistenza" soltanto con riferimento all'ipotesi che la varianza $\sigma_F^2 = \text{var}_F(\tilde{X}_i)$ esista. Si ricordi che in questo caso (per definizione) deve esistere anche il valore atteso $\mu_F = E_F(\tilde{X}_i)$ e che per le proprietà **M1** e **M2** devono esistere anche $E_F(\bar{X}_n)$ e $\text{var}_F(\bar{X}_n)$. Quindi possiamo applicare la disuguaglianza di Cebicef (proprietà **V5**) onde concludere che

$$P_F(\{|\bar{X}_n - E_F(\bar{X}_n)| > \epsilon\}) \leq \frac{\text{var}_F(\bar{X}_n)}{\epsilon^2} \quad \text{per ogni } \epsilon > 0.$$

Sostituendo

$$E_F(\bar{X}_n) = [[\text{proprietà M1}]] = \mu_F \quad \text{e} \quad \text{var}_F(\bar{X}_n) = [[\text{proprietà M2}]] = \sigma_F^2/n,$$

possiamo riscrivere la precedente disuguaglianza come

$$P_F(\{|\bar{X}_n - \mu_F| > \epsilon\}) \leq \frac{\sigma_F^2}{n\epsilon^2} \quad \text{per ogni } \epsilon > 0.$$

Ricordando che una probabilità non può mai essere negativa (assioma **K1**) e prendendo il limite per n che tende a infinito ad ambo i membri dell'ultima disuguaglianza, otteniamo dunque

$$0 \leq \lim_{n \rightarrow \infty} P_F(\{|\bar{X}_n - \mu_F| > \epsilon\}) \leq \lim_{n \rightarrow \infty} \frac{\sigma_F^2}{n\epsilon^2} = 0 \quad \text{per ogni } \epsilon > 0$$

come volevamo dimostrare. □

M4) (Normalità asintotica). Se la varianza $\sigma_F^2 = \text{var}_F(\tilde{X}_i)$ esiste, si ha

$$\lim_{n \rightarrow \infty} \sup_{x \in \mathbb{R}} \left| P_F(\{\bar{X}_n \leq x\}) - \Phi\left(\frac{x - \mu_F}{\sigma_F/\sqrt{n}}\right) \right| = 0$$

e per n sufficientemente elevato si ottiene quindi l'approssimazione

$$P_F(\{\bar{X}_n \leq x\}) \simeq \Phi\left(\frac{x - \mu_F}{\sigma_F/\sqrt{n}}\right). \quad (189)$$

Per esprimere la proprietà di normalità asintotica in modo meno formale si può dire che se la varianza delle variabili casuali campionarie esiste e la numerosità campionaria n è sufficientemente elevata, la distribuzione della media campionaria può essere approssimata attraverso la distribuzione normale con valore atteso e varianza identici a quelli della media campionaria.

La proprietà **M4** è un'immediata conseguenza del TLC (Teorema 8.13). Nel caso particolare in cui la distribuzione comune alle variabili casuali campionarie è normale, l'errore di approssimazione nella (189) è nullo:

M4*) (Normalità esatta). Se la distribuzione comune alle variabili casuali campionarie è normale, si ha

$$\bar{X}_n \sim \text{Normale} \left(\mu = \mu_F = E_F(\tilde{X}_i); \sigma = \frac{\sigma_F}{\sqrt{n}} = \sqrt{\frac{\text{var}_F(\tilde{X}_i)}{n}} \right).$$

M5) (Proprietà di ottimalità). Con riferimento

- alla famiglia \mathcal{F} delle distribuzioni normali
- e con riferimento alla famiglia \mathcal{F} che contiene tutte le distribuzioni $F(x)$ che sono dotate di varianza

la media campionaria è l'unico stimatore UMVUE (per il valore atteso $\mu = \mu_F = E_F(\tilde{X}_i)$).

Non dimostreremo la proprietà M5 perché una sua dimostrazione sarebbe troppo lunga e complicata per essere presentata in un testo introduttivo sull'inferenza statistica.

Con riferimento alla proprietà M5 conviene aggiungere che la media campionaria è anche l'unico stimatore UMVUE con riferimento a molte altre famiglie di distribuzioni. In particolare, la media campionaria è l'unico stimatore UMVUE con riferimento alla famiglia \mathcal{F} di tutte le distribuzioni bernoulliane. Come vedremo tra breve, problemi di stima dove a priori è noto che la distribuzione delle variabili casuali campionarie è bernoulliana si presentano in molte situazioni reali.

Esercizio 9.1. Si supponga che gli incassi giornalieri di un supermercato siano descritti da variabili casuali i.i.d. con scarto quadratico medio $\sigma = 0,030$ milioni di euro e con valore atteso μ ignoto.

- a) Per quanti giorni bisogna rilevare l'incasso giornaliero affinché lo scarto quadratico medio della media campionaria degli incassi giornalieri sia minore di 0,01 milioni di euro?
- b) Si supponga che l'ignota distribuzione degli incassi giornalieri sia normale con $\sigma = 0,030$ e μ ignoto. Tenendo conto di questa ipotesi, per quanti giorni bisogna rilevare l'incasso giornaliero affinché

$$P(\{|\bar{X}_n - \mu| > 0,015\}) \leq 0,01?$$

- c) Qual è la probabilità che la media campionaria calcolata su $n = 36$ incassi giornalieri si discosti dall'ignoto valore atteso μ per più di 0,015 milioni di euro?
- d) Qual è la probabilità che la differenza tra l'ignoto valore atteso μ e la media campionaria calcolata su $n = 36$ incassi giornalieri sia superiore a 0,015 milioni di euro?
- e) Si supponga che negli ultimi $n = 36$ giorni il supermercato abbia incassato complessivamente 21,6 milioni di euro. Alla luce del risultato ottenuto nella risposta al quesito precedente, si può considerare plausibile l'ipotesi secondo la quale l'ignoto valore atteso μ è superiore a 0,615 milioni di euro?

Risposte:

Nelle risposte indicheremo con $\tilde{X}_1, \tilde{X}_2, \dots$ le variabili casuali che rappresentano gli incassi giornalieri del supermercato. Per ipotesi, le variabili casuali \tilde{X}_i sono i.i.d. con scarto quadratico medio $\sigma = 0,030$ milioni di euro. Come nella teoria che abbiamo appena esposto, indicheremo con $F^*(x)$ l'ignota distribuzione comune agli incassi giornalieri. Per indicare il valore atteso e la varianza degli incassi giornalieri scriveremo dunque $\mu_{F^*} = E_{F^*}(\tilde{X}_i)$ e $\sigma_{F^*}^2 = \text{var}_{F^*}(\tilde{X}_i)$. Si noti che il valore numerico di μ_{F^*} è ignoto, mentre quello di $\sigma_{F^*}^2$ è noto e pari a $\sigma_{F^*}^2 = 0,030^2$.

- a) Per quanti giorni bisogna rilevare l'incasso giornaliero affinché lo scarto quadratico medio della media campionaria degli incassi giornalieri sia minore di 0,01 milioni di euro?

Lo scarto quadratico medio della media campionaria è dato da

$$\sigma_{F^*}(\bar{X}_n) = \sqrt{\text{var}_{F^*}(\bar{X}_n)} = [[\text{proprietà M2}]] = \frac{\sigma_{F^*}}{\sqrt{n}} = \frac{0,030}{\sqrt{n}}.$$

Il suo valore è minore di 0,01 milioni di euro se e solo se

$$\sigma_{F^*}(\bar{X}_n) < 0,01 \quad \Leftrightarrow \quad \frac{0,030}{\sqrt{n}} < 0,01 \quad \Leftrightarrow \quad \sqrt{n} > \frac{0,030}{0,010} \quad \Leftrightarrow \quad n > 9.$$

- b) Si supponga che l'ignota distribuzione degli incassi giornalieri sia normale con $\sigma = 0,030$ e μ ignoto. Tenendo conto di questa ipotesi, per quanti giorni bisogna rilevare l'incasso giornaliero affinché

$$P(\{|\bar{X}_n - \mu| > 0,015\}) \leq 0,01?$$

Per rispondere al quesito osserviamo innanzitutto che se $F(x)$ è una distribuzione normale con $\sigma = \sigma_F = 0,030$, allora si deve avere (proprietà M4*)

$$\bar{X}_n \sim \text{Normale}(\mu = \mu_F = ???; \sigma = \sigma_F = 0,030).$$

In virtù di questa osservazione possiamo scrivere

$$\begin{aligned} P_F(\{|\bar{X}_n - \mu_F| > 0,015\}) &= \\ &= P_F(\{\bar{X}_n > \mu_F + 0,015\}) + P_F(\{\bar{X}_n < \mu_F - 0,015\}) \\ &= 1 - \Phi\left(\frac{(\mu_F + 0,015) - \mu_F}{\sigma_F/\sqrt{n}}\right) + \Phi\left(\frac{(\mu_F - 0,015) - \mu_F}{\sigma_F/\sqrt{n}}\right) \\ &= 1 - \Phi\left(\frac{0,015}{\sigma_F/\sqrt{n}}\right) + \Phi\left(\frac{-0,015}{\sigma_F/\sqrt{n}}\right) \\ &[[\text{simmetria delle distribuzioni normali}]] \\ &= 2 \left[1 - \Phi\left(\frac{0,015}{\sigma_F/\sqrt{n}}\right) \right]. \end{aligned}$$

e questo risultato dimostra che

$$P_F(\{|\bar{X}_n - \mu_F| > 0,015\}) = 2 \left[1 - \Phi\left(\frac{0,015}{\sigma_F/\sqrt{n}}\right) \right] \quad (190)$$

non dipende dal valore di μ_F !!! Per rispondere al quesito dobbiamo quindi determinare il più piccolo valore di n tale che

$$2 \left[1 - \Phi\left(\frac{0,015}{\sigma_F/\sqrt{n}}\right) \right] \leq 0,01.$$

