

Proprietà campionarie ed asintotiche di alcune misure di concentrazione basate sul modello Gamma standardizzato (*)

Agostino Tarsitano
Università degli studi della Calabria
Dipartimento di Economia e Statistica
87030 Arcavacata di Rende (Cs)
agotar@unical.it

Riassunto

Il lavoro inizia con una panoramica sull'uso del modello Gamma nell'analisi della distribuzione dei redditi. Si prosegue con uno studio di vari indici di concentrazione: Gini, Giorgi, Lee, Pietra e Zenga. Si impostano poi le stime di massima verosimiglianza degli indici le cui manifestazioni campionarie ed asintotiche vengono studiate con delle simulazioni. Tali simulazioni saranno ripetute a diverse ampiezze campionarie e per diversi gradi di asimmetria della funzione di densità Gamma.

.

keywords: indici di concentrazione, simulazione di modelli, distribuzioni asintotiche

(*) *Statistica* anno XLVII, n.3, pp. 437-451

1. Introduzione

Questo lavoro riprende lo studio fatto da McDonald e Jensen (1979) sul comportamento di alcuni indici di concentrazione: Gini, Pietra, Theil quando la distribuzione del reddito è rappresentata dal modello Gamma. L'obiettivo principale è di comparare tali indici con altri recentemente proposti da Lee (1979), Giorgi (1981) e Zenga (1985).

Come prima anticipazione dei risultati ottenuti si può dire che gli stimatori di massima verosimiglianza sia per il parametro di forma del modello che degli indici di concentrazione, raggiungono rapidamente la loro soglia Cramer-Rao. Questo permette adeguate conclusioni inferenziali già per campioni di 250 redditi. La stima del parametro di forma presenta però una certa distorsione positiva che porta a sottostimare la concentrazione anche per campioni di ampiezza 2000. La distorsione si accentua quando la densità dei redditi è zeromodale. Per quanto attiene agli indici di concentrazione non sono emerse grandi differenze nelle manifestazioni campionarie (per qualsiasi ampiezza del campione) tra quelli esaminati. Ci sono analogie di comportamento tra l'indice del Pietra e quello di Gini; pure appaiati, ma leggermente discosti dai primi sono l'indice di Lee e quello di Giorgi. L'indice di Zenga, che ha la formula analitica più semplice, si colloca in una posizione intermedia fra i due gruppi.

Si è pure accertato che variazioni negli indici (con dati non raggruppati) relative alla terza o quarta cifra decimale sono piuttosto da attribuire a fluttuazioni campionarie che ad effettive modifiche nella concentrazione dei redditi.

Il presente lavoro è così articolato: nel prossimo paragrafo verrà fatta una breve panoramica sull'uso del modello Gamma nell'analisi della distribuzione dei redditi. In quello successivo si studierà il modello Gamma standardizzato e a partire da quest'ultimo, si otterranno le formule di vari indici di concentrazione: Gini, Pietra, Lee, Giorgi e Zenga. Nel quarto paragrafo si discuterà la stima di massima verosimiglianza del parametro di forma del modello Gamma ed infine, nel quinto paragrafo, si esamineranno i comportamenti campionari ed asintotici delle misure anzidette.

2. La funzione di densità Gamma come modello teorico per l'analisi della distribuzione dei redditi

La funzione di distribuzione Gamma nella versione più generale data da Amoroso (1924-25) ha formula

$$Am(x; a, b, c, d) = \frac{\left| d \left(\frac{x-c}{a} \right)^{db-1} e^{-\left(\frac{x-c}{a} \right)^d} \right|}{a \Gamma(b)} \quad (2.1)$$

dove $x > c$, $a, b > 0, d \neq 0$. Il parametro di locazione c indica il livello di reddito minimo preso in considerazione dal modello; il parametro a è di scala, mentre, ai parametri b e d , è legata la forma della curva di densità associata alla (2.1) (cfr. D'Addario, 1932).

L'applicazione di vari casi speciali del modello di Amoroso allo studio della distribuzione dei redditi ha lunga tradizione: Ammon (1895), March (1898), Vinci (1921). Lo stesso Pareto l'ha suggerita come seconda approssimazione.

