

Lo schema dell'urna di Polya per la distribuzione delle insolvenze.

Raffaella Calabrese.

25 settembre 2006

1 Introduzione.

Lo schema dell'urna di Polya é stato spesso impiegato per rappresentare sia fenomeni aventi caratteristiche simili a quelle del contagio delle malattie, per tale motivo il suddetto schema é noto anche come schema del contagio, oppure per rappresentare fenomeni empirici con caratteristiche simili a quelle dell'immunitá da malattie.

In uno schema probabilistico con correlazione lineare non nulla tra le prove di ogni serie, é possibile applicare lo schema dell'urna di Polya per rappresentare il numero di successi che si verificano in ciascuna serie. Si analizzerá un caso particolare in cui le k serie di uno schema probabilistico sono rappresentate da k urne aventi la medesima composizione di palline, per ciascuna delle quali il numero di palline s reimmesse nell'urna rimane costante da serie a serie. Si mostra in particolare che tale schema probabilistico appartiene alla classe degli schemi con correlazione uniforme tra le prove di ciascuna serie ed indipendenza tra le serie.

Confrontando infine le funzioni di ripartizione delle variabili casuali che rappresentano il numero di successi rilevati nello schema dell'urna di Polya, mantenendo costante sia il numero delle prove sia la probabilitá di successo e permettendo al parametro s di variare, si dimostra che le variabili casuali caratterizzate da un dato numero di palline s reimmesse nell'urna dominano strettamente secondo l'ordinamento convesso quelle variabili casuali caratterizzate da un numero di palline reimmesse nell'urna inferiore.

Lo schema dell'urna di Polya puó essere applicato nel rischio di credito

per rappresentare il fenomeno delle insolvenze, consentendo di valutare la dispersione di un portafoglio di crediti.

2 Lo schema dell'urna di Polya.

Un modello molto impiegato per rappresentare alcuni fenomeni dicotomici in cui esiste dipendenza tra le diverse prove consiste nello *schema dell'urna*: si consideri un'urna che contiene b palline nere (successo) e r palline rosse (insuccesso). Si estrae casualmente una pallina dall'urna, che viene in seguito reimmessa nell'urna insieme a s palline dello stesso colore della pallina estratta ed a d palline dell'altro colore. Si effettua in seguito una nuova estrazione dall'urna, che adesso contiene $r + b + s + d$ palline e seguendo tale procedimento si effettuano successivamente n estrazioni. I parametri s e d sono dei numeri interi arbitrari, che possono assumere anche valori negativi, ma in questo caso le estrazioni potrebbero terminare per mancanza di palline nell'urna.

Un particolare schema di urna con $d = 0$ é noto come *urna di Polya*, proposto in un articolo da Eggenberg e Polya nel 1923. In questo schema, dunque, se $s > 0$, dopo ciascuna estrazione aumenta il numero delle palline del colore estratto, mentre il numero delle palline dell'altro colore rimane invariato, per tale motivo l'urna di Polya é impiegata per rappresentare quei fenomeni che possiedono caratteristiche simili a quelle del contagio delle malattie.

Per comprendere meglio le proprietà di tale schema di urna conviene considerare le seguenti probabilità. L'evento 'pallina nera alla prima estrazione', indicato con b_1 , ha probabilità pari a $P(b_1) = b/(b+r)$. Dato che nella prima estrazione é uscita una pallina nera, la probabilità che nella seconda venga estratta una pallina nera é pari a

$$P(b_2|b_1) = \frac{b+s}{b+r+s},$$

nel caso in cui $s > 0$ tale probabilità risulta maggiore rispetto a quella di estrarre una pallina nera alla seconda estrazione, sapendo che nella prima sia uscita una pallina dell'altro colore

$$P(b_2|r_1) = \frac{b}{b+r+s},$$

che a sua volta é minore della probabilità che la prima pallina estratta sia nera.

Si noti che in tale schema la probabilità di estrarre una pallina nera rimane costantemente pari a $b/(b+r)$ in ogni estrazione, infatti per il teorema delle probabilità totali risulta

$$\begin{aligned} P(b_2) &= P(b_2|b_1)P(b_1) + P(b_2|r_1)P(r_1) = \frac{b+s}{(b+r+s)} \frac{b}{b+r} + \frac{b}{(b+r+s)} \frac{r}{(b+r)} \\ &= \frac{b(b+r+s)}{(b+r+s)(b+r)} = \frac{b}{b+r}. \end{aligned}$$

La probabilità che, per esempio, in 4 estrazioni si verifichino nelle prime due successi e nelle rimanenti due insuccessi é data da

$$\begin{aligned} P(b_1 b_2 r_3 r_4) &= P(b_1)P(b_2|b_1)P(r_3|b_1, b_2)P(r_4|b_1, b_2, r_3) \\ &= \frac{b}{(b+r)} \frac{b+s}{(b+r+s)} \frac{r}{(b+r+2s)} \frac{r+s}{(b+r+3s)}. \end{aligned}$$

Effettuando tale calcolo per ogni possibile allineamento delle 4 realizzazioni contenente due successi e due insuccessi, in ordine diverso, si ottiene sempre la stessa probabilità. Dal calcolo combinatorio risulta che il numero di tali allineamenti é pari al coefficiente binomiale di 4 su 2.

Per generalizzare il precedente risultato si consideri la variabile casuale X che rappresenta il numero di successi in n estrazioni. Dal momento che il numero dei possibili allineamenti dei risultati di n estrazioni che presentano k successi e $n-k$ insuccessi é pari al coefficiente binomiale di n su k e poiché ciascuno di tali allineamenti ha probabilità

$$\frac{b(b+s)\dots[b+(k-1)s]r(r+s)\dots[r+(n-k-1)s]}{(b+r)(b+r+s)\dots[b+r+(n-1)s]} \quad (2.1)$$

risulta dunque

$$P\{X = k\} = \binom{n}{k} \frac{b(b+s)\dots[b+(k-1)s]r(r+s)\dots[r+(n-k-1)s]}{(b+r)(b+r+s)\dots[b+r+(n-1)s]} \quad (2.2)$$

con $k = 0, 1, \dots, n$.

Esistono diverse espressioni per trascrivere in modo piú compatto la distribuzione di probabilità della variabile casuale X . Considerando il fattoriale ascendente

$$x^{[r]} = x(x+1)\dots(x+r-1),$$

Johnson e Kotz (1977), ad esempio, definiscono $\alpha = b/s$ e $\beta = r/s$, ottenendo

$$\begin{aligned} P\{X = k\} &= \binom{n}{k} \frac{\alpha(\alpha+1)\dots[\alpha+(k-1)]\beta(\beta+1)\dots[\beta+(n-k-1)]}{(\alpha+\beta)(\alpha+\beta+1)\dots[\alpha+\beta+(n-1)]} \\ &= \binom{n}{k} \frac{\alpha^{[k]}\beta^{[n-k]}}{(\alpha+\beta)^{[n]}}. \end{aligned} \quad (2.3)$$

Moltiplicando per $(\alpha-1)!$ e $(\beta-1)!$ sia numeratore che denominatore della prima espressione presente nella (2.3), si ricava

$$\begin{aligned} P\{X = k\} &= \binom{n}{k} \frac{(\alpha+k-1)\dots\alpha(\alpha-1)!(\beta+n-k-1)\dots(\beta+1)\beta(\beta-1)!}{(\beta-1)!(\alpha-1)!(\alpha+\beta)(\alpha+\beta+1)\dots[\alpha+\beta+(n-1)]} \\ &= \binom{n}{k} \frac{(\alpha+k-1)!(\beta+n-k-1)!}{(\beta-1)!(\alpha-1)![\alpha+\beta+(n-1)]\dots(\alpha+\beta+1)(\alpha+\beta)}, \end{aligned}$$

moltiplicando inoltre sia numeratore che denominatore per $(\alpha+\beta-1)!$ si ottiene

$$P\{X = k\} = \binom{n}{k} \frac{(\alpha+k-1)!(\beta+n-k-1)!(\alpha+\beta-1)!}{(\beta-1)!(\alpha-1)!(\alpha+\beta+n-1)!}. \quad (2.4)$$

Per comprendere il legame tra la funzione di probabilità di un'urna di Polya ed alcune note distribuzioni conviene ricondurre la precedente espressione al rapporto tra coefficienti binomiali

$$\begin{aligned} P\{X = k\} &= \frac{(-1)^k \frac{(\alpha+k-1)!}{k!(\alpha-1)!} (-1)^{n-k} \frac{(\beta+n-k-1)!}{(n-k)!(\beta-1)!}}{(-1)^n \frac{(\alpha+\beta+n-1)!}{n!(\alpha+\beta-1)!}} \\ &= \frac{(-1)^k \binom{\alpha+k-1}{k} (-1)^{n-k} \binom{\beta+n-k-1}{n-k}}{(-1)^n \binom{\alpha+\beta+n-1}{n}}. \end{aligned}$$

Sapendo che ¹

$$\binom{-n}{u} = (-1)^u \binom{n+u-1}{u} \quad \text{per } n < 0 \quad (2.5)$$

¹Si veda Riordan (1980), pp.5.

si ottiene che

$$P\{X = k\} = \frac{\binom{-\alpha}{k} \binom{-\beta}{n-k}}{\binom{-(\alpha+\beta)}{n}} = \frac{\binom{-b/s}{k} \binom{-r/s}{n-k}}{\binom{-(b+r)/s}{n}}, \quad (2.6)$$

con $k = 0, 1, \dots, n$. Dalla equazione (2.5) si deduce che la precedente espressione può essere applicata solamente nel caso in cui sia $s < 0$.