Dopo alcuni semplici ragionamenti ...

$$\begin{aligned} 2 \left[1 - \Phi\left(\frac{0,015}{\sigma_F/\sqrt{n}}\right) \right] \leq 0,01 &\Leftrightarrow \\ \Leftrightarrow \Phi\left(\frac{0,015}{\sigma_F/\sqrt{n}}\right) \geq 1 - \frac{0,01}{2} = 0,995 &\Leftrightarrow \\ \Leftrightarrow \frac{0,015}{\sigma_F/\sqrt{n}} = z_{0,995} = 2,575 &\Leftrightarrow \\ \Leftrightarrow n \geq 2,575^2 \frac{\sigma_F^2}{0,015^2} = 2,575^2 \frac{0,03^2}{0,015^2} = 26,52 \end{aligned}$$

... vediamo che

$$P(\{|\bar{X}_n - \mu| > 0,015\}) \leq 0,01$$

se e solo se la media campionaria \bar{X}_n è basata su almeno $n = [26,52] = 27$ incassi giornalieri.

- c) Qual è la probabilità che la media campionaria calcolata su $n = 36$ incassi giornalieri si discosti dall'ignoto valore atteso μ per più di 0,015 milioni di euro?

Per dare una risposta esatta a questa domanda bisognerebbe conoscere la distribuzione esatta della media campionaria che però è ignota perché dipende dall'ignota distribuzione comune agli incassi giornalieri \tilde{X}_i . Tuttavia, l'evento di cui è richiesta la probabilità si riferisce ad una media campionaria calcolata su un numero abbastanza elevato di osservazioni e quindi possiamo calcolare un'approssimazione alla probabilità richiesta facendo riferimento proprietà [M4](#). Così facendo otteniamo

$$\begin{aligned} P_{F^*}(\{|\bar{X}_n - \mu_{F^*}| > 0,015\}) &\simeq \\ &\simeq \text{[[proprietà M4 e formula (190)]]} \simeq \\ &\simeq 2 \left[1 - \Phi \left(\frac{0,015}{\sigma_{F^*}/\sqrt{n}} \right) \right] = \\ &= 2 \left[1 - \Phi \left(\frac{0,015}{0,03/\sqrt{36}} \right) \right] = 2[1 - \Phi(3)] = 2[1 - 0,9987] = 0,0026. \end{aligned}$$

Si noti che senza ulteriori ipotesi sull'ignota distribuzione $F^*(x)$ non possiamo valutare l'errore di approssimazione che stiamo commettendo. Possiamo soltanto dire che l'errore di approssimazione è nullo se la distribuzione comune alle variabili casuali campionarie è normale, e che l'errore di approssimazione è tanto più elevato quanto più la distribuzione comune alle variabili casuali campionarie è asimmetrica e/o quanto più una o entrambe le sue code sono "pesanti".

- d) Qual è la probabilità che la differenza tra l'ignoto valore atteso μ e la media campionaria calcolata su $n = 36$ incassi giornalieri sia superiore a 0,015 milioni di euro?

Anche in questo caso possiamo calcolare soltanto un'approssimazione alla probabilità richiesta:

$$\begin{aligned} P_{F^*}(\{\mu_{F^*} - \bar{X}_n > 0,015\}) &= P_{F^*}(\{\bar{X}_n < \mu_{F^*} - 0,015\}) = \\ &\text{[[proprietà M4*]]} \\ &\simeq \Phi \left(\frac{(\mu_{F^*} - 0,015) - \mu_{F^*}}{\sigma_{F^*}/\sqrt{n}} \right) = \Phi \left(\frac{-0,015}{\sigma_{F^*}/\sqrt{n}} \right) = \\ &\text{[[simmetria della curva normale]]} \\ &= 1 - \Phi \left(\frac{0,015}{\sigma_{F^*}/\sqrt{n}} \right) = 1 - \Phi \left(\frac{0,015}{0,03/\sqrt{36}} \right) = \\ &= 1 - \Phi(3) = 1 - 0,9987 = 0,0013. \end{aligned}$$

- e) Si supponga che negli ultimi $n = 36$ giorni il supermercato abbia incassato complessivamente 21,6 milioni di euro. Alla luce del risultato ottenuto nella risposta al quesito precedente, si può considerare plausibile l'ipotesi secondo la quale l'ignoto valore atteso μ è superiore a 0,615 milioni di euro?

La media (campionaria) degli ultimi $n = 36$ incassi giornalieri ammonta a

$$\bar{x}_n = \frac{21,6}{36} = 0,600 \text{ milioni di euro.}$$

Se l'ignoto valore atteso $\mu = \mu_{F^*}$ fosse superiore a 0,615 milioni di euro, allora si avrebbe

$$\mu_{F^*} - \bar{x}_n > 0,615 - 0,600 = 0,015 \text{ milioni di euro.}$$

Nella risposta al quesito precedente abbiamo visto che un errore di sottostima così elevato è molto improbabile dato che

$$P_{F^*}(\{\mu_{F^*} - \bar{X}_n > 0,015\}) \simeq 0,0013.$$

Dal punto di vista statistico possiamo quindi escludere che l'ignoto valore atteso $\mu = \mu_{F^*}$ sia superiore a 0,615 milioni di euro.

9.1.2 La frequenza relativa campionaria

In molti problemi di stima si incontrano **variabili casuali campionarie \tilde{X}_i che sono bernoulliane**, ovvero che possono assumere soltanto i valori 0 e 1 (si pensi per esempio al caso in cui si vuole stimare un'ignota probabilità di successo p e a tal fine si esegue un certo numero n di repliche di un esperimento casuale). Come vedremo tra breve, il caso di variabili casuali campionarie bernoulliane presenta alcune peculiarità che conviene analizzare in modo separato.

La prima peculiarità riguarda l'interpretazione della media campionaria

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n \tilde{X}_i.$$

Se le variabili casuali campionarie \tilde{X}_i sono bernoulliane, la media campionaria \bar{X}_n rappresenta una **frequenza relativa** (ovvero la frequenza relativa dei "successi" nel campione) e per questo motivo viene chiamata **"frequenza relativa campionaria"** (e non "media campionaria" come negli altri casi).

Per distinguere le frequenze relative campionarie dalle altre medie campionarie, in questa *dispensa* indicheremo le prime con \tilde{p}_n piuttosto che con \bar{X}_n . Per indicare la

realizzazione di una frequenza relativa campionaria scriveremo invece \hat{p}_n piuttosto che \bar{x}_n .

Siccome la probabilità di "successo" $p = P(\{\tilde{X}_i = 1\})$ determina univocamente una distribuzione bernoulliana, scriveremo inoltre $P_p(\cdot)$, $E_p(\cdot)$ e $var_p(\cdot)$ (con la "p" al posto della F al pedice) per indicare probabilità, valori attesi e varianze di funzioni di variabili casuali campionarie bernoulliane.

Trattandosi di un caso particolare di media campionaria, anche la frequenza relativa campionaria deve soddisfare le proprietà M1 - M4. Le proprietà M4* e M5 invece non riguardano la frequenza relativa campionaria perché si riferiscono al caso in cui la famiglia \mathcal{F} di tutte le possibili distribuzioni delle variabili casuali campionarie contiene anche (o solo) distribuzioni che non sono bernoulliane. Tuttavia, al posto della proprietà M5, la frequenza relativa campionaria soddisfa una proprietà analoga dato che (come abbiamo già anticipato) è l'unico stimatore UMVUE con riferimento alla famiglia \mathcal{F} che contiene tutte le distribuzioni bernoulliane. Per dare il giusto risalto alle principali peculiarità che distinguono la frequenza relativa campionaria da una generica media campionaria conviene riscrivere gli enunciati delle proprietà della media campionaria adattandoli al caso specifico che si ottiene quando a priori è noto che l'ignota distribuzione $F^*(x)$ è di tipo bernoulliano.

P1) (Correttezza). La frequenza relativa campionaria è uno stimatore corretto:

$$E_p(\tilde{p}_n) = p \quad \text{per ogni } p \in [0, 1].$$

P2) La varianza della frequenza relativa campionaria è data da

$$var_p(\tilde{p}_n) = \frac{p(1-p)}{n}.$$

Si noti che la proprietà P2 è una diretta conseguenza della proprietà M2 e del fatto che per variabili casuali campionarie bernoulliane si ha $var_p(\tilde{X}_i) = p(1-p)$.

