

La determinazione dei parametri di un nuovo modello
distributivo per variabili non negative:
aspetti metodologici e applicazioni

Alberto Arcagni

19 luglio 2011

Indice

1	Introduzione	4
2	Misure di ineguaglianza	6
2.1	Spezzata di Lorenz e curva di Lorenz	6
2.2	Indice di Gini	7
2.3	Indice di Bonferroni	8
2.4	Indice di Pietra	9
2.5	Curva $I_{(p)}$ e indice I di Zenga	9
2.6	Requisiti degli indici di concentrazione	12
3	Modelli distributivi	13
3.1	Variabile casuale di Dagum	13
3.2	Variabile casuale di Zenga	20
3.2.1	La distribuzione condizionata	20
3.2.2	Funzione di densità	22
3.2.3	Funzione di ripartizione	25
3.2.4	Momenti dall'origine e momenti centrali	28
3.2.5	Misure di ineguaglianza del modello di Zenga	31
4	Stima dei parametri	34
4.1	Metodo dei momenti	34
4.1.1	Stimatori ottenuti con il metodo dei momenti per la variabile casuale di Zenga	35
4.2	Minimizzazione di indici di accostamento	38
4.2.1	Stima non vincolata	40
4.2.2	Introduzione di vincoli	42
4.2.3	I valori iniziali	45
4.3	Distribuzioni degli stimatori e degli indici di accostamento della variabile casuale di Zenga mediante Bootstrap parametrico	46
4.3.1	Il Bootstrap parametrico	47
4.3.2	Generazione di valori dal modello di Zenga	48

<i>INDICE</i>	3
5 Applicazione	51
5.1 Descrizione dei dati	51
5.2 Descrizione dei risultati ottenuti dall'applicazione sui dati reali	52
5.3 Confronto fra modello di Zenga e modello di Dagum	57
5.4 Confronto fra i diversi metodi di stima impiegati per il modello di Zenga e discussione sui vincoli	69
5.5 Studio delle distribuzioni degli stimatori: primi risultati	73
6 Conclusioni	78

Capitolo 1

Introduzione

In questo lavoro si affronterà lo studio dei metodi di stima dei parametri di un nuovo modello distributivo proposto recentemente da Zenga (2010).

Il modello ha come supporto l'insieme dei numeri reali non negativi, ed è finalizzato a rappresentare distribuzioni per ordine di grandezza, quali ad esempio le distribuzioni dei tempi di attesa e distribuzioni dei redditi.

Con particolare riferimento alle distribuzioni dei redditi, in letteratura sono presenti diversi modelli atti a rappresentarle, ad esempio la distribuzione di Pareto, quella di Dagum, la Gamma, la log-Normale, la Singh-Maddala e la distribuzione Pareto-Lognormale (si vedano Johnson *et al.*, 1994; Dagum, 1977; Singh e Maddala, 1976; Colombi, 1990). Il nuovo modello proposto pertanto si inserisce in una varietà all'interno della quale è possibile scegliere quello più adatto a seconda della distribuzione di dati reali di cui si dispone, delle assunzioni e dei requisiti che i modelli devono rispettare, del metodo di stima e dei criteri per valutare la bontà di adattamento.

I passaggi che hanno portato alla definizione del modello illustrano i criteri che lo caratterizzano:

1. Zenga (2007) ha definito la curva $I_{(p)}$, indice di ineguaglianza puntuale, e il corrispettivo indice sintetico di ineguaglianza I quale media integrale dell'indice puntuale;
2. Poliscchio (2008) ha dimostrato che l'unico modello distributivo, con supporto non negativo valore atteso finito e curva $I_{(p)}$ costante, pertanto pienamente descritta dall'indice sintetico I , è una particolare Pareto troncata a due parametri, μ parametro di scala che ne definisce il valore atteso e k parametro di forma che ne definisce l'*uniformità* quale complemento a 1 dell'indice sintetico di ineguaglianza I ;
3. il nuovo modello distributivo è ottenuto come miscuglio di tali Pareto troncate con valore atteso costante e parametro di *uniformità* che varia secondo la legge di distribuzione Beta.

Ne risulta una distribuzione a tre parametri con coda Paretiana. Il parametro di scala determina il valore atteso, che conseguentemente è sempre finito. Gli altri due parametri ne determinano la forma, uno dei quali regola il peso delle code e l'esistenza dei momenti. Entrambi i parametri di forma sono indicatori di ineguaglianza.

Il nuovo modello proposto da Zenga, dopo aver stabilito la locazione, pone l'attenzione sull'ineguaglianza della distribuzione. Di conseguenza, per la determinazione della stima dei parametri

del modello, sembra opportuno impiegare metodi di stima che considerano i valori di statistiche campionarie relative alla locazione, quali media e mediana, e relative all'ineguaglianza. Per tali ragioni, in questo lavoro, si è dato ampio spazio a metodi di stima come il metodo dei momenti e ai metodi basati sulla minimizzazione di indici di accostamento, con l'introduzione di vincoli. Sono state effettuate delle applicazioni su dati forniti da Eurostat (2003) nelle quali si è confrontato l'adattamento del nuovo modello con quello della distribuzione di Dagum sulla base di indici di ineguaglianza e di scostamenti relativi fra valori campionari e valori teorici di media e mediana.

Nel secondo capitolo verranno illustrate le misure di ineguaglianza, tra le quali anche la curva $I_{(p)}$ e l'indice sintetico I .

Nel terzo capitolo si presenteranno i modelli distributivi oggetto di studio, richiamando le definizioni e le caratteristiche della variabile casuale di Dagum e del nuovo modello proposto da Zenga.

Nel quarto capitolo ci si occuperà della stima dei parametri della variabile casuale di Zenga. Verrà fornita la soluzione analitica degli stimatori dei parametri secondo il metodo dei momenti, verranno definiti gli indici di accostamento e discussa la stima mediante la minimizzazione di quest'ultimi e si valuterà la possibilità di introdurre opportuni vincoli. Poiché in questo capitolo verranno impiegati metodi numerici, verrà presentato anche il metodo Bootstrap al fine di avere uno strumento per una valutazione preliminare delle caratteristiche degli stimatori, e contestualmente ad esso si illustrerà un metodo di generazione di valori dal modello di Zenga.

Nel quinto capitolo verrà presentata un'applicazione sulle distribuzioni dei redditi familiari di 15 Stati Membri dell'Unione Europea fornite da Eurostat (2003).

Capitolo 2

Misure di ineguaglianza

Prima di intraprendere lo studio dei modelli atti a rappresentare la distribuzione dei redditi, si fa presente che uno dei principali aspetti di interesse di tali distribuzioni è l'analisi dell'ineguaglianza.

Dal 1895 al 1915 diversi studiosi italiani hanno fornito ampi contributi nella formalizzazione matematica dell'ineguaglianza e nella definizione delle prime misure (Zenga, 1987). Tra essi si ricordano Pareto, il cui modello è l'esempio dello stretto legame tra gli studi dell'ineguaglianza e delle distribuzioni dei redditi, Mortara, che attraverso la definizione di densità di reddito (rapporto fra reddito effettivo di un individuo e reddito medio, ovvero di equidistribuzione) fornì un indice basato sul rapporto fra scostamento medio assoluto e media aritmetica, e Pietra, che contribuì ampiamente allo studio della curva di Lorenz. Infine si ricorda Gini che fornì la definizione di concentrazione più accreditata e da essa ricavò l'indice ancora oggi di largo impiego.

Tra indici di ineguaglianza è necessario formulare una prima distinzione: indici puntuali e indici globali. Gli indici *puntuali* forniscono informazione sull'ineguaglianza al variare del livello di reddito considerato. Di seguito verranno descritte la curva di Lorenz e la curva $I(p)$ introdotta da Zenga (2007). Gli indici *globali* rappresentano una misura del livello di ineguaglianza relativa alla totalità della popolazione. Fra questi verranno presentati l'indice R di Gini, l'indice B di Bonferroni, l'indice P di Pietra e l'indice I di Zenga.

2.1 Spezzata di Lorenz e curva di Lorenz

Siano $0 \leq x_1 \leq x_2 \leq \dots \leq x_i \leq \dots \leq x_n$ n quantità ordinate che misurano l'intensità di un dato carattere trasferibile. La *spezzata* di Lorenz è il grafico costituito da segmenti che congiungono gli $n + 1$ punti di coordinate $(p_0, q_0) = (0, 0)$ e (p_i, q_i) per $i = 1, \dots, n$ con

$$\begin{aligned} p_i &= \frac{i}{n} \\ q_i &= \frac{\sum_{j=1}^i x_j}{T} \end{aligned}$$

dove $T = \sum_{i=1}^n x_i$ è l'ammontare totale delle quantità osservate. In termini di reddito, la coppia di valori (p_i, q_i) indica che la frazione di popolazione più povera (in quanto le osservazioni sono ordinate) p_i detiene la frazione di reddito q_i . Si osserva che i segmenti che compongono la spezzata

hanno inclinazione crescente, ovvero la spezzata è convessa, e $(p_N, q_N) = (1, 1)$, conseguentemente risiede nel quadrato unitario e sotto la bisettrice primo-terzo quadrante. Nel caso le n quantità x_i siano tutte uguali, la spezzata coincide con la bisettrice principale che viene definita retta di equiripartizione. Se $x_i = 0$ per $i = 1, 2, \dots, n-1$ e $x_n = T$, caso di massima concentrazione, la spezzata è composta da due segmenti congiungenti i punti $(0, 0)$, $(1 - 1/n, 0)$ e $(1, 1)$. L'area di concentrazione è rappresentata dall'area compresa fra la spezzata e la retta di equiripartizione. In figura 2.1.1 è riportato un esempio di spezzata di Lorenz per $n = 5$, nel quale sono evidenziate la retta di equiripartizione, l'area di concentrazione e l'area di massima concentrazione, quest'ultima corrisponde all'area di concentrazione nel caso una sola unità detiene tutta la quantità.

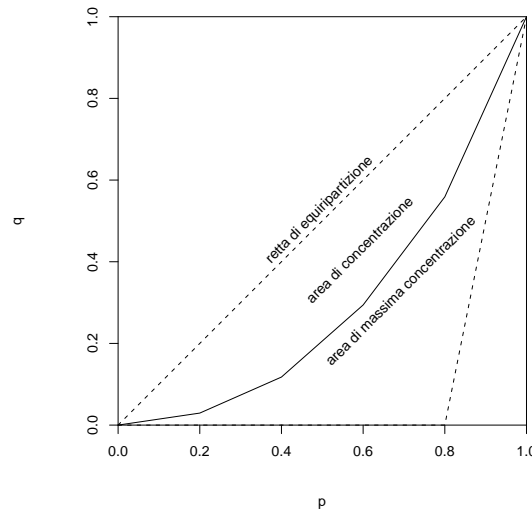


Figura 2.1.1: spezzata di Lorenz

Sostituendo le n osservazioni con una variabile casuale continua X con supporto (a, b) , con $0 \leq a < b \leq \infty$, funzione di densità $f(x)$, funzione di ripartizione $F(x)$ e valore atteso $\mu > 0$, la spezzata diviene la *curva di Lorenz*, i cui punti hanno per ascissa la funzione di ripartizione del modello e per ordinata il primo momento incompleto, così definito

$$H(x) = \begin{cases} 0, & \text{per } x < a \\ \frac{1}{\mu} \int_a^x t f(t) dt, & \text{per } x \geq a. \end{cases} \quad (2.1.1)$$

La definizione della curva di Lorenz come funzione in $0 < p < 1$ è $L(p) = H(x_{(p)})$, con $x_{(p)} = F^{-1}(p)$ funzione quantile del modello. L'area di massima concentrazione diviene l'intero triangolo sottostante la retta di equiripartizione.

2.2 Indice di Gini

L'indice di concentrazione di Gini, R , è il rapporto fra area di concentrazione e area di massima concentrazione. Suddividendo l'area di concentrazione in trapezi, dopo alcuni passaggi si giunge

alla seguente espressione

$$R = \frac{2}{n-1} \sum_{i=1}^{n-1} (p_i - q_i), \quad (2.2.1)$$

si osserva che i termini della sommatoria sono non negativi poiché $p_i \geq q_i$ per $i = 1, 2, \dots, n$. I valori assunti dall'indice di Gini sono compresi tra 0 e 1. Assume valore 0 nel caso di equidistribuzione, infatti la spezzata di Lorenz coincide con la bisettrice e l'area compresa è nulla. Assume valore 1 nel caso di massima concentrazione poiché le aree che vengono rapportate sono le medesime. Un'altra forma in cui può essere espressa l'equazione (2.2.1) è

$$R = 1 - 2 \frac{1}{n-1} \sum_{i=1}^{n-1} q_i. \quad (2.2.2)$$

Sostituendo le n osservazioni con una variabile casuale continua X l'indice di concentrazione di Gini può essere valutato come segue

$$R = 1 - 2 \int_0^1 L(p) dp. \quad (2.2.3)$$

2.3 Indice di Bonferroni

Ricordando che

$$\sum_{i=1}^{n-1} p_i = \frac{1}{n} \sum_{i=1}^{n-1} i = \frac{n-1}{2}$$

e definendo gli scarti relativi

$$s_i = \frac{p_i - q_i}{p_i}, \text{ per } i = 1, 2, \dots, n-1$$

si osserva che l'indice di Gini può essere definito come la media aritmetica degli s_i ponderata con pesi p_i . Partendo dall'equazione (2.2.1) infatti

$$R = \frac{2}{n-1} \sum_{i=1}^{n-1} \frac{p_i - q_i}{p_i} \cdot p_i = \frac{\sum_{i=1}^{n-1} s_i \cdot p_i}{\sum_{i=1}^{n-1} p_i} = M_1(s_i; p_i)$$

dove la funzione $M_1(\cdot; \cdot)$ rappresenta la media aritmetica del primo argomento con pesi al secondo argomento. Se al posto di utilizzare i pesi p_i si adottano dei pesi uniformi si ottiene l'indice di Bonferroni

$$B = M_1(s_i; 1) = \frac{1}{n-1} \sum_{i=1}^{n-1} \frac{p_i - q_i}{p_i} = \frac{\sum_{i=1}^{n-1} (M - M_i)}{(n-1) \cdot M},$$

dove le M_i sono le medie parziali definite come segue

$$M_i = \frac{\sum_{j=1}^i x_j}{i} = M \cdot \frac{q_i}{p_i}$$

e $M = M_n = \bar{x}$. Nel caso di equiripartizione, tutte le medie parziali M_i sono uguali a M e le differenze $M - M_i$ sono nulle, e di conseguenza $B = 0$. Nel caso di massima concentrazione invece $M_i = 0$ per $i = 1, 2, \dots, n-1$ e le differenze $M - M_i = M$, conseguentemente $B = 1$.

2.4 Indice di Pietra

In letteratura è nota la relazione che lega l'indice di Gini con la media aritmetica \bar{x} e la differenza media semplice Δ

$$R = \frac{\Delta}{2\bar{x}}$$

con

$$\Delta = \frac{\sum_{i=1}^n \sum_{j=1}^n |x_i - x_j|}{n(n-1)}.$$

Quindi l'indice di Gini può essere espresso come il rapporto fra un indice di variabilità e due volte la media aritmetica. Analogamente, se a numeratore si pone lo scostamento medio assoluto dalla media aritmetica si ottiene l'indice di Pietra P

$$P = \frac{\frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}|}{2\bar{x}}. \quad (2.4.1)$$

Nel caso di equidistribuzione gli scarti assoluti dalla media aritmetica sono tutti nulli e conseguentemente anche P è pari a 0. Nel caso in cui una sola unità detiene tutto il reddito, $x_n = T$, la media aritmetica è $\bar{x} = x_n/n$, conseguentemente lo scostamento medio assoluto dalla media aritmetica vale $2\frac{n-1}{n}\bar{x}$, che al divergere di n tende dal basso a $2\bar{x}$. Quindi nel caso di massima concentrazione, al crescere della numerosità delle osservazioni, l'indice di Pietra tende dal basso ad 1.

Nel caso di una variabile casuale X con funzione di densità $f(x)$ definita sul supporto (a, b) e con valore atteso μ , l'indice di Pietra è definito come

$$P = \frac{E(|X - \mu|)}{2\mu} = \frac{1}{2\mu} \int_a^b |x - \mu| f(x) dx.$$

Pietra (1915) ha mostrato che la massima distanza fra la curva di Lorenz $L(p)$ e la retta di equipartizione, $y(p) = p$ per $0 < p < 1$, si ha per $p^* = F(\mu)$ e che tale distanza coincide con l'indice P

$$P = p^* - L(p^*) = F(\mu) - H(\mu). \quad (2.4.2)$$

2.5 Curva $I_{(p)}$ e indice I di Zenga

Si considerino N osservazioni che presentano $s \leq N$ valori distinti $0 \leq x_1 \leq x_2 \leq \dots \leq x_j \leq \dots \leq x_s$ rispettivamente con frequenze n_j per $j = 1, 2, \dots, s$. Siano

$$\begin{aligned} N_j &= \sum_{i=1}^j n_i & j &= 1, 2, \dots, s \\ N &= N_s = \sum_{j=1}^s n_j \\ p_j &= \frac{N_j}{N} & j &= 1, 2, \dots, s \\ Q_j &= \sum_{i=1}^j x_i n_i & i &= 1, 2, \dots, s \\ T &= Q_s = \sum_{j=1}^s x_j n_j \end{aligned}$$

Siano $M = T/N$ la media totale e rispettivamente la media inferiore e la media superiore

$$\begin{aligned} \bar{M}_{(p_j)} &= \frac{Q_j}{N_j} & j &= 1, 2, \dots, s \\ {}^+M_{(p_j)} &= \begin{cases} \frac{T-Q_j}{N-N_j} & \text{per } j = 1, 2, \dots, s-1 \\ x_{s+1}^* & \text{per } j = s, \text{ dove } x_{s+1}^* \geq x_s \text{ può essere posto uguale a } x_s. \end{cases} \end{aligned}$$

Per misurare l'ineguaglianza è possibile usare l'indice puntuale costituito dalla variazione relativa della media inferiore rispetto quella superiore

$$I_{(p_j)} = \frac{\bar{M}_{(p_j)} - \bar{M}_{(p_j)}}{\bar{M}_{(p_j)}} = 1 - \frac{\bar{M}_{(p_j)}}{\bar{M}_{(p_j)}} = 1 - U_{(p_j)} \quad (2.5.1)$$

dove il rapporto $U_{(p_j)}$, in termini di redditi, misura l'uniformità fra il gruppo costituito dalla frazione di popolazione p_j più povera ed il gruppo costituito dalla frazione di popolazione $1 - p_j$ più ricca. Si osserva che $\bar{M}_{(p_j)} \geq \bar{M}_{(p_j)}$ e quindi $0 \leq I_{(p_j)} \leq 1$ per $j = 1, 2, \dots, s$. Nel caso le N osservazioni dovessero essere uguali, allora la media inferiore è uguale alla media superiore, il rapporto $U_{(p_j)}$ vale 1 e l'indice puntuale $I_{(p_j)}$ è nullo. Nel caso il reddito dovesse essere tutto detenuto dalla frazione di popolazione $1 - p_j$ più ricca allora la media inferiore sarebbe pari a 0 e l'indice puntuale $I_{(p_j)}$ sarebbe pari a 1. Il diagramma è rappresentato nel quadrato unitario, nelle ascisse sono riportate le frequenze cumulate e nelle ordinate i valori dell'indice puntuale, ed è costituito da s rettangoli affiancati. Il primo ha base $[0, p_1]$ e altezza $I_{(p_1)}$ e i restanti hanno base $[p_{j-1}, p_j]$ e altezza $I_{(p_j)}$ per $j = 2, 3, \dots, s$. L'ampiezza della base dei rettangoli è n_j e la somma delle aree di essi costituisce l'indice globale di ineguaglianza I di Zenga

$$I = \sum_{j=1}^s I_{(p_j)} \frac{n_j}{N}. \quad (2.5.2)$$

La linea egualitaria corrisponde con l'asse delle ascisse $I_{(p_j)} = 0$. Nel caso di massima ineguaglianza, nel quale un solo individuo detiene tutta la ricchezza, il diagramma è costituito da due rettangoli, il primo di base $(N - 1)/N$ e altezza 1 ed il secondo di base $1/N$ e altezza $(N - 1)/N^1$, conseguentemente l'indice globale nel caso di massima ineguaglianza è $I = 1 - 1/N^2$. In figura 2.5.1 è riportato un esempio del diagramma di Ineguaglianza. A differenza della spezzata di Lorenz tale diagramma non ha un comportamento predeterminato. La linea orizzontale che attraversa il diagramma rappresenta il valore dell'indice I .

¹in questo caso, $s = 2$ e $p_2 = 1$, quindi il gruppo più povero include anche l'unica unità che detiene tutto il reddito, conseguentemente la media inferiore è pari a $x_s/N > 0$

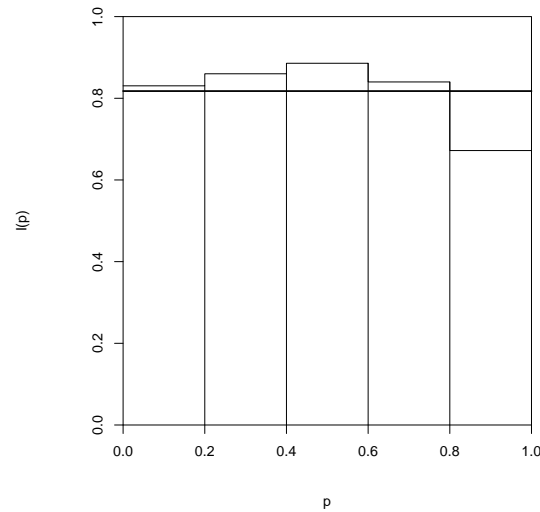


Figura 2.5.1: diagramma di Ineguaglianza

Sia X una variabile casuale non negativa e continua con supporto (a, b) , funzione di densità $f(x)$, funzione di ripartizione $F(x)$ e valore atteso $0 < \mu < \infty$, definendo la media inferiore e la media superiore rispettivamente come

$$\begin{aligned}\bar{\mu}_{(x)} &= \frac{1}{F(x)} \int_a^x t f(t) dt \\ \mu_{(x)}^+ &= \frac{1}{1 - F(x)} \int_x^b t f(t) dt\end{aligned}$$

e operando il cambio di variabile mediante $x_{(p)} = F^{-1}(p)$ per $0 < p < 1$, si ha

$$\begin{aligned}\bar{\mu}_{(F^{-1}(p))} &= \bar{M}_{(p)} = \frac{1}{p} \int_0^p x_{(t)} dt \\ \mu_{(F^{-1}(p))}^+ &= M_{(p)}^+ = \frac{1}{1-p} \int_p^1 x_{(t)} dt.\end{aligned}$$

La curva $I_{(p)}$ per variabili casuali continue è quindi definita come

$$I_{(p)} = 1 - \frac{\bar{M}_{(p)}}{M_{(p)}^+} = 1 - U_{(p)} \quad \text{per } 0 < p < 1$$

e l'indice di ineguaglianza globale I

$$I = \int_0^1 I_{(p)} dp = 1 - \int_0^1 U_{(p)} dp \quad (2.5.3)$$

che assume valori tra 0 (eguaglianza distributiva) e 1 (massima ineguaglianza). L'indice di ineguaglianza globale I , quale media integrale della curva $I_{(p)}$ rappresenta il valore costante che deve

assumere la curva affinché l'area sottesa, e quindi la misura di ineguaglianza complessiva, resti immutata. Se la curva $I_{(p)}$ è un valore costante $1 - k$, con $0 < k < 1$, su tutto l'intervallo di valori che può assumere p , l'indice puntuale $I_{(p)}$ fornisce la medesima informazione dell'indice globale I . In termini di redditi, qualunque sia la soglia di reddito che bipartisce la popolazione, la media dei redditi del gruppo più povero è sempre k volte la media dei redditi del gruppo più ricco (Polisicchio, 2008).

Zenga (2007) ha fornito la seguente relazione tra la curva di Lorenz $L(p)$ e la curva $I_{(p)}$

$$I_{(p)} = \frac{p - L(p)}{p[1 - L(p)]}, \quad (0 < p < 1). \quad (2.5.4)$$

Attraverso tale relazione, nei casi di modelli distributivi per i quali è nota la forma analitica della curva di Lorenz è facile ottenere la forma analitica della curva $I_{(p)}$. Sarabia (2008) ha fornito un'analisi delle curve di Lorenz note in forma parametrica. Nella tabella seguente vengono riportate le forme parametriche della curva di Lorenz e della curva $I_{(p)}$ ottenuta mediante la relazione (2.5.4) per quattro modelli distributivi per variabili non negative.

Distribuzione	Curva di Lorenz	Curva $I_{(p)}$
Uniforme $\mathcal{U}[a, b]$	$L(p) = \frac{2ap + (b-a)p^2}{a+b}$	$I_{(p)} = \frac{(1-p)(b-a)}{(a+b) - 2ap - (b-a)p^2}$
Esponenziale *	$L(p) = p + \left(1 + \frac{\mu}{\sigma}\right)^{-1} (1-p) \log(1-p)$	$I_{(p)} = \frac{1}{p} \left[1 - \frac{1 + \mu/\sigma}{\log(1-p)}\right]^{-1}$
Pareto	$L(p) = 1 - (1-p)^{1-1/\alpha}$	$I_{(p)} = \frac{1}{p} \left[1 - (1-p)^{1/\alpha}\right]$
Lognormale	$L(p) = \phi(\phi^{-1}(p) - \sigma)$	$I_{(p)} = \frac{p - \phi(\phi^{-1}(p) - \sigma)}{p[1 - \phi(\phi^{-1}(p) - \sigma)]}$

Tabella 2.1: Curve di Lorenz e curve $I_{(p)}$ di classici modelli distributivi per variabili non negative.

* distribuzione esponenziale con funzione di ripartizione $F(x) = 1 - e^{-(x-\mu)/\sigma}$ se $x > \mu$, con $\mu, \sigma > 0$.

2.6 Requisiti degli indici di concentrazione

L'indice globale di ineguaglianza di Zenga gode delle seguenti proprietà (si veda Zenga, 2010):

1. assume valore 0 in presenza di equiripartizione;
2. in caso di massima concentrazione, è uguale a una funzione C_n del numero n di osservazioni, crescente e tale che: $\lim_{n \rightarrow \infty} C_n = 1$;
3. è invariante alle trasformazioni di scala;
4. diminuisce se tutti i valori x_i vengono aumentati di una quantità $h > 0$;
5. è essere sensibile ai trasferimenti.

Tali proprietà sono requisiti che un indice di concentrazione deve rispettare. Tutti gli indici globali che sono stati presentati rispettano i primi quattro requisiti. Il quinto requisito necessita di ulteriori precisazioni. Bonferroni propose la seguente definizione di sensibilità ai trasferimenti: l'indice deve aumentare se, essendo $x_i < x_j$, si sostituiscono x_i e x_j con $x_i - h$ e $x_j + h$, posto $0 < h \leq x_j$. Data tale definizione tutti gli indici globali presentati, ad eccezione dell'indice di Pietra, rispettano la sensibilità ai trasferimenti.

Capitolo 3

Modelli distributivi

3.1 Variabile casuale di Dagum

Dagum (1977) ha formulato un modello per le distribuzioni del reddito partendo dalle seguenti assunzioni:

1. esse sono distribuzioni unimodali ad asimmetria positiva;
2. può esistere una percentuale α di unità economiche con reddito nullo;
3. il supporto delle distribuzioni del reddito è l'intervallo semiaperto $[x_0, \infty)$ dove x_0 rappresenta il reddito minimo;
4. l'elasticità rispetto all'origine α , con $\alpha < 1$, della funzione di ripartizione dei redditi dev'essere una funzione monotona decrescente rispetto alla probabilità cumulata e convergere a un valore finito positivo quando il reddito tende a zero.

L'elasticità di F rispetto ad x è il rapporto tra la variazione relativa di F e la variazione relativa di x

$$\eta_{Fx} = \frac{\Delta F/F}{\Delta x/x}$$

nel caso di variazioni infinitesimali

$$\eta_{Fx} = \frac{dF/F}{dx/x} = \frac{d \log F}{d \log x}.$$

Contestualmente alle distribuzioni dei redditi, x rappresenta il reddito e F la funzione di ripartizione del modello di distribuzione, quindi l'elasticità esprime l'incremento percentuale della probabilità cumulata all'aumentare dell'1% del reddito. In altri termini η_{Fx} esprime, per ogni quota di reddito x , l'incremento percentuale di popolazione che si otterrebbe se, invece di considerare gli individui che detengono un reddito inferiore o uguale a x , si dovessero considerare gli individui con un reddito inferiore o uguale a x incrementato dell'1%.

Dalle quattro assunzioni enunciate, Dagum ha definito la seguente equazione differenziale per l'elasticità

$$\frac{d \log [F(x) - \alpha]}{d \log x} = \beta \delta \left[1 - \left(\frac{F - \alpha}{1 - \alpha} \right)^{1/\beta} \right],$$

con $x > x_0 \geq 0$ e $\alpha < 1$, $\beta > 0$ e $\delta > 0$, dove x_0 è l'estremo inferiore del supporto. La soluzione dell'equazione differenziale è

$$F(x) = \alpha + \frac{1 - \alpha}{(1 + \lambda x^{-\delta})^\beta}, \quad \lambda > 0. \quad (3.1.1)$$

La seconda assunzione è verificata mediante il parametro α . Se $0 \leq \alpha < 1$, α può essere interpretato come tasso di disoccupazione per le unità economiche considerate, in questo caso il reddito minimo è $x_0 = 0$; in figura 3.1.1 è riportata la funzione di ripartizione al fine di osservare l'effetto del parametro α sul modello, nel caso della funzione di densità in corrispondenza del punto x_0 la massa di probabilità non è nulla.