Le motivazioni teoriche del modello di Amoroso sono svariate. Nell'approccio delle trasformazioni di variabili di Edgeworth-D'Addario (D'Addario, 1949) si perviene alla (2.1) quando la funzione generatrice $Am(x;1,b,0,1)$ è associata alla funzione trasformatrice

$$y = \left(\frac{x-c}{a} \right)^d \quad (2.2)$$

Secondo l'approccio dell'elasticità (Benini, 1905) la (2.1) si ottiene quando l'elasticità della funzione di ripartizione superiore $[1-F(y)]$ rispetto al reddito y è data da

$$\eta(y) = (1-db) + d \left(\frac{x-c}{a} \right)^d \quad (2.3)$$

La stessa (2.3) genera il modello di Amoroso negli schemi suggeriti da Singh e Maddala (1976). Mukerji (1967), sulla falsariga di Gibrat e Champenowne, ha presentato un processo stocastico che ha la distribuzione $Am(x;a,b,c,1)$ come soluzione di equilibrio.

Nell'approccio di Cantelli (1957) (Ripreso anche da Brush, 1961 e da Ord *et al.* 1981) il modello di Amoroso è quello più probabile, tra le infinite curve di densità possibili, quando della distribuzione del reddito siano prefissati il momento d-esimo centrato sul parametro di locazione e la media geometrica traslata

$$\int_{-\infty}^{+\infty} (x-c)^d Am(x;a,b,c,d) dx = costante \quad (2.4)$$

$$\int_{-\infty}^{+\infty} Ln(x-c) Am(x;a,b,c,d) dx = costante \quad (2.5)$$

In tempi più recenti Salem e Mount (1974) hanno trovato il modello $Am(x;a,b,0,1)$ più efficace, per l'interpolazione delle distribuzioni empiriche del reddito, del modello di Pareto e di quello lognormale. In altri lavori: Eltetö e Frigyes, (1968), Dagum, (1980), van Praag *et al.* (1983), la (2.1) è proposta solo come terza scelta, dopo i summenzionati modelli.

Kloek e van Dijk (1978) e (1979) hanno invece approfondito l'interpretazione dei parametri. Taguchi (1980) ha discusso la diversa parametrizzazione $Am(x;1,(b+1)/b,0,d)$ in cui il parametro d esprime il grado di asimmetria della curva di densità ed il parametro b il suo grado di appuntimento.

McDonald e Butler (1986) hanno notato che la densità di Amoroso è riconducibile a casi limite della densità beta generalizzata del secondo tipo.

$$GB2(x;d,a,c,r,s) = \frac{|d\left(\frac{x-c}{a}\right)^{dr-1}}{aB(r,s)\left[1+\left(\frac{x-c}{a}\right)^d\right]^{r+s}} \quad (2.6)$$

in particolare

$$Am(x;a,b,c,d) = \begin{cases} \lim_{s \rightarrow \infty} GB2\left(x;d,s^{\frac{1}{d}}a,c,b,s\right) & \text{per } d > 0 \\ \lim_{r \rightarrow \infty} GB2\left(x;d,r^{\frac{1}{d}}a,c,r,b\right) & \text{per } d < 0 \end{cases} \quad (2.7)$$

Gli stessi autori hanno poi dimostrato che la GB2 è esprimibile come il risultato della mistura

$$Am(x;\theta,b,c,d > 0) \cap_{\theta} Am(x;\theta,b,c,d < 0) \quad (2.8)$$

in cui la densità di Amoroso modella sia il reddito della popolazione che il parametro di scala del modello stesso.