Considerando $s < 0$, esiste l'eventualità di dover cessare le estrazioni per assenza di palline nell'urna. Questo significa che bisogna considerare come supporto della variabile casuale X solamente quei valori di k tali per cui i coefficienti binomiali, presenti al numeratore nell'espressione (2.6), sono diversi da zero. Questo si verifica se è soddisfatto il seguente sistema

$$\begin{cases} s < 0 \\ k \leq -\frac{b}{s} \\ n - k \leq -\frac{r}{s}. \end{cases}$$

Tale osservazione permette di trascrivere con maggior precisione la distribuzione di probabilità (2.6), che per quanto detto in precedenza può essere applicata solamente nel caso in cui sia $s < 0$, ottenendo dunque

$$P\{X = k\} = \begin{cases} \frac{\binom{-b/s}{k} \binom{-r/s}{n-k}}{\binom{-(b+r)/s}{n}} & \max(0, n + \frac{r}{s}) \leq k \leq \min(n, -\frac{b}{s}) \\ 0 & \text{altrove.} \end{cases} \quad (2.7)$$

Le funzioni di probabilità del numero di successi rilevati in n estrazioni seguendo lo schema dell'urna di Polya coincidono con alcune note distribuzioni per particolari valori di s . Considerando $s = -1$, questo significa che si effettuano estrazioni senza reimmissione, si riconosce nell'espressione (2.7) la funzione di probabilità di una variabile casuale ipergeometrica di parametri b , $r + b$ e n

$$P\{X = k\} = \begin{cases} \frac{\binom{b}{k} \binom{r}{n-k}}{\binom{b+r}{n}} & \max(0, n - r) \leq k \leq \min(n, b) \\ 0 & \text{altrove.} \end{cases}$$

Per $s = 0$ le estrazioni sono con reimmissione e dall'equazione (2.2) si ottiene la seguente distribuzione di probabilità

$$P\{X = k\} = \binom{n}{k} \left(\frac{b}{b+r}\right)^k \left(\frac{r}{b+r}\right)^{n-k}, \quad (2.8)$$

con $k = 0, 1, \dots, n$, che coincide con quella di una variabile casuale binomiale di parametri n e $p = b/(b+r)$. Sempre dall'equazione (2.2) per $b = r = s$ si ricava

$$P\{X = k\} = \frac{n!}{k!(n-k)!} \frac{b^k (k!) b^{n-k} (n-k)!}{b^n (n+1)!} = \frac{1}{n+1},$$

con $k = 0, 1, \dots, n$, la distribuzione di una variabile casuale rettangolare discreta di parametro $p = 1/(n+1)$.

Per calcolare i momenti dall'origine della variabile casuale X , avente distribuzione di probabilità data dall'equazione (2.3), conviene inizialmente determinare il momento fattoriale di ordine u . Definendo il fattoriale discendente

$$x^{(r)} = x(x-1)\dots(x-r+1),$$

si ottiene

$$\mu'_{(u)}(X) = E[X^{(u)}] = \frac{\sum_{k=u}^n \left[k^{(u)} \binom{n}{k} \alpha^{[k]} \beta^{[n-k]} \right]}{(\alpha + \beta)^{[n]}}. \quad (2.9)$$

Si considerino i singoli fattori della precedente sommatoria

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

$$\alpha^{[k]} = \alpha(\alpha+1)\dots(\alpha+u-1)(\alpha+u)\dots(\alpha+k-1) = \alpha^{[u]}(\alpha+u)^{[k-u]}$$

$$\beta^{[n-k]} = \beta^{[(n-u)-(k-u)].}$$

Sostituendo tali risultati nell'equazione (2.9) si ottiene

$$\mu'_{(u)}(X) = \frac{n^{(u)} \alpha^{[u]} \sum_{k=u}^n \left[\binom{n-u}{k-u} (\alpha+u)^{[k-u]} \beta^{[(n-u)-(k-u)]} \right]}{(\alpha + \beta)^{[n]}}.$$

Effettuando la seguente trasformazione $j = k - u$ nella precedente sommatoria ed adattando il binomio di Newton alla somma di fattoriali ascendenti (Riordan (1980), pp.9), si ottiene

$$\sum_{j=0}^{n-u} \left[\binom{n-u}{j} (\alpha+u)^{[j]} \beta^{[(n-u)-j]} \right] = (\alpha+\beta+u)^{[n-u]},$$

e dunque

$$\begin{aligned} \mu'_{(u)}(X) &= \frac{n^{(u)} \alpha^{[u]} (\alpha+\beta+u)(\alpha+\beta+u+1)\dots(\alpha+\beta+n-1)}{(\alpha+\beta)(\alpha+\beta+1)\dots(\alpha+\beta+u-1)(\alpha+\beta+u)\dots(\alpha+\beta+n-1)} \\ &= \frac{n^{(u)} \alpha^{[u]}}{(\alpha+\beta)^{[u]}}. \end{aligned} \tag{2.10}$$

Per $u = 1$ si ottiene

$$\mu'_1 = \mu_X = E(X) = \frac{n\alpha}{\alpha+\beta} = n \frac{b}{b+r}$$

e per $u = 2$

$$\mu'_{[2]} = E[X(X-1)] = \frac{n(n-1)\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)} = \frac{n(n-1)b(b+s)}{(r+b)(r+b+s)},$$

ricavando che²

$$\begin{aligned} \mu_2 = V(X) &= \mu'_{[2]} + \mu_X - \mu_X^2 = \frac{n(n-1)\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)} + \frac{n\alpha}{\alpha+\beta} - \frac{n^2\alpha^2}{(\alpha+\beta)^2} \\ &= \frac{n\alpha}{\alpha+\beta} \left[\frac{(n-1)(\alpha+1)}{\alpha+\beta+1} + 1 - \frac{n\alpha}{\alpha+\beta} \right] \\ &= \frac{n\alpha}{\alpha+\beta} \left[\frac{(n-1)(\alpha+1)(\alpha+\beta) + (\alpha+\beta)(\alpha+\beta+1) - n\alpha(\alpha+\beta+1)}{(\alpha+\beta+1)(\alpha+\beta)} \right] \\ &= \frac{n\alpha}{\alpha+\beta} \left[\frac{(\alpha+\beta)(n\alpha+n+\beta) - n\alpha(\alpha+\beta+1)}{(\alpha+\beta+1)(\alpha+\beta)} \right] \\ &= \frac{n\alpha}{\alpha+\beta} \left[\frac{n\alpha^2 + n\alpha\beta + n\alpha + n\beta + \alpha\beta + \beta^2 - n\alpha^2 - n\alpha\beta - n\alpha}{(\alpha+\beta+1)(\alpha+\beta)} \right] \\ &= \frac{n\alpha\beta}{(\alpha+\beta)^2} \frac{n+\alpha+\beta}{\alpha+\beta+1} = n \frac{b}{b+r} \frac{r}{b+r} \frac{b+r+ns}{b+r+s}. \end{aligned} \tag{2.11}$$

²Per la relazione tra momenti fattoriali e momenti dall'origine si veda Johnson, Kemp e Kotz (2005), pp. 53.

Considerando $s = 0$, si ottiene la varianza di una variabile casuale binomiale di parametri n e $p = b/(b+r)$. Questo significa che il rapporto $(b+r+ns)/(b+r+s)$, presente nell'espressione della varianza, rappresenta un fattore di correzione. Infatti per $s = -1$ si ottiene

$$V(X) = n \frac{b}{b+r} \frac{r}{b+r} \frac{b+r-n}{b+r-1}$$

la varianza di una variabile casuale ipergeometrica di parametri b , $r+b$ e n , avente fattore di correzione pari a $(b+r-n)/(b+r-1)$. Infine nel caso in cui risulta $s = r = b$, si ricava

$$V(X) = \frac{n^2 + 2n}{12} = \frac{(n+1)^2 - 1}{12}$$

la varianza di una variabile casuale discreta uniforme di parametro $p = 1/(n+1)$.

Polya propose nei suoi articoli un'espressione diversa della funzione di probabilità dello schema di contagio che da lui prese il nome. Impiegando la funzione gamma e sapendo che, per $m > 0$ intero, vale la formula di Eulero

$$\Gamma(m) = \int_0^\infty x^{m-1} e^{-x} dx = (m-1)!$$

l'espressione (2.4) diventa

$$P\{X = k\} = \binom{n}{k} \frac{\Gamma(\alpha+k)\Gamma(\beta+n-k)\Gamma(\alpha+\beta)}{\Gamma(\beta)\Gamma(\alpha)\Gamma(\alpha+\beta+n)}. \quad (2.12)$$

Dal momento che esiste la seguente relazione tra la funzione beta e la funzione gamma

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

l'espressione (2.12) diventa

$$P\{X = k\} = \binom{n}{k} \frac{B(\alpha+k, \beta+n-k)}{B(\alpha, \beta)}. \quad (2.13)$$

La precedente equazione permette di supporre l'esistenza di un legame tra la funzione di probabilità di un'urna di Polya e la distribuzione di probabilità

beta. Sia infatti A_i , con $i = 1, 2, \dots, n$, la variabile casuale indicatore così definita

$$A_i = \begin{cases} 1 & \text{se la } i\text{-ma pallina estratta é nera,} \\ 0 & \text{altrimenti.} \end{cases}$$