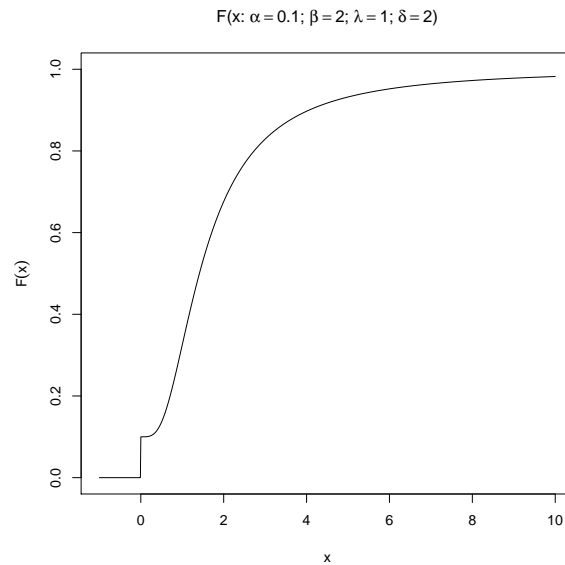


Figura 3.1.1: funzione di ripartizione della variabile casuale di Dagum nel caso in cui la proporzione α della popolazione non percepisce reddito

In questo caso il modello può essere scomposto come miscuglio

$$F(x) = \alpha F_1(x) + (1 - \alpha) F_2(x)$$

dove

$$F_1(x) = \begin{cases} 0, & \text{per } x < 0 \\ 1, & \text{per } x \geq 0 \end{cases}$$

e $F_2(x)$ è una funzione continua e differenziabile per ogni $x > 0$. Dopo aver definito il modello come un miscuglio è possibile derivare la funzione di ripartizione e ricavare la funzione di densità

$$f(x) = \begin{cases} \alpha, & \text{quando } x = 0 \\ (1 - \alpha)\beta\delta\lambda x^{-\delta-1}(1 + \lambda x^{-\delta})^{-\beta-1}, & \text{per ogni } x > 0 \\ 0, & \text{per ogni } x < 0. \end{cases} \quad (3.1.2)$$

Se il prodotto fra i parametri $\beta\delta > 1$ la distribuzione è unimodale, quindi è verificata la prima assunzione, altrimenti se $0 < \beta\delta \leq 1$ la distribuzione è zeromodale, ciò può verificarsi nel caso di un paese povero e sovrappopolato. Si dimostra che la moda x_M è

$$x_M = \lambda^{1/\delta} \left(\frac{\beta\delta - 1}{\delta + 1} \right)^{1/\delta}, \quad \beta\delta > 1.$$

In figura 3.1.2 è riportato il confronto fra le funzioni di densità.

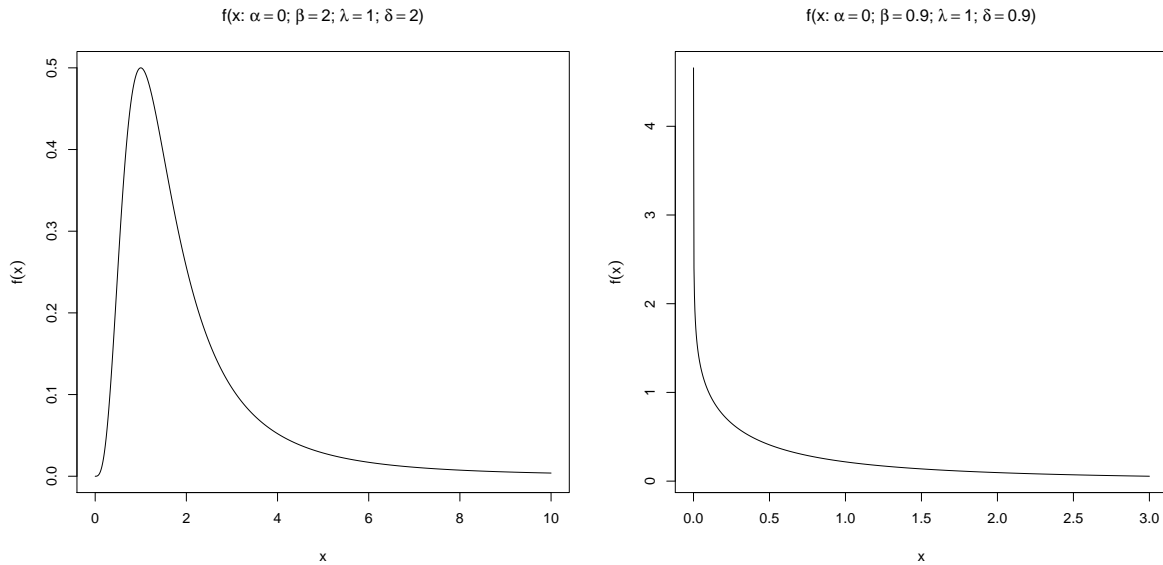


Figura 3.1.2: funzione di densità della variabile casuale di Dagum nel caso di forma unimodale e zeromodale

Con riferimento alla terza assunzione, il modello di Dagum può assumere anche un reddito minimo non nullo. $\alpha < 0$ implica $x_0 > 0$ il quale può essere ottenuto come soluzione dell'equazione $F(x_0) = 0$. A tal proposito si fa presente che la funzione di ripartizione è invertibile, conseguentemente la funzione quantile risulta

$$x_p = \lambda^{1/\delta} \left[\left(\frac{1 - \alpha}{p - \alpha} \right)^{1/\beta} - 1 \right]^{-1/\delta}, \quad p > \alpha \quad (3.1.3)$$

e x_0 è

$$x_0 = \lambda^{1/\delta} \left[\left(\frac{\alpha - 1}{\alpha} \right)^{1/\beta} - 1 \right]^{-1/\delta}.$$

In figura 3.1.3 è riportato un esempio della forma che il modello di Dagum assume nel caso di reddito minimo non nullo.

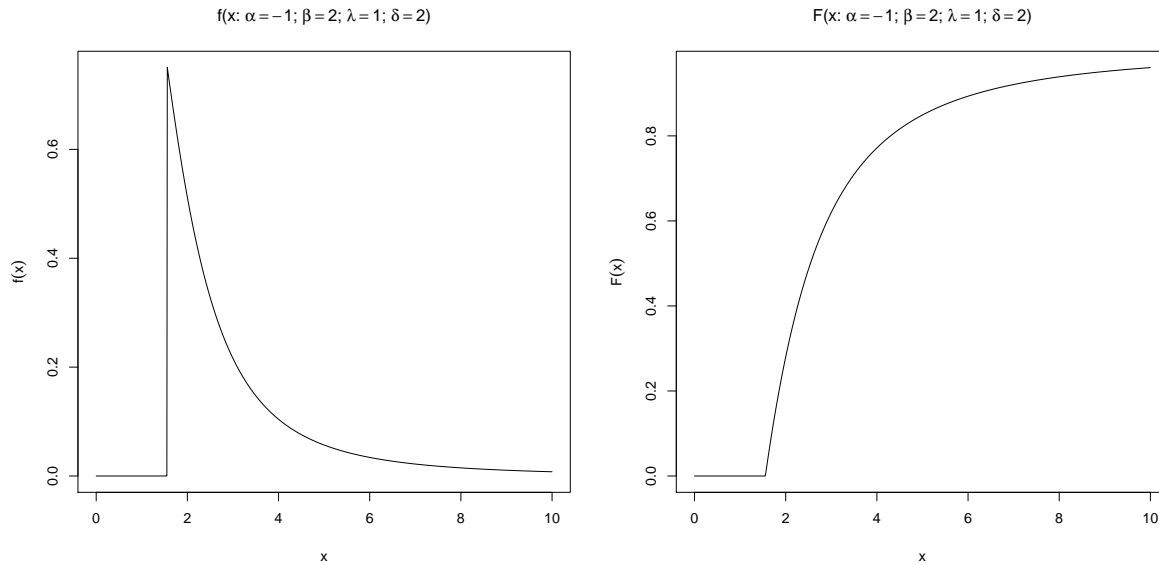


Figura 3.1.3: funzione di densità e funzione di ripartizione della variabile casuale di Dagum nel caso di reddito minimo non nullo

Dall'equazione (3.1.3) è possibile ricavare la mediana x_m

$$x_m = \begin{cases} \lambda^{1/\delta} \left[\left(\frac{1-\alpha}{0.5-\alpha} \right)^{1/\beta} - 1 \right]^{-1/\delta}, & \alpha < 0.5 \\ 0, & 0.5 \leq \alpha < 1. \end{cases}$$

Si può dimostrare che il momento di ordine r dall'origine è, per ogni intero positivo $r < \delta$

$$E(X^r) = (1 - \alpha)\beta\lambda^{r/\delta} B\left(1 - \frac{r}{\delta}; \beta + \frac{r}{\delta}\right), \text{ quando } 0 \leq \alpha < 1, \quad (3.1.4)$$

e

$$E(X^r) = (1 - \alpha)\beta\lambda^{r/\delta} BI\left(\frac{\lambda}{\lambda + x_0^\delta}; 1 - \frac{r}{\delta}; \beta + \frac{r}{\delta}\right), \text{ quando } \alpha < 0,$$

dove $B(a; b)$ è la funzione Beta e $BI(x; a; b)$ è la funzione Beta incompleta, definite come segue

$$\begin{aligned} B(a; b) &= \int_0^1 t^{a-1} (1-t)^{b-1} dt \\ BI(x; a; b) &= \int_0^x t^{a-1} (1-t)^{b-1} dt, \quad 0 \leq x < 1 \end{aligned}$$

e la funzione Beta incompleta normalizzata è definita da

$$B(x; a; b) = \frac{BI(x; a; b)}{B(a; b)}.$$

La curva di Lorenz associata al modello è

$$L(y) = B\left(y^{1/\beta}; \beta + \frac{1}{\delta}; 1 - \frac{1}{\delta}\right), \quad \delta > 1, \quad 0 \leq \alpha < 1, \quad (3.1.5)$$

con

$$y = \frac{p - \alpha}{1 - \alpha}, \quad \text{per } p \in [\alpha, 1]$$

e il corrispondente rapporto di concentrazione di Gini è

$$R = (2\alpha - 1) + (1 - \alpha) \frac{\Gamma(\beta)\Gamma(2\beta + \frac{1}{\delta})}{\Gamma(2\beta)\Gamma(\beta + \frac{1}{\delta})}$$

dove $\Gamma(a)$ è la funzione Gamma definita come segue

$$\Gamma(a) = \int_0^\infty t^{a-1} e^{-t} dt.$$

Si osserva che l'indice di Gini è funzione crescente di α e funzione decrescente di β e δ , inoltre non dipende da λ e si può facilmente riscontrare che $\lambda^{1/\delta}$ è *parametro di scala*. Sia $f(\cdot : a)$, $a > 0$, una famiglia di densità indicizzate da un parametro reale a . Il parametro a è definito *parametro di scala* se e solo se la densità $f(x : a)$ può essere scritta come $(1/a)f(x/a : 1)$ (Mood *et al.*, 1988). Conseguentemente la funzione di ripartizione $F(x; a)$ può essere scritta come $F(x/a; 1)$.

Per la relazione (2.5.4) la curva $I_{(p)}$ è data da

$$I_{(y)} = \frac{(1 - \alpha)y + \alpha - B(y^{1/\beta} : \beta + \frac{1}{\delta}; 1 - \frac{1}{\delta})}{[(1 - \alpha)y + \alpha] [1 - B(y^{1/\beta} : \beta + \frac{1}{\delta}; 1 - \frac{1}{\delta})]}, \quad \delta > 1, 0 \leq \alpha < 1, \quad (3.1.6)$$

con

$$y = \frac{p - \alpha}{1 - \alpha}, \quad \text{per } p \in [\alpha, 1].$$

Al fine di completare il quadro di forme che il modello può assumere, in figura 3.1.4 vengono riportati i grafici¹ delle funzioni di densità al variare di β e δ , ponendo $\alpha = 0$ e $\lambda = 1$, per cui anche il parametro di scala resta invariato nonostante δ assume valori differenti.

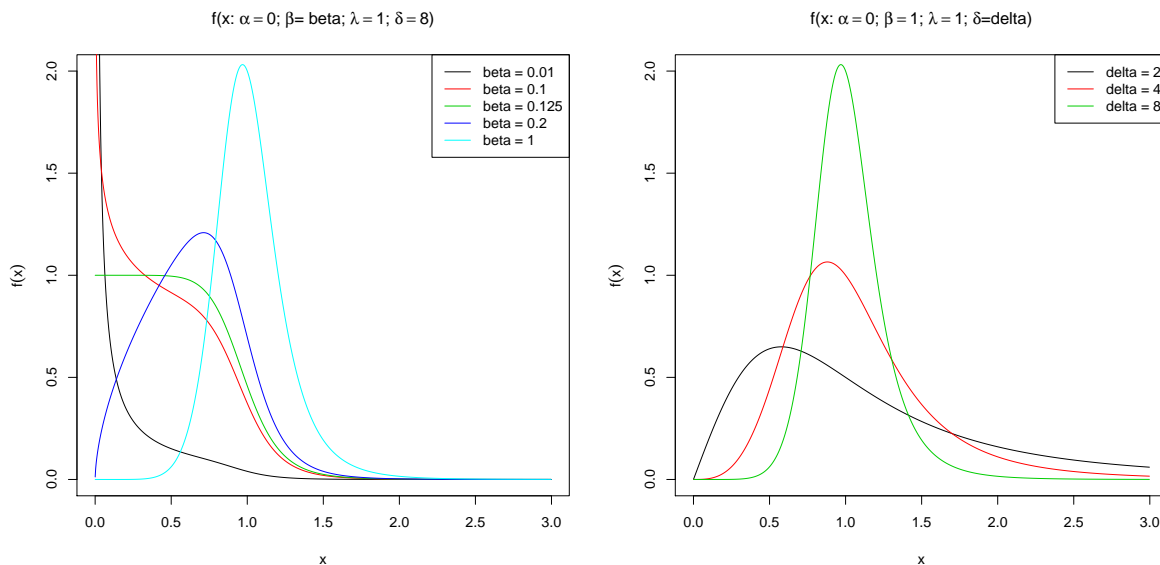


Figura 3.1.4: funzione di densità della variabile casuale di Dagum al variare dei parametri β e δ

¹i grafici riportati riproducono quelli pubblicati da Kleiber (2008)

Si può osservare che la funzione di densità per $\beta\delta = 1$ in $x = 0$ assume valore non nullo e finito, al variare di β e δ in $x = 0$ può avere derivata strettamente positiva e al variare del parametro β può presentare punti di flesso.

A seconda dei valori che può assumere α , si configurano tre tipologie di modelli di Dagum:

Dagum di tipo I: con $\alpha = 0$, è un modello a tre parametri con supporto $[0, \infty)$;

Dagum di tipo II: con $0 \leq \alpha < 1$, è un modello a quattro parametri con supporto $[0, \infty)$ e può essere espresso come miscuglio fra una variabile casuale degenere (con l'intera massa di probabilità in corrispondenza di $x = 0$) e una Dagum di tipo I;

Dagum di tipo III: con $\alpha < 0$, è un modello a quattro parametri con supporto $[x_0, \infty)$.

Di seguito sarà preso in considerazione solo il modello Dagum di tipo I.

Attraverso le equazioni (3.1.5) e (3.1.6) si forniscono alcuni grafici della curva di Lorenz e della curva $I_{(p)}$ per il modello di Dagum tipo I, al variare dei parametri β e δ , poiché esse non dipendono dal parametro λ .

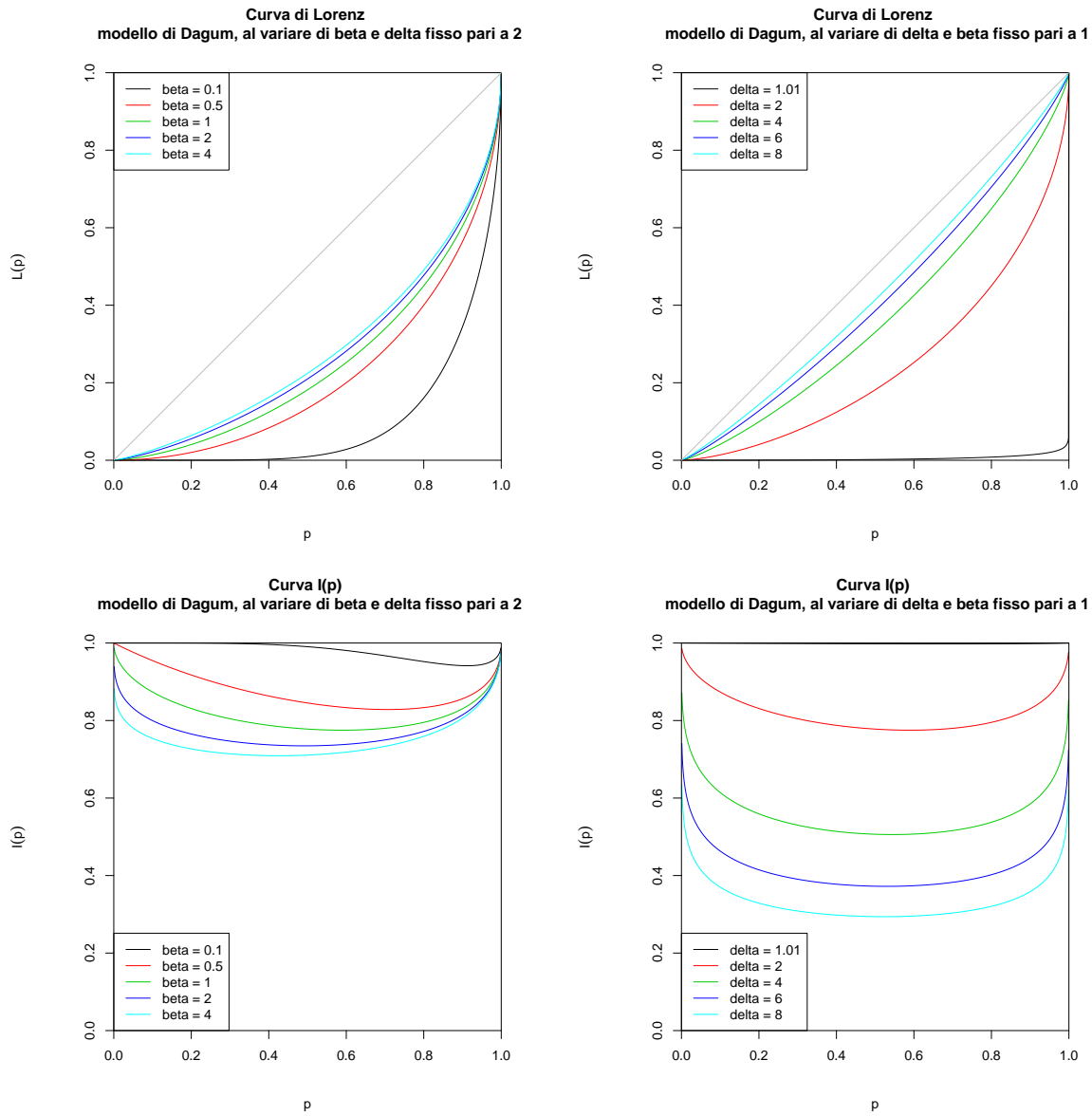


Figura 3.1.5: curva di Lorenz e curva $I(p)$ della variabile casuale di Dagum al variare dei parametri β e δ

Dalla figura 3.1.5 si può osservare che i parametri β e δ sono entrambi indici indiretti di concentrazione.

In conclusione si ricava la forma analitica dell'indice di Pietra per il modello di Dagum di tipo I. In primo luogo dall'equazione (3.1.4) si ricava il valore atteso (per $\alpha = 0$)

$$\mu = E(X) = \beta \lambda^{1/\delta} B\left(1 - \frac{1}{\delta}; \beta + \frac{1}{\delta}\right), \text{ per } \delta > 1 \quad (3.1.7)$$

quindi sostituendo il valore ottenuto nell'equazione (3.1.1) si ottiene $F(\mu)$

$$F(\mu) = \left(1 + \beta^{-\delta} B\left(1 - \frac{1}{\delta}; \beta + \frac{1}{\delta}\right)^{-\delta}\right)^{-\beta},$$

è rilevante osservare che tale valore non dipende dal parametro λ e quindi dal parametro di scala. Operando il cambio di variabile $p = F(x)$ nell'equazione (3.1.5) si ottiene il primo momento incompleto

$$H(x) = L(F(x)) = B\left(F(x)^{1/\beta} : \beta + \frac{1}{\delta}; 1 - \frac{1}{\delta}\right), \text{ per } \delta > 1.$$

Dall'equazione (2.4.2) si ricava la definizione dell'indice di Pietra per il modello di Dagum tipo I

$$P = \left(1 + \beta^{-\delta} B\left(1 - \frac{1}{\delta}; \beta + \frac{1}{\delta}\right)^{-\delta}\right)^{-\beta} - B\left(\frac{1}{1 + \beta^{-\delta} B\left(1 - \frac{1}{\delta}; \beta + \frac{1}{\delta}\right)^{-\delta}} : \beta + \frac{1}{\delta}; 1 - \frac{1}{\delta}\right)$$

3.2 Variabile casuale di Zenga

La variabile casuale di Zenga è un miscuglio con distribuzione condizionata una particolare Pareto troncata (determinata da Poliscchio, 2008) e distribuzione dei pesi Beta.

3.2.1 La distribuzione condizionata

Poliscchio (2008) ha dimostrato che l'unica variabile casuale, con valore atteso μ positivo e finito, per la quale la misura di ineguaglianza puntuale $I_{(p)}$ è costante pari a $0 < (1 - k) < 1$ per ogni $0 < p < 1$ è una Pareto troncata con tradizionale parametro di ineguaglianza $\theta = 0.5$.

Si ricorda che la funzione di ripartizione della Pareto di parametri $x_0, \theta > 0$ con supporto $[x_0, \infty)$ è

$$F(x : \theta; x_0) = 1 - \left(\frac{x_0}{x}\right)^\theta, \quad x_0 \leq x$$

e la funzione di densità è

$$f(x : \theta; x_0) = \theta x_0^\theta x^{-(\theta+1)}, \quad x_0 \leq x.$$

Limitando il supporto a destra a $x_1 < \infty$ e moltiplicando per un fattore di normalizzazione, pari al reciproco di $F(x_1 : \theta, x_0)$ in modo tale che la funzione di densità integri a 1, si ottengono la funzione di ripartizione della Pareto troncata con supporto $[x_0, x_1]$

$$F(x : \theta; x_0; x_1) = \frac{1 - \left(\frac{x_0}{x}\right)^\theta}{1 - \left(\frac{x_0}{x_1}\right)^\theta}, \quad x_0 \leq x \leq x_1$$

e la funzione di densità

$$f(x : \theta; x_0; x_1) = \frac{\theta x_0^\theta x^{-(\theta+1)}}{1 - \left(\frac{x_0}{x_1}\right)^\theta}, \quad x_0 \leq x \leq x_1.$$

Ponendo $\theta = 0.5$, $x_0 = \mu k$ e $x_1 = \mu/k$ si ottiene la Pareto troncata di Poliscchio. Il supporto è l'intervallo chiuso e limitato $[\mu k, \mu/k]$. La funzione di ripartizione è

$$V(x : \mu; k) = \begin{cases} 0, & x \leq \mu k \\ \frac{1}{1-k} \left[1 - \sqrt{\frac{\mu k}{x}}\right], & \mu k < x < \mu/k \\ 1, & x \geq \mu/k \end{cases} \quad (3.2.1)$$

e la funzione di densità

$$v(x; \mu; k) = \begin{cases} \frac{\sqrt{\mu k}}{2(1-k)} x^{-1.5}, & \mu k \leq x \leq \mu/k \\ 0, & \text{altrove.} \end{cases} \quad (3.2.2)$$

Il momento di ordine r dall'origine è

$$\begin{aligned} \mu'_r = E(X^r) &= \int_{\mu k}^{\mu/k} x^{0.5} \frac{\sqrt{\mu k}}{2(1-k)} x^{-1.5} dx = \\ &= \frac{\sqrt{\mu k}}{2(1-k)} \int_{\mu k}^{\mu/k} x^{r-1.5} dx = \\ &= \frac{\sqrt{\mu k}}{2(1-k)} \left[\frac{x^{r-0.5}}{r-0.5} \right]_{\mu k}^{\mu/k} = \quad \text{per } r \neq 0.5 \\ &= \frac{\sqrt{\mu k}}{2(1-k)} \frac{\mu^r}{2r-1} \left[k^{-r} \left(\frac{\mu}{k} \right)^{-0.5} - k^r (\mu k)^{-0.5} \right] = \\ &= \frac{1}{2(1-k)} \frac{\mu^r}{2r-1} [k^{-r} k - k^r] = \\ &= \frac{\mu^r}{2r-1} \frac{k^{1-r}(1-k^{2r-1})}{1-k} \quad 0 < k < 1, \mu > 0, r \neq 0.5 \end{aligned} \quad (3.2.3)$$

mentre per $r = 0.5$ è pari a

$$\begin{aligned} \mu'_{0.5} = E(\sqrt{X}) &= \int_{\mu k}^{\mu/k} x^{0.5} \frac{\sqrt{\mu k}}{2(1-k)} x^{-1.5} dx = \\ &= \frac{\sqrt{\mu k}}{2(1-k)} \int_{\mu k}^{\mu/k} \frac{1}{x} dx = \\ &= \frac{\sqrt{\mu k}}{2(1-k)} [\ln x]_{\mu k}^{\mu/k} = \\ &= \frac{\sqrt{\mu k}}{2(1-k)} [\ln \mu - \ln k - \ln \mu - \ln k] = \\ &= \sqrt{\mu k} \frac{\ln k}{k-1} \quad 0 < k < 1, \mu > 0. \end{aligned}$$

Zenga *et al.* (2010c) hanno mostrato che, per $r \in \mathbb{N}$, l'equazione (3.2.3) può essere scritta come segue

$$\mu'_r = \frac{\mu^r}{2r-1} \sum_{i=1}^{2r-1} k^{i-r}.$$

Dai precedenti risultati consegue che la varianza (momento centrale di ordine 2) di una variabile casuale X con funzione di densità (3.2.2) è

$$\mu_2 = \text{Var}(X) = E(X^2) - [E(X)]^2 = \frac{\mu^2(k-1)^2}{3k}, \quad (3.2.4)$$

e il momento centrale di ordine $r = 3$ è

$$\mu_3 = E[(X - \mu)^3] = E(X^3) - 3E(X)\text{Var}(X) - E(X)^3 = \frac{\mu^3}{5k^2}(1-k)^4. \quad (3.2.5)$$

Per costruzione $I(p) = 1 - k$ per ogni $0 < p < 1$, inoltre dall'inversione dell'equazione (2.5.4)

$$L(p) = \frac{p[1 - I(p)]}{1 - pI(p)}, \quad 0 < p < 1$$

si ottiene

$$L(p) = \frac{pk}{1 - p(1 - k)}, \quad 0 < p < 1.$$

3.2.2 Funzione di densità

Si consideri ora il modello di Zenga (2010). Esso è un miscuglio di variabili casuali di Poliscchio, con funzione di densità $v(x : \mu, k)$, al variare del parametro k , e la distribuzione dei pesi è data dalla funzione di densità della variabile casuale Beta, di parametri $\alpha > 0$ e $\theta > 0$, così definita

$$g(k : \alpha; \theta) = \frac{k^{\alpha-1}(1-k)^{\theta-1}}{B(\alpha; \theta)} \quad 0 < k < 1$$

conseguentemente la funzione di densità del miscuglio è

$$f(x : \mu; \alpha; \theta) = \int_0^1 v(x : \mu; k)g(k : \alpha; \theta) dk.$$

dalla quale risulta la seguente forma esplicita

$$f(x : \mu; \alpha; \theta) = \begin{cases} \frac{1}{2\mu B(\alpha; \theta)} \left(\frac{x}{\mu}\right)^{-1.5} \int_0^{\frac{x}{\mu}} k^{\alpha+0.5-1}(1-k)^{\theta-2} dk, & 0 < x < \mu \\ \frac{1}{2\mu B(\alpha; \theta)} \left(\frac{\mu}{x}\right)^{1.5} \int_0^{\frac{\mu}{x}} k^{\alpha+0.5-1}(1-k)^{\theta-2} dk, & \mu < x, \end{cases} \quad (3.2.6)$$

il valore della densità $f(x : \mu; \alpha; \theta)$ in $x = \mu$ per $0 < \theta \leq 1$ può essere valutato soltanto al limite quindi

$$\lim_{x \rightarrow \mu} f(x : \mu; \alpha; \theta) = \begin{cases} \frac{B(\alpha+0.5; \theta-1)}{2\mu B(\alpha; \theta)}, & \text{per } \theta > 1 \\ \infty, & \text{per } 0 < \theta \leq 1. \end{cases} \quad (3.2.7)$$

Il risultato precedente è stato ottenuto da Zenga (2010) specificando che per $\theta > 1$ la densità del modello può essere espressa come segue

$$f(x : \mu; \alpha; \theta) = \begin{cases} \frac{B(\alpha+0.5; \theta-1)}{2\mu B(\alpha; \theta)} \left(\frac{x}{\mu}\right)^{-1.5} B\left(\frac{x}{\mu} : \alpha + 0.5; \theta - 1\right), & 0 < x \leq \mu \\ \frac{B(\alpha+0.5; \theta-1)}{2\mu B(\alpha; \theta)} \left(\frac{\mu}{x}\right)^{1.5} B\left(\frac{\mu}{x} : \alpha + 0.5; \theta - 1\right), & \mu < x \end{cases}$$

e Zenga *et al.* (2010c) hanno fornito un'altra forma per rappresentare la funzione di densità nel caso $\theta > 0$

$$f(x : \mu; \alpha; \theta) = \begin{cases} \frac{1}{2\mu} \sum_{i=1}^{\infty} \frac{B(\alpha-0.5+i; \theta)}{B(\alpha; \theta)} \left(\frac{x}{\mu}\right)^{-1.5} B\left(\frac{x}{\mu} : \alpha - 0.5 + i; \theta\right), & 0 < x < \mu \\ \frac{1}{2\mu} \sum_{i=1}^{\infty} \frac{B(\alpha-0.5+i; \theta)}{B(\alpha; \theta)} \left(\frac{\mu}{x}\right)^{1.5} B\left(\frac{\mu}{x} : \alpha - 0.5 + i; \theta\right), & \mu < x. \end{cases}$$

Di seguito sono riportati dei grafici che riproducono alcune delle forme che può assumere la funzione di densità al variare dei parametri α e θ . Il parametro μ , come si mostrerà nella sezione successiva, è parametro di scala, conseguentemente non comporta cambiamenti nella forma.