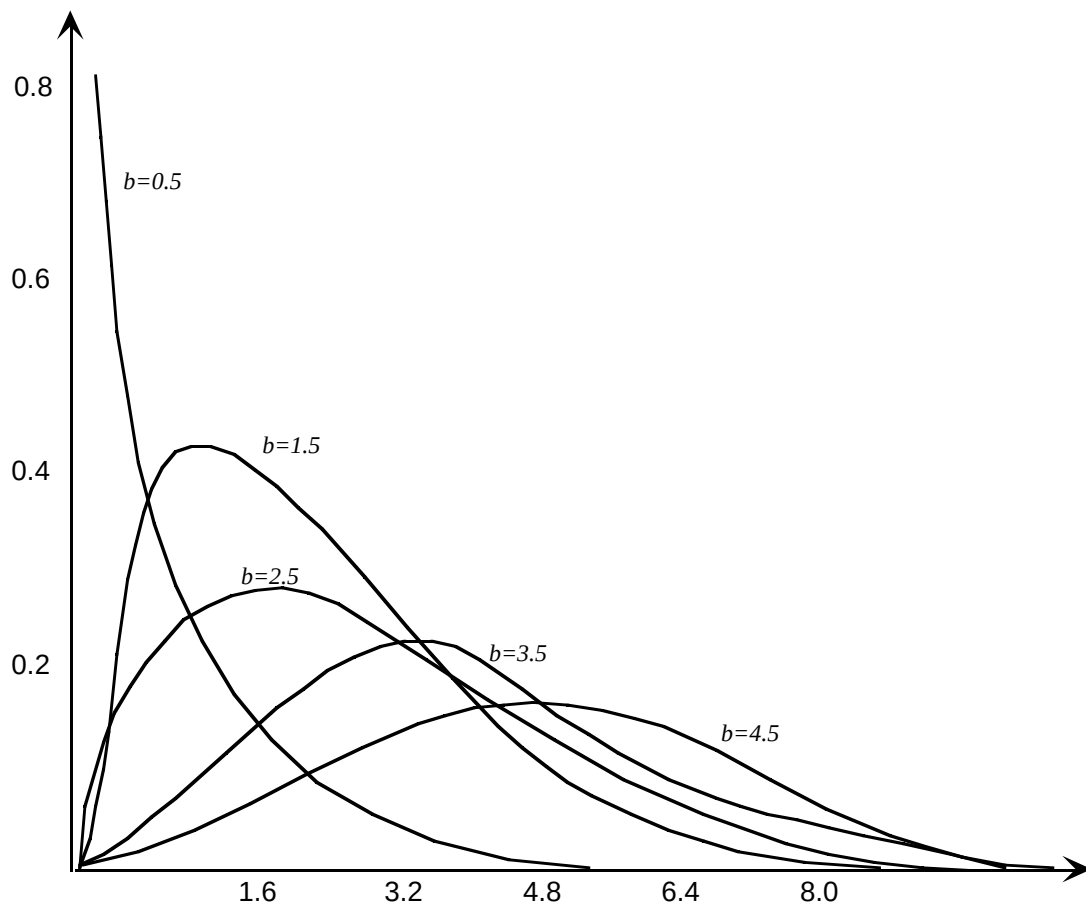
In una prospettiva analoga, Arnold e Press (1983) hanno suggerito la $Am(x;a,b,0,1)$ come distribuzione a priori naturale coniugata per il parametro a della distribuzione di Pareto.

3. Alcuni indici di concentrazione basati sul modello Gamma standardizzato

Il modello di Amoroso è eccessivamente complesso, almeno per i semplici scopi comparativi che ci si prefigge in questo lavoro. Si preferisce pertanto studiare gli indici di concentrazione nell'ambito della sua versione standardizzata. La trasformazione del reddito (2.2) dà la semplificazione voluta

$$Am(y;1,b,0,1) = GS(y,b) = \frac{y^{b-1}e^{-y}}{\Gamma(b)} \quad (3.1)$$

per $y > 0$. Tale modello ha media e varianza b , indice di asimmetria $\gamma_1(b) = 2/\sqrt{b}$ ed indice di curtosi $\gamma_2(b) = 3 + 6/b$. La curva di densità è asimmetrica positiva e leptocurtica. Entrambe queste caratteristiche si attenuano all'aumentare di b (si veda la figura 1). La curva è inoltre unimodale per $b > 1$ (con moda in $y = b - 1$) e zeromodale per $b \leq 1$.