Si noti che nello schema dell'urna di Polya le variabili casuali A_i sono dipendenti, dal momento che ad ogni estrazione cambia la composizione delle palline nell'urna in base ai risultati delle precedenti estrazioni e, quindi, cambia anche la probabilità condizionata di estrarre una pallina nera, date le precedenti realizzazioni. Per tale motivo conviene considerare la probabilità P di estrarre una pallina nera come una variabile casuale. Per un dato valore assunto dalla variabile casuale $P = p$, le variabili A_i sono indipendenti e quindi $X = \sum_{i=1}^n A_i$ si distribuisce come una binomiale di parametri n e p . Supponendo che P sia una variabile casuale beta con parametri α e β ³, risulta che

$$\begin{aligned} P\{X = k\} &= \int_0^1 P\{X = k, P = p\} dp = \int_0^1 P\{X = k | P = p\} P\{P = p\} dp \\ &= \int_0^1 \binom{n}{k} p^k (1-p)^{n-k} \frac{p^{\alpha-1} (1-p)^{\beta-1}}{B(\alpha, \beta)} dp. \end{aligned}$$

Sapendo che

$$B(a, b) = \int_0^1 t^{a-1} (1-t)^{b-1} dt \quad a > 0, b > 0$$

si ottiene

$$P\{X = k\} = \binom{n}{k} \frac{B(\alpha + k, \beta + n - k)}{B(\alpha, \beta)}$$

che coincide con l'equazione (2.13). Questo significa che la variabile X , che rappresenta il numero di palline nere estratte seguendo lo schema dell'urna di Polya in n estrazioni, é una variabile casuale miscuglio. In particolare é possibile ricavare la funzione di probabilità di X come distribuzione marginale della variabile casuale bidimensionale (X, P) , data dal prodotto tra una distribuzione binomiale, di parametri n e p , e una funzione di densità beta, di parametri $\alpha = b/s$ e $\beta = r/s$, dove quest'ultima rappresenta la distribuzione di probabilità del parametro P della binomiale.

³La distribuzione di P coincide con la distribuzione limite della frequenza relativa X/n per $n \rightarrow \infty$, con b, r e s costanti. Si é dimostrato che il momento di ordine h della frequenza relativa X/n per $n \rightarrow \infty$ coincide con quello di una distribuzione beta con parametri b/s e r/s (si veda Johnson e Kotz (1977), pp. 376-377).

3 Lo schema dell'urna di Polya visto come schema probabilistico con correlazione uniforme tra le prove di ciascuna serie.

Sia X la variabile casuale che rappresenta il numero di palline nere estratte seguendo lo schema dell'urna di Polya in n estrazioni. Si é dimostrato in precedenza che la probabilitá di estrarre una pallina nera da un'urna contenente b palline nere e r palline rosse rimane costante e pari a $b/(b+r)$ in ogni estrazione. Questo significa che é possibile rappresentare la variabile casuale X come somma di n variabili casuali indicatore identicamente distribuite

$$X = \sum_{i=1}^n A_i \quad \text{con } A_i = \begin{cases} 1 & \frac{b}{b+r} \\ 0 & \frac{r}{b+r} \end{cases} \text{ per ogni } i = 1, 2, \dots, n.$$

L'aspettativa della variabile casuale X risulta dunque

$$E(X) = E\left(\sum_{i=1}^n A_i\right) = \sum_{i=1}^n E(A_i) = n \frac{b}{b+r}.$$

Poiché le variabili casuali indicatore A_i , con $i = 1, 2, \dots, n$, sono dipendenti, la varianza della variabile casuale X risulta

$$V(X) = V\left(\sum_{i=1}^n A_i\right) = \sum_{i=1}^n V(A_i) + \sum_{i=1}^n \sum_{u=1, u \neq i}^n \text{cov}(A_i, A_u). \quad (3.1)$$

Per determinare la covarianza tra due generici indicatori A_i e A_u , con $i = 1, 2, \dots, n$, $u = 1, 2, \dots, n$ e $i \neq u$, é utile considerare il risultato fornito dal seguente teorema.

Teorema 3.1. *Da un'urna contenente b palline nere e r rosse si effettuano n estrazioni seguendo lo schema dell'urna di Polya, reimmettendo $s+1$ palline del colore estratto nell'urna dopo ogni estrazione. Si rappresenta l'esito di ciascuna estrazione mediante una variabile casuale indicatore*

$$A_i = \begin{cases} 1 & \frac{b}{b+r} \\ 0 & \frac{r}{b+r} \end{cases} \quad \text{per ogni } i = 1, 2, \dots, n.$$

La covarianza tra due variabili casuali A_i e A_u , con $i = 1, 2, \dots, n$ e $u = 1, 2, \dots, n$ e $i \neq u$, é costante per ogni coppia di estrazioni considerate e risulta pari a

$$\text{cov}(A_i, A_u) = \frac{b}{b+r} \frac{r}{b+r} \frac{s}{b+r+s} \quad (3.2)$$

per ogni $i = 1, 2, \dots, n$ e $u = 1, 2, \dots, n$, con $i \neq u$.

Dimostrazione. É noto che

$$\text{cov}(A_i, A_u) = E(A_i \cdot A_u) - E(A_i)E(A_u). \quad (3.3)$$

Essendo il valore atteso $E(A_i \cdot A_u)$ pari alla somma dei valori riportati nel seguente prospetto

$A_i \backslash A_u$	1	0
1	$P\{(A_i = 1) \cap (A_u = 1)\}$	0
0	0	0

e sapendo che le variabili casuali A_i e A_u sono identicamente distribuite, si ricava dall'equazione (3.3)

$$\begin{aligned} \text{cov}(A_i, A_u) &= P\{(A_i = 1) \cap (A_u = 1)\} - E(A_i)E(A_u) \\ &= P\{(A_i = 1) \cap (A_u = 1)\} - E^2(A_i) \end{aligned} \quad (3.4)$$

per $i = 1, 2, \dots, n$ e $u = 1, 2, \dots, n$, con $i \neq u$.

L'unico fattore non noto nella precedente equazione é rappresentato dalla probabilità congiunta $P\{(A_i = 1) \cap (A_u = 1)\}$, per il calcolo della quale conviene considerare la u -ma estrazione come l'estrazione di ordine $i+t$, con $1 \leq t \leq n-i$. Tale assunzione non limita l'applicabilità del teorema poiché risulta

$$\text{cov}(A_i, A_u) = \text{cov}(A_u, A_i).$$

Applicando il teorema delle probabilità totali si ottiene che

$$\begin{aligned} P\{(A_i = 1) \cap (A_u = 1)\} &= P\{(A_i = 1) \cap (A_{i+t} = 1)\} \\ &= \sum_{x=0,1} P\{(A_1 = x) \cap \dots \cap (A_i = 1) \cap \dots \cap (A_{i+t-1} = x) \cap (A_{i+t} = 1)\}. \end{aligned}$$

Da quanto affermato nel precedente paragrafo risulta che ogni possibile allineamento delle $i+t$ realizzazioni contenente il medesimo numero di successi presenta la stessa probabilità. Il numero di tali allineamenti é pari al coefficiente binomiale di $i+t-2$ sul numero di successi meno due rilevati nelle $i+t$ realizzazioni. Considerando l'espressione (2.3) della distribuzione di

probabilità del numero di successi rilevati in uno schema dell'urna di Polya si ricava che

$$\begin{aligned}
P\{(A_i = 1) \cap (A_{i+t} = 1)\} &= \sum_{k=2}^{i+t} \binom{i+t-2}{k-2} \frac{\alpha^{[k]} \beta^{[i+t-k]}}{(\alpha + \beta)^{[i+t]}} \\
&= \frac{1}{(\alpha + \beta)^{[i+t]}} \sum_{z=0}^{i+t-2} \binom{i+t-2}{z} \alpha^{[z+2]} \beta^{[i+t-z-2]}.
\end{aligned} \tag{3.5}$$

Nell'ultima uguaglianza si é effettuato un cambio di variabile considerando $z = k - 2$. Dalla definizione di fattoriale ascendente si deduce che

$$\alpha^{[z+2]} = \alpha(\alpha + 1) \dots (\alpha + z + 1) = \alpha(\alpha + 1)(\alpha + 2)^{[z]}.$$

Sostituendo il precedente risultato nell'equazione (3.5) si ottiene

$$P\{(A_i = 1) \cap (A_{i+t} = 1)\} = \frac{\alpha(\alpha + 1)}{(\alpha + \beta)^{[i+t]}} \sum_{z=0}^{i+t-2} \binom{i+t-2}{z} (\alpha + 2)^{[z]} \beta^{[i+t-z-2]}.$$

Il risultato della precedente sommatoria si ricava adattando il binomio di Newton alla somma di fattoriali ascendenti (Riordan (1980), pp.9), ottenendo dunque

$$\begin{aligned}
P\{(A_i = 1) \cap (A_{i+t} = 1)\} &= \frac{\alpha(\alpha + 1)}{(\alpha + \beta)^{[i+t]}} (\alpha + \beta + 2)^{[i+t-2]} \\
&= \frac{\alpha(\alpha + 1)}{(\alpha + \beta)(\alpha + \beta + 1)(\alpha + \beta + 2)^{[i+t-2]}} (\alpha + \beta + 2)^{[i+t-2]} \\
&= \frac{\alpha}{\alpha + \beta} \frac{\alpha + 1}{\alpha + \beta + 1} = \frac{b}{b + r} \frac{b + s}{b + r + s}.
\end{aligned}$$

Sostituendo il risultato ottenuto nell'equazione (3.4) si ricava

$$\begin{aligned}
cov(A_i, A_u) &= P\{(A_i = 1) \cap (A_u = 1)\} - E^2(A_i) \\
&= \frac{b}{b + r} \frac{b + s}{b + r + s} - \left(\frac{b}{b + r} \right)^2 \\
&= \frac{b}{(b + r)} \left[\frac{b^2 + sb + rb + sr - b^2 - rb - sb}{(b + r)(b + r + s)} \right] \\
&= \frac{b}{b + r} \frac{r}{b + r} \frac{s}{b + r + s}.
\end{aligned}$$

□

Si noti che per $s = -1$ l'equazione (3.2) diventa

$$\text{cov}(A_i, A_u) = -\frac{b}{b+r} \frac{r}{b+r} \frac{1}{b+r-1}$$

la covarianza tra due variabili casuali indicatore che rappresentano l'esito di due estrazioni effettuate senza reimmissione da un'urna contenente $b+r$ palline.