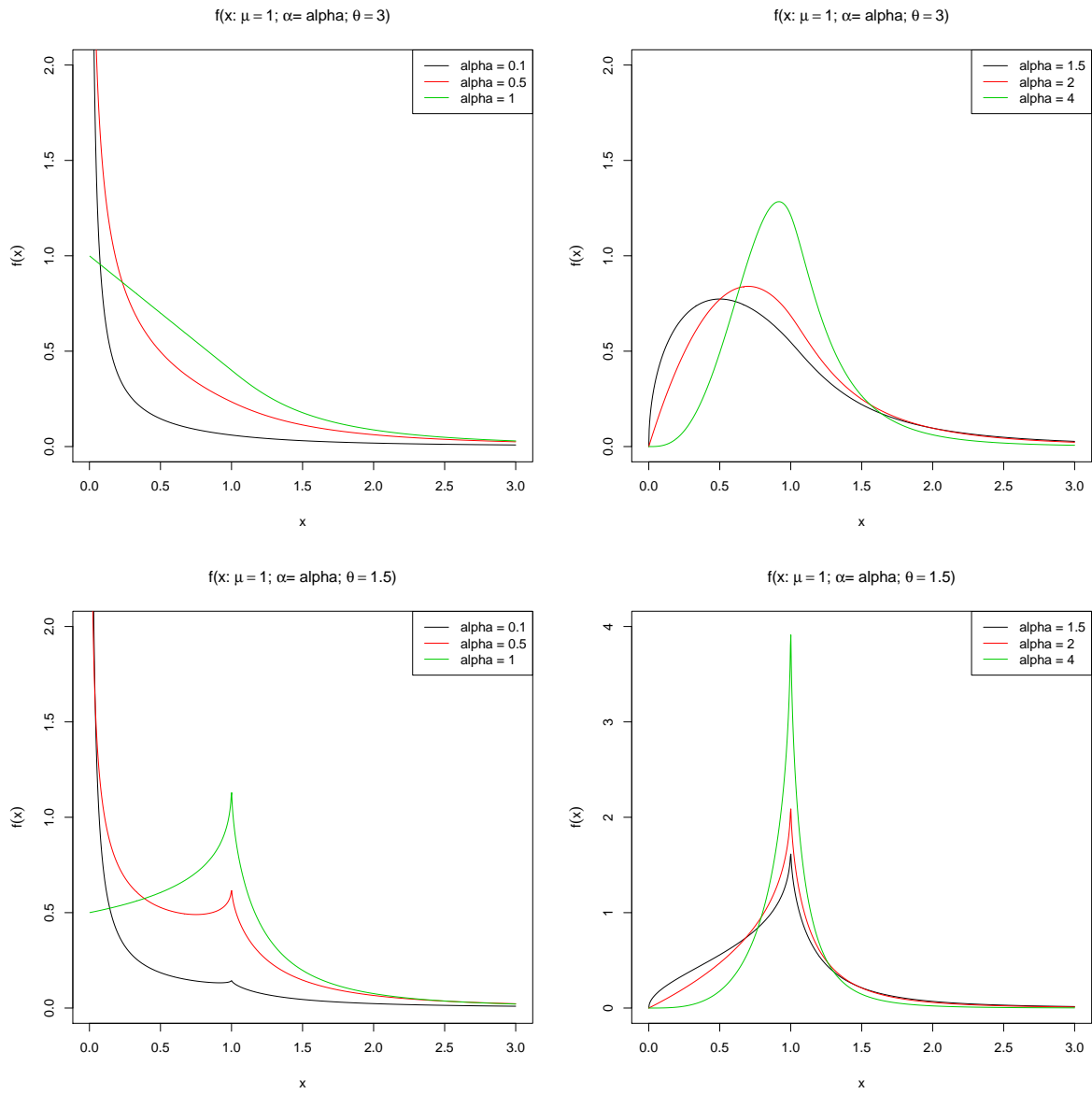


Figura 3.2.1: funzione di densità della variabile casuale di Zenga al variare del parametro α

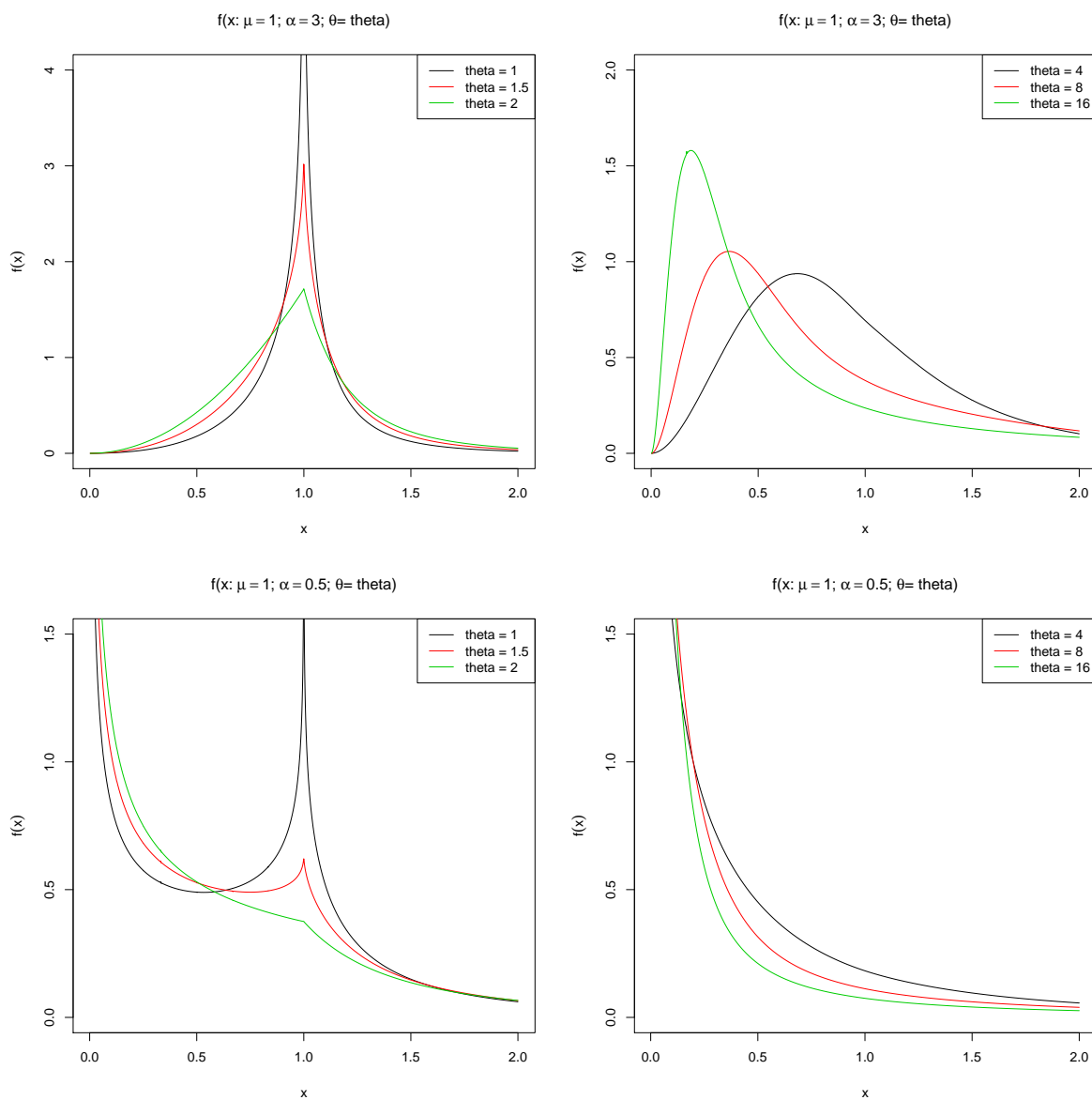


Figura 3.2.2: funzione di densità della variabile casuale di Zenga al variare del parametro θ

Dai grafici nelle figure 3.2.1 e 3.2.2 si può osservare che il parametro α influisce sulla forma della distribuzione nell'origine, mentre il parametro θ determina la forma della distribuzione per x pari al valore atteso μ . In particolare

$$\lim_{x \rightarrow 0} f(x : \mu; \alpha; \theta) = \begin{cases} \infty, & \text{per } 0 < \alpha < 1 \\ \frac{1}{3} \frac{\theta}{\mu}, & \text{per } \alpha = 1 \\ 0, & \text{per } \alpha > 1, \end{cases}$$

inoltre si osserva che al crescere di α , la derivata della funzione di densità in 0 è crescente. Al variare di θ il comportamento della distribuzione in μ è descritto dall'equazione (3.2.7).

3.2.3 Funzione di ripartizione

Zenga (2010) fornisce la funzione di ripartizione per $\theta > 1$ e per $0 < x \leq \mu$

$$\begin{aligned} F(x : \mu; \alpha; \theta) &= \int_0^{x/\mu} V(x : \mu; k)g(k : \alpha; \theta) dk = \\ &= \frac{BI\left(\frac{x}{\mu} : \alpha; \theta - 1\right)}{B(\alpha; \theta)} - \left(\frac{\mu}{x}\right)^{0.5} \frac{BI\left(\frac{x}{\mu} : \alpha + 0.5; \theta - 1\right)}{B(\alpha; \theta)}. \end{aligned} \quad (3.2.8)$$

Per l'asimmetria positiva, la mediana $x_{(0.5)}$, è minore del valore atteso, $0 < x_{(0.5)} < \mu$, per cui essa potrà essere valutata invertendo la precedente funzione posta uguale a 0.5

$$F(x_{(0.5)} : \mu; \alpha; \theta) = 0.5.$$

Come osservato per l'indice di Pietra, equazione (2.4.2), è rilevante determinare il valore della funzione di ripartizione in corrispondenza del valore atteso μ . Dall'equazione (3.2.8) si ottiene

$$F(\mu : \mu; \alpha; \theta) = \frac{B(\alpha; \theta - 1)}{B(\alpha; \theta)} - \frac{B(\alpha + 0.5; \theta - 1)}{B(\alpha, \theta)} \quad \text{per } \theta > 1.$$

Si osserva che il valore della funzione di ripartizione in corrispondenza del valore atteso non dipende dal parametro μ .

Si determina ora la forma analitica della funzione di ripartizione per valori di $x \geq \mu$. Per tali valori di x è necessario ricordare che $0 < k \leq \mu/x$ (per maggiori dettagli si vedano Zenga, 2010; Zenga *et al.*, 2010c). Inoltre è utile considerare la funzione di sopravvivenza della variabile casuale di Poliscchio, ottenuta come complemento a uno dell'equazione (3.2.1)

$$1 - V(x : \mu, k) = -k(1 - k)^{-1} + \left(\frac{\mu}{x}\right)^{0.5} k^{0.5}(1 - k)^{-1}.$$

Ne consegue che la funzione di sopravvivenza del miscuglio, per $x \geq \mu$ e $\theta > 1$ è

$$\begin{aligned} 1 - F(x : \mu; \alpha; \theta) &= \int_0^{\mu/x} \{1 - V(x : \mu; k)\} g(k : \alpha; \theta) dk = \\ &= \int_0^{\mu/x} \left\{ -k(1 - k)^{-1} + \left(\frac{\mu}{x}\right)^{0.5} k^{0.5}(1 - k)^{-1} \right\} \frac{k^{\alpha-1}(1 - k)^{\theta-1}}{B(\alpha; \theta)} dk = \\ &= \frac{1}{B(\alpha; \theta)} \left\{ - \int_0^{\mu/x} k^{(\alpha+1)-1}(1 - k)^{(\theta-1)-1} dk + \right. \\ &\quad \left. + \left(\frac{\mu}{x}\right)^{0.5} \int_0^{\mu/x} k^{(\alpha+0.5)-1}(1 - k)^{(\theta-1)-1} dk \right\} = \\ &= -\frac{BI\left(\frac{\mu}{x} : \alpha + 1; \theta - 1\right)}{B(\alpha; \theta)} + \left(\frac{\mu}{x}\right)^{0.5} \frac{BI\left(\frac{\mu}{x} : \alpha + 0.5; \theta - 1\right)}{B(\alpha; \theta)}. \end{aligned} \quad (3.2.9)$$

Si osserva che con il risultato appena ottenuto

$$F(\mu : \mu; \alpha; \theta) = \frac{B(\alpha + 1; \theta - 1) - B(\alpha + 0.5; \theta - 1) + B(\alpha; \theta)}{B(\alpha; \theta)}, \quad (3.2.10)$$

e ricordando la proprietà della funzione Beta

$$B(\alpha; \theta) = B(\alpha; \theta - 1) - B(\alpha + 1; \theta - 1)$$

si può concludere che anche mediante l'equazione (3.2.9) si giunge al medesimo valore in corrispondenza di μ dell'equazione (3.2.8). Riassumendo, per $\theta > 1$, la funzione di ripartizione del modello di Zenga è

$$F(x : \mu; \alpha; \theta) = \begin{cases} \frac{1}{B(\alpha; \theta)} \left\{ BI \left(\frac{x}{\mu}; \alpha; \theta - 1 \right) - \left(\frac{\mu}{x} \right)^{0.5} BI \left(\frac{x}{\mu}; \alpha + 0.5; \theta - 1 \right) \right\}, & 0 < x \leq \mu \\ 1 - \frac{1}{B(\alpha; \theta)} \left\{ \left(\frac{\mu}{x} \right)^{0.5} BI \left(\frac{\mu}{x}; \alpha + 0.5; \theta - 1 \right) - BI \left(\frac{\mu}{x}; \alpha + 1; \theta - 1 \right) \right\}, & \mu < x. \end{cases}$$

Zenga *et al.* (2010c) hanno fornito la forma analitica della funzione di ripartizione del modello per $\theta > 0$

$$F(x : \mu; \alpha; \theta) = \begin{cases} \frac{1}{B(\alpha; \theta)} \sum_{i=1}^{\infty} \left\{ BI \left(\frac{x}{\mu}; \alpha + i - 1; \theta \right) - \left(\frac{\mu}{x} \right)^{0.5} BI \left(\frac{x}{\mu}; \alpha + i - 0.5; \theta \right) \right\}, & 0 < x \leq \mu \\ 1 - \frac{1}{B(\alpha; \theta)} \sum_{i=1}^{\infty} \left\{ \left(\frac{\mu}{x} \right)^{0.5} BI \left(\frac{\mu}{x}; \alpha + i - 0.5; \theta \right) - BI \left(\frac{\mu}{x}; \alpha + i; \theta \right) \right\}, & \mu < x. \end{cases} \quad (3.2.11)$$

Per concludere la trattazione della funzione di ripartizione, si fa notare che, con entrambe le equazioni impiegate per definirla, è verificata la relazione

$$F(x : \mu; \alpha; \theta) = F(x/\mu : 1; \alpha; \theta)$$

conseguentemente μ è parametro di scala per il modello di Zenga e, come per il modello di Dagum, il valore della funzione di ripartizione in corrispondenza del valore atteso non dipende dal parametro di scala.

Di seguito sono riportati i grafici della funzione di ripartizione del modello di Zenga con gli stessi parametri dei grafici nelle figure 3.2.1 e 3.2.2.

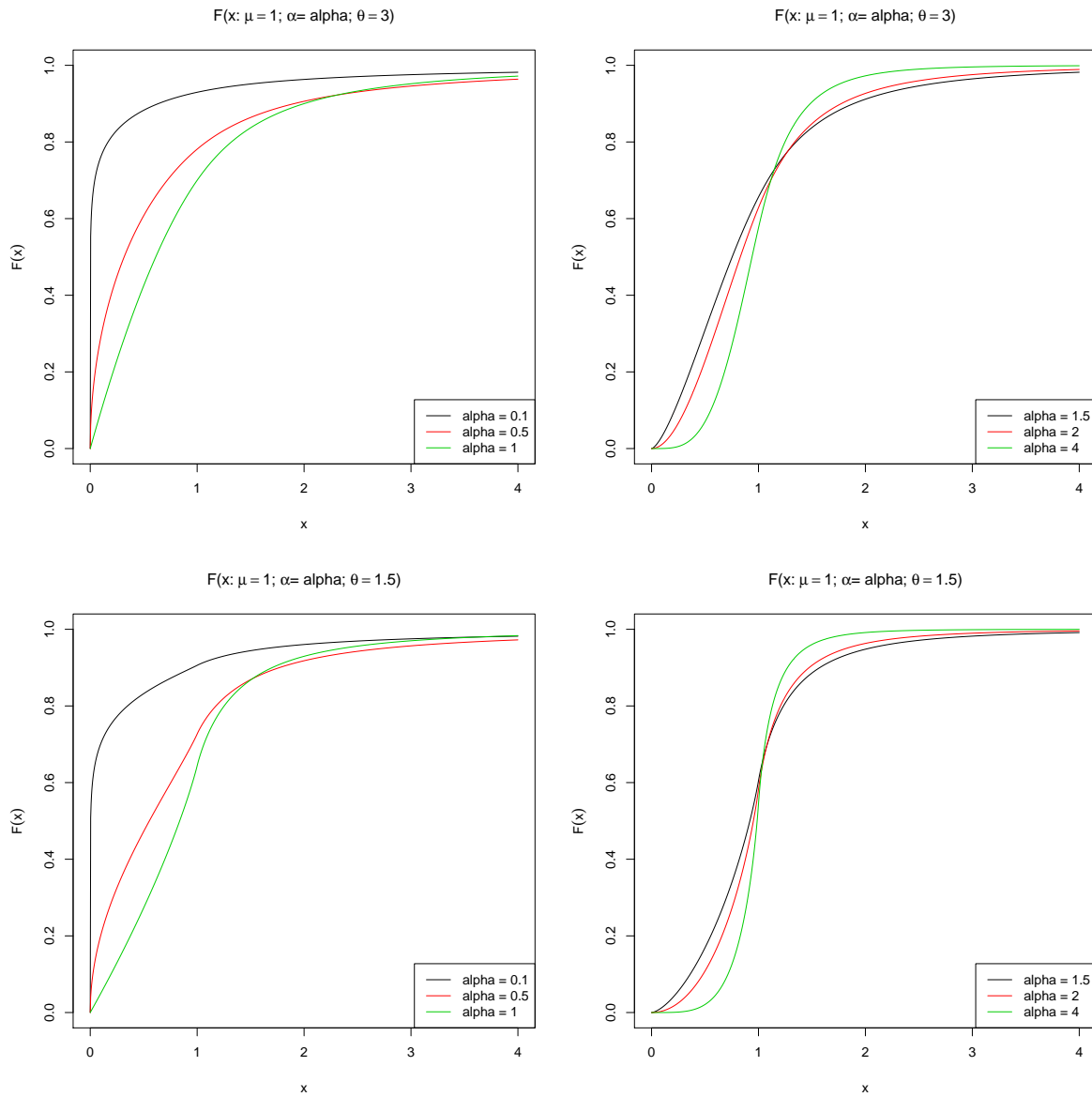


Figura 3.2.3: funzione di ripartizione della variabile casuale di Zenga al variare del parametro α

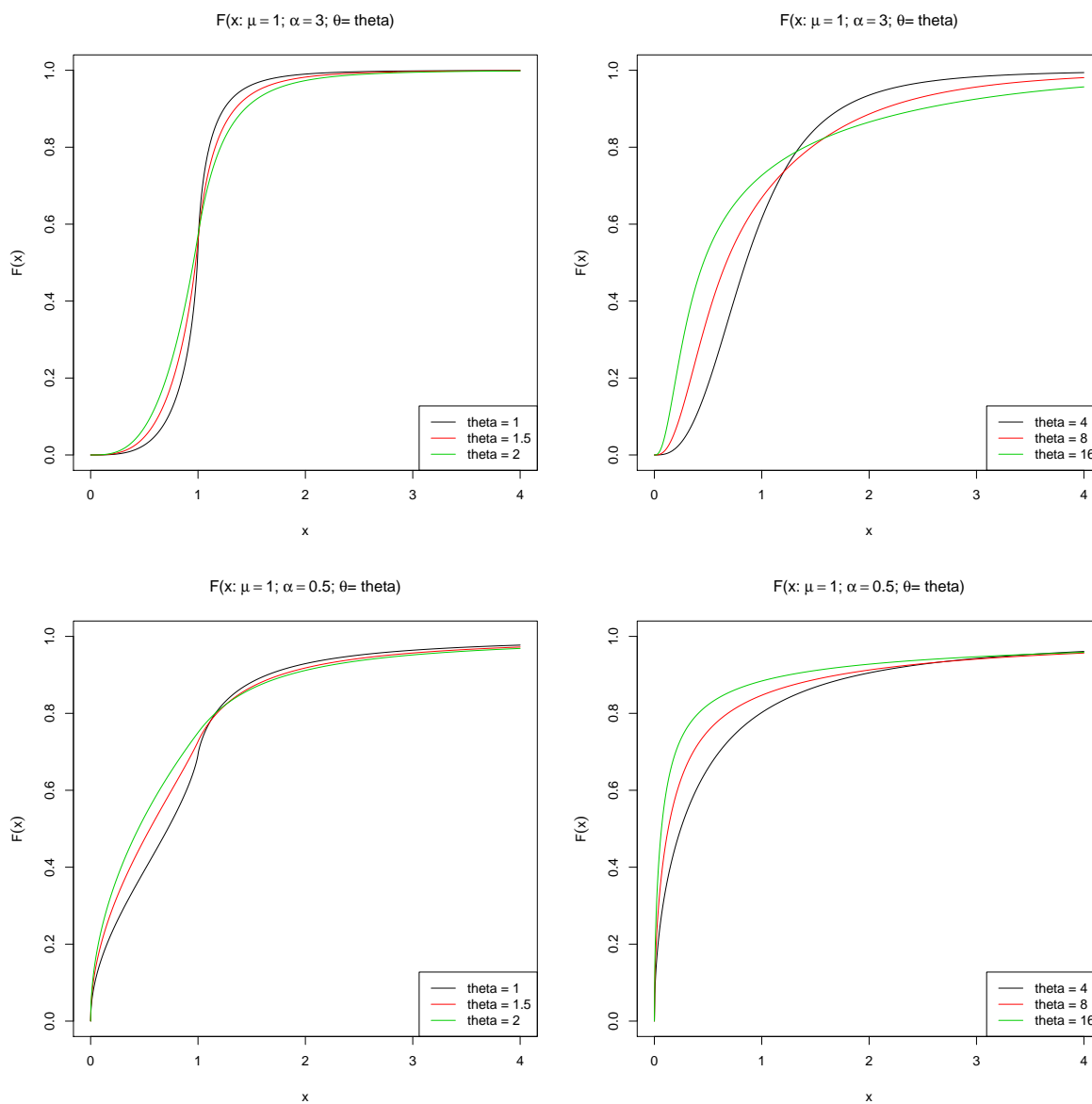


Figura 3.2.4: funzione di ripartizione della variabile casuale di Zenga al variare del parametro θ

3.2.4 Momenti dall'origine e momenti centrali

Il momento dall'origine di ordine r del modello di Zenga può essere ottenuto come momento centrale del miscuglio, ovvero, dato il momento dall'origine di ordine r della variabile casuale di Poliscchio

definito in equazione (3.2.3)

$$\begin{aligned}
\mu'_r = E(X^r) &= \int_0^1 \frac{\mu^r}{2r-1} \frac{k^{1-r}(1-k^{2r-1})}{1-k} g(k; \alpha; \theta) dk = \quad \text{per } r \neq 0.5 \\
&= \int_0^1 \frac{\mu^r}{2r-1} \frac{k^{1-r}(1-k^{2r-1})}{1-k} \frac{k^{\alpha-1}(1-k)^{\theta-1}}{B(\alpha; \theta)} dk = \\
&= \frac{\mu^r}{(2r-1)B(\alpha; \theta)} \int_0^1 (1-k^{2r-1})k^{\alpha-r}(1-k)^{\theta-2} dk = \\
&= \frac{\mu^r}{(2r-1)B(\alpha; \theta)} \left\{ \int_0^1 k^{(\alpha-r+1)-1}(1-k)^{(\theta-1)-1} dk - \int_0^1 k^{(\alpha+r)-1}(1-k)^{(\theta-1)-1} dk \right\}
\end{aligned}$$

che per $\theta > 1$ e $r < \alpha + 1$ e $r \neq 0.5$ può essere espresso come segue

$$\mu'_r = \frac{\mu^r}{(2r-1)B(\alpha; \theta)} \{B(\alpha-r+1; \theta-1) - B(\alpha+r; \theta-1)\}. \quad (3.2.12)$$

Si determina ora il momento dall'origine di ordine $r = 0.5^2$

$$\begin{aligned}
\mu'_{0.5} = E(\sqrt{X}) &= \int_0^1 \sqrt{\mu k} \frac{\ln k}{k-1} g(k; \alpha; \theta) dk = \\
&= \int_0^1 \sqrt{\mu k} \frac{\ln k}{k-1} \frac{k^{\alpha-1}(1-k)^{\theta-1}}{B(\alpha; \theta)} dk = \\
&= -\frac{\sqrt{\mu}}{B(\alpha; \theta)} \int_0^1 \ln k \cdot k^{(\alpha+0.5)-1}(1-k)^{(\theta-1)-1} dk = \\
&= -\frac{\sqrt{\mu}}{B(\alpha; \theta)} \frac{d}{d(\alpha+0.5)} B(\alpha+0.5; \theta-1) = \quad \text{per } \theta > 1 \\
&= \frac{\sqrt{\mu}}{B(\alpha; \theta)} B(\alpha+0.5; \theta-1) [\psi(\alpha+\theta-0.5) - \psi(\alpha+0.5)] \quad \text{per } \theta > 1,
\end{aligned}$$

²nei passaggi che seguono si fa uso delle seguenti relazioni

$$\begin{aligned}
\frac{d}{da} B(a, b) &= \frac{d}{da} \int_0^1 t^{a-1}(1-t)^{b-1} dt = \\
&= \int_0^1 (1-t)^{b-1} \left\{ \frac{d}{da} t^{a-1} \right\} dt = \\
&= \int_0^1 (1-t)^{b-1} \{t^{a-1} \ln t\} dt
\end{aligned}$$

poiché $|(1-t)^{b-1} \{t^{a-1} \ln t\}|$ esiste finito su $(0, 1)$ ed è integrabile su $[0, 1]$, e

$$\begin{aligned}
\frac{d}{da} B(a, b) &= \Gamma(b) \frac{d}{da} \frac{\Gamma(a)}{\Gamma(a+b)} = \\
&= \Gamma(b) \frac{\Gamma'(a)\Gamma(a+b) - \Gamma'(a+b)\Gamma(a)}{\Gamma^2(a+b)} = \\
&= \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \left[\frac{\Gamma'(a)}{\Gamma(a)} - \frac{\Gamma'(a+b)}{\Gamma(a+b)} \right] = \\
&= B(a, b) [\psi(a) - \psi(a+b)]
\end{aligned}$$

dove $\psi(z)$ è la funzione *Digamma*

$$\psi(z) = \frac{d}{dz} \ln \Gamma(z) = \frac{\Gamma'(z)}{\Gamma(z)}.$$

Zenga *et al.* (2010c) hanno fornito un'espressione per il momento dall'origine di ordine $r \in \mathbb{N}$ per $\theta > 0$

$$\mu'_r = \frac{\mu^r}{(2r-1)B(\alpha; \theta)} \sum_{i=1}^{2r-1} B(\alpha - r + i; \theta). \quad (3.2.13)$$

Dall'equazione (3.2.13) si possono ricavare i momenti dall'origine di ordine 1, 2 e 3 (Zenga *et al.*, 2010c):

$$\mu'_1 = \frac{\mu}{B(\alpha; \theta)} \cdot B(\alpha; \theta) = \mu \quad (3.2.14)$$

$$\begin{aligned} \mu'_2 &= \frac{\mu^2}{3B(\alpha; \theta)} \cdot [B(\alpha - 1; \theta) + B(\alpha; \theta) + B(\alpha + 1; \theta)] \\ &= \frac{\mu^2}{3} \cdot \left[\frac{\theta}{\alpha - 1} + \frac{\alpha}{\alpha + \theta} + 2 \right], \quad \text{per } \alpha > 1 \end{aligned} \quad (3.2.15)$$

$$\begin{aligned} \mu'_3 &= \frac{\mu^3}{5B(\alpha; \theta)} \cdot [B(\alpha - 2; \theta) + B(\alpha - 1; \theta) + B(\alpha; \theta) + B(\alpha + 1; \theta) + B(\alpha + 2; \theta)] = \\ &= \frac{\mu^3}{5} \cdot \left[\frac{\theta^4 + (5\alpha - 4)\theta^3 + (10\alpha^2 - 15\alpha + 1)\theta^2}{(\alpha - 2)(\alpha - 1)(\alpha + \theta)(\alpha + \theta + 1)} + \right. \\ &\quad \left. + \frac{2(5\alpha^3 - 10\alpha^2 + 3)\theta + 5\alpha(\alpha^3 - 2\alpha^2 - \alpha + 2)}{(\alpha - 2)(\alpha - 1)(\alpha + \theta)(\alpha + \theta + 1)} \right], \quad \text{per } \alpha > 2. \end{aligned} \quad (3.2.16)$$

L'equazione (3.2.14) avvalorata il procedimento con il quale è stata ottenuta la distribuzione, quale miscuglio di variabili casuali con stesso valore atteso pari a μ .

Nel caso le variabili componenti il miscuglio hanno lo stesso valore atteso, Zenga *et al.* (2010b) hanno mostrato che i momenti centrali del miscuglio possono essere ottenuti analogamente ai momenti dall'origine (ovvero mediante la media ponderata dei momenti centrali delle componenti del miscuglio). Poiché nel caso del modello di Zenga le variabili componenti il miscuglio hanno lo stesso valore atteso, per le equazioni (3.2.4) e (3.2.5), la varianza e il momento centrale di una variabile casuale X con funzione di densità (3.2.6) sono

$$\begin{aligned} \mu_2 &= \int_0^1 \frac{\mu^2(k-1)^2}{3k} \cdot g(k; \alpha; \theta) dk = \\ &= \frac{\mu^2}{3} \cdot \frac{\theta(\theta+1)}{(\alpha-1)(\alpha+\theta)}, \quad \text{per } \alpha > 1, \end{aligned} \quad (3.2.17)$$

$$\begin{aligned} \mu_3 &= \int_0^1 \frac{\mu^3}{5k^2} (1-k)^4 \cdot g(k; \alpha; \theta) dk = \\ &= \frac{\mu^3}{5} \cdot \frac{\theta(\theta+1)(\theta+2)(\theta+3)}{(\alpha-1)(\alpha-2)(\alpha+\theta+1)(\alpha+\theta)} = \\ &= \mu_2 \cdot \frac{3\mu(\theta+3)(\theta+2)}{5(\alpha+\theta+1)(\alpha-2)}, \quad \text{per } \alpha > 2. \end{aligned} \quad (3.2.18)$$

Esso è stato eguagliato al momento terzo centrale campionario per ottenere la soluzione analitica dello stimatore di θ secondo il metodo dei momenti. In tal modo è stata definita un'equazione di quarto grado in θ . Sfruttando però il fatto che essa è funzione della varianza è stato possibile ridurre la complessità dell'equazione riconducendola a una di secondo grado. Si osservi inoltre che il momento terzo centrale del modello assume solo valori positivi, questa considerazione permette di scegliere la soluzione ammissibile.