Siccome la funzione $f(p) = p(1-p)$ è una parabola rovesciata con punto di massimo in $p = 0,5$ (lasciamo al lettore il compito di verificare), possiamo concludere che

$$var_p(\tilde{p}_n) = [[\text{proprietà P2}]] = \frac{p(1-p)}{n} \leq \frac{0,5^2}{n} \quad \text{per ogni } p \in [0, 1]$$

In parole: la varianza della frequenza relativa campionaria non può mai essere maggiore di $0,5^2/n$.

P3) (Consistenza). La frequenza relativa campionaria è uno stimatore consistente per la probabilità di successo p :

$$\lim_{n \rightarrow \infty} P_p(\{|\tilde{p}_n - p| > \epsilon\}) = 0$$

per ogni $p \in [0, 1]$ e per ogni $\epsilon > 0$.

Come nota storica, vale la pena osservare che la proprietà di consistenza della frequenza relativa campionaria fu enunciata e pubblicata per la prima da Jacob Bernoulli (1604 - 1705) nel suo libro *Ars Conjectandi* che abbiamo già avuto modo di citare in precedenza.

P4) (Normalità asintotica). Per ogni $p \in (0, 1)$ si ha

$$\lim_{n \rightarrow \infty} \sup_{x \in \mathbb{R}} \left| P_p(\{\tilde{p}_n \leq x\}) - \Phi\left(\frac{x - p}{p(1-p)/\sqrt{n}}\right) \right| = 0$$

e per n sufficientemente elevato si ottiene quindi l'approssimazione

$$P_p(\{\tilde{p}_n \leq x\}) \simeq \Phi\left(\frac{x - \mu_F}{\sigma_F/\sqrt{n}}\right). \quad (191)$$

Per l'errore di approssimazione nella (191) valgono esattamente le stesse considerazioni che abbiamo già enunciato con riferimento all'errore di approssimazione che si commette quando si approssima una distribuzione binomiale con la distribuzione normale che ha lo stesso valore atteso e la stessa varianza: l'errore di approssimazione è trascurabile se $np(1-p) \geq 5$.

P5) (Proprietà di ottimalità). La frequenza relativa campionaria è l'unico stimatore UMVUE per la probabilità di successo p con riferimento alla famiglia \mathcal{F} che contiene tutte le distribuzioni bernoulliane.

Esercizio 9.2. Per stimare l'ignota proporzione p^* di elettori che sostengono il partito XXX, un sondaggista vuole intervistare un campione di n elettori.

- Quanti elettori si devono intervistare affinché lo scarto quadratico medio della frequenza relativa campionaria sia minore di 0,01?
- Si supponga che 24 elettori di un campione preliminare di $n = 100$ elettori dichiarino di sostenere il partito XXX. Si calcoli, sulla base di questo campione preliminare, il numero minimo di elettori che si devono intervistare affinché

lo scarto quadratico medio della frequenza relativa campionaria sia minore di 0,01.

- c) Si supponga che il valore dell'ignota proporzione p^* sia $p^* = 0,22$. Qual è la probabilità che in un campione di $n = 900$ elettori ci siano più di 220 elettori che dichiarano di sostenere il partito XXX?
- d) Si supponga che in un campione di $n = 900$ elettori ci siano 200 elettori che dichiarano di sostenere il partito XXX. Si stabilisca, alla luce della risposta al quesito precedente, se dal punto di vista statistico si può considerare plausibile l'ipotesi secondo la quale l'ignota proporzione p^* è uguale a 0,22.

Risposte:

Nelle risposte considereremo le preferenze degli elettori rispetto al partito XXX come realizzazioni di variabili casuali campionarie i.i.d. con distribuzione bernoulliana di parametro $p = p^*$ ignoto. L' i -esima variabile casuale campionaria \tilde{X}_i assume dunque il valore 1 se l' i -esimo elettore intervistato dichiara di sostenere il partito XXX, e assume il valore 0 altrimenti.

- a) Quanti elettori si devono intervistare affinché lo scarto quadratico medio della frequenza relativa campionaria sia minore di 0,01?

In generale, lo scarto quadratico medio della frequenza relativa campionaria è dato da

$$\sqrt{\text{var}_p(\tilde{p}_n)} = \text{[[proprietà P2]]} = \sqrt{\frac{p(1-p)}{n}}$$

e il suo valore è dunque minore di 0,01 se e solo se

$$n > \frac{p(1-p)}{0,01^2}.$$

Siccome non conosciamo il vero valore p^* della proporzione di elettori che sostengono il partito XXX, non possiamo calcolare il valore di

$$n^* = \left\lceil \frac{p^*(1-p^*)}{0,01^2} \right\rceil$$

che rappresenta la numerosità campionaria minima per ottenere $\sqrt{\text{var}_{p^*}(\tilde{p}_n)} < 0,01$. Tuttavia, siccome

$$n^* = \left\lceil \frac{p^*(1-p^*)}{0,01^2} \right\rceil \leq \left\lceil \frac{0,5(1-0,5)}{0,01^2} \right\rceil = 2500,$$

possiamo concludere che, **indipendentemente dall'ignoto valore di p^*** , con un campione di $n = 2500 + 1 = 2501$ elettori si ottiene $\sqrt{\text{var}_{p^*}(\tilde{p}_n)} < 0,01$.

- b) Si supponga che 24 elettori di un campione preliminare di $n = 100$ elettori dichiarino di sostenere il partito XXX. Si calcoli, sulla base di questo campione preliminare, una stima del numero minimo di elettori che si devono intervistare affinché lo scarto quadratico medio della frequenza relativa campionaria sia minore di 0,01.

Nella risposta al quesito a) abbiamo visto che il numero minimo di elettori da intervistare è dato da

$$n^* = \left\lceil \frac{p^*(1-p^*)}{0,01^2} \right\rceil,$$

ma che il valore di n^* è ignoto perché non conosciamo il valore di p^* . Tuttavia, se al posto di p^* sostituiamo la stima puntuale $\hat{p}_n = 24/100 = 0,24$, possiamo ottenere una **stima puntuale per l'ignoto valore di n^*** :

$$\hat{n}^* = \left\lceil \frac{\hat{p}_n(1-\hat{p}_n)}{0,01^2} \right\rceil = \left\lceil \frac{0,24(1-0,24)}{0,01^2} \right\rceil = [1824] = 1824.$$

[[A proposito di questa stima puntuale, conviene precisare che non si tratta della realizzazione di uno stimatore UMVUE. Infatti, il fatto che \hat{p}_n sia la realizzazione di uno stimatore UMVUE non implica che ciò sia vero anche per \hat{n}^* . Di fatto, si può dimostrare che lo stimatore associato alla stima puntuale \hat{n}^* non è nemmeno corretto (rispetto alla famiglia di tutte le distribuzioni bernoulliane).]]

- c) Si supponga che il valore dell'ignota proporzione p^* sia $p^* = 0,22$. Qual è la probabilità che in un campione di $n = 900$ elettori ci siano più di 200 elettori che dichiarano di sostenere il partito XXX?

Sia $\tilde{Y}_{900} = \sum_{i=1}^{900} \tilde{X}_i$ la variabile casuale che rappresenta il numero complessivo di elettori del partito XXX all'interno di un campione di $n = 900$ elettori. In virtù dell'ipotesi sulle variabili casuali campionarie (vedi la premessa alle risposte), possiamo affermare che

$$\tilde{Y}_{900} \sim \text{Binomiale}(n = 900; p = p^* = 0,22).$$

La probabilità richiesta è dunque data da

$$\begin{aligned} P(\{\tilde{Y}_{900} > 200\}) &= [[\text{TLC} - \text{Teorema 8.13}]] = \\ &= 1 - \Phi\left(\frac{200 - 900 \times 0,22}{\sqrt{900 \times 0,22(1-0,22)}}\right) = \\ &= 1 - \Phi(0,16) = 1 - 0,5636 = 0,4364. \end{aligned}$$

- d) Si supponga che in un campione di $n = 900$ elettori ci siano 200 elettori che dichiarano di sostenere il partito XXX. Si stabilisca, alla luce della risposta al

quesito precedente, se dal punto di vista statistico si può considerare plausibile l'ipotesi secondo la quale l'ignota proporzione p^* è uguale a 0,22.

Dalla risposta al quesito precedente desumiamo che con $p^* = 0,22$ è molto probabile che un campione di $n = 900$ elettori contenga più di 200 elettori che sostengono il partito XXX, e che allo stesso tempo è anche molto probabile che un campione di $n = 900$ contenga meno di 200 elettori che sostengono il partito XXX. Sulla base di queste considerazioni possiamo affermare che se in un campione di $n = 900$ elettori osserviamo 200 elettori che sostengono il partito XXX, allora (dal punto di vista statistico) non si può escludere l'ipotesi che l'ignoto valore di p^* sia uguale a 0,22.

9.1.3 La varianza campionaria e la varianza campionaria corretta

Come suggerisce il loro nome, la **varianza campionaria** e la **varianza campionaria corretta** sono due stimatori per la varianza $\sigma_F^2 = \text{var}_F(\tilde{X}_i)$ delle variabili casuali campionarie.