Calcolando il coefficiente di correlazione lineare tra due generiche variabili indicatore A_i e A_u si ottiene

$$r(A_i, A_u) = \frac{\text{cov}(A_i, A_u)}{\sqrt{V(A_i)V(A_u)}} = \frac{\text{cov}(A_i, A_u)}{V(A_i)} \quad (3.6)$$

per $i = 1, 2, \dots, n$ e $u = 1, 2, \dots, n$, con $i \neq u$

poiché le variabili casuali indicatore sono identicamente distribuite. Sostituendo nella precedente equazione il risultato fornito dal teorema (3.1) si ricava

$$\rho = r(A_i, A_u) = \frac{brs}{(b+r)^2(b+r+s)} \frac{(b+r)^2}{br} = \frac{s}{b+r+s} \quad (3.7)$$

per ogni $i = 1, 2, \dots, n$ e $u = 1, 2, \dots, n$, con $i \neq u$.

Poiché l'espressione presente nella precedente equazione non dipende dalle estrazioni (i -ma e u -ma) considerate, si deduce che tra ogni coppia di variabili casuali indicatore esiste una correlazione lineare uniforme. Essendo inoltre le variabili casuali indicatore identicamente distribuite, la probabilità di successo, pari a $b/(b+r)$, è costante da estrazione ad estrazione.

Questo significa che, se si considerano k urne da ciascuna delle quali si effettuano n_j estrazioni seguendo lo schema dell'urna di Polya, si ottiene uno schema probabilistico con indipendenza tra le serie. Mantenendo inoltre costante da serie a serie sia la probabilità di successo $p = b/(b+r)$ sia il numero di palline s reimmesse nell'urna, per quanto affermato in precedenza si ottiene dunque uno schema probabilistico con dipendenza lineare uniforme tra le prove di ciascuna serie ed indipendenza tra le serie⁴.

Tale considerazione permette di impiegare i risultati dello schema probabilistico con dipendenza (lineare) uniforme tra le prove di ciascuna serie ed

⁴Tale schema probabilistico è stato proposto da Zenga (2003) ed analizzato in seguito da Calabrese (2006).

indipendenza tra le serie nell'analisi di uno schema in cui si effettuano n_j estrazioni, seguendo lo schema dell'urna di Polya, da k urne aventi la medesima composizione di palline e considerando lo stesso numero di palline s reimmesse nell'urna. Si ottiene, dunque, anche in quest'ultimo schema una dispersione iper-binomiale, binomiale e ipo-binomiale a seconda che la correlazione lineare ρ tra le variabili casuali indicatore sia rispettivamente positiva, nulla e negativa.

Analogamente alla rappresentazione impiegata nell'analisi degli schemi probabilistici, si consideri A_{ji} la variabile casuale indicatore associata alla i -ma prova della j -ma serie, con $i = 1, 2, \dots, n_j$ e $j = 1, 2, \dots, k$

$$A_{ji} = \begin{cases} 1 & \frac{b}{b+r} \\ 0 & \frac{r}{b+r} \end{cases} \quad \text{per ogni } i = 1, 2, \dots, n_j \text{ e } j = 1, 2, \dots, k.$$

Dall'equazione (3.6) risulta che la covarianza tra due variabili casuali indicatore della j -ma serie coincide con la seguente espressione

$$\text{cov}(A_{ji}, A_{ju}) = V(A_{ji})\rho = \frac{b}{b+r} \frac{r}{b+r} \rho$$

per $i = 1, 2, \dots, n_j$ e $u = 1, 2, \dots, n_j$ con $i \neq u$.

Per analizzare la dispersione di uno schema probabilistico in cui le estrazioni di ciascuna serie sono effettuate seguendo lo schema dell'urna di Polya bisogna calcolare la varianza della variabile casuale X_j , che rappresenta il numero di successi nelle n_j prove della j -ma serie

$$\begin{aligned} V(X_j) &= V\left(\sum_{i=1}^{n_j} A_{ji}\right) = \sum_{i=1}^{n_j} V(A_{ji}) + \sum_{i=1}^{n_j} \sum_{u=1, u \neq i}^{n_j} \text{cov}(A_{ji}, A_{ju}) \\ &= n_j \frac{b}{b+r} \frac{r}{b+r} + n_j(n_j - 1) \frac{b}{b+r} \frac{r}{b+r} \rho. \end{aligned} \quad (3.8)$$

Si noti che tale equazione coincide con la varianza della variabile casuale X_j che rappresenta il numero di successi nella j -ma serie in uno schema probabilistico con correlazione uniforme tra le prove di ciascuna serie ed indipendenza tra le serie⁵, con probabilità di successo costante e pari a $b/(b+r)$.

⁵Tale risultato é stato ricavato da Zenga (2003) e riportato da Calabrese (2006).

Sostituendo nella seconda equazione della (3.8) il risultato (3.7) ricavato in precedenza, si ottiene⁶

$$V(X_j) = n_j \frac{b}{b+r} \frac{r}{b+r} + n_j(n_j - 1) \frac{b}{b+r} \frac{r}{b+r} \frac{s}{b+r+s}. \quad (3.9)$$

Si noti che, raccogliendo a fattor comune, tale espressione coincide con la varianza (2.11) ricavata nel precedente paragrafo mediante i momenti fattoriali.

Uno schema probabilistico con indipendenza da prova a prova e da serie a serie, nel quale le probabilità di successo rimangono costanti sia da prova a prova che da serie a serie é definito schema di Bernoulli⁷. Per verificare l'ipotesi che lo schema in esame presenti una dispersione binomiale, caratteristica dello schema probabilistico di Bernoulli, si impiega la seguente statistica test

$$L = \sum_{j=1}^k \frac{\left(\frac{X_j}{n_j} - p\right)^2 n_j}{pq} = \sum_{j=1}^k \frac{\left(\frac{X_j}{n_j} - \frac{b}{b+r}\right)^2 n_j}{\frac{b}{b+r} \frac{r}{b+r}}. \quad (3.10)$$

É noto in letteratura (Boldrini (1968), pp.1202) che, al divergere delle numerosità n_j , la statistica L tende in distribuzione ad una variabile casuale Chi-Quadrato con k gradi di libertà⁸.

Il valore atteso del numeratore della statistica L in uno schema di Bernoulli risulta⁹

$$\mathbb{E} \left[\sum_{j=1}^k \left(\frac{X_j}{n_j} - \frac{b}{b+r}\right)^2 n_j \right] = k \frac{b}{b+r} \frac{r}{b+r}. \quad (3.11)$$

Nello schema probabilistico in cui vengono effettuate n_j estrazioni in ciascuna

⁶L'espressione (3.9) coincide con quella ottenuta da Faleschini (1954).

⁷Per un approfondimento sullo schema di Bernoulli si veda Boldrini (1968), pp. 1199-1201 e Faleschini (1949).

⁸Nel caso in cui la probabilità $b/(b+r)$ non sia nota, é possibile considerare una trasformazione di scala del quoziente di divergenza di Lexis proposta da Zenga (2003) ed impiegata da Calabrese (2006).

⁹Si veda Faleschini (1949).

serie seguendo lo schema dell'urna di Polya si ottiene invece che

$$\begin{aligned}
\mathbb{E} \left[\left(\frac{X_j}{n_j} - \frac{b}{b+r} \right)^2 n_j \right] &= n_j \mathbb{V} \left(\frac{X_j}{n_j} \right) \\
&= n_j \frac{1}{n_j^2} \left(n_j \frac{b}{b+r} \frac{r}{b+r} + n_j(n_j - 1) \rho \frac{b}{b+r} \frac{r}{b+r} \right) \\
&= \frac{b}{b+r} \frac{r}{b+r} + \rho \frac{b}{b+r} \frac{r}{b+r} (n_j - 1) \\
&= \frac{b}{b+r} \frac{r}{b+r} + \frac{s}{b+r+s} \frac{b}{b+r} \frac{r}{b+r} (n_j - 1),
\end{aligned}$$

dove $\hat{p}_j = X_j/n_j$ rappresenta la frequenza relativa del numero di successi nelle n_j estrazioni. Considerando uno schema probabilistico con indipendenza tra le serie rappresentate da k urne aventi la medesima composizione di palline, da ciascuna delle quali si effettuano n_j estrazioni seguendo lo schema dell'urna di Polya mantenendo costante da estrazione ad estrazione il numero di palline s reimmesse nell'urna si ottiene

$$\begin{aligned}
\mathbb{E} \left[\sum_{j=1}^k \left(\frac{X_j}{n_j} - \frac{b}{b+r} \right)^2 n_j \right] &= \sum_{j=1}^k n_j \left(\frac{1}{n_j^2} \right) \mathbb{V}(X_j) \quad (3.12) \\
&= k \frac{b}{b+r} \frac{r}{b+r} + \rho \frac{b}{b+r} \frac{r}{b+r} \sum_{j=1}^k (n_j - 1) \\
&= k \frac{b}{b+r} \frac{r}{b+r} + \frac{s}{b+r+s} \frac{b}{b+r} \frac{r}{b+r} \left(\sum_{j=1}^k n_j - k \right).
\end{aligned}$$

Si vuole confrontare il precedente risultato con l'equazione (3.11) che rappresenta la dispersione binomiale di uno schema di Bernoulli con probabilità di successo pari a $b/(b+r)$. Tutte le frazioni presenti nella precedente equazione sono positive tranne quella che presenta al numeratore s ed al denominatore $b+r+s > 0$. Questo significa che il numero di palline s che vengono reimmesse nell'urna, che determina, in base all'equazione (3.7), il segno della dipendenza lineare ρ , individua la dispersione dello schema dell'urna di Polya con indipendenza ed identiche modalità da serie a serie:

- se $s > 0$, e quindi $\rho > 0$, la dispersione é iper-binomiale;

- se $s = 0$, e quindi $\rho = 0$, la dispersione é binomiale;
- se $s < 0$, e quindi $\rho < 0$, la dispersione é ipo-binomiale.