3.2.5 Misure di ineguaglianza del modello di Zenga

Per valutare gli indici puntuali di ineguaglianza è necessario determinare il momento primo incompleto, definito dall'equazione (2.1.1). Zenga *et al.* (2010c) hanno fornito il momento primo incompleto del modello di Zenga. Esso, come per la variabile casuale di Dagum, non dipende dal parametro di scala. Nel caso $\theta > 1$ risulta

$$H(x : 1; \alpha; \theta) = \begin{cases} \frac{1}{B(\alpha; \theta)} & [x^{0.5} IB(x : \alpha + 0.5; \theta - 1) - IB(x : \alpha + 1; \theta - 1)], & 0 < x \leq 1 \\ 1 - \frac{1}{B(\alpha; \theta)} & [IB(\frac{1}{x} : \alpha; \theta - 1) - x^{0.5} IB(\frac{1}{x} : \alpha + 0.5; \theta - 1)], & 1 < x. \end{cases}$$

Il caso più generale $\theta > 0$, è fornito mediante serie

$$H(x : 1; \alpha; \theta) = \begin{cases} \frac{1}{B(\alpha; \theta)} & \sum_{i=1}^{\infty} [x^{0.5} IB(x : \alpha + i - 0.5; \theta) - IB(x : \alpha + i; \theta)], & 0 < x \leq 1 \\ 1 - \frac{1}{B(\alpha; \theta)} & \sum_{i=1}^{\infty} [IB(\frac{1}{x} : \alpha + i - 1; \theta) - x^{0.5} IB(\frac{1}{x} : \alpha + i - 0.5; \theta)], & 1 < x. \end{cases}$$

La curva di Lorenz può essere rappresentata, al variare di x , ponendo in ascissa la funzione di ripartizione $F(x : 1; \alpha; \theta)$ e in ordinata i valori del momento primo incompleto valutato sui medesimi valori di x . Analogamente, per la relazione (2.5.4), la curva di ineguaglianza $I_{(p)}$ verrà rappresentata ponendo in ascissa la funzione di ripartizione $F(x : 1; \alpha; \theta)$ e in ordinata il corrispondente valore di

$$J(x : \alpha; \theta) = \frac{F(x : 1; \alpha; \theta) - H(x : 1; \alpha; \theta)}{F(x : 1; \alpha; \theta) [1 - H(x : 1; \alpha; \theta)]}.$$

Zenga *et al.* (2010c) hanno mostrato che la curva di ineguaglianza in corrispondenza del valore atteso μ è

$$J(\mu : \alpha; \theta) = 1 - \left\{ \frac{1 - F(1 : 1; \alpha; \theta)}{F(1 : 1; \alpha; \theta)} \right\}^2. \quad (3.2.19)$$

Di seguito sono forniti alcuni grafici della curva di Lorenz e della curva $I_{(p)}$ al variare dei parametri α e θ .

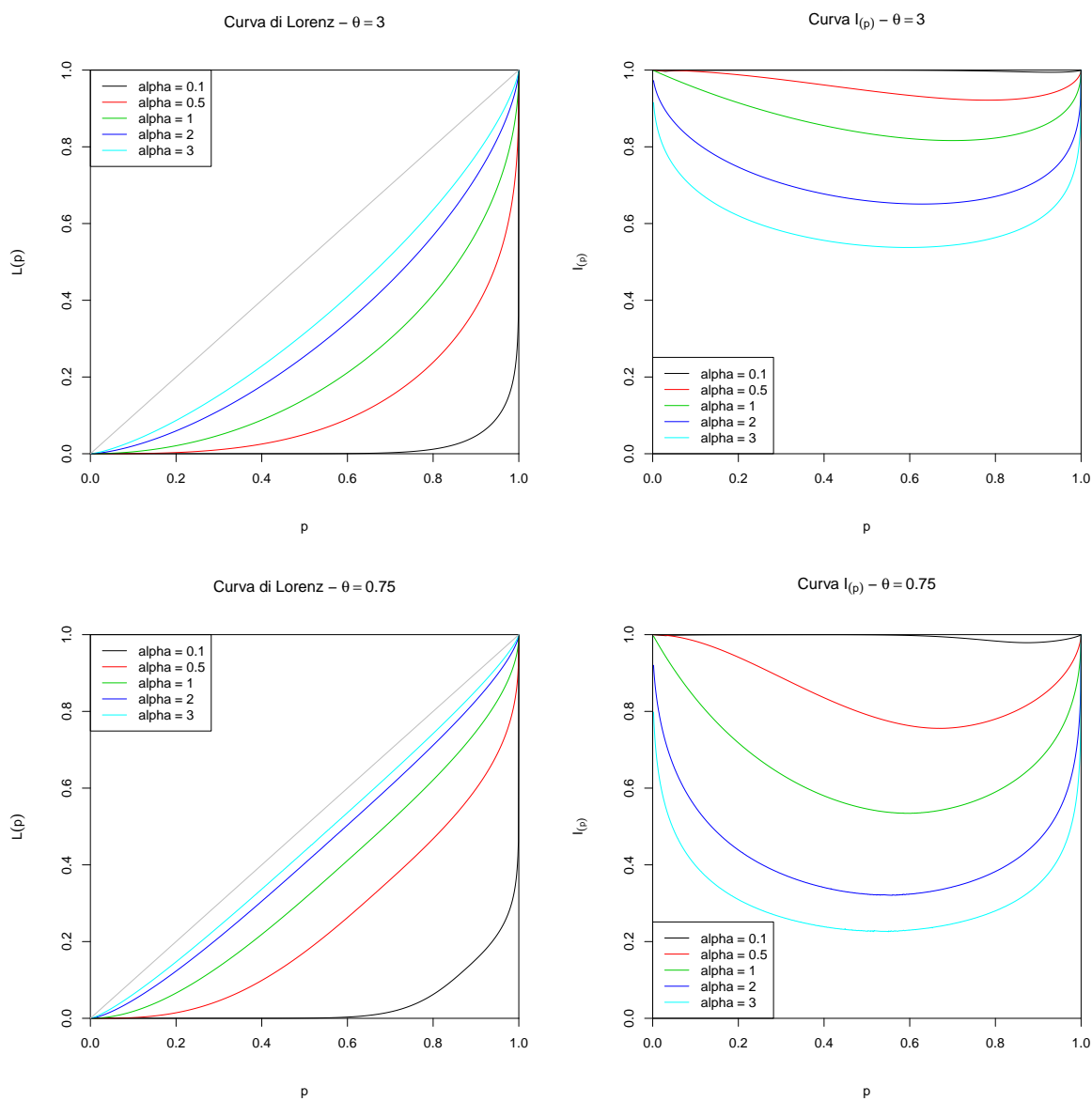


Figura 3.2.5: Curva di Lorenz e curva $I(p)$ della variabile casuale di Zenga al variare del parametro α

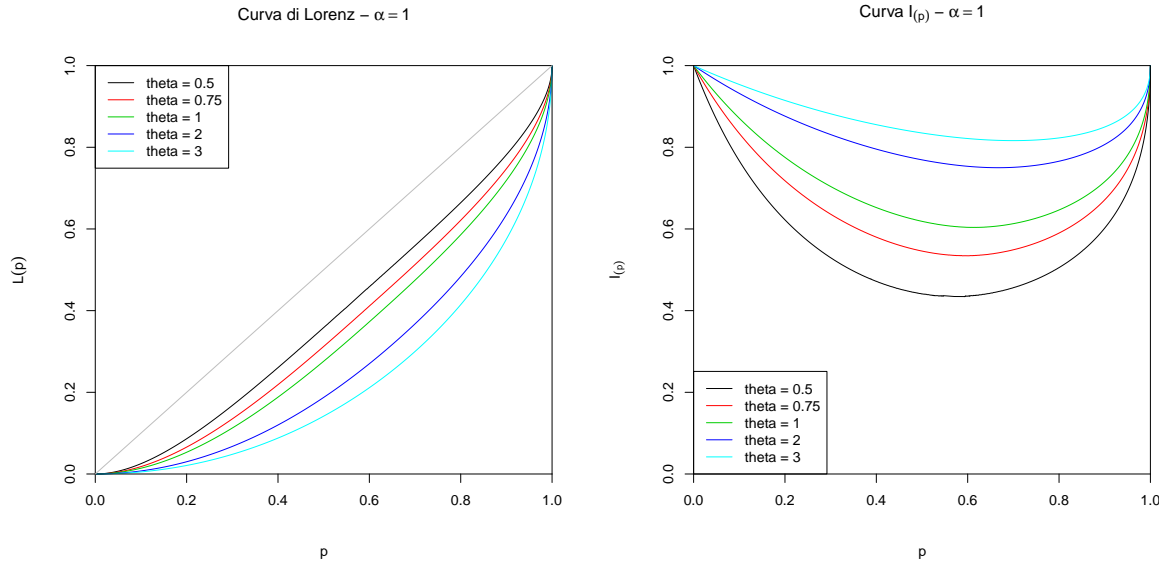


Figura 3.2.6: Curva di Lorenz e curva $I_{(p)}$ della variabile casuale di Zenga al variare del parametro θ

Dai grafici si può osservare che, mantenendo fisso θ , al crescere di α gli indici puntuali tendono al caso di equidistribuzione senza sovrapporsi. Viceversa, mantenendo fisso α , al crescere di θ gli indici puntuali tendono al caso di massima concentrazione. Ciò può essere espresso affermando che α è un indice indiretto di concentrazione e θ è un indice diretto di concentrazione.

L'indice R di Gini e l'indice globale di ineguaglianza I , devono essere valutati attraverso integrazione numerica delle equazioni (2.2.3) e (2.5.3).

La forma analitica dell'indice P di Pietra è (si vedano Zenga, 2010; Zenga *et al.*, 2010c)

$$P = 2F(1 : 1; \alpha, \theta) - 1. \quad (3.2.20)$$

Tale risultato è stato ottenuto dalla scomposizione dello scostamento medio assoluto per variabili casuali continue X

$$E(|X - \mu|) = \mu F(\mu) - \int_0^{\mu} x f(x) dx + \int_{\mu}^{\infty} x f(x) dx - \mu [1 - F(\mu)]$$

e dalle uguaglianze che sussistono per il modello di Zenga

$$\begin{aligned} H(1 : 1; \alpha; \theta) &= \int_0^1 x f(x : 1; \alpha; \theta) dx = \int_1^{\infty} f(x : 1; \alpha; \theta) dx = 1 - F(1 : 1; \alpha; \theta) \\ 1 - H(1 : 1; \alpha; \theta) &= \int_1^{\infty} x f(x : 1; \alpha; \theta) dx = \int_0^1 f(x : 1; \alpha; \theta) dx = F(1 : 1; \alpha; \theta). \end{aligned}$$

Si osservi che mediante le precedenti uguaglianze e l'equazione (2.4.2) si può giungere al medesimo risultato

$$P = F(1 : 1; \alpha; \theta) - H(1 : 1; \alpha; \theta) = F(1 : 1; \alpha; \theta) - [1 - F(1 : 1; \alpha; \theta)] = 2F(1 : 1; \alpha; \theta) - 1.$$

Capitolo 4

Stima dei parametri

In questo capitolo verrà affrontato il tema della stima dei parametri della variabile casuale di Zenga. In particolare si tratteranno due metodi di stima: il metodo dei momenti e la minimizzazione di indici di accostamento.

Lo scopo di questa analisi è valutare l'attitudine del nuovo modello distributivo a rappresentare distribuzioni di fenomeni socio-economici con particolare riferimento alle distribuzioni dei redditi. A tal fine si è preferito privilegiare, rispetto al metodo della massima verosimiglianza, il metodo dei momenti e la minimizzazione degli indici di accostamento. Con il primo metodo è possibile valutare la relazione tra le stime dei parametri e i valori assunti dai momenti campionari e con il secondo è possibile monitorare l'accostamento del modello alla distribuzione osservata. Si desidera inoltre che il modello distributivo, mediante le stime dei parametri, rispetti caratteristiche fondamentali per questo genere di fenomeni. Ciò è perseguito attraverso l'introduzione di vincoli nei metodi di stima. L'introduzione di vincoli restringe lo spazio dei valori che gli stimatori possono assumere, comportando un vantaggio nei metodi numerici di ricerca delle stime dei parametri, per contro, la bontà di adattamento del modello ai dati può diminuire.

4.1 Metodo dei momenti

Sia $f(x : \theta_1, \dots, \theta_k)$ la funzione di densità della variabile casuale X . I momenti dall'origine zero e i momenti centrali di ordine r della variabile casuale X sono funzioni dei k parametri

$$\begin{aligned}\mu'_r(\theta_1, \dots, \theta_k) &= \int_{-\infty}^{\infty} x^r f(x : \theta_1, \dots, \theta_k) dx \\ \mu_r(\theta_1, \dots, \theta_k) &= \int_{-\infty}^{\infty} (x - \mu)^r f(x : \theta_1, \dots, \theta_k) dx\end{aligned}$$

con $\mu = \mu'_1$.

Sia (X_1, \dots, X_n) un campione casuale di numerosità n , per esso si possono definire i momenti campionari dall'origine zero e i momenti campionari centrali di ordine r

$$\begin{aligned} M'_r(X_1, \dots, X_n) &= \frac{1}{n} \sum_{i=1}^n X_i^r \\ M_r(X_1, \dots, X_n) &= \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^r \end{aligned}$$

con $\bar{X} = M'_1$.

Mediante il metodo dei momenti classico, gli stimatori $\hat{\Theta}_1, \dots, \hat{\Theta}_k$, rispettivamente dei parametri $\theta_1, \dots, \theta_k$, sono ottenuti come soluzione del sistema di k vincoli che impongono l'uguaglianza fra i primi k momenti con i rispettivi momenti campionari

$$\begin{cases} \mu'_1(\theta_1, \dots, \theta_k) = M'_1(X_1, \dots, X_n) \\ \vdots \\ \mu'_k(\theta_1, \dots, \theta_k) = M'_k(X_1, \dots, X_n). \end{cases} \quad (4.1.1)$$

Mediante lo sviluppo binomiale

$$(a + b)^r = \sum_{j=0}^r \binom{r}{j} a^j b^{r-j}, \quad r \in \mathbb{N}$$

i momenti dall'origine zero di ordine $r \in \mathbb{N}$, campionari e non, possono essere espressi in funzione dei del momento dall'origine zero di ordine 1 e dei momenti centrali di ordine j con $j = 0, 1, \dots, r$ come segue

$$\begin{aligned} \mu'_r &= \sum_{j=0}^r \binom{r}{j} \mu^{r-j} \mu_j \\ M'_r &= \sum_{j=0}^r \binom{r}{j} \bar{X}^{r-j} M_j, \end{aligned}$$

pertanto il sistema

$$\begin{cases} \mu(\theta_1, \dots, \theta_k) = \bar{X}(X_1, \dots, X_n) \\ \mu_2(\theta_1, \dots, \theta_k) = M_2(X_1, \dots, X_n) \\ \vdots \\ \mu_k(\theta_1, \dots, \theta_k) = M_k(X_1, \dots, X_n) \end{cases} \quad (4.1.2)$$

è equivalente al sistema di vincoli costruito sui momenti dall'origine e conduce agli stessi stimatori $\hat{\Theta}_1, \dots, \hat{\Theta}_k$.

4.1.1 Stimatori ottenuti con il metodo dei momenti per la variabile casuale di Zenga

La variabile casuale di Zenga è caratterizzata da $k = 3$ parametri di conseguenza vengono imposti tre vincoli. Si richiede quindi l'esistenza del momento terzo e, per l'equazione (3.2.16), è necessario

che $\alpha > 2$. Da ciò deriva che le stime per α che si possono ottenere con il metodo dei momenti sono maggiori di $\alpha^* = 2$.

Zenga *et al.* (2010a) hanno esaminato la stima dei parametri della variabile casuale di Zenga mediante il metodo dei momenti. Le stime dei parametri sono state ottenute partendo dal sistema di vincoli basati sui momenti dall'origine zero, e sottraendo alla seconda equazione il quadrato della prima è stata introdotta la varianza. Per $\theta > 1$ e $\alpha > 2$ Zenga *et al.* (2010a) hanno impostato il seguente sistema di equazioni

$$\begin{cases} \mu = \bar{x} \\ \frac{\mu^2}{3} \frac{\theta(\theta+1)}{(\alpha-1)(\alpha+\theta)} = m_2 \\ \frac{\mu^3}{5} \frac{1}{B(\alpha;\theta)} [B(\alpha-2;\theta-1) - B(\alpha+3;\theta-1)] = m'_3 \end{cases} \quad (4.1.3)$$

dove \bar{x} , m_2 e m'_3 sono rispettivamente i valori assunti dagli stimatori media campionaria, varianza campionaria e terzo momento dall'origine zero campionario, per l'insieme di osservazioni x_1, \dots, x_n . La stima del parametro μ è quindi fornita da $\hat{\mu} = \bar{x}$. Dalla seconda equazione del sistema (4.1.3) è stata individuata la relazione fra il parametro α e il parametro θ dati i valori di \bar{x} e m_2 :

$$\tilde{\alpha}(\theta) = \frac{(\theta+1) \left\{ -1 + \sqrt{1 + \frac{4}{3} \frac{\bar{x}^2}{m_2} \frac{\theta}{\theta+1}} \right\}}{2} + 1.$$

Essa è una funzione crescente di θ , e poiché $\alpha > 2$ è possibile individuare il valore minimo che può assumere θ risolvendo l'equazione

$$\tilde{\alpha}(\theta) = 2,$$

da cui risulta che $\theta > \theta^*$ con

$$\theta^* = \frac{\left[1 + 3 \frac{m_2}{\bar{x}^2} \right] + \sqrt{\left[1 + 3 \frac{m_2}{\bar{x}^2} \right]^2 + 12 \frac{m_2}{\bar{x}^2}}}{2} - 1.$$

Definita

$$\begin{aligned} T(\alpha; \theta) &= \frac{1}{B(\alpha; \theta)} [B(\alpha-2; \theta-1) - B(\alpha+3; \theta-1)] = \\ &= \frac{1}{B(\alpha; \theta)} \left[\frac{\Gamma(\alpha-2)\Gamma(\theta-1)}{\Gamma(\alpha+\theta-3)} - \frac{\Gamma(\alpha+3)\Gamma(\theta-1)}{\Gamma(\alpha+\theta+2)} \right] = \\ &= \frac{(\alpha+\theta-3)(\alpha+\theta-2)(\alpha+\theta-1)}{(\alpha-2)(\alpha-1)(\theta-1)} - \frac{\alpha(\alpha+1)(\alpha+2)}{(\theta-1)(\alpha+\theta)(\alpha+\theta+1)} = \\ &= \frac{\theta^4 + (5\alpha-4)\theta^3 + (10\alpha^2-15\alpha+1)\theta^2}{(\alpha-2)(\alpha-1)(\alpha+\theta)(\alpha+\theta+1)} + \\ &\quad + \frac{2(5\alpha^3-10\alpha^2+3)\theta + 5\alpha(\alpha^3-2\alpha^2-\alpha+2)}{(\alpha-2)(\alpha-1)(\alpha+\theta)(\alpha+\theta+1)}, \end{aligned}$$

dalla terza equazione del sistema (4.1.3) si ha

$$T(\alpha; \theta) = 5 \frac{m'_3}{\bar{x}^3}. \quad (4.1.4)$$

A questo punto è complesso invertire la terza equazione e trovare una soluzione per il sistema (4.1.3), così Zenga *et al.* (2010a) hanno ottenuto le stime $\hat{\mu}$, $\hat{\alpha}$ e $\hat{\theta}$ con il seguenti passaggi:

1. $\hat{\mu} = \bar{x}$;
2. $\hat{\theta}$ è dato dalla soluzione numerica di $T(\tilde{\alpha}(\theta); \theta) = 5 \frac{m'_3}{\bar{x}^3}$;
3. $\hat{\alpha} = \tilde{\alpha}(\hat{\theta})$.

In figura 4.1.1 (Zenga *et al.*, 2010a) è riportata la rappresentazione grafica del metodo impiegato per $\bar{x} = 18502.7$, $m_2 = 341808922$ e $m'_3 = 1.163341 \times 10^{14}$, valori ottenuti dalla distribuzione non pesata dei redditi italiani del 2006, fornita da Banca D'Italia. In figura sono riportati anche i valori minimi che possono assumere le stime.

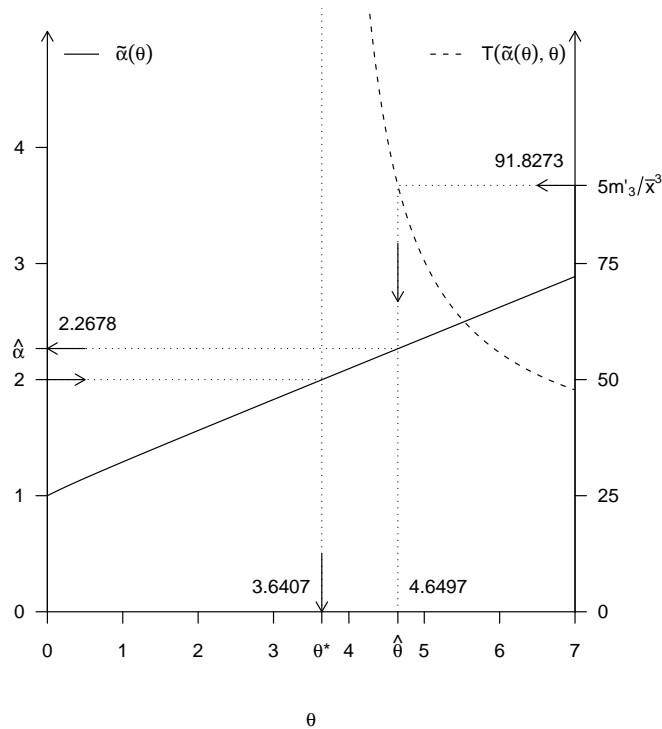


Figura 4.1.1: Grafici di $\tilde{\alpha}(\theta)$ e $T(\tilde{\alpha}(\theta); \theta)$

La soluzione del metodo dei momenti e la definizione analitica degli stimatori sono state ottenute da Zenga *et al.* (2010b). Il sistema di vincoli che è stato impiegato è basato sui momenti centrali e sul valore atteso che, come mostrato nell'introduzione della sezione 4.1, è equivalente a quello basato sui momenti dall'origine.

Sia (x_1, \dots, x_n) la realizzazione di un campione casuale dalla densità $f(x : \mu; \alpha; \theta)$ descritta in equazione (3.2.6). Siano rispettivamente \bar{x} , m_2 e m_3 i valori dalla media, dalla varianza e dal terzo momento centrale campionari, calcolati sulla realizzazione (x_1, \dots, x_n) . Quando X ha densità $f(x : \mu; \alpha; \theta)$ il valore atteso, la varianza e il terzo momento centrale sono descritti dalle equazioni (3.2.14), (3.2.17) e (3.2.18). Le stime dei parametri μ , α e θ possono quindi essere ottenute risolvendo il sistema

$$\begin{cases} \mu = \bar{x} \\ \frac{\bar{x}^2}{3} \cdot \frac{\theta(\theta+1)}{(\alpha-1)(\alpha+\theta)} = m_2 \\ m_2 \cdot \frac{3}{5} \cdot \bar{x} \cdot \frac{(\theta+3)(\theta+2)}{(\alpha+\theta+1)(\alpha-2)} = m_3. \end{cases} \quad (4.1.5)$$

Dalla prima equazione del sistema (4.1.5) si ha $\hat{\mu} = \bar{x}$, mentre le stime di α e θ sono ottenute dalla soluzione del sistema

$$\begin{cases} \frac{\bar{x}^2}{3} \cdot \theta(\theta + 1) = m_2 \cdot (\alpha - 1)(\alpha + \theta) \\ m_2 \cdot \frac{3}{5} \cdot \bar{x} \cdot (\theta + 3)(\theta + 2) = m_3 \cdot (\alpha + \theta + 1)(\alpha - 2). \end{cases} \quad (4.1.6)$$

Si osservi che l'impiego del momento terzo centrale, espresso come funzione della varianza, fa sì che la seconda equazione del precedente sistema sia un'equazione di secondo grado sia in α che in θ , risolvibile in modo più agevole rispetto all'equazione (4.1.4). Dopo alcuni passaggi sulle singole equazioni il sistema (4.1.6) può essere riscritto nel modo seguente

$$\begin{cases} \frac{\bar{x}^2}{3m_2} \theta(\theta + 1) + \theta = \alpha^2 + \alpha(\theta - 1) \\ \frac{3\bar{x}}{5} \cdot \frac{m_2}{m_3} (\theta + 3)(\theta + 2) + 2(\theta + 1) = \alpha^2 + \alpha(\theta - 1) \end{cases}$$

mediante il confronto delle due equazioni si ottiene

$$\begin{cases} \frac{\bar{x}^2}{3m_2} \theta(\theta + 1) + \theta = \alpha^2 + \alpha(\theta - 1) \\ \frac{3\bar{x}}{5} \cdot \frac{m_2}{m_3} (\theta + 3)(\theta + 2) + 2(\theta + 1) = \frac{\bar{x}^2}{3m_2} \theta(\theta + 1) + \theta \end{cases} \quad (4.1.7)$$

in quest'ultimo passaggio la seconda equazione è di secondo grado ed espressa in funzione del solo parametro θ . L'unica soluzione ammissibile della seconda equazione è

$$\hat{\theta} = \frac{-\left[\frac{1}{3} \frac{\bar{x}^2}{m_2} - 3 \frac{\bar{x}m_2}{m_3} - 1\right] + \sqrt{\left[\frac{1}{3} \frac{\bar{x}^2}{m_2} - 3 \frac{\bar{x}m_2}{m_3} - 1\right]^2 + 4 \left[\frac{1}{3} \frac{\bar{x}^2}{m_2} - \frac{3}{5} \frac{\bar{x}m_2}{m_3}\right] \left[\frac{18}{5} \frac{\bar{x}m_2}{m_3} + 2\right]}}{2 \left[\frac{1}{3} \frac{\bar{x}^2}{m_2} - \frac{3}{5} \frac{\bar{x}m_2}{m_3}\right]}$$

dato che la seconda soluzione sarebbe negativa e quindi non appartenente all'insieme dei valori che il parametro θ può assumere. La prima equazione del sistema (4.1.7) è un'equazione di secondo grado in α , e mediante la sostituzione di θ con $\hat{\theta}$, si giunge all'unica soluzione ammissibile

$$\hat{\alpha} = \frac{-(\hat{\theta} - 1) + \sqrt{(\hat{\theta} - 1)^2 + 4 \left[\frac{1}{3} \frac{\bar{x}^2}{m_2} \hat{\theta}(\hat{\theta} + 1) + \hat{\theta}\right]}}{2}$$

poiché, anche in questo caso, la seconda soluzione è esterna all'insieme di valori che la stima può assumere.

4.2 Minimizzazione di indici di accostamento

Sia $(X_1, \dots, X_i, \dots, X_n)$ un campione casuale di numerosità n proveniente da una variabile casuale X con funzione di densità $f(x : \theta_1, \dots, \theta_k)$. Siano $C_1, \dots, C_j, \dots, C_s$ s intervalli consecutivi che costituiscono una partizione del supporto di X . La probabilità che un osservazione cada nella classe C_j , $j = 1, \dots, s$, è data da

$$\hat{f}_j(\theta_1, \dots, \theta_k) = \int_{C_j} f(x : \theta_1, \dots, \theta_k) dx, \quad (4.2.1)$$

e poiché le classi costituiscono una partizione del supporto di X si ha

$$\sum_{j=1}^s \hat{f}_j = 1. \quad (4.2.2)$$

Si indichi con N_j la variabile casuale che descrive il numero di osservazioni del campione (X_1, \dots, X_n) che cadono nella classe C_j , per $j = 1, \dots, s$. Essa è una variabile casuale Binomiale

$$N_j \sim \text{Binom}(n, \hat{f}_j)$$

conseguentemente il numero atteso di osservazioni nella classe C_j è

$$\hat{n}_j(\theta_1, \dots, \theta_k) = E(N_j) = n \cdot \hat{f}_j.$$

Si osservi che per la (4.2.2) $\sum_{j=1}^s \hat{n}_j = \sum_{j=1}^s n_j = n$.

Sia $(x_1, \dots, x_i, \dots, x_n)$ la realizzazione del campione casuale e siano n_j , per $j = 1, \dots, s$, i valori assunti dalle variabili casuali N_j . Pertanto

$$|\rho_j| = \frac{|n_j - \hat{n}_j|}{\hat{n}_j}, \quad \text{per } j = 1, \dots, s$$

è lo *scostamento assoluto relativo* per la classe C_j fra la frequenza effettiva e la frequenza attesa. Si desidera che questi scostamenti siano minimi ogni classe C_j con $j = 1, \dots, s$. Una sintesi dell'accostamento globale, ovvero su tutto il supporto di X , può essere costruita attraverso una media ponderata dei singoli scostamenti relativi $|\rho_j|$. Conseguentemente si definiscono i due indici

$$A_1(\theta_1, \dots, \theta_k) = M_1(|\rho_j|; \hat{n}_j) = \frac{1}{n} \sum_{j=1}^s \frac{|n_j - \hat{n}_j|}{\hat{n}_j} \hat{n}_j = \frac{1}{n} \sum_{j=1}^s |n_j - \hat{n}_j| \quad (4.2.3)$$

$$A_2(\theta_1, \dots, \theta_k) = M_2(|\rho_j|; \hat{n}_j) = \sqrt{\frac{1}{n} \sum_{j=1}^s \left(\frac{|n_j - \hat{n}_j|}{\hat{n}_j} \right)^2 \hat{n}_j} = \sqrt{\frac{1}{n} \sum_{j=1}^s \frac{(n_j - \hat{n}_j)^2}{\hat{n}_j}}. \quad (4.2.4)$$

Si noti che $A_2 = \sqrt{\frac{1}{n} X^2}$ dove X^2 è l'indice di K. Pearson. Un'altra misura di accostamento globale nota in letteratura è l' X^2 *modificato*

$$X'^2 = \sum_{j=1}^s \frac{(n_j - \hat{n}_j)^2}{n_j}$$

conseguentemente si ridefinisce lo scostamento assoluto relativo per la classe C_j per ogni $j = 1, \dots, s$

$$|\rho'_j| = \frac{|n_j - \hat{n}_j|}{n_j}, \quad \text{per } j = 1, \dots, s$$

e l'indice di accostamento globale

$$A'_2(\theta_1, \dots, \theta_k) = M_2(|\rho'_j|; n_j) = \sqrt{\frac{1}{n} \sum_{j=1}^s \left(\frac{|n_j - \hat{n}_j|}{n_j} \right)^2 n_j} = \sqrt{\frac{1}{n} \sum_{j=1}^s \frac{(n_j - \hat{n}_j)^2}{n_j}}, \quad (4.2.5)$$

pari a $\sqrt{\frac{1}{n}X'^2}$. Si fa presente che la media aritmetica degli scostamenti $|\rho'_j|$, con pesi le frequenze empiriche n_j , conduce nuovamente alla definizione di A_1 .