La **varianza campionaria** è la varianza delle variabili casuali campionarie ed è dunque definita come

$$\tilde{V}_n^2 = \frac{1}{n} \sum_{i=1}^n (\tilde{X}_i - \bar{X}_n)^2,$$

mentre la **varianza campionaria corretta** è definita come

$$\tilde{S}_n^2 = \frac{1}{n-1} \sum_{i=1}^n (\tilde{X}_i - \bar{X}_n)^2.$$

Si noti che per $n = 1$ si ha $\tilde{V}_n^2 = \tilde{V}_1^2 = 0$ (infatti, con $n = 1$ si ha $\tilde{X}_1 = \bar{X}_1$) e che per $n = 1$ lo stimatore \tilde{S}_n^2 non è definito. Per valori elevati di n gli stimatori \tilde{V}_n^2 e \tilde{S}_n^2 sono invece praticamente identici perché

$$\tilde{S}_n^2 = \frac{n}{n-1} \tilde{V}_n^2 \quad (192)$$

e perché per n elevato il rapporto $\frac{n}{n-1}$ è molto prossimo a 1.

Il motivo per cui al posto dello stimatore \tilde{V}_n^2 (che a prima vista potrebbe sembrare quello più "naturale") spesso si considera lo stimatore \tilde{S}_n^2 è dovuto al fatto che quest'ultimo è uno stimatore corretto, mentre \tilde{V}_n^2 non lo è. Infatti, si può dimostrare che

$$E_F(\tilde{S}_n^2) = \sigma_F^2 = \text{var}_F(\tilde{X}_i)$$

per ogni distribuzione $F(x)$ tale che $\sigma_F^2 = \text{var}_F(\tilde{X}_i)$ esiste

mentre

$$E_F(\tilde{V}_n^2) = E_F\left(\frac{n-1}{n} \tilde{S}_n^2\right) = [[\text{proprietà E2}]] = \frac{n-1}{n} \sigma_F^2 \neq \sigma_F^2$$

per ogni distribuzione $F(x)$ tale che $\sigma_F^2 = \text{var}_F(\tilde{X}_i)$ esiste

Tranne per famiglie di distribuzioni \mathcal{F} che contengono soltanto distribuzioni degeneri (ovvero soltanto distribuzioni $F(x)$ con varianza $\sigma_F^2 = 0$), lo stimatore \tilde{V}_n^2 non è dunque uno stimatore corretto e non può dunque essere uno stimatore UMVUE. Come vedremo nel seguente elenco di proprietà, lo stimatore S_n^2 è invece UMVUE rispetto a due famiglie di distribuzioni che sono molto importanti per le applicazioni dell'inferenza statistica. A proposito dell'ottimalità dello stimatore S_n^2 conviene rimarcare una piccola curiosità che mette in luce i limiti del criterio della correttezza: nonostante il fatto che \tilde{S}_n^2 sia uno stimatore corretto e che \tilde{V}_n^2 non lo sia, si può dimostrare che per ogni $n \geq 2$ si ha

$$EQM_F(\tilde{V}_n^2) < EQM_F(\tilde{S}_n^2)$$

per ogni $F(x)$ tale che $\mu_4 = E_F(\tilde{X}_i^4)$ esiste

Tuttavia, come si può facilmente intuire in virtù della stretta relazione che sussiste tra questi due stimatori (vedi la formula (192)), il rapporto tra i loro EQM tende a 1 all'aumentare di n e per numerosità campionarie sufficientemente elevate è quindi indifferente quale dei due stimatori si utilizza.

Come per la media campionaria e per la frequenza relativa campionaria, riporteremo ora un elenco delle principali proprietà della varianza campionaria corretta (omettiamo le dimostrazioni). Si noti che queste proprietà sono perfettamente analoghe a quelle della media campionaria:

S1) (Correttezza). Se $\sigma_F^2 = \text{var}_F(\tilde{X}_i)$ esiste, si ha

$$E_F(\tilde{S}_n^2) = \sigma_F^2$$

S2) (Varianza). Se $\mu_{F,4} = E_F([\tilde{X}_i - \mu_F]^4)$ esiste (si può dimostrare che questa condizione è soddisfatta se e solo se esiste anche il momento quarto della distribuzione $F(x)$), si ha

$$\text{var}_F(\tilde{S}_n^2) = \begin{cases} \frac{1}{2}(\mu_{F,4} + \sigma_F^4) & \text{se } n = 2, \\ \frac{\mu_{F,4}}{n} - \frac{\sigma_F^4(n-3)}{n(n-1)} & \text{se } n \geq 3. \end{cases}$$

S3) (Consistenza). Se $\sigma_F^2 = \text{var}_F(\tilde{X}_i)$ esiste, si ha

$$\lim_{n \rightarrow \infty} P_F(\{|\tilde{S}_n^2 - \sigma_F^2| > \epsilon\}) = 0 \quad \text{per ogni } \epsilon > 0.$$

Con riferimento a qualsiasi famiglia di distribuzioni che contiene soltanto distribuzioni $F(x)$ che sono dotate di varianza $\sigma_F^2 = \text{var}_F(\tilde{X}_i)$, la successione delle varianze campionarie corrette \tilde{S}_n^2 è dunque una successione di stimatori consistente.

S4) (Normalità asintotica). Se $\mu_{F,4} = E_F([\tilde{X}_i - \mu_F]^4)$ esiste, si ha

$$\lim_{n \rightarrow \infty} \sup_{x \in \mathbb{R}} \left| P_F(\{\tilde{S}_n \leq x\}) - \Phi\left(\frac{x - \sigma_F^2}{\sqrt{\text{var}_F(\tilde{S}_n^2)}}\right) \right| = 0$$

e per n sufficientemente elevato si ottiene quindi l'approssimazione

$$P_F(\{\tilde{S}_n \leq x\}) \simeq \Phi\left(\frac{x - \sigma_F^2}{\sqrt{\text{var}_F(\tilde{S}_n^2)}}\right). \quad (193)$$

La proprietà della normalità asintotica può essere espressa in modo meno formale dicendo che se $\mu_{F,4} = E_F([\tilde{X}_i - \mu_F]^4)$ esiste e la numerosità campionaria n è sufficientemente elevata, allora la distribuzione della varianza campionaria corretta può essere approssimata attraverso la distribuzione normale con valore atteso e varianza identici a quelli della varianza campionaria corretta.

S4*) (Distribuzione esatta sotto l'ipotesi di normalità). Se la distribuzione comune alle variabili casuali campionarie è normale, si ha

$$\frac{(n-1)\tilde{S}_n^2}{\sigma_F^2} \sim \text{Gamma}(\alpha = (n-1)/2; \theta = 1/2).$$

Si ricordi che le distribuzioni $\text{Gamma}(\alpha = k/2; \theta = 1/2)$ con $k = 1, 2, \dots$ sono note come **distribuzioni chi-quadrato** e che il valore del parametro k che le identifica è il cosiddetto numero di gradi di libertà.

S5) (Proprietà di ottimalità). Con riferimento

- alla famiglia \mathcal{F} delle distribuzioni normali
- e con riferimento alla famiglia \mathcal{F} di tutte le distribuzioni $F(x)$ che sono dotate di momento quarto,

la varianza campionaria corretta \tilde{S}_n^2 è l'unico stimatore UMVUE per la varianza $\sigma^2 = \sigma_F^2 = \text{var}_F(\tilde{X}_i)$.

Esercizio 9.3. Si supponga che gli incassi giornalieri di un supermercato siano descritti da variabili casuali i.i.d., che l'incasso complessivo degli ultimi $n = 36$ giorni di apertura ammonti a $\sum_{i=1}^{36} x_i = 21,571$ milioni di euro e che la somma dei quadrati di tali incassi ammonti a $\sum_{i=1}^{36} x_i^2 = 14,122$ (milioni di euro al quadrato).

- a) Si calcoli una stima per l'ignota varianza $\sigma_{F^*}^2 = \text{var}_{F^*}(\tilde{X}_i)$ degli incassi giornalieri.
- b) Si calcoli una stima per il numero minimo di osservazioni dell'incasso giornaliero che sono necessarie affinché lo scarto quadratico medio della media

campionaria degli incassi giornalieri sia minore di 0,01 milioni di euro.

- c) Si supponga che l'ignota distribuzione degli incassi giornalieri sia normale con parametri μ e σ che sono entrambi ignoti. Si calcoli una stima per il numero minimo di osservazioni dell'incasso giornaliero che sono necessarie affinché

$$P(\{|\bar{X}_n - \mu| > 0,015\}) \leq 0,01?$$

- d) Si calcoli una stima per la probabilità che la differenza tra l'ignoto valore atteso μ e la media campionaria calcolata su $n = 36$ incassi giornalieri sia superiore a 0,015 milioni di euro.
- e) Alla luce del risultato ottenuto nella risposta al quesito precedente, si può giudicare plausibile l'ipotesi secondo la quale l'ignoto valore atteso $\mu = \mu_{F^*}$ è superiore a 0,610 milioni di euro?

Risposte:

- a) Si calcoli una stima per l'ignota varianza $\sigma_{F^*}^2 = \text{var}_{F^*}(\tilde{X}_i)$ degli incassi giornalieri.