Atkinson e Brown (1954) hanno notato che il momento primo incompleto

$$F_1(y) = \frac{\int_{-\infty}^y t dF(t)}{\int_{-\infty}^{+\infty} t dF(t)} \quad (3.2)$$

della distribuzione di Pareto e della lognormale è una funzione dello stesso tipo della $F(y)$. Anche la distribuzione di Amoroso, come è facile provare, gode di questa forma di chiusura. In particolare, per la versione standardizzata (3.1), se si pone

$$F(y) = \int_0^y \frac{t^{b-1} e^{-t}}{\Gamma(b)} dt = IG(y, b) \quad (3.3)$$

si ottiene

$$F_1(y) = \int_0^y \frac{t^b e^{-t}}{b\Gamma(b)} dt = IG(y, b+1) \quad (3.4)$$

Eliminando la y tra la (3.4) e la (3.3) si definisce l'insieme, indicizzato dal parametro b , delle curve di Lorenz associate al modello (2.1). Tali curve, come quelle associate alle versioni usuali del modello di Pareto e di quello lognormale, non si intersecano (Taillie, 1981). Una limitazione di cui, nel presente lavoro, si può non tenere conto poiché l'interesse è rivolto al solo confronto delle misure di concentrazione.

Gli indici presi in considerazione sono quello di Gini, Pietra, Lee (1979), Giorgi (1981), Zenga (1985), le cui espressioni analitiche sono, rispettivamente

$$R = 1 - \frac{\int_0^{\infty} [1 - F(t)]^2 dt}{\mu} \quad (3.5)$$

$$P = \frac{\int_0^{\infty} |t - \mu| dF(t)}{2\mu} \quad (3.6)$$

$$L = \frac{P + F(\mu)}{2} \quad (3.7)$$

$$G = P F(\mu) \quad (3.8)$$

$$Z = 1 - e^{-\frac{\int_0^{\infty} (t - \mu) \ln(t) dF(t)}{\mu}} \quad (3.9)$$

in cui $\mu > 0$ indica il reddito medio della popolazione. Gli indici assumono tutti valori nell'intervallo unitario (non così l'indice di Theil studiato da McDonald e Jensen (1979) e che perciò non è stato incluso in questo studio).

Come è noto l'indice (3.5) è pari al valore atteso della differenza media tra i redditi rapportata al suo massimo. L'indice (3.6) è dato dalla differenza massima tra la $F(y)$ e la $F_1(y)$; gli indici (3.7) e (3.8) sono stati proposti per superare, includendo la frazione di redditieri con reddito inferiore alla media, certe ambiguità riscontrate in R e P . Infine, l'indice (3.9) è una trasformata del tipo $Z = 1 - e^{-A}$ dove A è l'area compresa tra le curve $F(y)$ ed $F_1(y)$ misurate in scala logaritmica. Nel modello Gamma standardizzato gli indici valgono

$$R(b) = \frac{\Gamma(2b+1)}{[2^b \Gamma(b+1)]^2} \quad (3.10)$$

$$P(b) = \frac{e^{b[Ln(b)-1]}}{\Gamma(b+1)} \quad (3.11)$$

$$L(b) = \frac{P(b) + IG(b,b)}{2} \quad (3.12)$$

$$G(b) = P(b)IG(b,b) \quad (3.13)$$

$$Z(b) = 1 - e^{-\frac{1}{b}} \quad (3.14)$$

La (3.10) è stata data da Amoroso ed è del tutto analoga a quella presentata da McDonald e Jensen (1979). Questi ultimi però, per l'indice del Pietra, danno una formula poco trattabile. La (3.11) si ricava invece facilmente con la formula ricorsiva

$$IG(y,b+1) = IG(y,b) - y^b e^{-y} / \Gamma(b+1) \quad (3.15)$$

L'espressione dell'indice di Zenga si ottiene ricordando che l'area compresa tra $F(y)$ ed $F_1(y)$, misurate su scala logaritmica, nel modello Gamma standardizzato, è pari ad $(1/b)$ (Salem e Mount, 1973). Da notare che le espressioni (3.10)-(3.14) rimarrebbero le stesse se si operasse con il modello $GS(y/a;b)/a$ ovvero se vi si introducesse un parametro di scala

La figura 2 mette in evidenza l'andamento iperbolico delle relazioni (3.10)-(3.14) a significare che maggiore è il parametro b minore è la concentrazione dei redditi.