4 L'ordinamento convesso nello schema dell'urna di Polya.

Si consideri il seguente ordinamento definito tra le funzioni di ripartizione delle variabili casuali aventi lo stesso valore atteso.

Definizione 4.1 (Ordinamento concavo ed ordinamento convesso). Siano F_X e F_Y le funzioni di ripartizione delle rispettive variabili X e Y . Si dice che Y *domina strettamente* X secondo l'ordinamento concavo ($Y \succ_{cv} X$) se

- $E(X) = E(Y)$
- $\int_{-\infty}^x F_Y(t) dt \leq \int_{-\infty}^x F_X(t) dt \quad \forall x \in \mathbb{R}$

con la disuguaglianza che vale in senso stretto per almeno un punto $x \in \mathbb{R}$. Si dice anche che X *domina strettamente* Y secondo l'ordinamento convesso ($X \succ_{cx} Y$).

Si consideri un'urna avente un numero di palline equamente ripartite tra i due colori, ad esempio si sceglie $b = r = 90$. Supponendo che da tale urna si effettuino n estrazioni, seguendo lo schema dell'urna di Polya, reimmettendo nell'urna un numero di palline pari a $s + 1$ dello stesso colore della pallina estratta dopo ciascuna estrazione. Si calcolano le funzioni di ripartizione mantenendo costante il numero di estrazioni n e permettendo al parametro s di variare. Si sono considerati in particolare i valori $s = -2, -1, 0, 1, 2$. Si effettua tale calcolo per diversi valori del numero di estrazioni n , ad esempio $n = 20, 40, 60$.

Per comprendere se esiste un ordinamento tra le funzioni di ripartizione delle variabili casuali che rappresentano il numero di successi rilevati sul medesimo numero di estrazioni n effettuate dalla stessa urna, variando per ciascuna serie di estrazioni il numero di palline s reimmesse nell'urna, si rappresentano le funzioni cumulate delle probabilità di tali variabili casuali.

numero di estrazioni n=20

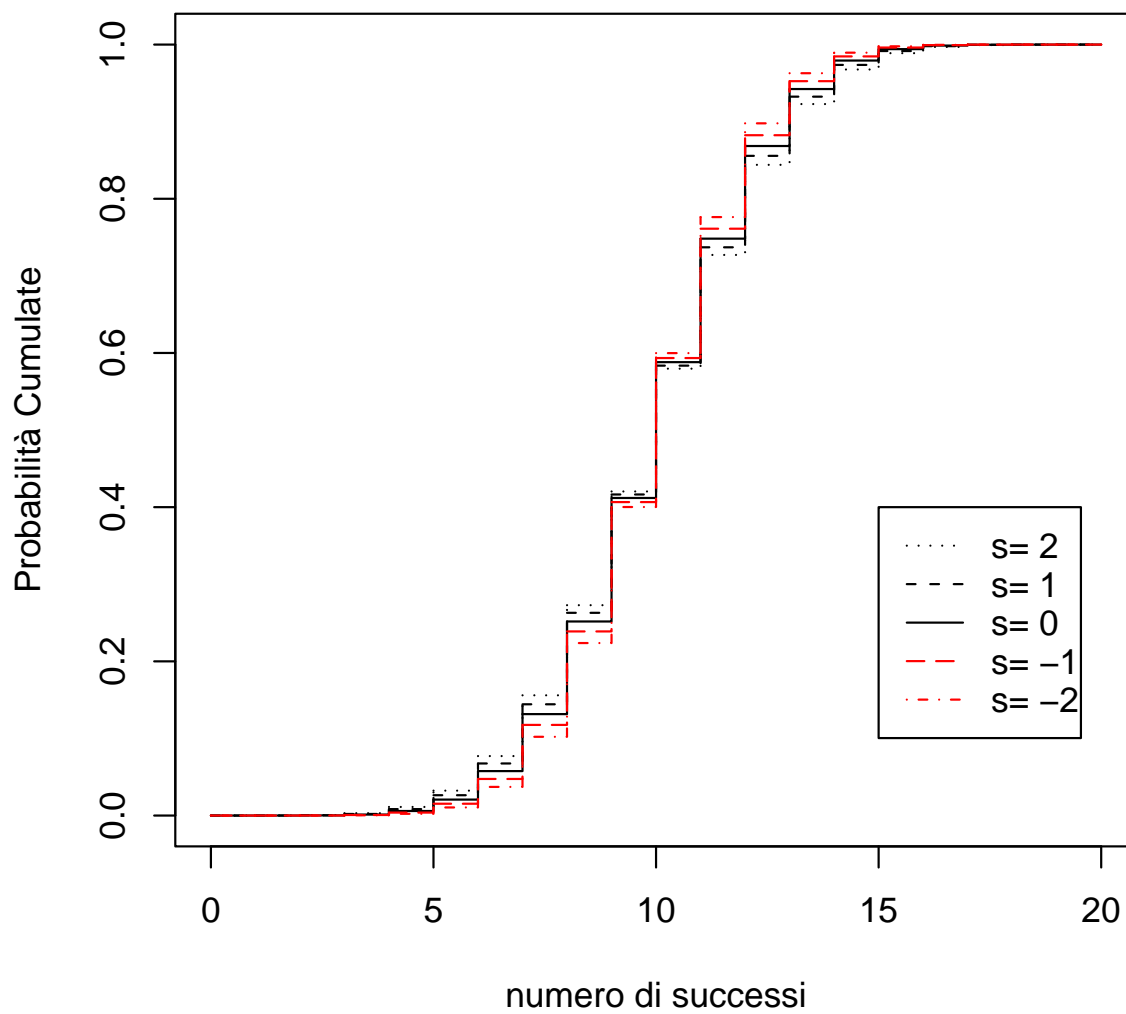


Figura 1: La funzione cumulata delle probabilità del numero di successi rilevati in uno schema dell'urna di Polya con $b = r = 90$ palline.

numero di estrazioni $n=40$

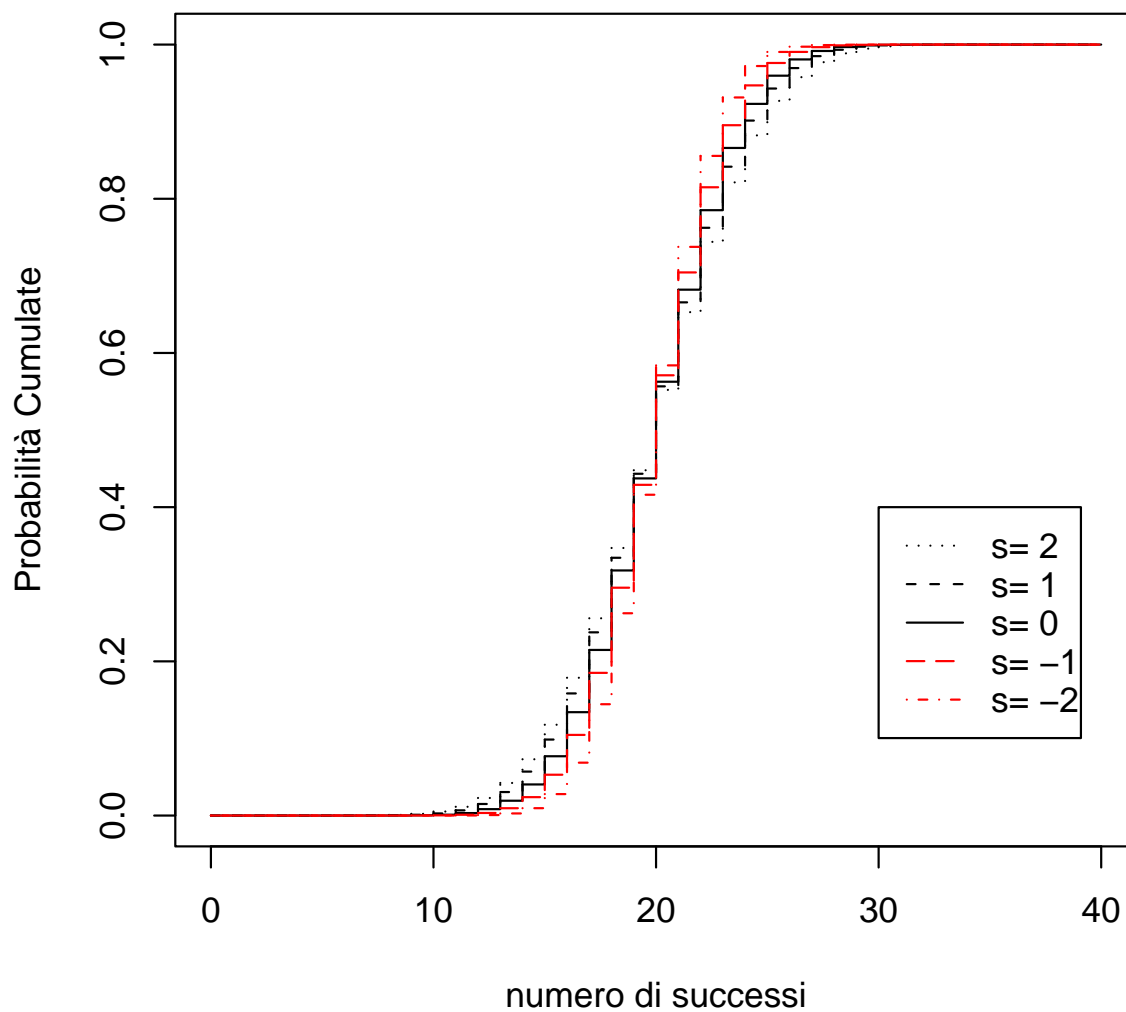


Figura 2: La funzione cumulata delle probabilità del numero di successi rilevati in uno schema dell'urna di Polya con $b = r = 90$ palline.