Usando la formula indiretta per il calcolo della devianza otteniamo la stima puntuale

$$\begin{aligned} s_n^2 &= \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^2 = \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - n\bar{x}_n^2 \right) = \\ &= \frac{1}{36-1} \left(14,122 - 36 \left(\frac{21,571}{36} \right)^2 \right) = \frac{1}{36-1} (14,122 - 36 \times 0,599^2) = \\ &= \frac{1,205}{35} = 0,034. \end{aligned}$$

- b) Si calcoli una stima per il numero minimo di osservazioni dell'incasso giornaliero che sono necessarie affinché lo scarto quadratico medio della media campionaria degli incassi giornalieri sia minore di 0,01 milioni di euro.

Siccome

$$\text{var}_{F^*}(\tilde{X}_n) = [[\text{proprietà M2}]] = \frac{\sigma_{F^*}^2}{n},$$

l'ignoto valore della numerosità campionaria n^* che ci interessa deve essere dato dal più piccolo valore intero di n tale che

$$\sqrt{\text{var}_{F^*}(\tilde{X}_n)} = \frac{\sigma_{F^*}}{\sqrt{n}} < 0,01 \quad \Rightarrow \quad n > \frac{\sigma_{F^*}^2}{0,01^2}.$$

Non conoscendo il valore di $\sigma_{F^*}^2$, non possiamo calcolare il "vero" valore di

$$n^* = \left\lceil \frac{\sigma_{F^*}^2}{0,01^2} \right\rceil.$$

Tuttavia, usando la stima puntuale $s_n^2 = 0,034$ che abbiamo calcolato nella risposta al quesito precedente, possiamo **stimare** l'ignoto valore di n^* attraverso

$$\hat{n}^* = \left\lceil \frac{s_n^2}{0,01^2} \right\rceil = \left\lceil \frac{0,034}{0,01^2} \right\rceil = 340.$$

- c) Si supponga che l'ignota distribuzione degli incassi giornalieri sia normale con parametri μ e σ che sono entrambi ignoti. Si calcoli una stima per il numero minimo di osservazioni dell'incasso giornaliero che sono necessarie affinché

$$P(\{|\bar{X}_n - \mu| > 0,015\}) \leq 0,01?$$

Come abbiamo già visto nella risposta al quesito b) dell'Esercizio 9.1, se l'ignota distribuzione degli incassi giornalieri è normale, allora il numero minimo di osservazioni affinché

$$P_{F^*}(\{|\bar{X}_n - \mu_{F^*}| > 0,015\}) \leq 0,01$$

deve essere dato da

$$n^* = \left\lceil 2,575^2 \frac{\sigma_{F^*}^2}{0,015^2} \right\rceil.$$

Non conoscendo il valore di $\sigma_{F^*}^2$, non possiamo calcolare il valore vero di n^* . Tuttavia, sostituendo al posto di $\sigma_{F^*}^2$ la stima puntuale $s_n^2 = 0,034$ che abbiamo calcolato nella risposta al quesito a), possiamo ottenere una **stima puntuale** per l'ignoto valore di n^* :

$$\hat{n}^* = \left\lceil 2,575^2 \frac{s_n^2}{0,015^2} \right\rceil = \left\lceil 2,575^2 \frac{0,034}{0,015^2} \right\rceil = [1001,961] = 1002.$$

- d) Si calcoli una stima per la probabilità che la differenza tra l'ignoto valore atteso μ e la media campionaria calcolata su $n = 36$ incassi giornalieri sia superiore a 0,015 milioni di euro.

Come abbiamo già visto nella risposta al quesito d) dell'Esercizio 9.1, per n sufficientemente elevato si ha

$$P_{F^*}(\{\mu_{F^*} - \bar{X}_n > 0,015\}) \simeq 1 - \Phi\left(\frac{0,015}{\sigma_{F^*}/\sqrt{n}}\right),$$

ma siccome non conosciamo il valore di $\sigma_{F^*}^2$, non possiamo calcolare il valore di questa approssimazione. Tuttavia, usando la stima puntuale $s_n^2 = 0,034$,

possiamo ottenere una **stima puntuale** per la suddetta approssimazione, e questa stima puntuale può allo stesso tempo essere considerata come una stima puntuale per la probabilità richiesta nel quesito:

$$\begin{aligned}\hat{P}_{F^*}(\{\mu_{F^*} - \bar{X}_n > 0,015\}) &= 1 - \Phi\left(\frac{0,015}{\sqrt{s_n^2/n}}\right) = 1 - \Phi\left(\frac{0,015}{\sqrt{0,034/36}}\right) = \\ &= 1 - \Phi(0,488) \simeq 1 - \Phi(0,49) = 1 - 0,6879 = \mathbf{0,3121}.\end{aligned}$$

- e) Alla luce del risultato ottenuto nella risposta al quesito precedente si può giudicare plausibile l'ipotesi secondo la quale l'ignoto valore atteso $\mu = \mu_{F^*}$ è superiore a 0,610 milioni di euro?

La media (campionaria) degli ultimi $n = 36$ incassi giornalieri ammonta a

$$\bar{x}_n = \frac{21,571}{26} = 0,599 \text{ milioni di euro.}$$

Se l'ignoto valore atteso $\mu = \mu_{F^*}$ fosse superiore a 0,615 milioni di euro, si avrebbe

$$\mu_{F^*} - \bar{x}_n > 0,610 - 0,599 = \mathbf{0,011} \text{ milioni di euro}$$

e questo significherebbe che si è verificato l'evento $\{\mu_{F^*} - \bar{X}_n > \mathbf{0,011}\}$. Alla luce della risposta al quesito precedente possiamo dire che questo evento **non è particolarmente raro**, avendo **stimato** che

$$\hat{P}_{F^*}(\{\mu_{F^*} - \bar{X}_n > \mathbf{0,015}\}) \simeq 0,3121$$

ovvero che con probabilità pari a 0,3121 si ottiene anche una differenza **più grande di 0,011** ($0,015 > 0,011$). Dal punto di vista statistico **non** possiamo dunque escludere che l'ignoto valore atteso $\mu = \mu_{F^*}$ sia superiore a 0,610 milioni di euro.

9.2 Stima intervallare

Come abbiamo visto poc'anzi, stime puntuali $\hat{t} = t(x_1, \dots, x_n)$ sono realizzazioni di variabili casuali che vengono chiamate stimatori e pertanto potrebbero differire (e di solito differiscono) dall'ignoto valore del parametro θ che si vuole stimare. Una stima puntuale sprovvista di una qualche misura per la precisione dello stimatore che l'ha generata è quindi poco utile. Per sopperire a questo problema, la stima puntuale per un dato parametro di interesse potrebbe essere affiancata da una stima puntuale per l'EQM dello stimatore che l'ha generata. Tuttavia, nelle applicazioni solitamente si preferisce una soluzione più elegante: quella di corredare una stima puntuale degli estremi di un cosiddetto **"intervallo di confidenza"** (d'ora in poi IC), ovvero degli estremi di un intervallo che

contiene l'ignoto valore del parametro θ che si vuole stimare con un determinato **"livello di confidenza"** che solitamente è prossimo al 100%.

Ma che cos'è esattamente un IC??? La seguente definizione e i successivi commenti risponderanno a questa domanda.

Definizione 9.1 (Intervallo di confidenza). Se \mathcal{F} è una famiglia di distribuzioni di cui a priori è noto che contiene l'ignota distribuzione $F^*(x)$ comune alle variabili casuali campionarie e θ è un parametro che si vuole stimare, allora si definisce come **"intervallo di confidenza per θ con livello di confidenza $1 - \alpha$ "** un qualunque intervallo che è delimitato dalle realizzazioni di due variabili casuali \tilde{L}_α e \tilde{U}_α che soddisfano entrambe le seguenti condizioni:

- \tilde{L}_α e \tilde{U}_α devono essere variabili casuali osservabili nel senso che le loro realizzazioni devono poter essere determinate una volta che sono note le realizzazioni delle variabili casuali campionarie;
- \tilde{L}_α e \tilde{U}_α devono essere definite in modo tale che

$$P_F(\{\tilde{L}_\alpha \leq \theta_F \leq \tilde{U}_\alpha\}) = 1 - \alpha \quad \text{per ogni } F(x) \in \mathcal{F}.$$

La definizione di IC impone soltanto delle condizioni che le variabili casuali \tilde{L}_α e \tilde{U}_α devono soddisfare, ma non chiarisce come in un problema di stima concreto si calcolano i valori assunti da queste due variabili casuali. Questa indeterminatezza è dovuta alla generalità della definizione di IC che abbiamo dato. A seconda del problema di stima si possono considerare diverse formule per definire le variabili casuali \tilde{L}_α e \tilde{U}_α che delimitano un IC. In questa dispensa vedremo soltanto come si calcolano gli estremi degli IC che di solito vengono insegnati per primi in un corso introduttivo all'inferenza statistica, ovvero mostreremo solo come di solito si calcolano gli estremi di un IC per un ignoto valore atteso $\mu_{F^*} = E_{F^*}(\tilde{X}_i)$ oppure per un'ignota "probabilità di successo" $p^* = E_{p^*}(\tilde{X}_i)$.