Gli indici A_1 , A_2 e A'_2 sono funzioni dei parametri $\theta_1, \dots, \theta_k$ poiché dipendono delle frequenze attese \hat{n}_i . Di conseguenza la minimizzazione dei tali indici rappresenta la ricerca congiunta di quei valori che, attribuiti ai parametri, minimizzano lo scostamento fra modello e distribuzione osservata. Tali valori costituiscono le stime dei parametri

$$\begin{aligned} (\hat{\theta}_1, \dots, \hat{\theta}_k)_{A_1} &= \operatorname{argmin}_{\theta_1, \dots, \theta_k} A_1(\theta_1, \dots, \theta_k) \\ (\hat{\theta}_1, \dots, \hat{\theta}_k)_{A_2} &= \operatorname{argmin}_{\theta_1, \dots, \theta_k} A_2(\theta_1, \dots, \theta_k) \\ (\hat{\theta}_1, \dots, \hat{\theta}_k)_{A'_2} &= \operatorname{argmin}_{\theta_1, \dots, \theta_k} A'_2(\theta_1, \dots, \theta_k). \end{aligned}$$

4.2.1 Stima non vincolata

Solo in pochi casi si riesce a ricavare la soluzione analitica dei metodi che minimizzano gli indici di accostamento a causa della complessità della funzione che deve essere minimizzata. La definizione articolata della funzione di densità e della funzione di ripartizione del modello di Zenga comporta un ulteriore ostacolo alla ricerca della forma analitica di tali stimatori. Conseguentemente si abbandona l'intento di definire in modo esplicito gli stimatori e si procede con la ricerca numerica delle stime caso per caso a seconda dei dati di cui si dispone.

La stima non vincolata mediante la minimizzazione di indici di accostamento è stata impiegata sia per il modello di Zenga sia per il modello di Dagum a tre parametri.

La suddivisione in classi e l'attribuzione delle frequenze osservate

I dati sono stati suddivisi in classi seguendo il procedimento proposto da Zenga *et al.* (2010a). I primo luogo vengono prefissati i valori delle probabilità di inclusione f'_j per $j = 1, \dots, s$. In questo modo è possibile controllare il numero di osservazioni appartenenti alle classi estreme.

Sia

$$0 \leq x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(i)} \leq \dots \leq x_{(n)}$$

l'insieme di n osservazioni ordinate della variabile X non negativa.

Siano $p'_1, \dots, p'_j, \dots, p'_s$ i valori cumulati delle probabilità di inclusione scelte f'_j . Le probabilità di inclusione f'_j , riportate in tabella 4.1 con i rispettivi valori cumulati p'_j , sono state poste uguali a quelle adottate da Zenga *et al.* (2010a).

j	f'_j	p'_j	j	f'_j	p'_j	j	f'_j	p'_j	j	f'_j	p'_j	j	f'_j	p'_j
1	0.01	0.01	6	0.05	0.15	11	0.1	0.5	16	0.05	0.85	21	0.01	0.96
2	0.01	0.02	7	0.05	0.2	12	0.1	0.6	17	0.05	0.9	22	0.01	0.97
3	0.015	0.035	8	0.05	0.25	13	0.1	0.7	18	0.02	0.92	23	0.01	0.98
4	0.015	0.05	9	0.05	0.3	14	0.05	0.75	19	0.015	0.935	24	0.01	0.99
5	0.05	0.1	10	0.1	0.4	15	0.05	0.8	20	0.015	0.95	25	0.01	1.00

Tabella 4.1: probabilità di inclusione nelle classi e rispettive cumulate

Dalla tabella 4.1 sono stati determinati gli s valori $n \cdot p'_j$. Quindi ogni valore $n \cdot p'_j$ è stato approssimato all'intero più vicino $n \cdot p_j$. Gli estremi superiori delle classi C_j sono posti pari a

$$x_j = x_{(np_j)}; \quad h = 1, 2, \dots, s-1; \quad x_s = \infty$$

dove $x_{(np_j)}$ è la statistica d'ordine $n \cdot p_j$. Le frequenze empiriche delle classi C_j sono pari ai valori interi n_j

$$n_j = n \cdot (p_j - p_{j-1}); \quad j = 1, 2, \dots, s; \quad p_0 = 0.$$

Poiché per la variabile casuale di Zenga è di particolare interesse la funzione di ripartizione in corrispondenza del valore atteso μ , si ritiene utile sostituire con la media aritmetica \bar{x} l'estremo x_j più prossimo ad essa.

Le frequenze attese

In generale indicando con $F(x : \theta_1, \dots, \theta_k)$ la funzione di ripartizione della variabile casuale X , e definendo con $x_{j-1} < x_j$ rispettivamente l'estremo inferiore e l'estremo superiore della classe C_j per $j = 1, \dots, s$, la probabilità $\hat{f}_j(\theta_1, \dots, \theta_k)$ definita nell'equazione (4.2.1) può essere determinata come

$$\hat{f}_j(\theta_1, \dots, \theta_k) = F(x_j : \theta_1, \dots, \theta_k) - F(x_{j-1} : \theta_1, \dots, \theta_k).$$

La variabile casuale di Dagum di tipo I ha funzione di ripartizione descritta dall'equazione (3.1.1) ponendo $\alpha = 0$.

Per la variabile casuale di Zenga si impiega la funzione di ripartizione del miscuglio con parametro di scala μ

$$F(x : \mu; \alpha; \theta) = F(x/\mu : 1; \alpha; \theta) = \int_0^1 V(x/\mu : 1; k)g(k : \alpha; \theta) dk \quad (4.2.6)$$

dove $V(x/\mu : 1; k) = V(x : \mu; k)$ è la funzione di ripartizione della variabile casuale di Poliscchio definita dall'equazione (3.2.1) e $g(k : \alpha; \theta)$ è la funzione di densità della variabile casuale Beta. Si noti che la funzione (4.2.6) è definita per ogni $x \geq 0$, per $\alpha > 0$ e per $\theta > 0$; non solo per $\theta > 1$. La scelta di impiegare l'equazione (4.2.6) è motivata dal fatto che, trattandosi di un metodo di stima non si vuole restringere l'insieme di valori che può assumere il parametro θ e non si è adottata l'equazione (3.2.11) perché a livello computazionale si è ritenuto più opportuno valutare numericamente un integrale su un intervallo limitato, piuttosto che determinare il valore di una serie e monitorarne la velocità di convergenza al variare dei parametri α e θ . Infatti l'integrale in equazione (4.2.6) si estende dall'intervallo $(0, 1)$, estremi di integrazione, all'intervallo $(0, 1)$, in quanto si tratta della valutazione di una funzione di ripartizione, quindi il grado di approssimazione può essere facilmente determinato a priori, e indipendentemente dai valori assunti da α e θ , sulla base del numero di punti generati, se si impiega il metodo Monte Carlo, o sul numero di parti in cui viene suddiviso il quadrato unitario.

Come illustrato in precedenza, una volta che si dispone della probabilità di inclusione del modello nella classe, moltiplicando per la numerosità delle osservazioni, si determina la frequenza attesa della classe C_i

$$\hat{n}_j = n \cdot \hat{f}_j.$$

La stima numerica

Una volta definite le classi C_j e determinate le frequenze effettive n_j e le frequenze attese dal modello oggetto di stima parametrica \hat{n}_j si hanno gli elementi per valutare le funzioni obiettivo descritte dalle equazioni (4.2.3), (4.2.4) e (4.2.5). Poiché le procedure sono state applicate in egual modo ai tre indici di accostamento, nella discussione seguente si farà riferimento al generico indice di accostamento A .

L'indice di accostamento A , in quanto dipendente dalle frequenze attese \hat{n}_j , è funzione dei parametri del modello $\theta_1, \dots, \theta_k$. La ricerca delle stime $\hat{\theta}_1, \dots, \hat{\theta}_k$ sulla base dell'indice di accostamento A è un problema di minimizzazione di una funzione in k variabili. Per ottenere tali risultati numerici è stato impiegato l'ambiente di sviluppo R (R Development Core Team, 2011). Per la minimizzazione numerica in più variabili, R mette a disposizione la funzione `optim` che utilizza il metodo del simplesso descritto da Nelder e Mead (1965).

4.2.2 Introduzione di vincoli

In accordo con quanto proposto da D'Addario (1934, 1939), si intende ricercare un insieme di vincoli per la stima parametrica, in modo tale che le stime ottenute garantiscano determinate caratteristiche del modello. In particolare si vuole che indici di locazione e indici di concentrazione del modello corrispondano alle rispettive statistiche valutate sul campione.

Per la variabile casuale di Dagum e la variabile casuale di Zenga sono state effettuate le stime introducendo il vincolo alla mediana e alla media aritmetica.

Vincolo alla mediana per la variabile casuale di Dagum a tre parametri

Dall'equazione (3.1.3), ponendo $p = 0.5$ si ottiene la forma analitica della mediana del modello

$$x_{0,5} = \lambda^{1/\delta} \left[\left(\frac{1}{0.5} \right)^{1/\beta} - 1 \right]^{-1/\delta},$$

imponendo l'uguaglianza

$$x_{0,5} = \text{Me} \tag{4.2.7}$$

dove Me è la mediana campionaria, e risolvendo in λ si ottiene

$$\tilde{\lambda}(\beta, \delta : \text{Me}) = \text{Me}^\delta [2^{1/\beta} - 1].$$

Sia $A(\beta, \lambda, \delta)$ la funzione obiettivo da minimizzare per la variabile casuale di Dagum a tre parametri. Sostituendo in essa $\lambda = \tilde{\lambda}(\beta, \delta)$ si ha

$$A_{\text{Me}}(\beta, \delta) = A(\beta, \tilde{\lambda}(\beta, \delta : \text{Me}), \delta)$$

per cui il problema di minimizzazione in $k = 3$ variabili si riduce a un problema di minimizzazione in $k - l = 2$ con $l = 1$ numero di vincoli. Si osserva che

$$\min_{\beta, \lambda, \delta} A(\beta, \lambda, \delta) \leq \min_{\beta, \delta} A_{\text{Me}}(\beta, \delta)$$

a causa della riduzione dello spazio dei valori assumibili dai parametri, e l'uguaglianza è verificata solo se le stime non vincolate rispettano l'equazione (4.2.7).

Vincolo alla media aritmetica per la variabile casuale di Dagum a tre parametri

Il vincolo alla media aritmetica viene posto uguagliando la media aritmetica campionaria \bar{x} al valore atteso della variabile di Dagum a tre parametri, descritto dall'equazione (3.1.7). Affinché si possa imporre tale vincolo è necessario che il valore atteso esista finito, per cui deve essere $\delta > 1$. Il vincolo viene nuovamente posto sui valori che può assumere il parametro λ

$$\begin{aligned}\bar{x} &= \beta\lambda^{1/\delta} B\left(1 - \frac{1}{\delta}; \beta + \frac{1}{\delta}\right), \quad \text{per } \delta > 1 \\ &\Downarrow \\ \tilde{\lambda}(\beta, \delta : \bar{x}) &= \left(\frac{\bar{x}}{\beta}\right)^\delta B\left(1 - \frac{1}{\delta}; \beta + \frac{1}{\delta}\right)^{-1/\delta}, \quad \text{per } \delta > 1.\end{aligned}$$

La funzione obiettivo che dev'essere minimizzata è di conseguenza

$$A_{\bar{x}}(\beta, \delta) = A\left(\beta, \tilde{\lambda}(\beta, \delta : \bar{x}), \delta\right)$$

per la quale valgono le stesse considerazioni espote per il vincolo precedente, ovvero

$$\min_{\beta, \lambda, \delta} A(\beta, \lambda, \delta) \leq \min_{\beta, \delta} A_{\bar{x}}(\beta, \delta)$$

Vincolo alla mediana per la variabile casuale di Zenga

Il vincolo alla mediana viene introdotto risolvendo in μ numericamente l'equazione

$$F(\text{Me} : \mu; \alpha; \theta) = 0.5 \tag{4.2.8}$$

con $F(\cdot : \mu; \alpha; \theta)$ funzione di ripartizione del modello di Zenga, definita come in equazione (4.2.6), che viene valutata in corrispondenza della mediana campionaria Me. Il valore di μ che rispetta il vincolo viene definito come funzione implicita in α e θ

$$\tilde{\mu}(\alpha, \theta : \text{Me})$$

quale soluzione dell'equazione (4.2.8) per valori di Me dati. Sia $A(\mu, \alpha, \theta)$ la funzione obiettivo da minimizzare per ottenere le stime dei parametri del modello di Zenga senza l'imposizione di vincoli. Allora la funzione che dev'essere minimizzata per ottenere le stime vincolate alla mediana è

$$A_{\text{Me}}(\alpha, \theta) = A(\tilde{\mu}(\alpha, \theta : \text{Me}), \alpha, \theta).$$

Poiché le triplette di valori (μ, α, θ) che rispettano il vincolo (4.2.8) costituisce un sottoinsieme dell'insieme di valori che i tre parametri possono assumere si ha

$$\min_{\mu, \alpha, \theta} A(\mu, \alpha, \theta) \leq \min_{\alpha, \theta} A_{\text{Me}}(\alpha, \theta)$$

con l'uguaglianza verificata solo se le stime non vincolate soddisfano l'equazione (4.2.8).

Vincolo alla media aritmetica per la variabile casuale di Zenga

Il vincolo alla media aritmetica viene posto uguagliando la media aritmetica campionaria \bar{x} al valore atteso della variabile di Zenga, ovvero il parametro μ . Conseguentemente la funzione da minimizzare è

$$A_{\bar{x}}(\alpha, \theta) = A(\bar{x}, \alpha, \theta).$$

L'imposizione del vincolo sul parametro μ implica

$$\min_{\mu, \alpha, \theta} A(\mu, \alpha, \theta) \leq \min_{\alpha, \theta} A_{\bar{x}}(\alpha, \theta).$$

Imposizione di due vincoli per le stime dei parametri della variabile casuale di Zenga

Per la variabile casuale di Zenga si impongono anche vincoli sulle misure di ineguaglianza, congiuntamente al vincolo sulla media aritmetica.

Le misure di ineguaglianza che sono state utilizzate sono la curva di ineguaglianza di Zenga in corrispondenza del valore atteso $J(\mu)$, e l'indice di Pietra P . Per la variabile casuale di Zenga entrambi dipendono dalla funzione di ripartizione valutata in corrispondenza del valore atteso $F(\mu : \mu; \alpha; \theta)$ definita in equazione (3.2.10) per $\theta > 1$. Inoltre, poiché μ è parametro di scala, al fine di evidenziare l'autonomia del vincolo alla misura di ineguaglianza dal vincolo alla media aritmetica, si fa riferimento a $F(1 : 1; \alpha; \theta)$. Siano quindi $\bar{J}(\bar{x})$ la curva di ineguaglianza osservata sui dati in corrispondenza della media aritmetica e \bar{P} l'indice di Pietra osservato. La conformità a tali vincoli viene valutata per via numerica poiché non è possibile esplicitare α o θ da $F(1 : 1; \alpha; \theta)$. Di conseguenza i vincoli vengono definiti in forma implicita.

Dall'equazione (3.2.19) si ricava il vincolo sulla curva di ineguaglianza

$$\begin{aligned} \bar{J}(\bar{x}) &= J(\mu : \alpha; \theta) \\ \bar{J}(\bar{x}) &= 1 - \left\{ \frac{1 - F(1 : 1; \alpha; \theta)}{F(1 : 1; \alpha; \theta)} \right\}^2 \\ &\Downarrow \\ F(1 : 1; \alpha; \theta) &= \left(1 + \sqrt{1 - \bar{J}(\bar{x})} \right)^{-1} \end{aligned} \quad (4.2.9)$$

di conseguenza, sotto il vincolo, il parametro α può essere definito come funzione implicita in θ per valori dati di $\bar{J}(\bar{x})$

$$\tilde{\alpha}(\theta : \bar{J}(\bar{x}))$$

quale soluzione dell'equazione (4.2.9). La funzione obiettivo da minimizzare è così definita

$$A_{\bar{x}, \bar{J}(\bar{x})}(\theta) = A_{\bar{x}}(\tilde{\alpha}(\theta : \bar{J}(\bar{x})), \theta).$$

Poiché viene ristretto ulteriormente l'insieme di valori che i parametri possono assumere si ha

$$\min_{\mu, \alpha, \theta} A(\mu, \alpha, \theta) \leq \min_{\alpha, \theta} A_{\bar{x}}(\alpha, \theta) \leq \min_{\theta} A_{\bar{x}, \bar{J}(\bar{x})}(\theta).$$

Dall'equazione (3.2.20) si ricava il vincolo all'indice di Pietra

$$\begin{aligned}
 \bar{P} &= P(\alpha; \theta) \\
 \bar{P} &= 2F(1 : 1; \alpha; \theta) - 1 \\
 &\Downarrow \\
 F(1 : 1; \alpha; \theta) &= \frac{\bar{P} + 1}{2}
 \end{aligned} \tag{4.2.10}$$

da cui il valore di α viene definito come funzione implicita in θ per valori dati di \bar{P}

$$\tilde{\alpha}(\theta : \bar{P})$$

tale che verifica l'equazione (4.2.10), e la funzione da minimizzare

$$A_{\bar{x}, \bar{P}}(\theta) = A_{\bar{x}}(\tilde{\alpha}(\theta : \bar{P}), \theta).$$

Le stime che si ottengono imponendo i vincoli su media aritmetica e indice di Pietra sono tali che

$$\min_{\mu, \alpha, \theta} A(\mu, \alpha, \theta) \leq \min_{\alpha, \theta} A_{\bar{x}}(\alpha, \theta) \leq \min_{\theta} A_{\bar{x}, \bar{P}}(\theta).$$

Si osservi che le coppie di vincoli scelte hanno delle interpretazioni in termini di curva $I_{(p)}$ e di curva di Lorenz.

La coppia di vincoli su \bar{x} e $\bar{J}(\bar{x})$ fa sì che la curva $I_{(p)}$ del modello in corrispondenza di

$$p = F(\mu : \mu; \alpha; \theta)$$

assuma lo stesso valore della curva $I_{(p)}$ valutata sui dati in corrispondenza di

$$p = \hat{F}(\bar{x} : x_1, \dots, x_n),$$

con $\hat{F}(\cdot : x_1, \dots, x_n)$ funzione di ripartizione empirica determinata sull'insieme di dati x_1, \dots, x_n .

La coppia di vincoli su \bar{x} e \bar{P} , fa sì che la massima distanza fra retta di equiripartizione e curva di Lorenz del modello, corrisponda alla massima distanza fra retta di equiripartizione e spezzata di Lorenz valutata sull'insieme di dati x_1, \dots, x_n .

L'imposizione di $l = 2$ vincoli ha ridotto il numero di variabili di minimizzazione a $k - l = 1$.

4.2.3 I valori iniziali

La stima per minimizzazione fa ampio uso di metodi numerici per i quali è necessario individuare in modo opportuno i valori iniziali necessari per l'inizializzazione della procedura iterativa. L'imposizione di due vincoli nella stima dei parametri della variabile di Zenga, comporta che il numero di variabili di minimizzazione sia 1, ovvero solo il parametro θ , dando la possibilità di usare altri algoritmi più efficienti rispetto a quello del semplice di Nelder e Mead (1965). R mette a disposizione per la minimizzazione univariata la funzione `optimize` che impiega una combinazione di algoritmi. A tale funzione, invece di un valore iniziale è necessario passare un intervallo all'interno del quale si può trovare il punto di minimo. Come estremo inferiore è stato adottato il valore

minimo che il parametro può assumere. Come estremo superiore si è scelto un valore pari a due volte la stima di θ ottenuta con il metodo dei momenti.

Una volta ottenute le stime con due vincoli, come valore iniziale per le stime con un vincolo e senza vincoli, sono stati impiegati i risultati che hanno restituito l'indice di accostamento più basso.

Si fa presente che l'esistenza di una soluzione analitica, fa sì che il metodo dei momenti per il modello di Zenga costituisca un valido criterio per la determinazione di valori iniziali per i metodi numerici.

Per la variabile di Dagum i valori iniziali delle stime con un vincolo sono stati ottenuti attraverso alcune prove, e per le stime senza vincoli sono stati impiegati i risultati che hanno restituito il valore più basso del rispettivo indice di accostamento.

4.3 Distribuzioni degli stimatori e degli indici di accostamento della variabile casuale di Zenga mediante Bootstrap parametrico

Nella sezione precedente è stata descritta la procedura numerica mediante la quale ottenere le stime dei parametri del modello di Zenga minimizzando gli indici di accostamento. Sebbene le numerosità campionarie nel campo delle distribuzioni dei redditi possano ritenersi elevate, invece di considerare le proprietà asintotiche, si è scelto l'impiego del Bootstrap parametrico per lo studio delle distribuzioni degli stimatori basati sul minimo A_1 , il minimo A_2' e il minimo A_2 . Inoltre attraverso il Bootstrap parametrico è stato possibile stimare le distribuzioni degli indici di accostamento. Le distribuzioni così ottenute rappresentano un riferimento per valutare le performance degli stimatori e i valori delle funzioni obiettivo ottenuti sui dati reali.

Il Bootstrap fu introdotto da Efron (1979) e attualmente, insieme al Jackknife (Quenouille, 1956), costituisce uno dei metodi di ricampionamento più usati in analisi statistica.

Si illustra brevemente il procedimento, descritto da Efron (1979), sul quale si basano i metodi Bootstrap.

Sia $\mathbf{X} = (X_1, \dots, X_n)$ un campione casuale di dimensione n a componenti indipendenti e con funzione di ripartizione F , e sia $\mathbf{x} = (x_1, \dots, x_n)$ una sua realizzazione. Si desidera ricavare la distribuzione della variabile casuale $T_n = t(X_1, \dots, X_n)$. L'algoritmo base Bootstrap di tipo non parametrico si compone dei seguenti passaggi:

1. costruzione della funzione di ripartizione empirica \hat{F} , attribuendo massa $1/n$ a ogni osservazione x_1, \dots, x_n ;
2. costruzione del *campione Bootstrap* $\mathbf{X}^* = (X_1^*, \dots, X_n^*)$ come campione casuale di dimensione n , costituito da variabili casuali iid X_l^* , $l = 1, \dots, n$, con funzione di ripartizione \hat{F} ; si indica con $\mathbf{x}^* = (x_1^*, \dots, x_n^*)$ una sua realizzazione, che rappresenta un ricampionamento con reimmissione da $\mathbf{x} = (x_1, \dots, x_n)$;
3. approssimazione della distribuzione di T_n attraverso la distribuzione Bootstrap di $T_n^* = t(X_1^*, \dots, X_n^*)$; per ogni realizzazione \mathbf{x}_j^* , $j = 1, \dots, B$, è possibile determinare il valore di $t_j^* = t(x_{j,1}^*, \dots, x_{j,n}^*)$ ottenendo una distribuzione simulata di T_n^* , con B scelto sufficientemente grande.

Quindi la distribuzione di T_n^* viene ottenuta attraverso un procedimento di simulazione da \hat{F} . Sotto opportune condizioni di regolarità, T_n e T_n^* hanno la medesima legge asintotica, garantendo la consistenza del metodo Bootstrap.

4.3.1 Il Bootstrap parametrico

Descrizioni e applicazioni del Bootstrap in ambito parametrico possono essere trovate in Efron (1982), Efron (1985) e Davison e Hinkley (1997).

Sia $\mathbf{X} = (X_1, \dots, X_n)$ un campione casuale di dimensione n a componenti indipendenti e con funzione di ripartizione F_θ nota a meno del parametro θ (scalare o vettore), e sia $T_n = t(X_1, \dots, X_n)$ una variabile casuale campionaria. Da tale campione si osservano i valori $\mathbf{x} = (x_1, \dots, x_n)$. Quando θ viene stimato con $\hat{\theta}$ sulla base dei valori \mathbf{x} , la sua sostituzione nella legge di distribuzione definisce il *modello adattato* con funzione di ripartizione $\hat{F}(x) = F_{\hat{\theta}}(x)$. Così come nel Bootstrap non parametrico, viene definito il disegno di campionamento $\mathbf{X}^* = (X_1^*, \dots, X_n^*)$ di dimensione n , a componenti indipendenti e con funzione di ripartizione \hat{F} , dal quale si generano le simulazioni \mathbf{x}_j^* , $j = 1, \dots, B$, e si ottiene l'approssimazione della distribuzione di T_n attraverso quella di $T_n^* = t(X_1^*, \dots, X_n^*)$.

Quindi il Bootstrap parametrico differisce da quello non parametrico per l'assunzione del modello distributivo F_θ dal quale provengono le osservazioni, e per il conseguente impiego della funzione di ripartizione del modello adattato per definire \hat{F} al posto della funzione di ripartizione empirica. Se lo stimatore di θ è consistente, al divergere di n , \hat{F} tende a F_θ .

Il Bootstrap parametrico, pertanto, consiste nel ricampionamento dal modello adattato, simulando B campioni \mathbf{x}^* di numerosità n pari a quello da cui sono state ricavate le stime $\hat{\theta}$. Su tali campioni vengono determinate le repliche Bootstrap $t_j^* = t(x_{j,1}^*, \dots, x_{j,n}^*)$, $j = 1, \dots, B$, della statistica T_n . In questo modo si ottiene la distribuzione Bootstrap di T_n^* .

Attraverso la distribuzione simulata è possibile realizzare intervalli di confidenza per il parametro ignoto $\tau(\theta)$ sotto l'assunzione di distribuzione F_θ delle osservazioni. Inoltre se T_n è stimatore per $\tau(\theta)$ e la distorsione per la data ampiezza campionaria n è

$$B(F_\theta) = E[T_n] - \tau(\theta)$$

lo stimatore Bootstrap della distorsione è

$$B(\hat{F}) = E[T_n^*] - \tau(\hat{\theta})$$

che a sua volta, in seguito alle B simulazioni, è stimato da

$$b_B = \frac{1}{B} \sum_{j=1}^B t_j^* - \tau(\hat{\theta}) = \bar{t}^* - \tau(\hat{\theta}).$$

Così come \bar{t}^* è stima del valore atteso di T_n^* , e quindi del valore atteso di T_n , è possibile valutare la variabilità della statistica attraverso lo stimatore corretto della varianza applicato sulle B repliche Bootstrap t_j^* .

A questo punto è opportuno sottolineare la differenza tra l'approccio parametrico e quello non parametrico con lo scopo di chiarire la finalità per la quale si è scelto di adottare il Bootstrap parametrico.

L'approccio non parametrico non avanza ipotesi circa la distribuzione F dalla quale provengono le osservazioni, e l'impiego del Bootstrap non parametrico fornisce, mediante ricampionamento dal campione originario, le distribuzioni simulate degli stimatori e degli indici di accostamento sulla base delle quali è possibile costruire intervalli di confidenza.

L'approccio parametrico, invece, parte dall'assunzione che le osservazioni provengano dal modello F_θ noto a meno del parametro θ e i campioni Bootstrap sono estratti dal modello "stimato" $F_{\hat{\theta}}$ dalla distribuzione $F_{\hat{\theta}}$. Quindi le distribuzioni delle statistiche simulate possono essere impiegate per la costruzione di intervalli di confidenza per i parametri sotto l'assunto che $F_{\hat{\theta}}$ sia una buona approssimazione di F_θ . Gli intervalli così ottenuti rappresentano un punto di riferimento per valutare se i risultati ottenuti sui dati reali sono coerenti, oppure se la distribuzione delle osservazioni è difforme dal modello assunto al punto tale da ottenere dei valori esterni a tali intervalli.

Il procedimento descritto prevede la stima dei parametri sulla base dei dati simulati per ricampionamento \mathbf{x}^* , che per omogeneità di notazione può essere definita $\hat{\theta}^*$, la quale è tale che

$$\hat{\theta}^* = \underset{\theta}{\operatorname{argmin}} A(\theta | \mathbf{x}^*) \quad (4.3.1)$$

per cui la replicazione Bootstrap dell'indice di accostamento, ovvero i valori assunti dalla funzione obiettivo

$$A(\hat{\theta}^* | \mathbf{x}^*) = \min_{\theta} A(\theta | \mathbf{x}^*) \quad (4.3.2)$$

possono essere maggiori di 0. Da tali valori si ottiene la distribuzione simulata di A sotto l'assunto che $F_{\hat{\theta}}$ sia una buona approssimazione di F_θ . Attraverso la distribuzione Bootstrap è possibile individuare un punto di riferimento per valutare la conformità con il modello dei risultati ottenuti sui dati reali.

Dalla (4.3.2) può sorgere il dubbio sul fatto che l'indice di accostamento, per come è stato impiegato, non costituisca una statistica del campione \mathbf{X}^* . A tal proposito si ricorda che i parametri che caratterizzano la distribuzione delle osservazioni del campione \mathbf{X}^* sono $\hat{\theta}$. I valori che variano nelle simulazioni sono $\hat{\theta}^*$, ovvero i valori assunti dagli stimatori dei parametri che, in quanto tali, sono statistiche del campione \mathbf{X}^* . Quindi l'indice di accostamento A , nella simulazione parametrica così definita, è funzione esclusivamente dei dati campionari simulati.

Si fa presente che per le statistiche (4.3.1) e (4.3.2), in quanto espresse in forma implicita, non sono state analizzate le condizioni di regolarità di cui si è fatto cenno nella parte introduttiva sul Bootstrap. Tali condizioni sono necessarie per garantire la convergenza delle distribuzioni simulate alla distribuzione reale della statistica di interesse, pertanto ci si astiene dal fare valutazioni asintotiche sulle distribuzioni simulate, e ci si limita a confrontare i valori ottenuti sui dati reali con le distribuzioni simulate a parità di numerosità campionaria.

4.3.2 Generazione di valori dal modello di Zenga

La generazione di numeri casuali da una distribuzione di probabilità univariata non uniforme, quale il modello di Zenga, può essere ottenuta mediante il *metodo della trasformazione inversa* (si veda, ad esempio, Devroye, 1986) basato sul seguente teorema:

Teorema Sia F una funzione di ripartizione continua su \mathbb{R} con inversa F^{-1} definita da

$$F^{-1}(u) = \inf \{x : F(x) = u, 0 < u < 1\}.$$

Se U è una variabile casuale Uniforme $[0, 1]$, allora $F^{-1}(U)$ ha funzione di ripartizione F . Inoltre, se X ha funzione di ripartizione F , allora $F(X)$ è uniformemente distribuita su $[0, 1]$.

In forza di tale teorema, il metodo della trasformazione inversa consiste nel generare valori casuali dalla distribuzione Uniforme $[0, 1]$ e trasformarli mediante l'inversa della funzione di ripartizione del modello che si intende simulare.

L'inversa della funzione di ripartizione della variabile casuale di Zenga non è nota in forma esplicita. Di conseguenza tale metodo non può essere impiegato direttamente. Le simulazioni dal modello di Zenga possono essere ottenute attraverso la sua definizione di miscuglio.

Le simulazioni dal modello di Zenga possono essere ottenute attraverso due stadi:

1. generazione del valore del parametro $K = k$ dalla distribuzione Beta di parametri α e θ ;
2. generazione del valore del miscuglio $X = x$ attraverso la distribuzione condizionata di $X|K = k$.

La distribuzione condizionata di $X|K = k$ è nota dalla definizione del miscuglio ed è la distribuzione Pareto troncata di Poliscchio, con funzione di ripartizione (3.2.1) la cui inversa è

$$V^{-1}(p : \mu; k) = \mu k [1 - (1 - k)p]^{-2}.$$

Le simulazioni unidimensionali di K e $X|K = k$ possono essere così ottenute attraverso il metodo della trasformazione inversa.

Per maggiore chiarezza si esemplifica il procedimento in figura 4.3.1.