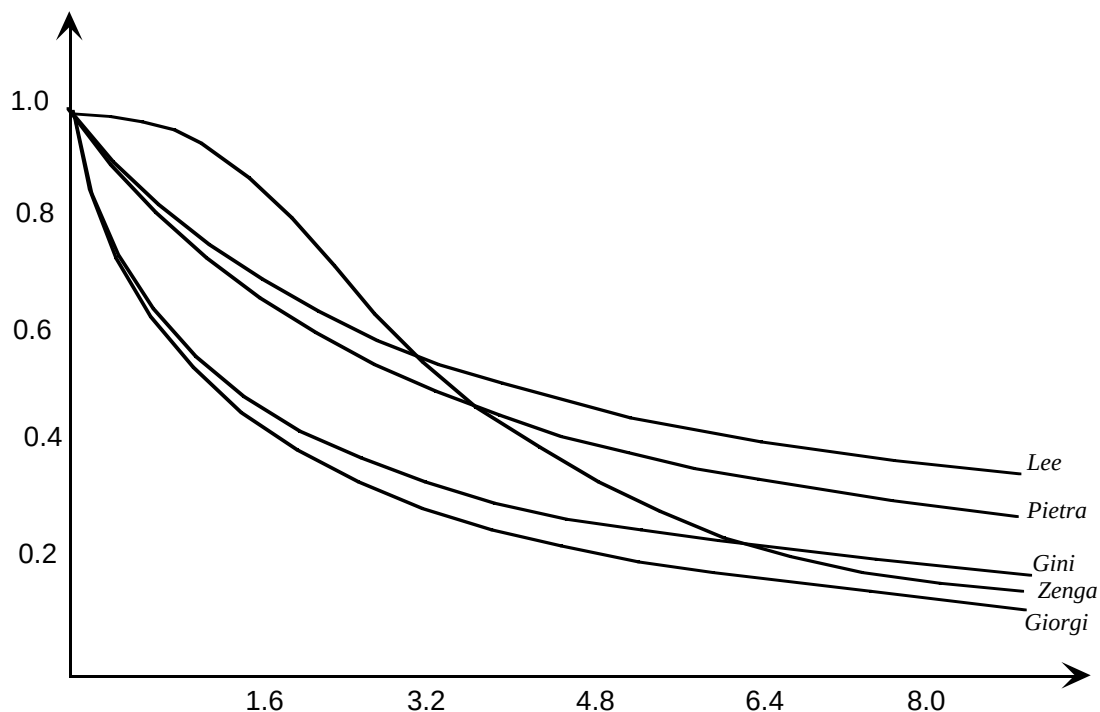


Figura 2: indici di concentrazione in funzione del parametro di forma

4. Funzione di verosimiglianza, distribuzione ed intervalli asintotici per il parametro di forma.

Il ruolo di primo piano che ha il parametro di forma nelle formule degli indici di concentrazione impone un approfondimento sulla sua stima. Non che la letteratura manchi di interventi in questo senso: Stacy e Mirham (1965), Choi e Wette (1969) Shenton e Bowman, (1969) e (1972), Cohen (1975), Glaser (1976). Si tratta tuttavia di studi che non fanno riferimento alle ampiezze campionarie che ricorrono nell'analisi delle distribuzioni del reddito. In questo paragrafo si cercherà di colmare tale lacuna