Per cominciare con il caso più semplice, consideriamo il problema di stima dove a priori è noto che la distribuzione comune alle variabili casuali campionarie è normale e dove a priori è anche noto il valore σ_0^2 della loro varianza. Il nostro obiettivo è quello di trovare delle formule che forniscono gli estremi di un IC per l'ignoto valore atteso μ . A tal fine osserviamo innanzitutto che nel problema di stima che stiamo considerando, la famiglia \mathcal{F} di tutte le possibili distribuzioni $F(x)$ è la famiglia delle distribuzioni normali con $\sigma = \sigma_0$. In quanto segue indicheremo questa famiglia con $\mathcal{F}_{N(\sigma_0)}$.

Ora, per trovare due variabili casuali \tilde{L}_α e \tilde{U}_α che delimitano un intervallo di confidenza con livello di confidenza $1 - \alpha$ per il valore atteso $\mu_{F^*} = E_{F^*}(\tilde{X}_i)$, partiamo dalla considerazione che (proprietà M4*)

$$\bar{X}_n \sim \text{Normale}(\mu = \mu_F; \sigma = \sigma_F/\sqrt{n} = \sigma_0/\sqrt{n}) \quad \text{per ogni } F(x) \in \mathcal{F}_{N(\sigma_0)},$$

e da questa considerazione deduciamo che per ogni $F(x) \in \mathcal{F}_{N(\sigma_0)}$ la probabilità di commettere un errore di stima minore di un dato valore positivo m sia data da

$$\begin{aligned} P_F(\{|\bar{X}_n - \mu_F| < m\}) &= P_F(\{\bar{X}_n < \mu_F + m\}) - P_F(\{\bar{X}_n < \mu_F - m\}) = \\ &= \Phi\left(\frac{\mu_F + m - \mu_F}{\sigma_0/\sqrt{n}}\right) - \Phi\left(\frac{\mu_F - m - \mu_F}{\sigma_0/\sqrt{n}}\right) = \\ &= \Phi\left(\frac{m}{\sigma_0/\sqrt{n}}\right) - \Phi\left(-\frac{m}{\sigma_0/\sqrt{n}}\right) = \Phi\left(\frac{m}{\sigma_0/\sqrt{n}}\right) - \left[1 - \Phi\left(\frac{m}{\sigma_0/\sqrt{n}}\right)\right] = \\ &= 2\Phi\left(\frac{m}{\sigma_0/\sqrt{n}}\right) - 1. \end{aligned}$$

Dall'uguaglianza

$$P_F(\{|\bar{X}_n - \mu_F| < m\}) = 2\Phi\left(\frac{m}{\sigma_0/\sqrt{n}}\right) - 1,$$

che, ricordiamolo, è valida per ogni $F(x) \in \mathcal{F}_{N(\sigma_0)}$, deduciamo che

$$\begin{aligned} P_F(\{|\bar{X}_n - \mu_F| < m\}) &= 1 - \alpha \quad \text{per ogni } F(x) \in \mathcal{F}_{N(\sigma_0)} \quad \Leftrightarrow \\ \Leftrightarrow 2\Phi\left(\frac{m}{\sigma_0/\sqrt{n}}\right) - 1 &= 1 - \alpha \quad \Leftrightarrow \Phi\left(\frac{m}{\sigma_0/\sqrt{n}}\right) = 1 - \alpha/2 \quad \Leftrightarrow \\ \Leftrightarrow \frac{m}{\sigma_0/\sqrt{n}} &= z_{1-\alpha/2} \quad \Leftrightarrow m = z_{1-\alpha/2} \frac{\sigma_0}{\sqrt{n}}. \end{aligned}$$

Quindi possiamo concludere che

$$P_F\left(\left\{|\bar{X}_n - \mu_F| < z_{1-\alpha/2} \frac{\sigma_0}{\sqrt{n}}\right\}\right) = 1 - \alpha \quad \text{per ogni } F(x) \in \mathcal{F}_{N(\sigma_0)}.$$

Siccome l'evento

$$\left\{|\bar{X}_n - \mu_F| < z_{1-\alpha/2} \frac{\sigma_0}{\sqrt{n}}\right\}$$

e l'evento

$$\left\{\bar{X}_n - z_{1-\alpha/2} \frac{\sigma_0}{\sqrt{n}} < \mu_F < \bar{X}_n + z_{1-\alpha/2} \frac{\sigma_0}{\sqrt{n}}\right\}$$

sono eventi equivalenti, possiamo anche concludere che

$$\begin{aligned} P_F\left(\left\{\bar{X}_n - z_{1-\alpha/2} \frac{\sigma_0}{\sqrt{n}} < \mu_F < \bar{X}_n + z_{1-\alpha/2} \frac{\sigma_0}{\sqrt{n}}\right\}\right) &= 1 - \alpha \\ \text{per ogni } F(x) \in \mathcal{F}_{N(\sigma_0)} \end{aligned}$$

e questo significa che

$$\tilde{L}_\alpha = \bar{X}_n - z_{1-\alpha/2} \frac{\sigma_0}{\sqrt{n}} \quad \text{e} \quad \tilde{U}_\alpha = \bar{X}_n + z_{1-\alpha/2} \frac{\sigma_0}{\sqrt{n}} \quad (194)$$

sono gli estremi di un IC per l'ignoto valore atteso $\mu_{F^*} = E_{F^*}(\tilde{X}_i)$ il cui livello di confidenza è uguale a $1 - \alpha$.

Esercizio 9.4. Si supponga che gli incassi giornalieri di un supermercato siano descritti da variabili casuali i.i.d. con distribuzione normale di parametro $\sigma = 0,030$ milioni di euro e di parametro μ ignoto (si confronti con la consegna dell'Esercizio 9.1).

- Per quanti giorni bisogna rilevare l'incasso giornaliero per poter ottenere un intervallo di confidenza per μ che abbia livello di confidenza pari al 99% e che abbia lunghezza minore di 0,005 milioni di euro.
- Si supponga che negli ultimi $n = 36$ giorni il supermercato abbia incassato complessivamente 21,6 milioni di euro. Si calcoli un IC al 99% per l'ignoto valore atteso degli incassi giornalieri.
- Di quanto bisogna aumentare il numero di osservazioni dell'incasso giornaliero per ridurre di un terzo la lunghezza dell'IC calcolato al punto precedente?

Risposte:

- Per quanti giorni bisogna rilevare l'incasso giornaliero per poter ottenere un intervallo di confidenza per μ che abbia livello di confidenza pari al 99% e che abbia lunghezza minore di 0,005 milioni euro.

La lunghezza dell'IC basato sulle variabili casuali

$$\tilde{L}_\alpha = \bar{X}_n - z_{1-\alpha/2} \frac{\sigma_0}{\sqrt{n}} \quad \text{e} \quad \tilde{U}_\alpha = \bar{X}_n + z_{1-\alpha/2} \frac{\sigma_0}{\sqrt{n}}$$

è data da

$$2 \times z_{1-\alpha/2} \frac{\sigma_0}{\sqrt{n}}. \quad (195)$$

Se vogliamo che in corrispondenza del livello di confidenza $1 - \alpha = 0,99 = 99\%$ questa lunghezza sia minore di 0,005 milioni di euro, allora dobbiamo scegliere la numerosità campionaria in modo tale che

$$2 \times z_{0,995} \frac{\sigma_0}{\sqrt{n}} < 0,005 \quad \Leftrightarrow$$

$$\Leftrightarrow n > 2^2 \times z_{0,995}^2 \frac{\sigma_0^2}{0,005^2} = 2^2 \times 2,575^2 \frac{0,030^2}{0,005^2} = 954,81$$

- Si supponga che negli ultimi $n = 36$ giorni il supermercato abbia incassato complessivamente 21,6 milioni di euro. Si calcoli un intervallo di confidenza al 99% per l'ignoto valore atteso degli incassi giornalieri.

Un intervallo di confidenza per μ con livello di confidenza $1 - \alpha = 0,99$ è dato da

$$\bar{x}_n \pm z_{0,995} \frac{\sigma_0}{\sqrt{n}} = 0,6 \pm 2,575 \frac{0,030}{\sqrt{36}} = \begin{cases} 0,613 \\ 0,587. \end{cases}$$

- c) Di quanto bisogna aumentare il numero di osservazioni dell'incasso giornaliero per ridurre di un terzo la lunghezza dell'IC calcolato al punto precedente?

Dalla risposta al quesito a) deduciamo che la lunghezza di un IC al 99% che è basato su un numero generico n di osservazioni è data da (formula (195) con $1 - \alpha = 0,99$)

$$2 \times z_{0,995} \frac{\sigma_0}{\sqrt{n}}.$$

La lunghezza dell'IC calcolato al punto precedente (dove $n = 36$) è dunque data da

$$2 \times z_{0,995} \frac{\sigma_0}{\sqrt{36}}.$$

Affinché

$$2 \times z_{0,995} \frac{\sigma_0}{\sqrt{n}} = \left(1 - \frac{1}{3}\right) \times 2 \times z_{0,995} \frac{\sigma_0}{\sqrt{36}},$$

bisogna quindi avere

$$n = \left(1 - \frac{1}{3}\right)^{-2} \times 36 = 81.$$

Con ulteriori $n' = 81 - 36 = 45$ osservazioni dell'incasso giornaliero potremmo quindi calcolare un IC al 99% la cui lunghezza è di un terzo inferiore rispetto a quella dell'IC che abbiamo calcolato nella risposta al quesito b).