numero di estrazioni $n=60$

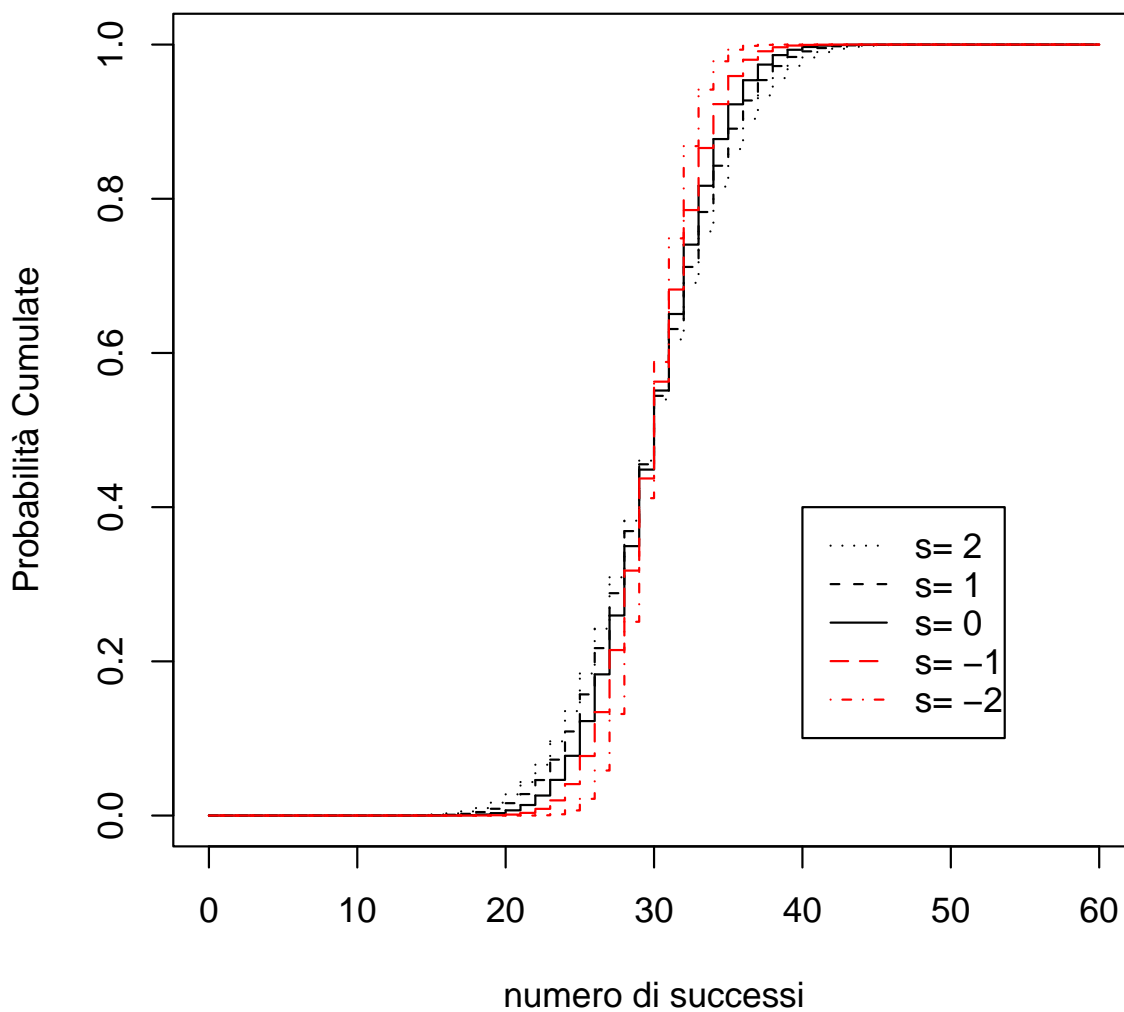


Figura 3: La funzione cumulata delle probabilità del numero di successi rilevati in uno schema dell'urna di Polya con $b = r = 90$ palline.

Dai grafici riportati nelle precedenti pagine si nota inizialmente che, avendo scelto $b = r$, la distribuzione di probabilità risulta simmetrica rispetto al valore atteso.

Per ogni fissato numero di estrazioni n , le aspettative delle variabili casuali, definite al variare del numero delle palline s reimmesse nell'urna, coincidono

e risultano pari a $n \frac{b}{b+r} = \frac{n}{2}$.

Confrontando le funzioni di ripartizione delle variabili casuali che rappresentano il numero di successi rilevati sul medesimo numero di estrazioni n , dai grafici riportati in precedenza si intuisce che le variabili casuali caratterizzate da un dato numero di palline s dominano strettamente secondo l'ordinamento convesso le variabili casuali aventi un valore di s inferiore. Essendo infatti le distribuzioni di probabilità simmetriche, l'area dei rettangoli delimitati dalle funzioni di ripartizione per numero di successi maggiore o uguale al valore atteso e per numero di successi minori dell'aspettativa coincidono a coppie. Si dimostrerà in seguito che è possibile estendere tale risultato a generiche urne contenenti un numero di palline nere b diverso dal numero di palline rosse r .

Teorema 4.1. *Siano X e Y due variabili casuali che rappresentano il numero di successi rilevati in n estrazioni effettuate seguendo lo schema dell'urna di Polya da due urne aventi la medesima composizione di palline, caratterizzate rispettivamente da un numero di palline reimmesse nell'urna pari a s e s^* , con $s > s^*$.*

Il rapporto tra le seguenti probabilità

$$\frac{P\{Y = k\}}{P\{X = k\}} \quad (4.1)$$

è una funzione strettamente monotona crescente di k

- per $0 \leq k < n \frac{b}{b+r} + \frac{r}{b+r}$ se $s > 0$ e $s^* > 0$;
- per $\max\left(0, n + \frac{r}{s^*}\right) \leq k < n \frac{b}{b+r} + \frac{r}{b+r}$
se $s > 0$ e $s^* < 0$ oppure se $s < 0$ e $s^* < 0$.

ed è una funzione strettamente monotona decrescente di k

- per $n \geq k > n \frac{b}{b+r} + \frac{r}{b+r}$ se $s > 0$ e $s^* > 0$;

- per $\min\left(n, -\frac{b}{s^*}\right) \geq k > n\frac{b}{b+r} + \frac{r}{b+r}$
se $s > 0$ e $s^* < 0$ oppure se $s < 0$ e $s^* < 0$.

Nel caso in cui $n\frac{b}{b+r} + \frac{r}{b+r}$ é un numero intero, risulta

$$\frac{P\{Y = k\}}{P\{X = k\}} = \frac{P\{Y = k - 1\}}{P\{X = k - 1\}} \quad \text{per } k = n\frac{b}{b+r} + \frac{r}{b+r}.$$

Dimostrazione. Dimostrare che il rapporto (4.1) é una funzione strettamente monotona crescente di k per $k < n\frac{b}{b+r} + \frac{r}{b+r}$ equivale a dimostrare che

$$\frac{P\{Y = k\}}{P\{X = k\}} > \frac{P\{Y = k - 1\}}{P\{X = k - 1\}} \quad \text{per } k < n\frac{b}{b+r} + \frac{r}{b+r},$$

che a sua volta coincide con la seguente condizione

$$\frac{P\{Y = k\}}{P\{Y = k - 1\}} > \frac{P\{X = k\}}{P\{X = k - 1\}} \quad \text{per } k < n\frac{b}{b+r} + \frac{r}{b+r}.$$

Dall'espressione (2.2) della distribuzione di probabilitá del numero di successi rilevati in n estrazioni effettuate secondo lo schema dell'urna di Polya da un'urna che contiene b palline nere e r rosse, si deduce che

$$\begin{aligned} \frac{P\{X = k\}}{P\{X = k - 1\}} &= \frac{\binom{n}{k} b(b+s) \dots [b+(k-1)s] r(r+s) \dots [r+(n-k-1)s]}{\binom{n}{k-1} b(b+s) \dots [b+(k-2)s] r(r+s) \dots [r+(n-k)s]} \\ &= \frac{\binom{n}{k} b+(k-1)s}{\binom{n}{k-1} r+(n-k)s}. \end{aligned}$$

Il precedente risultato permette di calcolare il seguente rapporto

$$\frac{\frac{P\{Y=k\}}{P\{Y=k-1\}}}{\frac{P\{X=k\}}{P\{X=k-1\}}} = \frac{\frac{\binom{n}{k}}{\binom{n}{k-1}} \frac{b+(k-1)s^*}{r+(n-k)s^*}}{\frac{\binom{n}{k}}{\binom{n}{k-1}} \frac{b+(k-1)s}{r+(n-k)s}} = \frac{\frac{b+(k-1)s^*}{r+(n-k)s^*}}{\frac{b+(k-1)s}{r+(n-k)s}}. \quad (4.2)$$

Sapendo che $s > s^*$, affinché le probabilità presenti nella precedente equazione siano non nulle si deve limitare l'analisi ai seguenti intervalli

- $0 < k < n$ se $s > 0$ e $s^* > 0$;
- $\max\left(0, n + \frac{r}{s^*}\right) < k < \min\left(n, -\frac{b}{s^*}\right)$

se $s > 0$ e $s^* < 0$ oppure se $s < 0$ e $s^* < 0$.