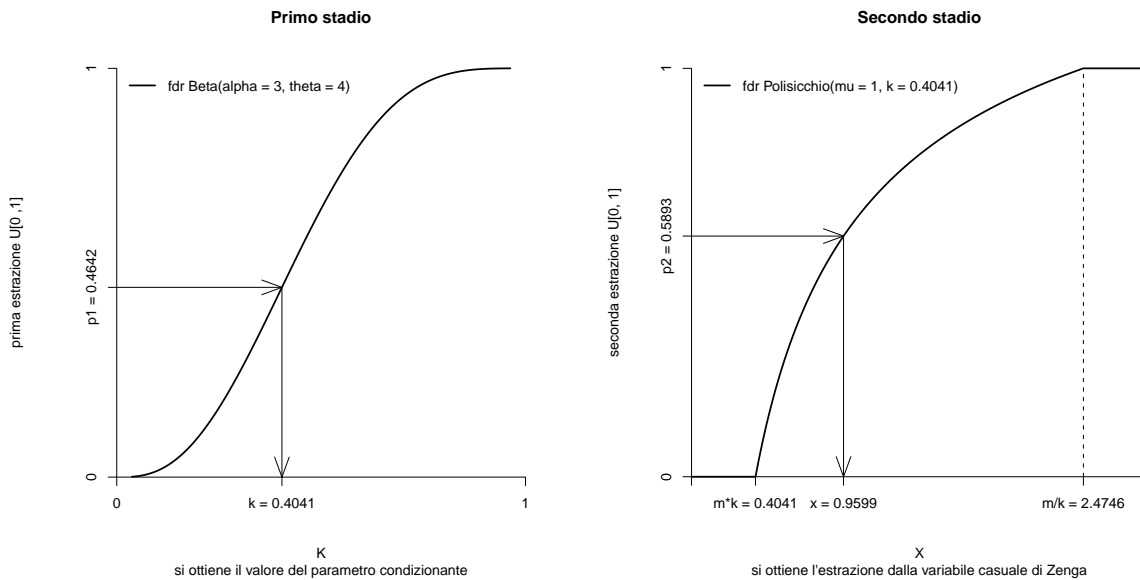


Figura 4.3.1: Rappresentazione del metodo di simulazione in due stadi da una variabile casuale di Zenga di parametri $\mu = 1$, $\alpha = 3$ e $\theta = 4$.

Un valore x viene generato dalla distribuzione di Zenga mediante il procedimento in due stadi.

Primo stadio

Nel primo stadio viene generato un valore da una variabile casuale $U[0, 1]$, in figura 4.3.1 è il valore $p_1 = 0.4642$ in ordinata del grafico a sinistra.

Si determina il corrispondente quantile della variabile casuale Beta di parametri α e θ , posti pari ai due parametri di forma della distribuzione di Zenga dalla quale si vuole simulare. Tale risultato, per il metodo della trasformazione inversa, corrisponde al valore k generato dalla rispettiva distribuzione Beta. In figura 4.3.1, a sinistra, si può osservare che il corrispondente quantile di $p_1 = 0.4642$, valutato sulla funzione di ripartizione di una Beta di parametri $\alpha = 3$ e $\theta = 4$, è $k = 0.4041$.

Con il primo stadio si è così determinato il valore del parametro di forma k da attribuire alla distribuzione condizionata $X|K = k$.

Secondo stadio

Una volta determinato il valore del parametro di forma della distribuzione condizionata, la Pareto troncata di Poliscchio, si genera un valore da essa, nuovamente con il metodo della trasformazione inversa.

Ciò viene ottenuto generando *un nuovo* valore dalla distribuzione $U[0, 1]$ e determinando il rispettivo quantile della distribuzione condizionata $X|K = k$, con parametro di scala μ , pari a quello della distribuzione di Zenga dalla quale si vuole generare, e parametro di forma k posto uguale al valore ottenuto nello stadio precedente.

Ad esempio, si può osservare, nella parte destra di figura 4.3.1, che il nuovo valore generato dalla distribuzione uniforme è $p_2 = 0.5893$, il parametro di scala μ è stato posto pari a 1 e il quantile in corrispondenza di p_2 è $x = 0.9599$.

Il valore x così ottenuto, in forza della definizione del miscuglio, dell'approccio della distribuzione condizionata e del metodo della trasformazione inversa risulta essere un valore generato dal modello di Zenga di parametri $\mu = 1$, $\alpha = 3$ e $\theta = 4$.

Capitolo 5

Applicazione

In questo capitolo si presentano le applicazioni di quanto descritto nei capitoli precedenti. Si fa presente che Zenga *et al.* (2010a) e Zenga *et al.* (2010b) forniscono già delle prime applicazioni su dati reali circa la stima dei parametri del modello di Zenga. In questo lavoro si intende estendere le applicazioni su nuove distribuzioni di dati, e si vuole verificare il comportamento del nuovo modello proposto e dei relativi metodi di stima. A tale scopo è stato necessario individuare un ampio numero distribuzioni empiriche.

In prima analisi si confronta il nuovo modello distributivo con la variabile casuale di Dagum. Tale scelta è motivata dal fatto che è un modello ampiamente utilizzato per la rappresentazione delle distribuzioni dei redditi e perché condivide diverse caratteristiche con la distribuzione di Zenga. In particolare entrambi i modelli hanno:

- supporto $(0, \infty)$;
- coda paretiana;
- dipendono da tre parametri.

Entrambi i modelli presentano diverse forme pertanto possono adattarsi alle distribuzioni anche per valori bassi, dovrebbero avere la stessa attitudine a rappresentare valori elevati, e hanno lo stesso numero di parametri sui quali operare per adattarli alle distribuzioni.

Inoltre, per il solo modello di Zenga, verrà presentato uno studio delle distribuzioni degli stimatori dei parametri e degli indici di accostamento, per mezzo del Bootstrap parametrico.

5.1 Descrizione dei dati

I dati reali sui quali sono state effettuate le applicazioni provengono dall'indagine European Community Household Panel (Eurostat, 2003). L'ECHP è un'indagine panel relativa agli anni dal 1994 al 2001 svolta nei 12 Stati Membri che costituivano l'Unione Europea fino al 1994, ai quali si sono aggiunti Austria (nel 1995), Finlandia (nel 1996) e Svezia (nel 1997). Le unità statistiche sono circa 130 000 individui raggruppati approssimativamente in 60 500 famiglie. L'indagine riguarda le condizioni di vita, in particolare include informazioni sul reddito, sulla situazione finanziaria, la condizione abitativa, le relazioni sociali e la salute.

La variabile sulla quale è stata posta l'attenzione è "Reddito totale familiare netto", ovvero il reddito percepito nell'anno precedente l'indagine, espresso in moneta corrente, per l'Italia in migliaia di Lire. I dati rilevati possono assumere valori interi da 1 a 9 999 999 990. Si considerano quindi le distribuzioni dei redditi per ogni Stato (15 stati) e per ogni anno (8 anni) di rilevazione. Poiché non per tutti gli Stati si dispongono rilevazioni per tutti gli 8 anni, il totale delle distribuzioni considerate non è 120 ma 114.

Infine si fa presente che nell'indagine ECHP sono disponibili i pesi relativi alla probabilità di inclusione e i dati relativi alla scala OECD (Organisation for Economic Co-operation and Development) per la determinazione dei redditi equivalenti. Nessuna delle due informazioni è stata impiegata in questo lavoro. La valutazione degli effetti del loro impiego è di indubbio interesse, ma viene lasciata a sviluppi futuri.

5.2 Descrizione dei risultati ottenuti dall'applicazione sui dati reali

Per ogni distribuzione estratta dall'indagine ECHP è stata condotta un'ampia analisi dei metodi di stima sia per il modello di Zenga che per il modello di Dagum, così come descritta nelle sezioni 4.1 e 4.2. A titolo esemplificativo si riportano i risultati ottenuti per la distribuzione dei redditi in Italia nell'anno 2001.

Statistiche

Statistiche	
N	5525.0000
media	40182.7470
mediana	34560.0483
varianza	718364413.5570
coef. var.	0.6670
m3 cent.	45476618318449.2891
Indice di Gini	0.3381
I	0.6855
Indice di Pietra	0.2426
F(media)	0.5897
I(media)	0.6301

Tabella 5.1: Statistiche Italia 2001(osservazioni in migliaia di Lire italiane)

In primo luogo sono state raccolte le statistiche campionarie di interesse, in modo tale da avere una descrizione dei dati sui quali è stata condotta l'analisi. Oltre ai tradizionali indicatori di posizione e variabilità sono state rilevate le statistiche necessarie per l'imposizione dei vincoli ai metodi di stima. In tabella 5.1, con riferimento ai dati italiani per l'anno 2001, sono riportate:

N: numerosità campionaria;

media: media aritmetica campionaria;

mediana: mediana campionaria;

varianza: varianza campionaria;

coef. var.: coefficiente di variazione, rapporto fra lo scarto quadratico medio e la media aritmetica;

m3 cent: momento terzo centrale campionario;

Indice di Gini: calcolato mediante l'equazione (2.2.2);

I: indice globale di ineguaglianza di Zenga, calcolato mediante l'equazione (2.5.2) attribuendo frequenza unitaria a ogni singola osservazione;

Indice di Pietra: calcolato mediante l'equazione (2.4.1);

F(media): funzione di ripartizione empirica, ovvero frequenza relativa cumulata, valutata in corrispondenza della media aritmetica campionaria $\hat{F}(\bar{x})$;

I(media): curva $I_{(p)}$ valutata in corrispondenza della frequenza relativa cumulata associata media campionaria $p_j = \hat{F}(\bar{x})$, secondo l'equazione (2.5.1).

Stime dei parametri del modello di Zenga

Stime Miscuglio	$\hat{\mu}$	$\hat{\alpha}$	$\hat{\theta}$	A_1	A'_2	A_2	Me
metodo dei momenti	40182.7470	4.8222	7.4180	0.1456	0.1982	0.7799	32782.0207
min A1	42412.8005	3.0158	4.7501	0.1003	0.1838	0.2571	34522.0828
min A'2	40682.3945	3.1650	4.7010	0.1133	0.1641	0.2836	33652.7559
min A2	42117.7259	2.2464	3.4981	0.1347	0.2079	0.1855	34521.7879
min A1 v Me	42519.8189	3.0304	4.7951	0.1005	0.1863	0.2598	34560.0483
min A'2 v Me	41527.0512	3.0363	4.4197	0.1198	0.1686	0.2755	34560.0483
min A2 v Me	42135.2517	2.2437	3.4861	0.1353	0.2076	0.1855	34560.0483
min A1 v M1	40182.7470	3.4652	5.1139	0.1179	0.1679	0.3456	33262.3171
min A'2 v M1	40182.7470	3.1987	4.7433	0.1203	0.1661	0.2875	33250.1277
min A2 v M1	40182.7470	2.3397	3.5757	0.1387	0.1996	0.1971	33101.2178
min A1 v M1, I(M1)	40182.7470	3.2774	4.8802	0.1195	0.1663	0.3012	33202.3109
min A'2 v M1, I(M1)	40182.7470	3.2048	4.7578	0.1203	0.1661	0.2882	33238.9136
min A2 v M1, I(M1)	40182.7470	2.3562	3.3311	0.1554	0.1942	0.2071	33829.0674
min A1 v M1, P	40182.7470	3.3570	4.9794	0.1189	0.1667	0.3192	33228.5376
min A'2 v M1, P	40182.7470	3.2109	4.7345	0.1204	0.1662	0.2919	33300.4325
min A2 v M1, P	40182.7470	2.3564	3.3080	0.1572	0.1949	0.2089	33893.2512

Tabella 5.2: Stime dei parametri del modello di Zenga per la distribuzione Italia 2001 (media: 40 182.7470 mil.Lit.; mediana: 34 560.0483 mil.Lit.)

In tabella 5.2 sono riportate le stime ottenute per i parametri del modello di Zenga, valutate sulla distribuzione dei redditi italiani per l'anno 2001.

Nella prima colonna vi sono i metodi di stima impiegati. Dalla seconda alla quarta colonna vi sono le stime dei parametri ottenute con i diversi metodi. Dalla quinta alla settima colonna sono riportati i valori degli indici di accostamento calcolati sulla base delle equazioni (4.2.3), (4.2.5) e (4.2.4). Nell'ultima colonna è riportato il valore della mediana del modello di Zenga adattato, ovvero con i parametri posti pari alle stime. Essa è stata determinata numericamente sulla base dell'equazione implicita (4.2.8), in tal modo è possibile verificare il vincolo sulla mediana, per i metodi che ne fanno uso. Il vincolo sulla media aritmetica, per il modello di Zenga, può essere controllato osservando l'uguaglianza fra la stima del parametro μ , che corrisponde al valore atteso, e la media aritmetica della popolazione.

Il primo metodo è quello dei momenti, valutato mediante le equazioni esplicite ottenute nella sezione 4.1.1.

I restanti metodi impiegati sono basati sulla minimizzazione di un indice di accostamento. Tali metodi, per il generico indice di accostamento A sono descritti nella sezione 4.2. Per questo tipo di metodi, la colonna del rispettivo indice di accostamento costituisce il valore assunto dalla funzione obiettivo.

È opportuno osservare che, ad esempio, A_1 nell'intera tabella assume valore minimo in corrispondenza della riga *min A1* e nella classe dei metodi di stima vincolati alla media aritmetica (righe caratterizzate dalla dicitura *v M1*) assume valore minimo in corrispondenza della riga *min A1 v M1*. Le stesse considerazioni possono essere effettuate per le altre classi di metodi di stima al variare dei vincoli impiegati o della funzione obiettivo.

Infine si osserva che per ogni riga l'indice di accostamento che assume valori più bassi è A_1 , ciò non è dovuto a una maggiore "efficienza" del metodo ma all'ordinamento delle medie potenziate.

Stime dei parametri del modello di Dagum

Stime Dagum	$\hat{\beta}$	$\hat{\lambda}$	$\hat{\delta}$	$\hat{\lambda}^{1/\hat{\delta}}$	A_1	A'_2	A_2	$E(X)$	Me
min A1	0.7772	3534094445217.93	2.7321	39166.1836	0.1266	0.2522	0.2010	43562.9112	34276.2810
min A'2	0.5567	17044379840399936.00	3.4836	45653.3154	0.1392	0.1678	0.1674	40850.1046	35204.3278
min A2	0.5171	19249219075370220.00	3.4852	47037.8908	0.1328	0.1757	0.1601	40641.4696	34931.4556
min A1 v Me	0.8191	1941973990560.33	2.6802	38450.1063	0.1269	0.2660	0.2094	44269.9913	34560.0483
min A'2 v Me	0.5805	6473730235661652.00	3.4040	44144.6314	0.1395	0.1696	0.1694	40375.7414	34560.0483
min A2 v Me	0.5249	11945026916593756.00	3.4457	46330.0675	0.1332	0.1781	0.1609	40351.6808	34560.0483
min A1 v M1	0.5931	176695976143184.38	3.0761	42820.7917	0.1293	0.2198	0.1857	40182.7470	33052.8942
min A'2 v M1	0.5515	21638299840201596.00	3.5095	45134.7240	0.1403	0.1697	0.1683	40182.7470	34707.7561
min A2 v M1	0.5041	35473593667109152.00	3.5414	47120.0258	0.1344	0.1776	0.1609	40182.7470	34698.8664

Tabella 5.3: Stime dei parametri del modello di Dagum per la distribuzione Italia 2001 (media: 40 182.7470 mil.Lit.; mediana: 34 560.0483 mil.Lit.)

In tabella 5.3 è riportato l'esempio dei valori osservati per il modello di Dagum. I metodi di stima impiegati sono solo quelli basati sulla minimizzazione degli indici di accostamento, non vincolati e con un vincolo. Non disponendo della soluzione analitica del metodo dei momenti tale metodo è stato escluso. Inoltre, mentre per il modello distributivo di Zenga, l'introduzione del vincolo sulla media aritmetica comporta, a livello pratico, una riduzione di complessità, non è così per il modello di Dagum. Di conseguenza per il modello di Dagum non è stato introdotto, congiuntamente al vincolo sulla media aritmetica, un secondo vincolo sulle misure di ineguaglianza.

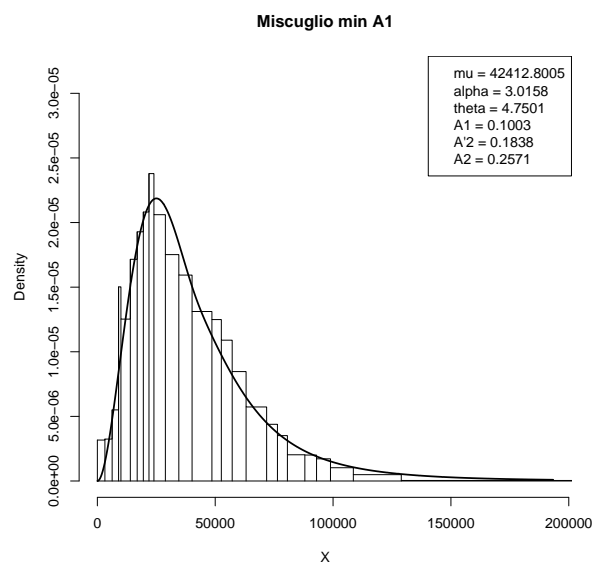


Figura 5.2.1: Istogramma della distribuzione Italia 2001 con sovrapposizione del modello di Zenga con parametri determinati attraverso il metodo del minimo A_1 .

Nel caso della variabile di Dagum non c'è un parametro che corrisponde al valore atteso, per cui è stato necessario aggiungere una colonna indicante il valore atteso del modello con parametri posti pari alle rispettive stime.

La parametrizzazione impiegata per il modello di Dagum, non presenta il parametro di scala, per cui è stata aggiunta la colonna $\hat{\lambda}^{1/\delta}$ al fine di avere un'indicazione sull'ordine di grandezza del fenomeno descritto dal modello (a tal proposito si veda la sezione 3.1).

Relazione per ogni metodo di stima

Per ogni metodo di stima, sia per il modello di Zenga che per quello di Dagum, è stata predisposta una relazione, composta di rappresentazione grafica e tabella delle classi. Di seguito si esemplifica attraverso la distribuzione dei redditi italiani e il metodo del minimo A_1 applicato al modello di Zenga.

Dalla rappresentazione grafica in figura 5.2.1 è possibile compiere una prima valutazione qualitativa d'insieme, circa il criterio di suddivisione in classi e il grado di accostamento del modello ai dati.

Per un'analisi quantitativa e più dettagliata è possibile fare riferimento alla tabella 5.4.

Nella prima colonna sono riportati gli estremi delle classi. Per tutte le distribuzioni considerate, le classi e le rispettive frequenze sono state determinate attraverso il procedimento descritto nella sezione 4.2.1, ponendo il numero s di classi pari a 25 e le probabilità di inclusione nelle classi pari a f'_j come riportati in tabella 4.1.

Tali probabilità sono state scelte, data la tipologia di caratteri studiati, al fine di ridurre l'ampiezza delle classi sulle code, rispetto al caso di equi-probabilità di inclusione delle classi. I valori cumulati delle probabilità di inclusione p'_j mostrano che tra gli estremi delle classi è compresa anche la mediana.

	p_j	$F(x; \hat{\mu}, \hat{\alpha}, \hat{\theta})$	n_j	\hat{n}_j	$\hat{n}_j - n_j$	$\frac{\hat{n}_j - n_j}{n_j}$	$\frac{\hat{n}_j - n_j}{\hat{n}_j}$
0 – 3200	0.0101	0.0016	56	8.8994	-47.1006	-0.8411	-5.2926
3200 – 6270	0.0201	0.0108	55	50.9573	-4.0427	-0.0735	-0.0793
6270 – 9000	0.0351	0.0288	83	99.2484	16.2484	0.1958	0.1637
9000 – 10000	0.0501	0.0379	83	50.5161	-32.4839	-0.3914	-0.6430
10000 – 13988	0.1001	0.0879	276	276.0000	0.0000	0.0000	0.0000
13988 – 16900	0.1500	0.1365	276	268.4868	-7.5132	-0.0272	-0.0280
16900 – 19500	0.2002	0.1865	277	276.2825	-0.7175	-0.0026	-0.0026
19500 – 21900	0.2501	0.2365	276	276.0000	0.0000	0.0000	0.0000
21900 – 24000	0.3001	0.2819	276	251.1435	-24.8565	-0.0901	-0.0990
24000 – 28857	0.4002	0.3871	553	581.4437	28.4437	0.0514	0.0489
28857 – 34560.0483	0.5001	0.5007	552	627.3811	75.3811	0.1366	0.1202
34560.0483 – 40182.747	0.5897	0.5946	495	518.7093	23.7093	0.0479	0.0457
40182.747 – 48600	0.7001	0.7042	610	605.5369	-4.4631	-0.0073	-0.0074
48600 – 52599	0.7500	0.7461	276	231.6198	-44.3802	-0.1608	-0.1916
52599 – 57200	0.8002	0.7873	277	227.7404	-49.2596	-0.1778	-0.2163
57200 – 63100	0.8501	0.8304	276	238.1303	-37.8697	-0.1372	-0.1590
63100 – 71820	0.9001	0.8778	276	261.6268	-14.3732	-0.0521	-0.0549
71820 – 76400	0.9202	0.8966	111	103.7352	-7.2648	-0.0654	-0.0700
76400 – 80626	0.9350	0.9110	82	79.7091	-2.2909	-0.0279	-0.0287
80626 – 88033	0.9500	0.9310	83	110.4215	27.4215	0.3304	0.2483
88033 – 92993	0.9600	0.9414	55	57.6167	2.6167	0.0476	0.0454
92993 – 98900	0.9701	0.9515	56	55.5954	-0.4046	-0.0072	-0.0073
98900 – 108600	0.9801	0.9638	55	68.3518	13.3518	0.2428	0.1953
108600 – 128939	0.9900	0.9794	55	85.7673	30.7673	0.5594	0.3587
>128939	1.0000	1.0000	55	114.0807	59.0807	1.0742	0.5179
Tot.			5525	5525.0000	0.0000		

Tabella 5.4: Tabella delle classi Italia 2001 e frequenze teoriche \hat{n}_j del modello di Zenga con parametri determinati attraverso il metodo del minimo A_1 .

I p'_j possono essere confrontati con i valori riportati nella seconda colonna della tabella 5.4, le p_j che costituiscono le frequenze relative cumulate di ogni classe. In questo modo è possibile valutare come il procedimento, di suddivisione in classi e attribuzione delle frequenze, ha operato sui dati osservati. In particolare si evidenzia che, nel caso mostrato in tabella, alla classe che avrebbe dovuto avere frequenza relativa cumulata pari a 0.6 è stato sostituito l'estremo superiore con la media aritmetica, e conseguentemente la frequenza relativa della classe corrisponde al valore di $F(\text{media})$ della tabella 5.1 delle statistiche dei dati.

Qualora l'introduzione della media aritmetica fra gli estremi delle classi comporti la costituzione di una classe con frequenza nulla, si è proceduto riducendo il numero di classi s a 24 lasciando tra gli estremi la media aritmetica.

Dalla differenza del valore assunto da p_j per l' j -esima riga con il valore precedente si ricavano le frequenze relative, moltiplicando queste per la numerosità campionaria N si ottengono le frequenze assolute n_j riportate in tabella.

Segue la colonna $F(x : \hat{\mu}, \hat{\alpha}, \hat{\theta})$ la quale mostra le probabilità cumulate, previste dal modello

adattato, in corrispondenza dell'estremo superiore della rispettiva classe. Operando la differenza fra il valore dell' i -esima riga con quello precedente si ottiene la probabilità che il modello adattato attribuisce all' j -esima classe, la quale, moltiplicata per la numerosità campionaria N determina la frequenza attesa dell' j -esima classe \hat{n}_j , riportata anch'essa in tabella.

L'intestazione delle ultime tre colonne è sufficiente a indicare ciò che vi è riportato. Esse rappresentano passaggi necessari per la determinazione degli indici di accostamento A_1 , A'_2 e A_2 . Costituiscono anche un'informazione rilevante per valutare dove e con quale grado il modello si allontana maggiormente dalla distribuzione campionaria. Considerando che

- la somma della colonna $\hat{n}_j - n_j$ deve essere nulla, e quindi gli scostamenti positivi e negativi devono compensarsi;
- gli scostamenti vengono sommati in valore assoluto al fine di valutare l'indice di adattamento;

lo studio dell'alternanza dei segni evidenzia le regioni di maggiore criticità. Ad esempio, dalla tabella riportata è possibile osservare che per i redditi compresi tra 40 182.747 e 80 626 (in migliaia di Lit), ovvero la regione dei redditi non eccessivamente superiori al reddito medio, il modello adattato sottostima la distribuzione.

5.3 Confronto fra modello di Zenga e modello di Dagum

Come descritto in precedenza, per il modello di Dagum è stato impiegato un numero inferiore di metodi di stima dei parametri. Questi metodi sono quelli basati sulla minimizzazione di un indice di accostamento, senza vincoli e con l'introduzione di un vincolo (sulla mediana o sulla media aritmetica). Sono stati impiegati quindi 9 metodi di stima per ognuna delle 114 distribuzioni, sia per il modello di Dagum che per il modello di Zenga. Conseguentemente è possibile confrontare le due distribuzioni sulla base di 1026 procedimenti di stima parametrica, valutando il valore assunto dall'indice di accostamento oggetto di minimizzazione.

Da tali confronti si evince che per 314 volte sulle 1026 stime, circa il 30%, il modello di Zenga ha dato luogo ad un indice di accostamento inferiore rispetto al modello di Dagum, è però necessario valutare l'entità di tali differenze e le cause.

Innanzitutto si analizza l'influenza delle diverse distribuzioni.

Nei diversi anni di rilevazione, dal 1994 al 2001, non si osserva una differenza significativa dei risultati.

Se si considerano i diversi Stati Membri oggetto dell'indagine, si osservano numerose disparità. Per Finlandia, Germania, Olanda, Regno Unito e Svezia il modello di Dagum ha sempre ottenuto un miglior adattamento rispetto al modello di Zenga. Per Austria, Grecia, Irlanda, Francia e Italia in un numero minore di casi il modello di Zenga è risultato maggiormente efficiente del modello di Dagum. Infine, per Danimarca, Portogallo, Belgio, Spagna e Lussemburgo, almeno nella metà dei confronti il modello di Zenga è risultato più efficiente del modello di Dagum.

Ora si valuta l'influenza dei metodi di stima e dei vincoli.

In figura 5.3.1 e in tabella 5.5 è riportata una sintesi dei risultati ottenuti con il metodo del minimo A_1 nell'applicazione sulle 114 distribuzioni. Nel grafico, in ascissa sono riportati i valori dell'indice

A_1 ottenuti nella stima dei parametri del modello di Dagum (A_1^D), in ordinata i rispettivi valori ottenuti per il modello di Zenga (A_1^Z). La linea diagonale continua rappresenta il luogo dei punti nei quali $A_1^D = A_1^Z$, pertanto più un punto si allontana da tale linea maggiore è la differenza del livello di accostamento ai dati fra il modello di Dagum e il modello di Zenga. Al di sopra di tale linea vi è un maggiore accostamento per il modello di Dagum, al di sotto un maggiore accostamento per il modello di Zenga. Nel grafico è rappresentata una quadri-partizione del piano. Scelto γ quale livello soglia di accettabilità dell'accostamento dei modelli, nei quadranti di sinistra vi sono le distribuzioni per le quali si ritiene che l'accostamento del modello di Dagum sia adeguato ($A_1^D \leq \gamma$), nei quadranti in basso le distribuzioni per le quali si ritiene che l'accostamento del modello di Zenga sia adeguato ($A_1^Z \leq \gamma$). In tabella sono riportate le frequenze, assolute e relative, delle distribuzioni suddivise per quadrante per diversi valori di γ .

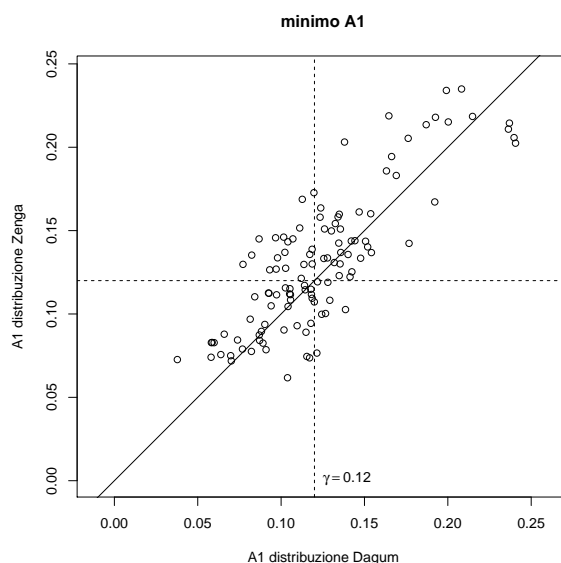


Figura 5.3.1: Grafico a dispersione per il confronto dell'accostamento fra il modello di Dagum e il modello di Zenga con stime ottenute con il minimo A_1

γ	$A_1^D \leq \gamma$ e $A_1^Z \leq \gamma$	$A_1^D \leq \gamma$ e $A_1^Z > \gamma$	$A_1^D > \gamma$ e $A_1^Z \leq \gamma$	$A_1^D > \gamma$ e $A_1^Z > \gamma$
0.1	19 (17%)	12 (11%)	9 (8%)	74 (65%)
0.11	25 (22%)	20 (18%)	11 (10%)	58 (51%)
0.12	43 (38%)	20 (18%)	7 (6%)	44 (39%)
0.13	56 (49%)	18 (16%)	4 (4%)	36 (32%)
0.14	68 (60%)	18 (16%)	5 (4%)	23 (20%)
0.15	81 (71%)	12 (11%)	4 (4%)	17 (15%)
0.16	91 (80%)	6 (5%)	1 (1%)	16 (14%)

Tabella 5.5: Tabella per il confronto dell'accostamento fra il modello di Dagum e il modello di Zenga con stime ottenute con il minimo A_1

Si ritiene che, più che la differenza fra i livelli di accostamento, sia rilevante valutare quante distribuzioni ottengono un livello di accostamento accettabile solo per uno dei due modelli, attribuendo meno importanza alle distribuzioni per le quali entrambi i modelli non sono stati in

grado di rappresentarle adeguatamente. Seppur è sempre maggiore il numero di distribuzioni per le quali si è osservato un migliore accostamento per il modello di Dagum, dal grafico 5.3.1, dalla terza e dalla quarta colonna della tabella 5.5, non si osserva una significativa superiorità di un modello sull'altro.

Si seguito sono riportati i grafici e le tabelle relative alla minimizzazione degli indici di accostamento A'_2 e A_2 . Nella lettura dei risultati è necessario ricordare che tali indici sono medie quadratiche, pertanto assumono valori più elevati dell'indice A_1 .