Finora abbiamo visto soltanto come si calcolano IC per il valore atteso delle variabili casuali campionarie quando a priori è noto che la loro distribuzione è normale e che la loro varianza è uguale ad un determinato valore σ_0^2 . Tuttavia, per molte applicazioni queste ipotesi sono troppo restrittive e quindi conviene analizzare anche situazioni dove non sono soddisfatte.

Chiaramente si potrebbero considerare molte situazioni dove le suddette ipotesi non sono soddisfatte, ma per non appesantire troppo la trattazione ne considereremo soltanto due molto generali:

- la situazione dove a priori è soltanto noto che l'ignota distribuzione $F^*(x)$ appartiene alla famiglia di tutte le distribuzioni che sono dotate di varianza;
- la situazione dove a priori è noto che l'ignota distribuzione $F^*(x)$ è una distribuzione bernoulliana, ovvero la situazione dove vogliamo calcolare un IC per il parametro p che identifica una distribuzione bernoulliana.

Per entrambe queste situazioni presenteremo soltanto delle formule per calcolare gli estremi di un IC per il valore atteso (ovvero la probabilità di successo p) senza esporre i dettagli tecnici di una loro deduzione.

Cominciamo con il caso dove a priori è soltanto noto che l'ignota distribuzione $F^*(x)$ appartiene alla famiglia di tutte le distribuzioni che sono dotate di varianza. In quanto segue indicheremo questa famiglia di distribuzioni con \mathcal{F}_{var} . In questo caso si possono ottenere gli estremi di un IC per l'ignoto valore atteso $\mu = \mu_{F^*} = E_{F^*}(\tilde{X}_i)$ attraverso le formule

$$\tilde{L}'_{\alpha,n} = \bar{X}_n - z_{1-\alpha/2} \sqrt{\frac{\tilde{S}_n^2}{n}} \quad \text{e} \quad \tilde{U}'_{\alpha,n} = \bar{X}_n + z_{1-\alpha/2} \sqrt{\frac{\tilde{S}_n^2}{n}} \quad (196)$$

dove \tilde{S}_n^2 è lo stimatore corretto per l'ignota varianza $\sigma^2 = \sigma_{F^*}^2 = var_{F^*}(\tilde{X}_i)$. Si noti che le variabili casuali $\tilde{L}'_{\alpha,n}$ e $\tilde{U}'_{\alpha,n}$ definite nella (196) sono identiche a quelle definite nella (194) tranne per il fatto che al posto dello scarto quadratico medio σ_0 compare lo stimatore $\tilde{S}_n = \sqrt{\tilde{S}_n^2}$.

Con riferimento all'IC delimitato dalle variabili casuali (196) è importante tenere presente che si tratta soltanto di un **IC "asintotico"** e non di un **IC "esatto"** come quello che otteniamo attraverso le formule (194) quando a priori è noto che $F^*(x) \in \mathcal{F}_{N(\sigma_0)}$. La differenza consiste nel fatto che sotto l'ipotesi $F^*(x) \in \mathcal{F}_{var}$ il valore del **livello di confidenza effettivo** dell'IC delimitato dalle variabili casuali (196), ovvero il valore di

$$P_{F^*} \left(\left\{ \tilde{L}'_{\alpha,n} < \mu_{F^*} < \tilde{U}'_{\alpha,n} \right\} \right),$$

non è noto e potrebbe differire dal **livello di confidenza nominale**, ovvero da $1 - \alpha$. Con riferimento alle variabili casuali (196) si può infatti soltanto dimostrare che

$$\lim_{n \rightarrow \infty} P_F \left(\left\{ \tilde{L}'_{\alpha,n} < \mu_F < \tilde{U}'_{\alpha,n} \right\} \right) = 1 - \alpha \quad \text{per ogni } F(x) \in \mathcal{F}_{var}$$

e da questa condizione possiamo soltanto dedurre che

$$P_{F^*} \left(\left\{ \tilde{L}'_{\alpha,n} < \mu_{F^*} < \tilde{U}'_{\alpha,n} \right\} \right) \simeq 1 - \alpha$$

per n sufficientemente elevato, ma l'errore di approssimazione dipende sia dal valore di n sia dall'ignota distribuzione $F^*(x)$. Alcune considerazioni teoriche (delle quali non daremo conto) suggeriscono che parità di n l'errore di approssimazione sia tanto più elevato quanto più l'ignota distribuzione $F^*(x)$ è asimmetrica e/o quanto più una o entrambe le sue code sono "pesanti".

Consideriamo ora invece il caso dove a priori è noto che $F^*(x)$ è una distribuzione bernoulliana, ovvero il caso dove vogliamo calcolare un IC per il parametro p che identifica una distribuzione bernoulliana. In questo caso possiamo considerare come estremi di un **IC asintotico per l'ignoto valore di p** le realizzazioni delle variabili casuali

$$\tilde{L}''_{\alpha} = \tilde{p}_n - z_{1-\alpha/2} \sqrt{\frac{\tilde{p}_n(1-\tilde{p}_n)}{n}} \quad \text{e} \quad \tilde{U}''_{\alpha} = \tilde{p}_n + z_{1-\alpha/2} \sqrt{\frac{\tilde{p}_n(1-\tilde{p}_n)}{n}}.$$

Si noti che anche le variabili casuali \tilde{L}''_{α} e \tilde{U}''_{α} sono molto simili a quelle definite nella (194). Infatti, \tilde{p}_n è solo un modo diverso per indicare la media campionaria \bar{X}_n quando a priori è noto che le variabili casuali campionarie sono bernoulliane, e siccome per variabili casuali campionarie bernoulliane si ha $\sigma^2 = \sigma_p^2 = \text{var}_p(\tilde{X}_i) = p(1-p)$, il prodotto $\tilde{p}_n(1-\tilde{p}_n)$ può essere considerato come uno stimatore per l'ignota varianza delle variabili casuali campionarie.

Ovviamente, trattandosi degli estremi di un IC asintotico, non conosciamo il valore esatto del livello di confidenza effettivo

$$P_{p^*} \left(\left\{ \tilde{L}''_{\alpha,n} < p^* < \tilde{U}''_{\alpha,n} \right\} \right),$$

ma sappiamo soltanto che all'aumentare della numerosità campionaria n il suo valore si avvicina al livello di confidenza nominale $1-\alpha$. Per valutare se l'errore di approssimazione può essere giudicato trascurabile, possiamo fare riferimento al valore di $n\hat{p}_n(1-\hat{p}_n)$. Se $n\hat{p}_n(1-\hat{p}_n) > 5$ possiamo ritenere che la differenza tra il livello di confidenza effettivo e quello nominale sia trascurabile (si noti che questo criterio è analogo a quello che abbiamo già visto per valutare se l'errore di approssimazione nell'approssimazione normale ad una distribuzione binomiale è trascurabile).

Esercizio 9.5. Si supponga che gli incassi giornalieri di un supermercato siano descritti da variabili casuali i.i.d., che l'incasso complessivo degli ultimi $n = 36$ giorni di apertura ammonti a $\sum_{i=1}^{36} x_i = 21,6$ milioni di euro e che la somma dei quadrati di tali incassi ammonti a $\sum_{i=1}^{36} x_i^2 = 14,122$ (si noti che la consegna è identica a quella dell'Esercizio 9.4)

- a) Si calcoli un IC al 98% per l'ignoto valore atteso della distribuzione degli incassi giornalieri.
- b) Qual è il livello di confidenza dell'IC delimitato da $\bar{x}_n \pm 0,040$?

Risposte:

- a) Si calcoli un IC al 98% per l'ignoto valore atteso della distribuzione degli incassi giornalieri.

Dai dati campionari deduciamo che

$$\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i = \frac{21,6}{36} = 0,6$$

e che

$$\begin{aligned} s_n^2 &= \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^2 = \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - n\bar{x}_n^2 \right) \\ &= \frac{1}{36-1} (14,122 - 36 \times 0,6^2) = \frac{1,162}{35} = 0,033. \end{aligned}$$

Gli estremi di un IC per $\mu^* = E_{F^*}(\tilde{X}_i)$ con livello di confidenza che dovrebbe essere prossimo a 98% sono quindi dati da

$$\bar{x}_n \pm z_{0,99} \sqrt{\frac{s_n^2}{n}} = 0,6 \pm 2,33 \sqrt{\frac{0,033}{36}} = \begin{cases} 0,671 \\ 0,529. \end{cases}$$

Il livello di confidenza effettivo di questo IC dipende dall'ignota distribuzione degli incassi giornalieri. Se la distribuzione degli incassi giornalieri è molto asimmetrica e/o ha code molto pesanti, il livello di confidenza effettivo potrebbe essere molto più piccolo di $0,98 = 98\%$.

- b) Qual è il livello di confidenza dell'IC delimitato da $\bar{x}_n \pm 0,040$?