Determinare per quali valori di k $P\{Y = k\}/P\{X = k\}$ é una funzione strettamente monotona crescente di k equivale a determinare per quali valori di k il rapporto definito nell'equazione (4.2) risulta maggiore di uno, in altri termini deve risultare

$$\frac{b + (k-1)s^*}{r + (n-k)s^*} > \frac{b + (k-1)s}{r + (n-k)s}. \quad (4.3)$$

Negli intervalli considerati in precedenza i denominatori presenti nella disuguaglianza (4.3) sono sempre positivi, si ricava quindi

$$[b + (k-1)s^*][r + (n-k)s] > [b + (k-1)s][r + (n-k)s^*]$$

$$br + (k-1)rs^* + bs(n-k) + (k-1)(n-k)ss^* > br + (k-1)rs + bs^*(n-k) + (k-1)(n-k)ss^*$$

$$s[b(n-k) - r(k-1)] > s^*[b(n-k) - r(k-1)].$$

Essendo $s > s^*$ per ipotesi, deve risultare che

$$b(n-k) - r(k-1) > 0$$

$$-k(b+r) > -bn - r$$

$$k(b+r) < bn+r$$

$$k < n\frac{b}{b+r} + \frac{r}{b+r}.$$

Avendo limitato la precedente analisi a probabilità non nulle, si é ricavato dunque che $P\{Y = k\}/P\{X = k\}$ é una funzione strettamente monotona crescente di k

- per $0 \leq k < n\frac{b}{b+r} + \frac{r}{b+r}$ se $s > 0$ e $s^* > 0$;
- per $\max\left(0, n + \frac{r}{s^*}\right) \leq k < n\frac{b}{b+r} + \frac{b}{b+r}$
se $s > 0$ e $s^* < 0$ oppure se $s < 0$ e $s^* < 0$.

Si dimostra analogamente che il rapporto definito nell'equazione (4.2) é minore di uno

- per $n \geq k > n\frac{b}{b+r} + \frac{r}{b+r}$ se $s > 0$ e $s^* > 0$;
- per $\min\left(n, -\frac{b}{s^*}\right) \geq k > n\frac{b}{b+r} + \frac{r}{b+r}$
se $s > 0$ e $s^* < 0$ oppure se $s < 0$ e $s^* < 0$,

quindi in tali intervalli $P\{Y = k\}/P\{X = k\}$ é una funzione strettamente monotona decrescente di k .

Infine nel caso in cui $n\frac{b}{b+r} + \frac{r}{b+r}$ é un numero intero, il rapporto definito nell'equazione (4.2) risulta pari a uno per $k = n\frac{b}{b+r} + \frac{r}{b+r}$. \square

Si noti nel precedente teorema che $n\frac{b}{b+r}$ rappresenta il valore atteso delle variabili casuali X e Y .

Per dimostrare che la variabile casuale X domina strettamente Y secondo l'ordinamento convesso si impiega il seguente risultato.

Teorema 4.2. *Siano X e Y due variabili casuali aventi valori attesi uguali. Se $P\{X = k\} - P\{Y = k\}$ presenta due cambi di segno con tale sequenza $+, -, +$, allora X domina strettamente Y secondo l'ordinamento convesso ($X \succ_{cx} Y$).*

Dimostrazione. Per la dimostrazione si veda Shaked e Shanthikumar (1994), pp. 66. \square

Per dimostrare che le variabili casuali X e Y che rappresentano il numero di successi nello schema dell'urna di Polya soddisfano le ipotesi del precedente teorema é necessario considerare il seguente lemma¹⁰.

Lemma 4.3. *Siano X e Y due variabili casuali discrete non negative aventi il medesimo supporto $D = \{0, 1, \dots, n\}$. Se le funzioni di ripartizione differiscono*

$$F_X(k) \neq F_Y(k) \quad \text{per } k \in I \subseteq D$$

e i valori attesi coincidono $E(X) = E(Y)$, allora la funzione $P\{X = k\} - P\{Y = k\}$ presenta almeno due cambi di segno nell'insieme D .

Dimostrazione. Sapendo che

$$\sum_{k \in D} P\{X = k\} = \sum_{k \in D} P\{Y = k\} = 1,$$

ed essendo $F_X(k) \neq F_Y(k)$ per $k \in I \subseteq D$, allora esiste un punto k_0 in cui la funzione $P\{X = k\} - P\{Y = k\}$ ha un cambio di segno.

Si supponga che la funzione $P\{X = k\} - P\{Y = k\}$ presenti un solo cambio di segno nell'insieme D . Se risulta che

$$\begin{aligned} P\{X = k\} - P\{Y = k\} &\geq 0 && \text{per } k \leq k_0 \\ P\{X = k\} - P\{Y = k\} &\leq 0 && \text{per } k > k_0, \end{aligned}$$

allora si ottiene che

$$(k - k_0)[P\{X = k\} - P\{Y = k\}] \leq 0 \quad \forall k \in D$$

ed esiste almeno un punto $k \in D$ tale che

$$(k - k_0)[P\{X = k\} - P\{Y = k\}] < 0.$$

Dalla precedente disuguaglianza si deduce che

$$\sum_{k=0}^n (k - k_0)[P\{X = k\} - P\{Y = k\}] = E(X) - E(Y) < 0.$$

¹⁰Per una generalizzazione del lemma (4.3) si veda Denuit e Lefèvre (1997).

Tale risultato é in contraddizione con l'ipotesi di uguaglianza dei valori attesi $E(X) = E(Y)$.

Analogamente se risulta che

$$\begin{aligned} P\{X = k\} - P\{Y = k\} &\leq 0 && \text{per } k \leq k_0 \\ P\{X = k\} - P\{Y = k\} &\geq 0 && \text{per } k > k_0, \end{aligned}$$

allora si ottiene che

$$(k - k_0)[P\{X = k\} - P\{Y = k\}] \geq 0 \quad \forall k \in D$$

ed esiste almeno un punto $k \in D$ tale che

$$(k - k_0)[P\{X = k\} - P\{Y = k\}] > 0.$$

Dalla precedente disuguaglianza si deduce che

$$\sum_{k=0}^n (k - k_0)[P\{X = k\} - P\{Y = k\}] = E(X) - E(Y) > 0.$$

Tale risultato é in contraddizione con l'ipotesi di uguaglianza dei valori attesi $E(X) = E(Y)$.

Si deduce dunque che la funzione $P\{X = k\} - P\{Y = k\}$ presenta almeno due cambi di segno nell'insieme D . \square

Tale risultato permette di dimostrare il seguente teorema.

Teorema 4.4. *Siano X e Y due variabili casuali che rappresentano il numero di successi rilevati in n estrazioni effettuate seguendo lo schema dell'urna di Polya da due urne aventi la medesima composizione di palline, caratterizzate rispettivamente da un numero di palline reimmesse nell'urna pari a s e s^* , con $s > s^*$. La funzione $P\{X = k\} - P\{Y = k\}$ ha due cambi di segno e presenta la sequenza di segni $+, -, +$ nei seguenti insiemi*

- $0 \leq k \leq n$ se $s > 0$ e $s^* > 0$ oppure se $s > 0$ e $s^* < 0$;
- $\max\left(0, n + \frac{r}{s}\right) \leq k \leq \min\left(n, -\frac{b}{s}\right)$ se $s < 0$ e $s^* < 0$.

Dimostrazione. Si consideri inizialmente il caso in cui le variabili casuali X e Y hanno supporti diversi. Essendo per ipotesi $s > s^*$, tale condizione si verifica solamente se $s > 0$ e $s^* < 0$ oppure $s < 0$ e $s^* < 0$. In entrambi i casi il supporto della variabile casuale Y risulta incluso o al piú coincide con quello della variabile casuale X . Questo significa che $P\{X = k\} - P\{Y = k\} > 0$ nei seguenti intervalli

- per $0 \leq k < \max\left(0, n + \frac{r}{s^*}\right)$ e $\min\left(n, -\frac{b}{s^*}\right) < k \leq n$
se $s > 0$ e $s^* < 0$;
- per $\max\left(0, n + \frac{r}{s}\right) \leq k < \max\left(0, n + \frac{r}{s^*}\right)$ e
 $\min\left(n, -\frac{b}{s^*}\right) < k \leq \min\left(n, -\frac{b}{s}\right)$
se $s < 0$ e $s^* < 0$.

Indicando con D_1 e D_2 i supporti delle rispettive variabili casuali X e Y , é noto che

$$\sum_{k \in D_1} P\{X = k\} = \sum_{k \in D_2} P\{Y = k\} = 1.$$

Deve dunque risultare $P\{X = k\} - P\{Y = k\} < 0$ per alcuni valori di $k \in D_1 \cap D_2$.

Poiché per il teorema (4.1) $P\{Y = k\}/P\{X = k\}$ é una funzione strettamente monotona crescente di k per $\max\left(0, n + \frac{r}{s^*}\right) \leq k < n\frac{b}{b+r} + \frac{r}{b+r}$ e strettamente monotona decrescente di k per $\min\left(n, -\frac{b}{s^*}\right) \geq k > n\frac{b}{b+r} + \frac{r}{b+r}$, esiste un solo intervallo I tale che

$$\frac{P\{Y = k\}}{P\{X = k\}} > 1 \quad \text{per } k \in I \subseteq \left[\max\left(0, n + \frac{r}{s^*}\right), \min\left(n, -\frac{b}{s^*}\right) \right],$$

che coincide con la seguente condizione

$$P\{X = k\} - P\{Y = k\} < 0 \quad \text{per } k \in I \subseteq \left[\max\left(0, n + \frac{r}{s^*}\right), \min\left(n, -\frac{b}{s^*}\right) \right].$$

Questo significa che la funzione $P\{X = k\} - P\{Y = k\}$ ha due cambi di segno nel supporto D_1 della variabile casuale X , con la seguente sequenza di

segni $+$, $-$, $+$.