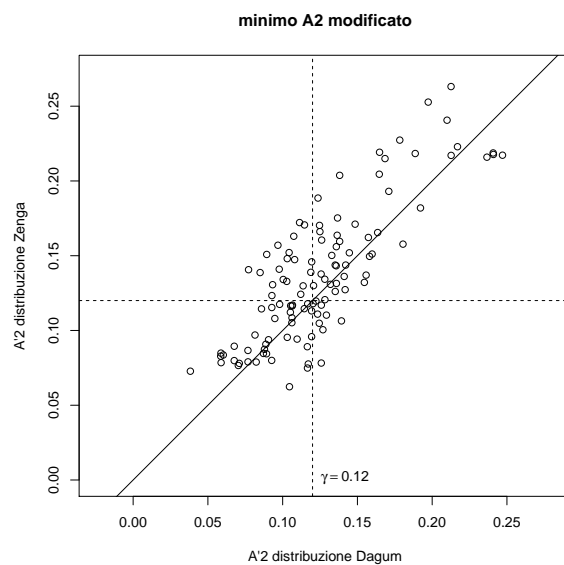


Figura 5.3.2: Grafico a dispersione per il confronto dell'accostamento fra il modello di Dagum e il modello di Zenga con stime ottenute con il minimo A'_2

γ	$A_2^{D} \leq \gamma$ e $A_2^{Z} \leq \gamma$	$A_2^{D} \leq \gamma$ e $A_2^{Z} > \gamma$	$A_2^{D} > \gamma$ e $A_2^{Z} \leq \gamma$	$A_1^{D} > \gamma$ e $A_1^{Z} > \gamma$
0.1	19 (17%)	11 (10%)	8 (7%)	76 (67%)
0.11	25 (22%)	20 (18%)	8 (7%)	61 (54%)
0.12	39 (34%)	19 (17%)	9 (8%)	47 (41%)
0.13	52 (46%)	22 (19%)	3 (3%)	37 (32%)
0.14	63 (55%)	23 (20%)	4 (4%)	24 (21%)
0.15	73 (64%)	18 (16%)	3 (3%)	20 (18%)
0.16	84 (74%)	12 (11%)	1 (1%)	17 (15%)

Tabella 5.6: Tabella per il confronto dell'accostamento fra il modello di Dagum e il modello di Zenga con stime ottenute con il minimo A'_2

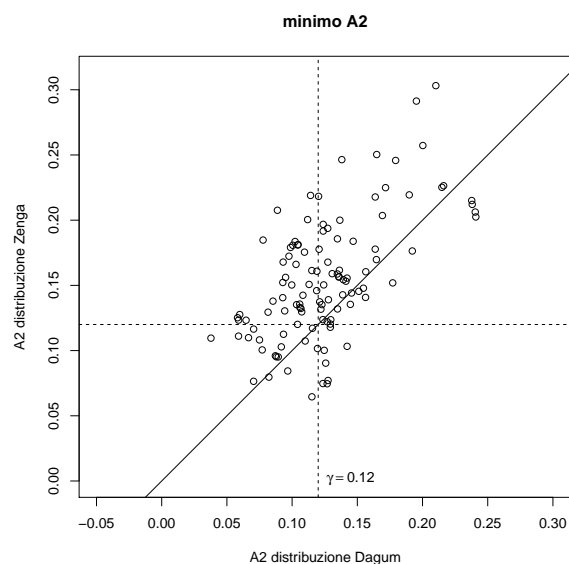


Figura 5.3.3: Grafico a dispersione per il confronto dell'accostamento fra il modello di Dagum e il modello di Zenga con stime ottenute con il minimo A_2

γ	$A_2^D \leq \gamma$ e $A_2^Z \leq \gamma$	$A_2^D \leq \gamma$ e $A_2^Z > \gamma$	$A_2^D > \gamma$ e $A_2^Z \leq \gamma$	$A_2^D > \gamma$ e $A_2^Z > \gamma$
0.1	6 (5%)	24 (21%)	5 (4%)	79 (69%)
0.11	11 (10%)	32 (28%)	9 (8%)	62 (54%)
0.12	18 (16%)	35 (31%)	7 (6%)	54 (47%)
0.13	35 (31%)	39 (34%)	1 (1%)	39 (34%)
0.14	46 (40%)	39 (34%)	2 (2%)	27 (24%)
0.15	53 (46%)	38 (33%)	3 (3%)	20 (18%)
0.16	68 (60%)	27 (24%)	1 (1%)	18 (16%)

Tabella 5.7: Tabella per il confronto dell'accostamento fra il modello di Dagum e il modello di Zenga con stime ottenute con il minimo A_2

Vengono forniti i grafici e le tabelle per il confronto dei due modelli nel caso della minimizzazione degli indici di accostamento vincolata alla mediana. Per l'introduzione del vincolo gli indici di accostamento possono assumere valori maggiori.

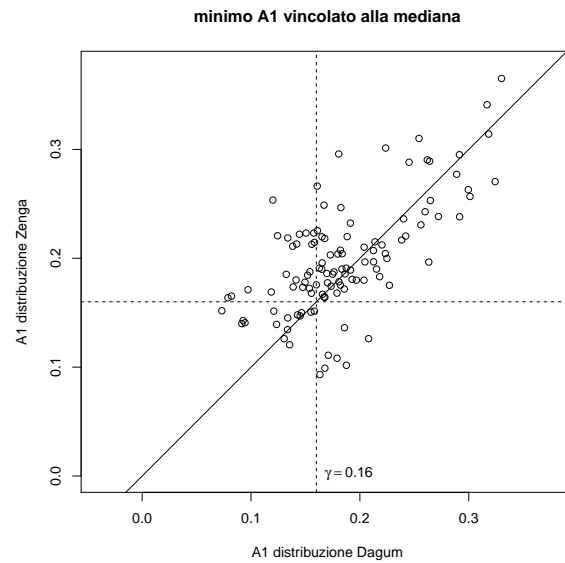


Figura 5.3.4: Grafico a dispersione per il confronto dell'accostamento fra il modello di Dagum e il modello di Zenga con stime ottenute con il minimo A_1 vincolato alla mediana

γ	$A_1^D \leq \gamma$ e $A_1^Z \leq \gamma$	$A_1^D \leq \gamma$ e $A_1^Z > \gamma$	$A_1^D > \gamma$ e $A_1^Z \leq \gamma$	$A_1^D > \gamma$ e $A_1^Z > \gamma$
0.14	5 (4%)	15 (13%)	7 (6%)	87 (76%)
0.15	11 (10%)	17 (15%)	7 (6%)	79 (69%)
0.16	15 (13%)	23 (20%)	7 (6%)	69 (61%)
0.17	24 (21%)	29 (25%)	6 (5%)	55 (48%)
0.18	35 (31%)	27 (24%)	9 (8%)	43 (38%)
0.19	49 (43%)	26 (23%)	7 (6%)	32 (28%)
0.2	56 (49%)	23 (20%)	9 (8%)	26 (23%)

Tabella 5.8: Tabella per il confronto dell'accostamento fra il modello di Dagum e il modello di Zenga con stime ottenute con il minimo A_1 vincolato alla mediana

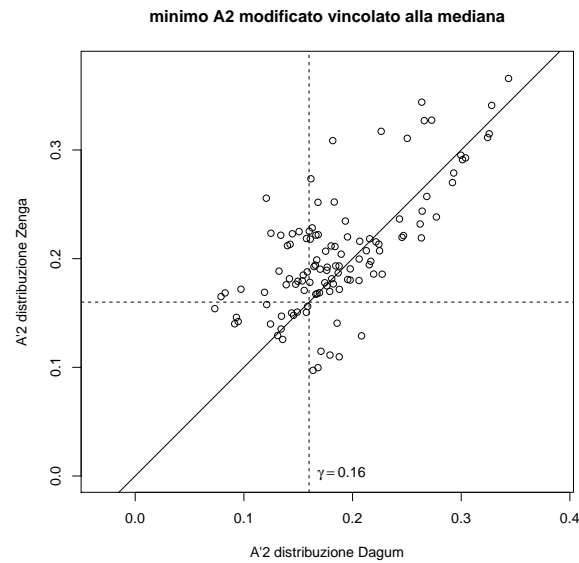


Figura 5.3.5: Grafico a dispersione per il confronto dell'accostamento fra il modello di Dagum e il modello di Zenga con stime ottenute con il minimo A_2' vincolato alla mediana

γ	$A_2'^D \leq \gamma$ e $A_2'^Z \leq \gamma$	$A_2'^D \leq \gamma$ e $A_2'^Z > \gamma$	$A_2'^D > \gamma$ e $A_2'^Z \leq \gamma$	$A_1'^D > \gamma$ e $A_1'^Z > \gamma$
0.14	4 (4%)	15 (13%)	6 (5%)	89 (78%)
0.15	9 (8%)	19 (17%)	7 (6%)	79 (69%)
0.16	15 (13%)	21 (18%)	7 (6%)	71 (62%)
0.17	23 (20%)	29 (25%)	6 (5%)	56 (49%)
0.18	35 (31%)	26 (23%)	6 (5%)	47 (41%)
0.19	46 (40%)	28 (25%)	6 (5%)	34 (30%)
0.2	56 (49%)	23 (20%)	7 (6%)	28 (25%)

Tabella 5.9: Tabella per il confronto dell'accostamento fra il modello di Dagum e il modello di Zenga con stime ottenute con il minimo A_2' vincolato alla mediana

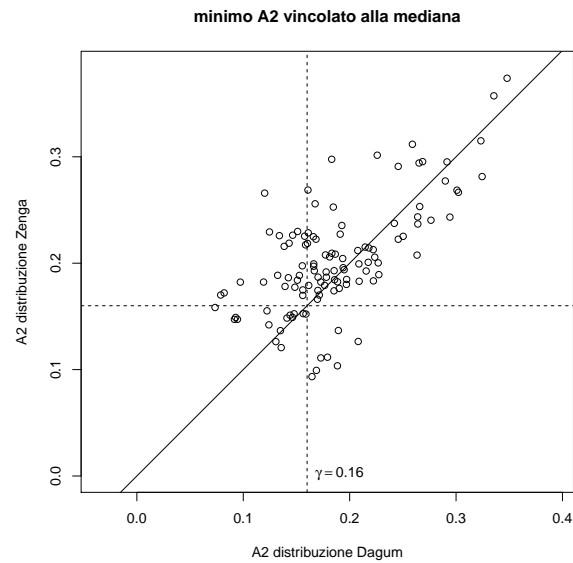


Figura 5.3.6: Grafico a dispersione per il confronto dell'accostamento fra il modello di Dagum e il modello di Zenga con stime ottenute con il minimo A_2 vincolato alla mediana

γ	$A_2^D \leq \gamma$ e $A_2^Z \leq \gamma$	$A_2^D \leq \gamma$ e $A_2^Z > \gamma$	$A_2^D > \gamma$ e $A_2^Z \leq \gamma$	$A_2^D > \gamma$ e $A_2^Z > \gamma$
0.14	3 (3%)	16 (14%)	7 (6%)	88 (77%)
0.15	9 (8%)	18 (16%)	7 (6%)	80 (70%)
0.16	15 (13%)	22 (19%)	7 (6%)	70 (61%)
0.17	19 (17%)	31 (27%)	5 (4%)	59 (52%)
0.18	30 (26%)	30 (26%)	5 (4%)	49 (43%)
0.19	44 (39%)	27 (24%)	7 (6%)	36 (32%)
0.2	55 (48%)	24 (21%)	6 (5%)	29 (25%)

Tabella 5.10: Tabella per il confronto dell'accostamento fra il modello di Dagum e il modello di Zenga con stime ottenute con il minimo A_2 vincolato alla mediana

Infine si riportano grafici e tabelle relativi alla minimizzazione degli indici di accostamento vincolata alla media aritmetica.

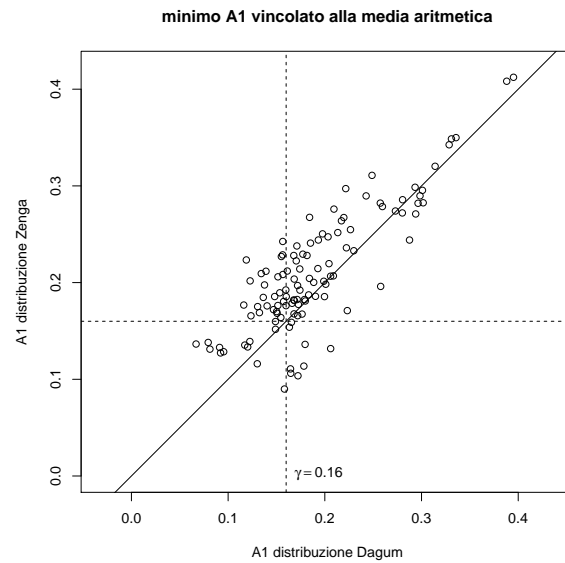


Figura 5.3.7: Grafico a dispersione per il confronto dell'accostamento fra il modello di Dagum e il modello di Zenga con stime ottenute con il minimo A_1 vincolato alla media aritmetica

γ	$A_1^D \leq \gamma$ e $A_1^Z \leq \gamma$	$A_1^D \leq \gamma$ e $A_1^Z > \gamma$	$A_1^D > \gamma$ e $A_1^Z \leq \gamma$	$A_1^D > \gamma$ e $A_1^Z > \gamma$
0.14	10 (9%)	10 (9%)	7 (6%)	87 (76%)
0.15	10 (9%)	15 (13%)	7 (6%)	82 (72%)
0.16	13 (11%)	26 (23%)	8 (7%)	67 (59%)
0.17	22 (19%)	29 (25%)	6 (5%)	57 (50%)
0.18	36 (32%)	30 (26%)	2 (2%)	46 (40%)
0.19	47 (41%)	25 (22%)	4 (4%)	38 (33%)
0.2	53 (46%)	25 (22%)	4 (4%)	32 (28%)

Tabella 5.11: Tabella per il confronto dell'accostamento fra il modello di Dagum e il modello di Zenga con stime ottenute con il minimo A_1 vincolato alla media aritmetica

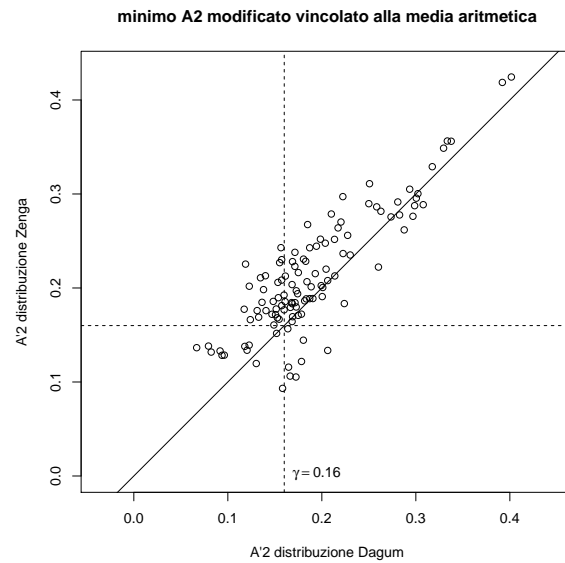


Figura 5.3.8: Grafico a dispersione per il confronto dell'accostamento fra il modello di Dagum e il modello di Zenga con stime ottenute con il minimo A_2' vincolato alla media aritmetica

γ	$A_2'^D \leq \gamma$ e $A_2'^Z \leq \gamma$	$A_2'^D \leq \gamma$ e $A_2'^Z > \gamma$	$A_2'^D > \gamma$ e $A_2'^Z \leq \gamma$	$A_1'^D > \gamma$ e $A_1'^Z > \gamma$
0.14	10 (9%)	9 (8%)	6 (5%)	89 (78%)
0.15	10 (9%)	14 (12%)	7 (6%)	83 (73%)
0.16	12 (11%)	27 (24%)	7 (6%)	68 (60%)
0.17	22 (19%)	29 (25%)	4 (4%)	59 (52%)
0.18	35 (31%)	27 (24%)	2 (2%)	50 (44%)
0.19	47 (41%)	25 (22%)	3 (3%)	39 (34%)
0.2	52 (46%)	25 (22%)	3 (3%)	34 (30%)

Tabella 5.12: Tabella per il confronto dell'accostamento fra il modello di Dagum e il modello di Zenga con stime ottenute con il minimo A_2' vincolato alla media aritmetica

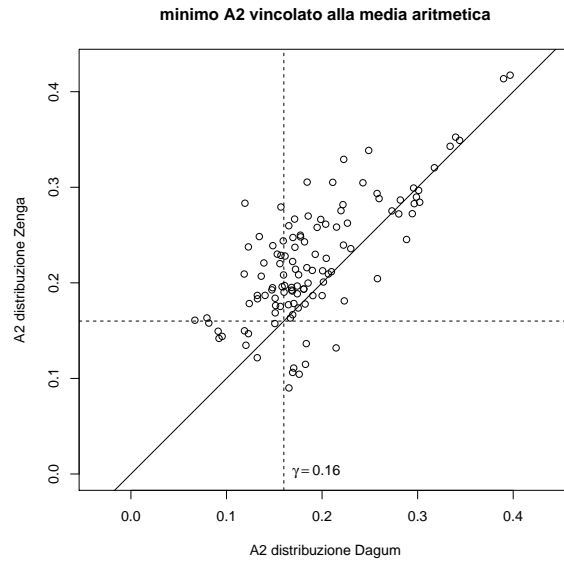


Figura 5.3.9: Grafico a dispersione per il confronto dell'accostamento fra il modello di Dagum e il modello di Zenga con stime ottenute con il minimo A_2 vincolato alla media aritmetica

γ	$A_2^D \leq \gamma$ e $A_2^Z \leq \gamma$	$A_2^D \leq \gamma$ e $A_2^Z > \gamma$	$A_2^D > \gamma$ e $A_2^Z \leq \gamma$	$A_2^D > \gamma$ e $A_2^Z > \gamma$
0.14	2 (2%)	17 (15%)	7 (6%)	88 (77%)
0.15	7 (6%)	16 (14%)	7 (6%)	84 (74%)
0.16	9 (8%)	26 (23%)	7 (6%)	72 (63%)
0.17	16 (14%)	33 (29%)	5 (4%)	60 (53%)
0.18	24 (21%)	37 (32%)	4 (4%)	49 (43%)
0.19	32 (28%)	40 (35%)	4 (4%)	38 (33%)
0.2	45 (39%)	31 (27%)	3 (3%)	35 (31%)

Tabella 5.13: Tabella per il confronto dell'accostamento fra il modello di Dagum e il modello di Zenga con stime ottenute con il minimo A_2 vincolato alla media aritmetica

Un altro criterio di confronto dei due modelli è la verifica di quanto alcuni indici del modello si accostano alle rispettive statistiche campionarie. A tale scopo si definiscono lo *scostamento assoluto relativo del valore atteso del modello stimato dalla media aritmetica campionaria*

$$\rho_E = \frac{|\bar{x} - \hat{E}(X)|}{\bar{x}}$$

dove \bar{x} è la media aritmetica campionaria e $\hat{E}(X)$ è il valore atteso del modello stimato, e lo *scostamento assoluto relativo della mediana del modello stimato dalla mediana campionaria*

$$\rho_{Me} = \frac{|Me - \hat{Me}|}{Me}$$

con Me mediana campionaria e \hat{Me} mediana del modello adattato. Di seguito sono riportati i grafici, nei casi di diversi metodi di stima, che pongono in ascissa ρ_E valutato per il modello di Dagum

e in ordinata ρ_{Me} valutato per il modello di Zenga. I punti sopra la diagonale rappresentano le distribuzioni per le quali lo scostamento relativo è maggiore per la distribuzione di Zenga, mentre sotto la diagonale vi sono le distribuzioni per le quali lo scostamento relativo è maggiore per la distribuzione di Dagum. Come si è fatto per gli indici di accostamento si può fissare uno scostamento relativo γ sotto il quale si può ritenere ammissibile la differenza fra media aritmetica campionaria e valore atteso del modello.

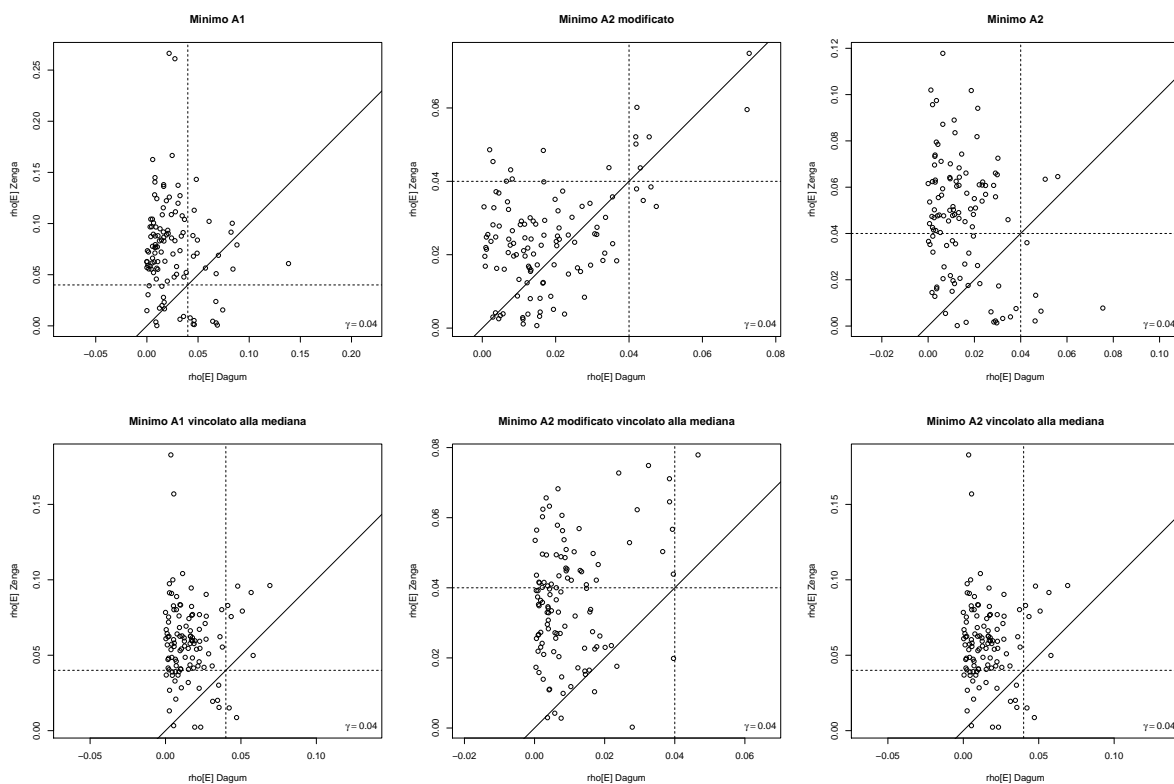


Figura 5.3.10: Grafici a dispersione di ρ_E per il confronto fra il modello di Dagum e il modello di Zenga per diversi metodi di stima

Dai grafici in figura 5.3.10 si può osservare un minore scostamento relativo del valore atteso dalla media aritmetica campionaria per il modello di Dagum. Sono in proporzione elevata le distribuzioni per le quali, scelta una soglia γ pari a 0.04, $\rho_E \leq \gamma$ per il modello di Dagum e $\rho_E > \gamma$ per il modello di Zenga. Si osserva che ρ_E si riduce impiegando il metodo del minimo A'_2 .

Di seguito sono riportati i grafici per ρ_{Me} .

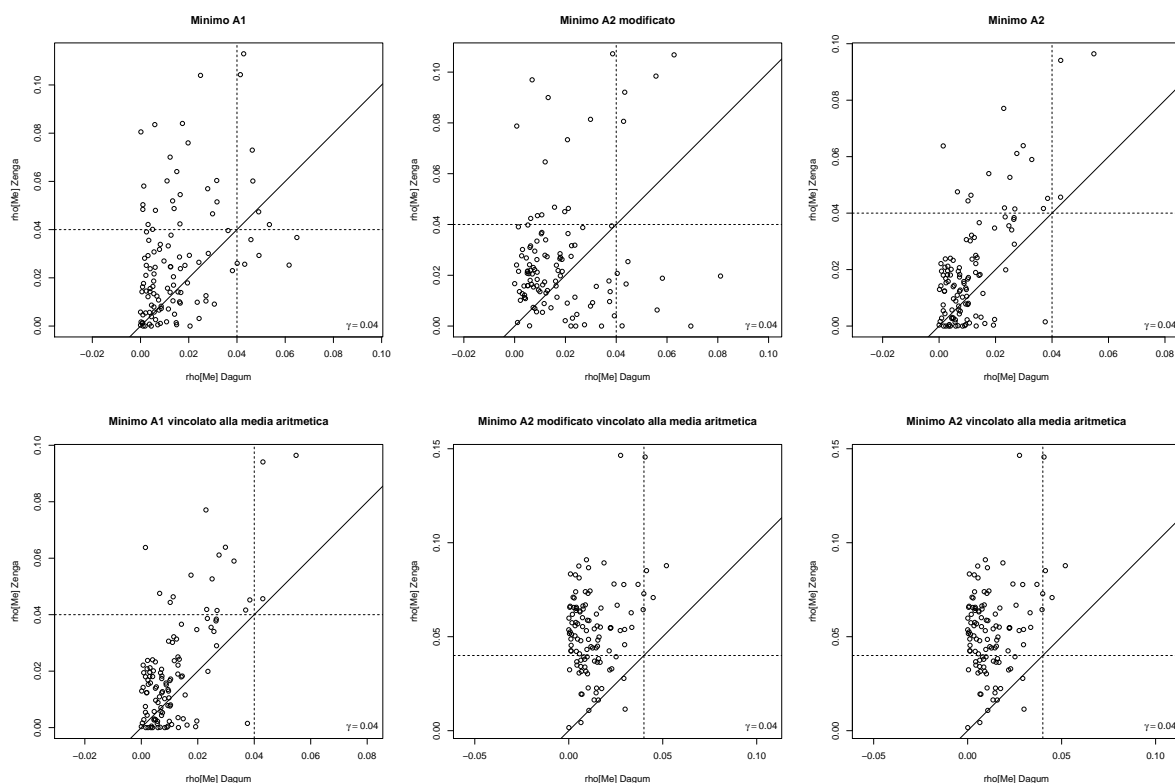


Figura 5.3.11: Grafici a dispersione di ρ_{Me} per il confronto fra il modello di Dagum e il modello di Zenga per diversi metodi di stima.

Scegliendo come soglia di accettabilità di ρ_{Me} il valore $\gamma = 0.04$ si osserva che aumentano le distribuzioni che rientrano in tale soglia, sia per il modello di Zenga che per il modello di Dagum. L'introduzione del vincolo alla media aritmetica si osserva che comporta valori di ρ_{Me} più elevati.

A conclusione della sezione di confronto fra il modello di Zenga e il modello di Dagum, in figura 5.3.12 si riportano i grafici che comparano le code dei due modelli, sulla base delle stime ottenute dei parametri che le regolano. A tal proposito si ricorda che per il modello di Zenga il parametro che regola la coda destra è α e i momenti, di ordine r del modello di Zenga, esistono finiti se $r < \alpha + 1$. Il parametro che regola la coda destra del modello di Dagum è δ e i momenti, di ordine r del modello di Dagum, esistono finiti se $r < \delta$. In figura è evidenziata la differenza fra il numero di momenti finiti di ordine intero r che si verifica fra il modello di Zenga e il modello di Dagum. Sulla diagonale più scura i modelli stimati di Zenga e Dagum hanno lo stesso numero di momenti, spostandosi sulle diagonali più elevate aumenta il numero di momenti finiti del modello di Zenga, viceversa spostandosi sulle diagonali più basse è il modello di Dagum ad avere più momenti finiti.

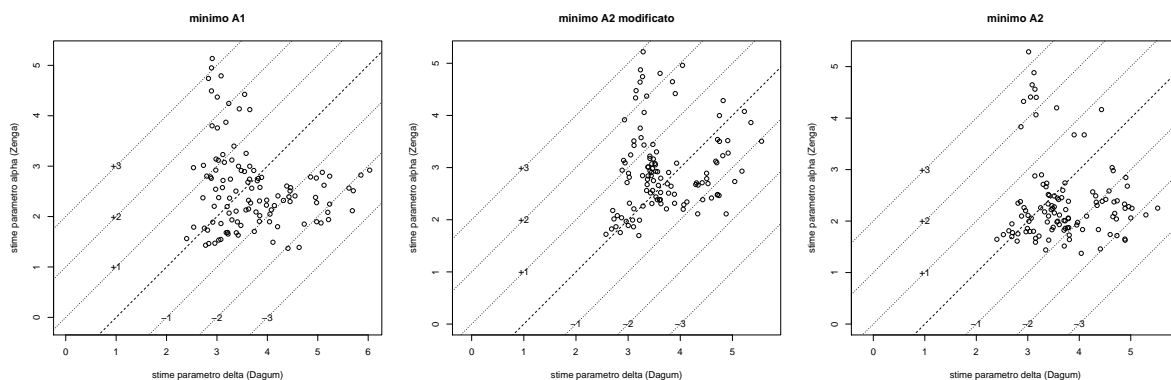


Figura 5.3.12: Grafici a dispersione che contrappongono i valori di $\hat{\delta}$ del modello di Dagum e i valori $\hat{\alpha}$ del modello di Zenga per il confronto fra le code dei modelli stimati

5.4 Confronto fra i diversi metodi di stima impiegati per il modello di Zenga e discussione sui vincoli

In tutte le distribuzioni empiriche analizzate, le stime dei parametri del modello di Zenga, ottenute con il metodo dei momenti comportano valori particolarmente elevati degli indici di accostamento, se confrontati con gli altri metodi applicati alla medesima distribuzione. La disponibilità di una definizione esplicita, fa sì che siano di facile determinazione, pertanto costituiscono un fondamentale ausilio, non sempre disponibile per altri modelli distributivi, per la determinazione dei valori iniziali di altri metodi di stima che fanno uso di procedure numeriche.

Si forniranno ora grafici e tabelle circa l'introduzione di vincoli nel metodo stima con il minimo A_1 . In ascissa e in ordinata verranno posti i valori di A_1 , la stessa analisi può essere condotta nei casi di minimizzazione degli indici A'_2 e A_2 .