Supponiamo che

$$\bar{x}_n \pm 0,040 = \bar{x}_n \pm z_{1-\alpha/2} \sqrt{\frac{s_n^2}{n}}.$$

In questo caso si ha

$$\begin{aligned} z_{1-\alpha/2} \sqrt{\frac{s_n^2}{n}} = 0,040 &\Leftrightarrow z_{1-\alpha/2} = \frac{0,040}{\sqrt{\frac{s_n^2}{n}}} \\ \Leftrightarrow z_{1-\alpha/2} = \frac{0,040}{\sqrt{\frac{0,033}{36}}} = 1,32 &\Leftrightarrow 1 - \alpha/2 = \Phi(1,32) = 0,9066 \end{aligned}$$

e quindi possiamo concludere che il livello di confidenza nominale dell'IC delimitato da $\bar{x}_n \pm 0,040$ sia dato da $1 - \alpha = 0,8132$. Siccome questo valore è molto lontano da $1 = 100\%$, possiamo dire che i dati campionari non forniscano alcuna garanzia che l'ignoto valore atteso $\mu_{F^*} = E_{F^*}(\tilde{X}_i)$ si trovi all'interno dell'intervallo delimitato da

$$\bar{x}_n \pm 0,040 = 0,60 \pm 0,040 = \begin{cases} 0,640 \\ 0,560 \end{cases}$$

Esercizio 9.6. Per stimare l'ignota proporzione p^* di elettori che sostengono il partito XXX, un sondaggista ha intervistato un campione di $n = 900$ elettori. 225 degli elettori intervistati dichiarano di sostenere il partito XXX.

- Si calcoli un IC al 99% per l'ignota proporzione p^* .
- A quanto ammonta il livello di confidenza dell'IC che ha per estremi i valori $0,25 \pm 0,03$?

Risposte:

- Si calcoli un IC al 99% per l'ignota proporzione p^* .

Con riferimento al campione, la proporzione di elettori che sostengono il partito XXX è data da

$$\hat{p}_n = \frac{225}{900} = 0,25.$$

Gli estremi di un IC con livello di confidenza prossimo a $1 - \alpha = 0,99$ sono dunque dati da

$$\begin{aligned} \hat{p}_n \pm z_{0,995} \sqrt{\frac{\hat{p}_n(1 - \hat{p}_n)}{n}} &= \\ = 0,25 \pm 2,575 \sqrt{\frac{0,25(1 - 0,25)}{900}} &= 0,25 \pm 0,037 = \begin{cases} 0,287 \\ 0,213. \end{cases} \end{aligned}$$

Siccome il valore di

$$n\hat{p}_n(1 - \hat{p}_n) = 900 \times 0,25(1 - 0,25) = 168,75$$

è di gran lunga maggiore di 5, possiamo ritenere che il livello di confidenza di questo IC sia molto prossimo a $1 - \alpha = 0,99$.

- b) A quanto ammonta il livello di confidenza dell'IC che ha per estremi i valori $0,25 \pm 0,03$?

Supponiamo che

$$0,25 \pm 0,03 = 0,25 \pm z_{1-\alpha/2} \sqrt{\frac{0,25(1-0,25)}{900}}.$$

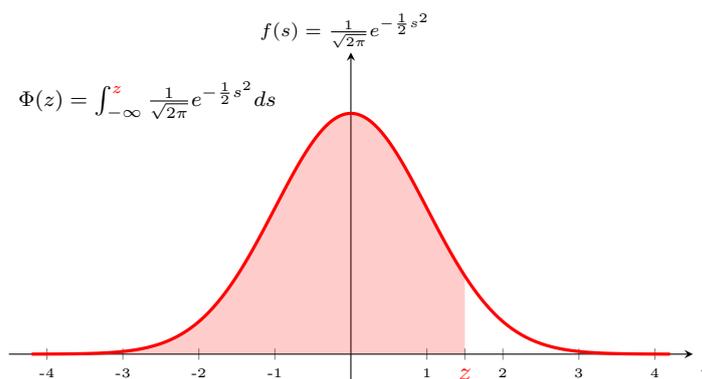
In questo caso si ha

$$\begin{aligned} z_{1-\alpha/2} \sqrt{\frac{0,25(1-0,25)}{900}} = 0,03 &\Leftrightarrow z_{1-\alpha/2} = \frac{0,03}{\sqrt{\frac{0,25(1-0,25)}{900}}} \simeq 2,08 \Rightarrow \\ \Rightarrow 1 - \alpha/2 = \Phi(2,08) = 0,9812 \end{aligned}$$

e il livello di confidenza dell'IC delimitato dagli estremi $0,25 \pm 0,03$ è quindi dato da

$$1 - \alpha = 1 - 2(1 - 0,9812) = 0,9624 = 96,24\%.$$

Tavola della distribuzione normale standard



z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,10	0,53983	0,54380	0,54776	0,55172	0,55567	0,55962	0,56356	0,56749	0,57142	0,57535
0,20	0,57926	0,58317	0,58706	0,59095	0,59483	0,59871	0,60257	0,60642	0,61026	0,61409
0,30	0,61791	0,62172	0,62552	0,62930	0,63307	0,63683	0,64058	0,64431	0,64803	0,65173
0,40	0,65542	0,65910	0,66276	0,66640	0,67003	0,67364	0,67724	0,68082	0,68439	0,68793
0,50	0,69146	0,69497	0,69847	0,70194	0,70540	0,70884	0,71226	0,71566	0,71904	0,72240
0,60	0,72575	0,72907	0,73237	0,73565	0,73891	0,74215	0,74537	0,74857	0,75175	0,75490
0,70	0,75804	0,76115	0,76424	0,76730	0,77035	0,77337	0,77637	0,77935	0,78230	0,78524
0,80	0,78814	0,79103	0,79389	0,79673	0,79955	0,80234	0,80511	0,80785	0,81057	0,81327
0,90	0,81594	0,81859	0,82121	0,82381	0,82639	0,82894	0,83147	0,83398	0,83646	0,83891
1,00	0,84134	0,84375	0,84614	0,84849	0,85083	0,85314	0,85543	0,85769	0,85993	0,86214
1,10	0,86433	0,86650	0,86864	0,87076	0,87286	0,87493	0,87698	0,87900	0,88100	0,88298
1,20	0,88493	0,88686	0,88877	0,89065	0,89251	0,89435	0,89617	0,89796	0,89973	0,90147
1,30	0,90320	0,90490	0,90658	0,90824	0,90988	0,91149	0,91309	0,91466	0,91621	0,91774
1,40	0,91924	0,92073	0,92220	0,92364	0,92507	0,92647	0,92785	0,92922	0,93056	0,93189
1,50	0,93319	0,93448	0,93574	0,93699	0,93822	0,93943	0,94062	0,94179	0,94295	0,94408
1,60	0,94520	0,94630	0,94738	0,94845	0,94950	0,95053	0,95154	0,95254	0,95352	0,95449
1,70	0,95543	0,95637	0,95728	0,95818	0,95907	0,95994	0,96080	0,96164	0,96246	0,96327
1,80	0,96407	0,96485	0,96562	0,96638	0,96712	0,96784	0,96856	0,96926	0,96995	0,97062
1,90	0,97128	0,97193	0,97257	0,97320	0,97381	0,97441	0,97500	0,97558	0,97615	0,97670
2,00	0,97725	0,97778	0,97831	0,97882	0,97932	0,97982	0,98030	0,98077	0,98124	0,98169
2,10	0,98214	0,98257	0,98300	0,98341	0,98382	0,98422	0,98461	0,98500	0,98537	0,98574
2,20	0,98610	0,98645	0,98679	0,98713	0,98745	0,98778	0,98809	0,98840	0,98870	0,98899
2,30	0,98928	0,98956	0,98983	0,99010	0,99036	0,99061	0,99086	0,99111	0,99134	0,99158
2,40	0,99180	0,99202	0,99224	0,99245	0,99266	0,99286	0,99305	0,99324	0,99343	0,99361
2,50	0,99379	0,99396	0,99413	0,99430	0,99446	0,99461	0,99477	0,99492	0,99506	0,99520
2,60	0,99534	0,99547	0,99560	0,99573	0,99585	0,99598	0,99609	0,99621	0,99632	0,99643
2,70	0,99653	0,99664	0,99674	0,99683	0,99693	0,99702	0,99711	0,99720	0,99728	0,99736
2,80	0,99744	0,99752	0,99760	0,99767	0,99774	0,99781	0,99788	0,99795	0,99801	0,99807
2,90	0,99813	0,99819	0,99825	0,99831	0,99836	0,99841	0,99846	0,99851	0,99856	0,99861
3,00	0,99865	0,99869	0,99874	0,99878	0,99882	0,99886	0,99889	0,99893	0,99896	0,99900
3,10	0,99903	0,99906	0,99910	0,99913	0,99916	0,99918	0,99921	0,99924	0,99926	0,99929
3,20	0,99931	0,99934	0,99936	0,99938	0,99940	0,99942	0,99944	0,99946	0,99948	0,99950
3,30	0,99952	0,99953	0,99955	0,99957	0,99958	0,99960	0,99961	0,99962	0,99964	0,99965
3,40	0,99966	0,99968	0,99969	0,99970	0,99971	0,99972	0,99973	0,99974	0,99975	0,99976
3,50	0,99977	0,99978	0,99978	0,99979	0,99980	0,99981	0,99981	0,99982	0,99983	0,99983