Si consideri il caso in cui le variabili casuali X e Y hanno il medesimo supporto $D = \{0, 1, \dots, n\}$. Per quanto dimostrato nel primo paragrafo risulta che $\mathbb{E}(X) = \mathbb{E}(Y)$ ed essendo per ipotesi $s \neq s^*$ dalla distribuzione di probabilità (2.2) risulta che $F_X(k) \neq F_Y(k)$ per $k \in B \subseteq D$. Questo significa che sono soddisfatte le ipotesi del lemma (4.3), dal quale si ottiene che la funzione $P\{X = k\} - P\{Y = k\}$ presenta almeno due cambi di segno nell'insieme D . Poiché esistono le seguenti corrispondenze biunivoche

- $P\{X = k\} - P\{Y = k\} > 0 \Leftrightarrow \frac{P\{Y=k\}}{P\{X=k\}} < 1 \quad \text{con } k \in I_1 \subset D$
- $P\{X = k\} - P\{Y = k\} < 0 \Leftrightarrow \frac{P\{Y=k\}}{P\{X=k\}} > 1 \quad \text{con } k \in I_2 \subset D,$

e sapendo inoltre dal lemma (4.1) che $P\{Y = k\}/P\{X = k\}$ é una funzione strettamente monotona crescente di k per $0 \leq k < n\frac{b}{b+r} + \frac{r}{b+r}$ e strettamente monotona decrescente di k per $n \geq k > n\frac{b}{b+r} + \frac{r}{b+r}$, si deduce che la funzione $P\{X = k\} - P\{Y = k\}$ presenta due cambi di segno nell'insieme D , che corrispondono ai punti $k_0, k_1 \in D$.

Essendo $P\{Y = k\}/P\{X = k\}$ una funzione strettamente monotona crescente di k per $k < n\frac{b}{b+r} + \frac{r}{b+r}$ e strettamente monotona decrescente per $k > n\frac{b}{b+r} + \frac{r}{b+r}$, si deduce che

$$\frac{P\{Y = k\}}{P\{X = k\}} \leq 1 \quad \text{per } k \leq k_0 \text{ e } k \geq k_1,$$

con $0 \leq k_0 < n\frac{b}{b+r} + \frac{r}{b+r}$ e $n \geq k_1 > n\frac{b}{b+r} + \frac{r}{b+r}$.

Sapendo che

$$\sum_{k=0}^n P\{X = k\} = \sum_{k=0}^n P\{Y = k\} = 1, \tag{4.4}$$

deve esistere almeno un punto $k \in D$ tale che

$$\frac{P\{Y = k\}}{P\{X = k\}} > 1 \quad \text{per } k_0 < k < k_1.$$

Questo significa che

$$\begin{aligned} P\{X = k\} - P\{Y = k\} &\geq 0 \quad \text{per } k \leq k_0 \text{ e } k \geq k_1 \\ P\{X = k\} - P\{Y = k\} &< 0 \quad \text{per } k_0 < k < k_1. \end{aligned} \quad (4.5)$$

Dal lemma (4.1) oppure dalla condizione (4.4) risulta che la prima disuguaglianza della (4.5) deve essere verificata in senso stretto per almeno un $k \in D$. \square

Dal risultato fornito dal precedente teorema e sapendo che le variabili casuali X e Y hanno la stessa aspettativa, dal teorema (4.2) si deduce che X domina strettamente Y secondo l'ordinamento convesso.

Tale risultato é coerente con quanto dimostrato nel precedente paragrafo, infatti é noto in letteratura (pp.56, Shaked e Shanthikumar (1994)) che se X domina strettamente Y secondo l'ordinamento convesso ($X \succ_{cx} Y$), allora risulta che la varianza della variabile casuale X é maggiore della varianza di Y

$$V(X) > V(Y).$$

Per comprendere tale coerenza si consideri la variabile casuale X_j che rappresenta il numero di successi (palline nere estratte) rilevati nelle n_j prove della j -ma serie, dove n_j rappresenta il numero di estrazioni effettuate seguendo lo schema dell'urna di Polya da un'urna che contiene b palline nere e r palline rosse.

Si é dimostrato in precedenza ricavando l'espressione (2.8) che, se si considera un numero di palline reimmesse nell'urna $s = 0$, allora X_j é una variabile casuale binomiale di parametri n_j e $b/(b+r)$. Se $s > 0$ la variabile casuale X_j domina strettamente secondo l'ordinamento convesso la variabile casuale binomiale di parametri n_j e $b/(b+r)$, quindi X_j ha una dispersione iperbinomiale. Se infine $s < 0$ la variabile casuale binomiale di parametri n_j e $b/(b+r)$ domina strettamente secondo l'ordinamento convesso X_j , quindi X_j ha dispersione ipo-binomiale.

É importante evidenziare che per poter confrontare tali dispersioni le variabili casuali esaminate devono avere lo stesso valore atteso, che dipende dal numero di prove n_j e dalla probabilitá di successo $b/(b+r)$.

Considerando k urne di eguale composizione di palline ed effettuando n_j estrazioni da ciascuna urna, mantenendo costante da serie e serie il numero di palline s reimmesse nell'urna, si ottiene uno schema probabilistico con indipendenza tra le serie, avente dispersione:

- ipo-binomiale se $s < 0$;
- binomiale se $s = 0$;
- iper-binomiale se $s > 0$.

Si noti che tale risultato coincide con quello ricavato in precedenza mediante la rappresentazione dello schema probabilistico con correlazione uniforme tra le prove di ciascuna serie e indipendenza tra le serie.

5 Conclusioni.

Poiché lo schema dell'urna di Polya rappresenta fenomeni dicotomici, è possibile applicare tale metodologia per analizzare il fenomeno delle insolvenze nel rischio di credito.

Considerando un portafoglio di crediti, si raggruppano quei clienti che presentano caratteristiche simili, come ad esempio la medesima classe di rating, criterio impiegato nel modello CreditMetrics (J. P. Morgan (1997)), oppure perdite simili in caso di insolvenza, principio impiegato nel modello CreditRisk+ (Credit Suisse Financial Products (1997)). Si ottengono dunque k classi composte da n_j crediti ciascuna, con $j = 1, 2, \dots, k$. La variabile casuale indicatore A_{ji} rappresenta quindi lo stato di insolvenza o solvenza dell' i -mo credito appartenente alla j -ma classe.

Il vantaggio dello schema dell'urna di Polya consiste nel poter rappresentare l'influenza che il ciclo economico esercita sul fenomeno dell'insolvenza. Tale relazione induce a supporre che tra i crediti appartenenti alla medesima classe esista una correlazione lineare positiva, che è possibile rappresentare assumendo che il numero di palline s reimmesse nell'urna dopo ciascuna estrazione sia positivo. Tale schema dell'urna di Polya è noto come schema del contagio. La dispersione del portafoglio che presenta correlazione positiva tra le variabili casuali indicatore appartenenti alla medesima classe è iper-binomiale.

Alla banca che concede i crediti conviene invece raggruppare nelle medesime classi quei clienti che presentano correlazione lineare negativa, ottenendo infatti un portafoglio con dispersione ipo-binomiale.

Dopo aver costituito un portafoglio di crediti in modo che le insolvenze tra le diverse fasce siano tra loro indipendenti, la banca può verificare tramite la statistica test (3.10) se è soddisfatta l'ipotesi di dispersione binomiale.

6 Bibliografia.

Boldrini M. (1968). *Statistica: Teoria e Metodi*. Giuffré, Milano, cap. XVIII.

Calabrese R. (2006). *I principali schemi delle prove ripetute per rappresentare le insolvenze*. Rapporto di Ricerca del Dip.to di Metodi Quantitativi per le Scienze Economiche ed Aziendali, Università degli Studi di Milano-Bicocca.

Credit Suisse Financial Products (1997). *CreditRisk+*. A Credit Risk Management Framework. Technical Document, Londra.

Denuit M. e Lefèvre C. (1997). *Some new classes of stochastic order relations among arithmetic random variables, with applications in actuarial sciences*. Insurance: Mathematics & Economics **20**, pp. 197-213.

Faleschini L. (1949). *Sullo schema generale del problema delle prove ripetute con probabilità indipendente*. Estratto dalla 'Rivista Italiana di demografia e Statistica', Vol **III**, pp. 1-25.

Faleschini L. (1954). *Sullo schema generale del problema delle prove ripetute con probabilità dipendenti secondo lo schema di contagio (o immunità)*. 'Bulletin de l'Institut International de Statistique', tome **XXXIV**, 2^{ème} livraison.

Johnson N. L. e Kotz S. (1977). *Urn Models and Their Application*. John Wiley & Sons, New York.

J. P. Morgan (1997). *CreditMetrics*. Technical Document, New York.

Riordan J. (1980). *An Introduction to Combinatorial Analysis*. Princeton University Press, Princeton.

Shaked M. e Shanthikumar J. G. (1994). *Stochastic Orders and Their Applications*. Academic Press, San Diego.

Zenga M. (2003). *Schemi probabilistici con indipendenza fra le prove di una serie e dipendenza fra le prove di una serie*. Seminario tenuto presso il Dipartimento di Metodi Quantitativi per le Scienze Economiche ed Aziendali, Università degli Studi di Milano-Bicocca.