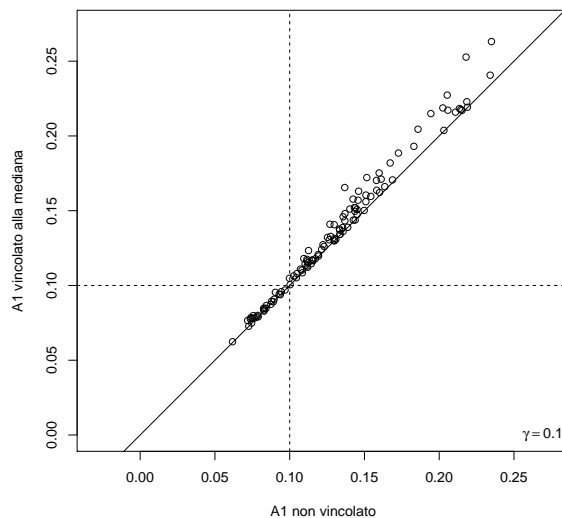


Figura 5.4.1: Grafico a dispersione per il confronto dell'accostamento del modello di Zenga con stime ottenute con il minimo A_1 non vincolato e il minimo A_1 vincolato alla mediana

γ	$A_1 \leq \gamma$ e $A_1 Me \leq \gamma$	$A_1 \leq \gamma$ e $A_1 Me > \gamma$	$A_1 > \gamma$ e $A_1 Me \leq \gamma$	$A_1 > \gamma$ e $A_1 Me > \gamma$
0.075	3 (3%)	4 (4%)	0 (0%)	107 (94%)
0.1	27 (24%)	1 (1%)	0 (0%)	86 (75%)
0.125	51 (45%)	2 (2%)	0 (0%)	61 (54%)
0.15	76 (67%)	9 (8%)	0 (0%)	29 (25%)
0.175	95 (83%)	3 (3%)	0 (0%)	16 (14%)

Tabella 5.14: Tabella per il confronto dell'accostamento del modello di Zenga con stime ottenute con il minimo A_1 non vincolato e il minimo A_1 vincolato alla mediana

Dal grafico in figura 5.4.1 e dalla tabella 5.14 emerge che l'aumento dell'indice di accostamento causato dall'introduzione del vincolo alla mediana è esiguo. I punti giacciono intorno alla diagonale che rappresenta l'uguaglianza degli indici di accostamento, allontanandosi da essa per valori elevati dell'indice. Ciò può essere riscontrato anche dalla terza colonna della tabella, che mostra per diverse soglie γ , che il numero di distribuzioni per le quali l'indice di accostamento supera γ in seguito all'introduzione del vincolo è esiguo. Si precisa che tutti i punti del grafico giacciono sopra la diagonale, e la quarta colonna della tabella contiene tutti valori nulli, poiché non si può ottenere una riduzione dell'indice di accostamento in seguito all'introduzione di un vincolo.

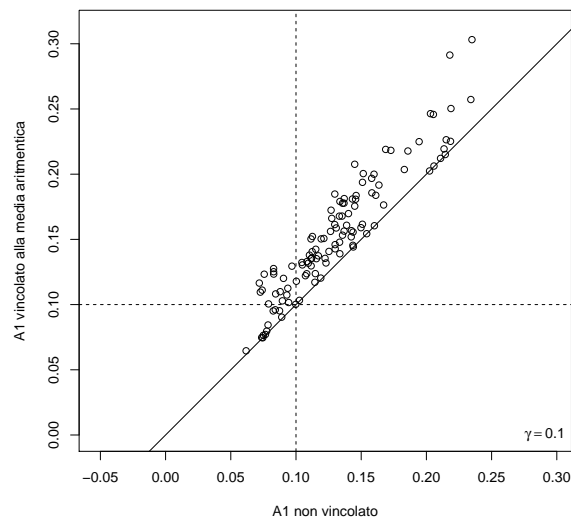


Figura 5.4.2: Grafico a dispersione per il confronto dell'accostamento del modello di Zenga con stime ottenute con il minimo A_1 non vincolato e il minimo A_1 vincolato alla media aritmetica

γ	$A_1 \leq \gamma$ e $A_1 \bar{x} \leq \gamma$	$A_1 \leq \gamma$ e $A_1 \bar{x} > \gamma$	$A_1 > \gamma$ e $A_1 \bar{x} \leq \gamma$	$A_1 > \gamma$ e $A_1 \bar{x} > \gamma$
0.075	3 (3%)	4 (4%)	0 (0%)	107 (94%)
0.1	11 (10%)	17 (15%)	0 (0%)	86 (75%)
0.125	32 (28%)	21 (18%)	0 (0%)	61 (54%)
0.15	56 (49%)	29 (25%)	0 (0%)	29 (25%)
0.175	78 (68%)	20 (18%)	0 (0%)	16 (14%)

Tabella 5.15: Tabella per il confronto dell'accostamento del modello di Zenga con stime ottenute con il minimo A_1 non vincolato e il minimo A_1 vincolato alla media aritmetica

Dal grafico in figura 5.4.2 e dalla tabella 5.15 si può osservare un incremento più significativo, rispetto al caso precedente, dell'indice di accostamento A_1 con l'introduzione del vincolo alla media aritmetica.

Il vincolo alla media aritmetica risulta però di più semplice applicazione del vincolo alla mediana, in quanto il valore atteso dipende esclusivamente dal parametro μ e la mediana dev'essere determinata numericamente e dipende da tutti e tre i parametri del modello. Pertanto la stima con due vincoli è stata ottenuta vincolando alla media aritmetica e a un indice di ineguaglianza.

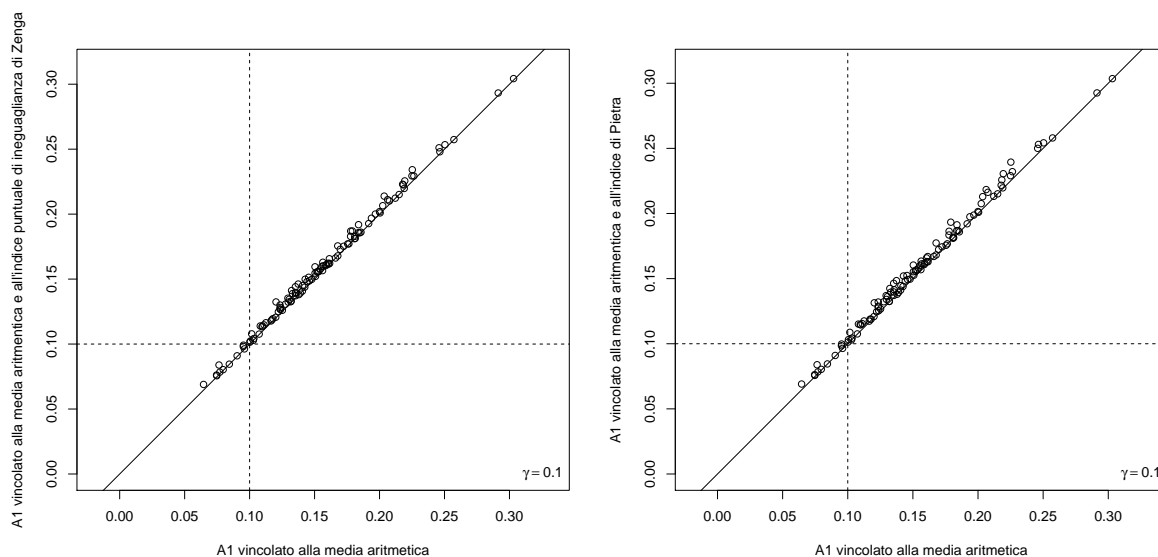


Figura 5.4.3: Grafici a dispersione per il confronto dell'accostamento del modello di Zenga con stime ottenute con il minimo A_1 con un vincolo e il minimo A_1 con due vincoli

Dai grafici in figura 5.4.3 si osserva che l'introduzione del vincolo all'indice di Pietra, P , piuttosto che del vincolo di ineguaglianza puntuale di Zenga in corrispondenza della media aritmetica, $\bar{J}(\bar{x})$, non comportano un significativo incremento nell'indice di accostamento A_1 . Pertanto si contrappongono i due metodi di stima con due vincoli.

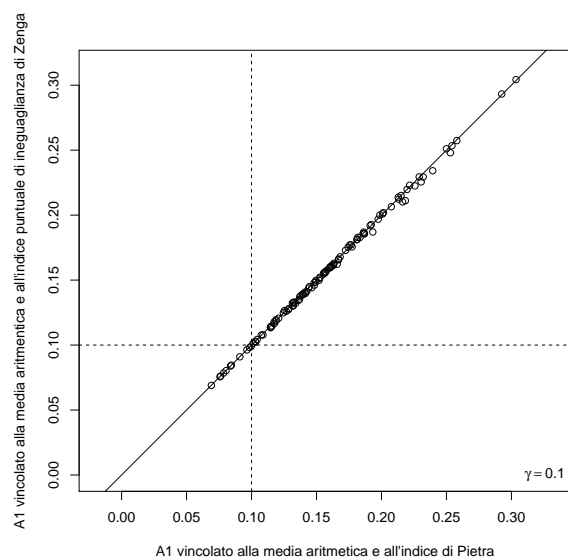


Figura 5.4.4: Grafico a dispersione per il confronto dell'accostamento del modello di Zenga con stime ottenute con il minimo A_1 vincolato alla media aritmetica e all'indice di Pietra e il minimo A_1 vincolato alla media aritmetica e all'indice di ineguaglianza puntuale di Zenga

γ	$A_1 \bar{x}, P \leq \gamma$ e $A_1 \bar{x}, J(\bar{x}) \leq \gamma$	$A_1 \bar{x}, P \leq \gamma$ e $A_1 \bar{x}, J(\bar{x}) > \gamma$	$A_1 \bar{x}, P > \gamma$ e $A_1 \bar{x}, J(\bar{x}) \leq \gamma$	$A_1 \bar{x}, P > \gamma$ e $A_1 \bar{x}, J(\bar{x}) > \gamma$
0.075	1 (1%)	0 (0%)	0 (0%)	113 (99%)
0.1	11 (10%)	0 (0%)	0 (0%)	103 (90%)
0.125	27 (24%)	0 (0%)	0 (0%)	87 (76%)
0.15	54 (47%)	0 (0%)	1 (1%)	59 (52%)
0.175	76 (67%)	1 (1%)	0 (0%)	37 (32%)

Tabella 5.16: Tabella per il confronto dell'accostamento del modello di Zenga con stime ottenute con il minimo A_1 vincolato alla media aritmetica e all'indice di Pietra e il minimo A_1 vincolato alla media aritmetica e all'indice di ineguaglianza puntuale di Zenga

Dal grafico 5.4.4 e dalla terza e dalla quarta colonna della tabella 5.16 si evince che i due metodi hanno restituito risultati equivalenti per le distribuzioni considerate.

5.5 Studio delle distribuzioni degli stimatori: primi risultati

In conclusione delle applicazioni si presentano gli studi circa le distribuzioni degli stimatori dei parametri del modello distributivo di Zenga. A tale scopo è stato impiegato il metodo Bootstrap parametrico descritto nella sezione 4.3. I risultati riportati sono soltanto preliminari a causa della complessità computazionale.

Attraverso il Bootstrap parametrico si intende approssimare le distribuzioni degli stimatori e degli indici di accostamento, al variare dei metodi di stima impiegati e delle distribuzioni reali sui quali sono stati impiegati.

Per ogni distribuzione reale, in questo lavoro, sono stati usati 16 metodi di stima dei parametri del modello distributivo di Zenga (a titolo esemplificativo si veda la tabella 5.2). Per ogni metodo di stima sono stati generati B campioni Bootstrap, dai quali sono state ricavate altrettante repliche Bootstrap delle stime dei parametri del modello di Zenga e dei rispettivi indici di accostamento.

Il numero B di campioni Bootstrap che è stato scelto è 1000. Considerando l'asimmetria del modello, il numero di repliche Bootstrap impiegato per approssimare le distribuzioni degli stimatori e degli indici di accostamento è certamente basso. Nonostante ciò il tempo necessario per ottenere tali risultati, relativi a una sola distribuzione reale, impiegando un calcolatore di media potenza, è di circa 10 ore. Di conseguenza questo studio è stato condotto su un numero limitato di distribuzioni.

Le distribuzioni simulate degli indici di accostamento sono state usate per costruire intervalli sulla base dei quali valutare i risultati ottenuti sui dati reali. Quindi se l'indice di accostamento ottenuto sui dati reali dovesse appartenere a tali intervalli, si è propensi a considerare che il modello di Zenga sia adeguato a descrivere la distribuzione reale. A causa del ridotto numero di replicazioni Bootstrap il livello nominale di copertura degli intervalli può essere inferiore al livello di copertura effettivo, facilitando il rifiuto del modello. Per tale motivo si è scelto di illustrare il procedimento in un caso particolarmente favorevole al modello di Zenga, la distribuzione del Lussemburgo 1998, rinviando l'approfondimento della metodologia e del livello di copertura effettivo di tali intervalli a lavori futuri.

Prima di presentare i risultati del Bootstrap Parametrico è opportuno mostrare le statistiche relative alla distribuzione reale e le stime dei parametri del modello.

Statistiche	
N	2.521000000000000e + 03
media	1.683882858800000e + 06
mediana	1.457700594300000e + 06
varianza	9.62991415340592e + 11
coef. var.	5.828000000000000e - 01
m3 cent.	1.70263106934764e + 18
Indice di Gini	3.010000000000000e - 01
I	6.293000000000000e - 01
Indice di Pietra	2.166000000000000e - 01
F(media)	5.970000000000000e - 01
I(media)	5.856000000000000e - 01

Tabella 5.17: Statistiche Lussemburgo 1998, moneta corrente Franco lussemburghese (sostituito nel 1999 con l'Euro al tasso di cambio di 1 Euro per 40.3399 BEF)

Innanzitutto si osserva che la numerosità campionaria, 2521, risulta essere un valore relativamente basso se confrontato con le numerosità campionarie di cui si dispone per le altre distribuzioni. Ciò avvalorava la scelta di impiegare le distribuzioni approssimate mediante Bootstrap parametrico invece delle distribuzioni asintotiche.

Seguono, in tabella 5.18, gli esiti dei diversi metodi di stima impiegati per questa distribuzione.

Stime Miscuglio	$\hat{\mu}$	$\hat{\alpha}$	$\hat{\theta}$	A_1	A'_2	A_2	Me
metodo dei momenti	1683882.8588	6.3769	8.5598	0.0927	0.1294	0.1327	1420181.2784
min A1	1710084.5009	4.4931	6.0160	0.0618	0.1049	0.0946	1447373.7454
min A'2	1677463.2409	4.6392	5.9513	0.0704	0.0932	0.0913	1433803.9974
min A2	1690568.5586	4.6462	6.0705	0.0654	0.0951	0.0899	1438800.4445
min A1 v Me	1722982.3907	4.6636	6.2453	0.0624	0.1088	0.0969	1457700.5943
min A'2 v Me	1701313.8298	4.5000	5.7207	0.0675	0.0972	0.0949	1457700.5943
min A2 v Me	1709836.6276	4.5395	5.8939	0.0642	0.0992	0.0932	1457700.5943
min A1 v M1	1683882.8588	4.6684	6.2152	0.0646	0.0987	0.0921	1426616.7051
min A'2 v M1	1683882.8588	4.6236	5.9400	0.0689	0.0934	0.0909	1438852.8149
min A2 v M1	1683882.8588	4.6736	6.0903	0.0667	0.0943	0.0900	1433905.6925
min A1 v M1, I(M1)	1683882.8588	4.6537	5.9737	0.0689	0.0934	0.0911	1439036.0582
min A'2 v M1, I(M1)	1683882.8588	4.6238	5.9315	0.0691	0.0934	0.0911	1439340.5744
min A2 v M1, I(M1)	1683882.8588	4.6686	5.9946	0.0690	0.0935	0.0911	1438886.3680
min A1 v M1, P	1683882.8588	4.6482	5.9609	0.0691	0.0934	0.0912	1439372.6599
min A'2 v M1, P	1683882.8588	4.6240	5.9268	0.0692	0.0934	0.0912	1439619.5977
min A2 v M1, P	1683882.8588	4.6682	5.9889	0.0691	0.0935	0.0912	1439171.7082

Tabella 5.18: Stime dei parametri del modello di Zenga per la distribuzione Lussemburgo 1998 (media: 1 683 882.8588 BEF; mediana: 1 457 700.5943 BEF)

I valori degli indici di accostamento risultano bassi se confrontati con quelli di tutte le altre distribuzioni osservate. Al fine di mostrare l'adattamento ottenuto dal modello di Zenga, in figura 5.5.1 è riportato il confronto fra il modello di Zenga e il modello di Dagum con stime dei parametri ottenute con il metodo del minimo A_1 .

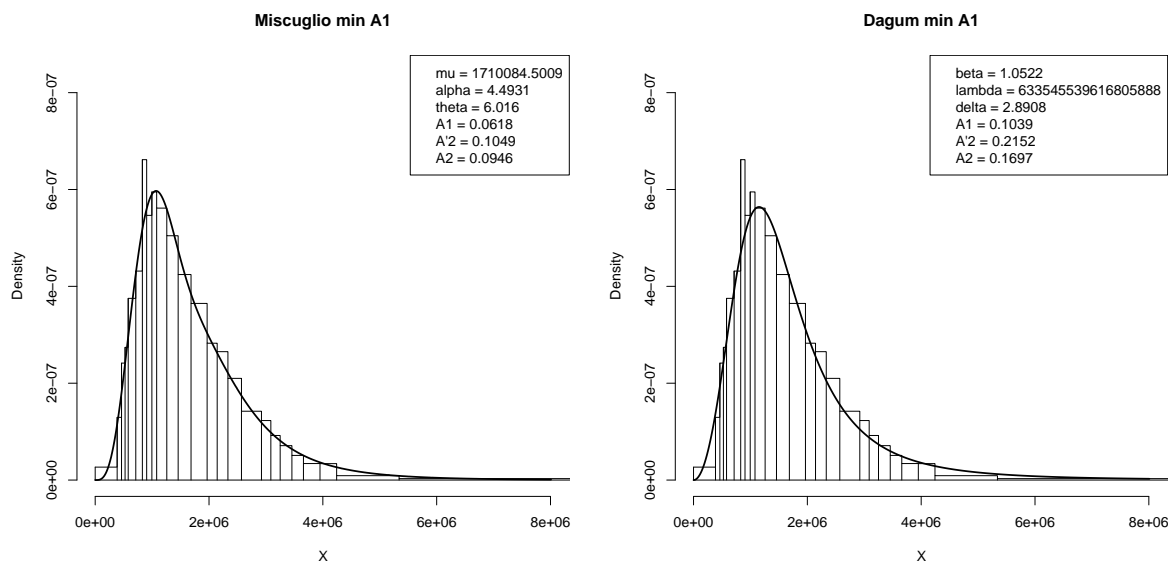


Figura 5.5.1: Modello di Zenga (a sinistra) e il modello di Dagum (a destra) con stime dei parametri ottenute con il metodo del minimo A_1 sovrapposti all'istogramma raffigurante la distribuzione dei redditi del Lussemburgo 1998

Ora si mostrano i risultati ottenuti a seguito dell'impiego del Bootstrap parametrico.

	$\hat{\mu}$	$\hat{\alpha}$	$\hat{\theta}$	A_1	A'_2	A_2
min	1620857.8230	3.2849	3.9167	0.0435	0.0576	0.0579
1stQ	1670796.5345	5.9877	8.0354	0.0674	0.0926	0.0936
me	1683921.4434	6.9757	9.4825	0.0760	0.1050	0.1059
3rdQ	1695951.4484	7.9897	11.0309	0.0868	0.1200	0.1241
max	1740813.9496	14.5924	20.6765	0.1845	0.2454	0.3489
inf int 0.025	1644813.9490	4.2693	5.4139	0.0515	0.0724	0.0712
sup int 0.025	1721702.2247	10.2321	14.1999	0.1246	0.1703	0.1849
mean	1683325.9379	7.0574	9.6062	0.0790	0.1085	0.1122
sd	19268.2972	1.5347	2.2817	0.0178	0.0241	0.0305
mad	15338.5744	1.1932	1.7756	0.0133	0.0178	0.0214

Tabella 5.19: Distribuzione approssimata degli stimatori del metodo dei momenti e dei rispettivi indici di accostamento

In tabella 5.19 sono riportati i risultati ottenuti circa il metodo dei momenti. Questi risultati devono essere confrontati con quelli della prima riga della tabella 5.18. Si illustra il significato dei valori riportati per riga:

min valore minimo assunto dalle repliche Bootstrap degli stimatori e degli indici di accostamento;

1stQ primo quartile delle distribuzioni Bootstrap degli stimatori e degli indici di accostamento;

me mediana;

3rdQ terzo quartile;

max valore massimo assunto dalle repliche Bootstrap;

inf int 0.025 estremo inferiore dell'intervallo bilaterale con livello di copertura nominale 0.95;

sup int 0.025 estremo superiore dell'intervallo bilaterale con livello di copertura nominale 0.95;

mean media;

sd deviazione standard;

mad mean absolute deviation, scostamento medio assoluto dalla media aritmetica.

Si ricorda che nella distribuzione Normale il rapporto tra scostamento medio assoluto e deviazione standard è costante e vale $\sqrt{2/\pi} = 0.79788456$. Valutando tale rapporto per le diverse distribuzioni Bootstrap ottenute è possibile formulare una prima considerazione sul grado di vicinanza al modello Normale. Ad esempio, la stima del parametro μ con il metodo dei momenti, è stata determinata per mezzo dello stimatore media campionaria. Tale stimatore ha distribuzione asintotica Normale. Valutando il rapporto fra scostamento medio assoluto e deviazione standard per la distribuzione di $\hat{\mu}$ si ottiene

$$\frac{15338.5744}{19268.2972} = 0.7960524$$

che è un valore molto vicino a quello previsto dal modello Normale.

Il confronto delle stime sui dati reali con la media della distribuzione Bootstrap degli stimatori dei parametri, permette di valutare la correttezza di tali stimatori. Si osserva che nel caso del metodo dei momenti, il valore atteso dello stimatore di μ , 1683325.9379, è prossimo in termini relativi, a 1683882.8588 che è il valore impiegato per ricampionare dal modello. Ciò non è verificato invece per gli stimatori di α e di θ . Lo stimatore di α assume mediamente il valore 7.0574 mentre il valore che è stato usato per ricampionare è 6.3769. Il valore medio dello stimatore di θ è 9.6062 mentre il valore usato nel ricampionamento è 8.5598.

Per il metodo dei momenti si è ottenuto un valore di A_1 pari a 0.0927 che risulta essere il più elevato fra tutti i metodi impiegati. L'estremo superiore dell'intervallo con livello di copertura nominale al 95% è pari a 0.1246. Con ciò si osserva che, seppur è elevato il valore di A_1 ottenuto sui dati reali, non appartiene al 2.5% che eccede l'intervallo. Le medesime considerazioni possono essere valutate anche per gli indici A_2 e A_2' .

	$\hat{\mu}$	$\hat{\alpha}$	$\hat{\theta}$	A_1	A_2'	A_2
min	1625679.2722	3.2880	4.1387	0.0296	0.0536	0.0529
1stQ	1689230.0543	4.2743	5.6690	0.0557	0.0827	0.0836
me	1707372.9768	4.5026	6.0289	0.0632	0.0924	0.0949
3rdQ	1725782.8852	4.7654	6.3895	0.0709	0.1025	0.1049
max	1808140.9726	6.1865	8.9659	0.1096	0.1646	0.1908
inf int 0.025	1656641.4480	3.8669	5.0040	0.0415	0.0654	0.0667
sup int 0.025	1760092.8270	5.3542	7.3468	0.0871	0.1247	0.1291
mean	1707430.7980	4.5357	6.0528	0.0633	0.0930	0.0951
sd	27170.7961	0.3883	0.5881	0.0114	0.0153	0.0164
mad	21723.8446	0.3032	0.4568	0.0091	0.0121	0.0128

Tabella 5.20: Distribuzione approssimata degli stimatori del minimo A_1 e dei rispettivi indici di accostamento

In tabella 5.20 sono riportati i risultati delle simulazioni applicate sul metodo del minimo A_1 . I valori delle stime dei parametri ottenuti sui dati reali e che sono stati usati per ricampionare dal modello distributivo sono:

$$\begin{aligned}\hat{\mu} &= 1710084.5009 \\ \hat{\alpha} &= 4.4931 \\ \hat{\theta} &= 6.0160\end{aligned}$$

e il livello di accostamento sulla distribuzione reale raggiunto dal modello è

$$A_1 = 0.0618.$$

Si osserva che i valori attesi delle distribuzioni simulate degli stimatori sono prossimi ai valori attribuiti ai parametri per il ricampionamento. Si evidenzia quindi la differenza di quanto ottenuto con il metodo dei momenti, che per la numerosità campionaria 2521 gli stimatori dei parametri α e θ sono risultati distorti. Per quanto riguarda il valore di A_1 ottenuto sui dati reali, esso appartiene all'intervallo con livello di copertura nominale del 95%. Da ciò si può dedurre che il modello distributivo di Zenga e le stime dei parametri ottenute con il minimo A_1 descrivono in modo adeguato la distribuzione reale.

Capitolo 6

Conclusioni

In questo lavoro sono stati analizzati alcuni metodi di stima dei parametri della variabile casuale di Zenga con particolare attenzione al metodo dei momenti e alla minimizzazione degli indici di accostamento non vincolata e con l'introduzione di vincoli. Sono state fornite delle applicazioni dei diversi metodi di stima presentati e per alcuni di essi è stato effettuato il confronto dell'adattamento tra la distribuzione di Zenga e la distribuzione di Dagum. Infine sono stati forniti dei primi studi numerici circa la distribuzione degli stimatori attraverso il metodo del Bootstrap parametrico.

Il metodo dei momenti pur fornendo valori dell'indice di accostamento nella maggior parte dei casi particolarmente elevati, e quindi non adatti, rappresenta un punto di riferimento nella stima dei parametri del modello di Zenga, poiché di esso si dispone della soluzione analitica. Pertanto le stime ottenute con il metodo dei momenti possono essere determinate facilmente e possono essere impiegate come valori iniziali per altri metodi di stima che impiegano metodologie numeriche. I possibili sviluppi circa questo metodo di stima possono riguardare l'impiego di momenti di ordine inferiore (nel lavoro si è determinato il momento dall'origine di ordine $1/2$) in modo tale da attenuare la restrizione che condiziona le possibili stime del parametro α e per attribuire più rilevanza alle osservazioni che assumono valori prossimi allo zero.

Per la stima dei parametri del modello di Zenga attraverso metodi che prevedono la minimizzazione di indici di accostamento si osserva che l'introduzione di un vincolo relativo alla locazione in generale comporta una significativa riduzione dell'accostamento. L'introduzione di un ulteriore vincolo relativo a un indice puntuale di ineguaglianza comporta un'ulteriore riduzione dell'accostamento ma di minore entità. I possibili sviluppi possono riguardare l'impiego di un indice sintetico di ineguaglianza quale secondo vincolo, in luogo degli indici puntuali che sono stati adottati. Va però precisato che per gli indici sintetici di ineguaglianza del modello di Zenga, allo stato attuale non si dispone di soluzione analitica, pertanto andrebbero determinati attraverso integrazione numerica, aumentando la complessità computazionale.

Bibliografia

- Colombi R. (1990). A new model of income distribution: The pareto-lognormal distribution. *Income and wealth distribution, inequality and poverty: Studies in Contemporary Economics series*, pp. 18–32.
- D’Addario R. (1934). Sulla misura della concentrazione dei redditi. *Roma: Poligrafo dello Stato*.
- D’Addario R. (1939). La curva dei redditi: sulla determinazione numerica dei parametri della seconda equazione paretiana. *Annali dell’Istituto di Statistica dell’Università di Bari*, **XII**.
- Dagum C. (1977). A new model of personal income distribution: specification and estimation. *Economie Appliquee*, **30**, 413–437. (Ristampato in: Duangkamon Chotikapanich Editor, *Modeling Income Distributions and Lorenz Curves*, Springer, 2008, pp. 3–25.).
- Davison A.; Hinkley D. (1997). *Bootstrap methods and their application*. Cambridge Univ Pr.
- Devroye L. (1986). *Non-uniform random variate generation*. Springer-Verlag.
- Efron B. (1979). Bootstrap methods: another look at the Jackknife. *The annals of Statistics*, **7**(1), 1–26.
- Efron B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans*. Society for Industrial and Applied Mathematics [SIAM], Philadelphia.
- Efron B. (1985). Bootstrap confidence intervals for a class of parametric problems. *Biometrika*, **72**(1), 45.
- Eurostat (2003). ECHP UDB - version of December 2003.
- Johnson N.; Kotz S.; Balakrishnan N. (1994). Continuous univariate distributions, vol. 1-2.
- Kleiber C. (2008). A Guide to the Dagum Distributions. *Modeling Income Distributions and Lorenz Curves*, pp. 97–117.
- Landenna G.; Marasini D.; Ferrari P. (1997). *Probabilità e variabili casuali*. Il Mulino, Bologna.
- Mood A.; Graybill F.; Boes D. (1988). *Introduzione alla statistica*. McGraw-Hill Libri Italia.
- Nelder J. A.; Mead R. (1965). A simplex method for function minimization. *The Computer Journal*, **7**(4), 308–313.
- Piccolo D. (1998). *Statistica*. Il Mulino, Bologna.

- Pietra G. (1915). Delle relazioni fra indici di variabilità, note I e II. *Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti*, **74**, 775–804.
- Polisicchio M. (2008). The continuous random variable with uniform point inequality measure $I(p)$. *Statistica & Applicazioni*, **VI**(2), 137–151.
- Quenouille M. (1956). Notes on bias in estimation. *Biometrika*, **43**(3/4), 353–360.
- R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
- Sarabia J. (2008). Parametric Lorenz Curves: Models and Applications. *Modeling Income Distributions and Lorenz Curves*, pp. 167–190.
- Singh S.; Maddala G. (1976). A function for size distribution of incomes. *Econometrica*, **44**, 963–970. (Ristampato in: Duangkamon Chotikapanich Editor, *Modeling Income Distributions and Lorenz Curves*, Sptinger, 2008, pp. 3–25.).
- Zenga M. (1987). Il contributo degli italiani allo studio della concentrazione (Prima parte: dal 1895 al 1915). *La distribuzione personale del reddito: problemi di formazione, di ripartizione e di misurazione*, pp. 307–328.
- Zenga M. (2007). Inequality curve and inequality index based on the ratios between lower and upper arithmetic means. *Statistica & Applicazioni*, **V**(1), 3–27.
- Zenga M. (2010). Mixture of Polisicchio's truncated Pareto distributions with beta weights. *Statistica & Applicazioni*, **VIII**(1), 3–25.
- Zenga M. M.; Pasquazzi L.; Zenga, Ma. (2010a). First applications of a new three parameter distribution for non-negative variables. Relazione Tecnica 188, Dipartimento di Metodi Quantitativi per le Scienze Economiche Aziendali, Università degli Studi di Milano - Bicocca.
- Zenga M. M.; Porro F.; Arcagni A. (2010b). Method of moments for Zenga's distribution. Relazione Tecnica 193, Dipartimento di Metodi Quantitativi per le Scienze Economiche Aziendali, Università degli Studi di Milano - Bicocca.
- Zenga M. M.; Polisicchio M.; Zenga, Ma.; Pasquazzi L. (2010c). More on M. M. Zenga's new three-parameters distribution for non-negative variables. Relazione Tecnica 187, Dipartimento di Metodi Quantitativi per le Scienze Economiche Aziendali, Università degli Studi di Milano - Bicocca.