Il logaritmo della funzione di verosimiglianza di un campione di redditi $S=(y_1, y_2, \dots, y_n)$ dalla (2.1) è

$$Ln(S; b) = -nLn[\Gamma(b)] + b \sum_{i=1}^n Ln(y_i) - \sum_{i=1}^n [y_i + Ln(y_i)] \quad (4.1)$$

le cui prime due derivate rispetto al parametro b sono

$$\frac{\partial Ln[L(S; b)]}{\partial b} = \sum_{i=1}^n Ln(y_i) - n\psi(b) \quad (4.2)$$

$$\frac{\partial^2 Ln[L(S; b)]}{\partial b^2} = -n\psi'(b) \quad (4.3)$$

dove $\psi(b)$ e $\psi'(b)$ rappresentano le funzioni Digamma e Trigamma. Poiché $y'(b)$ è sempre positiva, la soluzione $b = \beta_n$ di

$$\psi(b) = Ln(M) \quad (4.4)$$

con M media geometrica campionaria, è lo stimatore di massima verosimiglianza del parametro b. La (4.4) suggerisce una interessante analogia con il metodo dei momenti. La $\psi(b)$ è il logaritmo della media geometrica della $GS(y; b)$ e quindi la (4.4) non fa che eguagliare la media geometrica del campione a quella della popolazione. Ne consegue che per la densità Gamma standardizzata, la stima di massima verosimiglianza del parametro di forma coincide con quella ottenibile con il metodo dei momenti.

Il calcolo di β_n può essere realizzato con lo schema iterativo Newton-Raphson

$$\beta_n^{k+1} = \beta_n^k - \frac{[\psi(\beta_n^k) - Ln(M)]}{\psi'(\beta_n^k)} \quad (4.5)$$

schema che è sempre convergente se $\beta_n^k > 0$ (cfr. Choi e Wette, 1969). Un valore iniziale piuttosto buono (migliore ad esempio di quello suggerito da Greenwood e Durand, 1960) è il seguente

$$\beta_n^0 = \frac{[M + \sqrt{M(M+2)}]}{2} \quad (4.6)$$

La (4.6) deriva dall'approssimazione $\psi(x) = \text{Log}(x) - [1/(2x)] + O(x^2)$ e dal troncamento della risultante esponenziale dopo il termine lineare. Per il calcolo delle funzioni digamma e Trigamma si possono adoperare gli algoritmi riportati in Griffiths e Hill (1985).

Poiché la (2.1) appartiene alla classe di densità esponenziali

$$f(y; b) = g(b)h(y)e^{bt(y)} \quad (4.7)$$

dove $g(b) = \Gamma(b)^{-1}$, $h(y) = \exp\{-[y + \text{Ln}(y)]\}$ e $t(y) = \text{Ln}(y)$, la distribuzione asintotica di β_n è la normale $N[b, s^2(b)]$ con $s^2(b) = 1/[n\psi'(b)]$ (Cfr. ad esempio Piegorsch, 1983).

Nella tabella che segue sono riassunti i risultati di un piano di simulazioni riguardante il parametro b . Il piano di simulazione è consistito in $L=2000$ campioni di varia ampiezza ($n=250, 500, 1000, 2000$) estratti dalla Gamma standardizzata. Per la generazione dalla Gamma si è adoperato l'algoritmo proposto in Atkinson e Pierce (1976) e per la uniforme l'algoritmo di Wichmann ed Hill, pubblicato in Griffiths ed Hill (1985). I valori teorici del parametro di forma sono ($b=0.5, 1.5, 2.5, 3.5, 4.5$).

Tabella 1: Risultati delle simulazioni per il parametro b .

n	b	$m(\beta_n)$	$s(b)$	$s(\beta_n)$	g1	g2	K	α	5%	1%
250	0.5	0.5012	0.0285	0.0290	0.2466	0.1250	0.8748	0.4285	1895	1977
	1.5	1.5023	0.0655	0.0649	0.1406	0.0945	0.7627	0.6059	1918	1981
	2.5	2.5009	0.0903	0.0898	0.0614	-0.2314	0.9091	0.3802	1912	1983
	3.5	3.5036	0.1100	0.1107	0.0953	-0.1875	0.6834	0.7387	1900	1989
	4.5	4.5009	0.1268	0.1245	0.0838	0.0743	1.0041	0.2656	1903	1982
500	0.5	0.5005	0.0201	0.0202	0.2739	0.0661	1.0329	0.2364	1903	1978
	1.5	1.5011	0.0462	0.0456	0.0212	-0.1513	0.8592	0.4515	1919	1982
	2.5	2.5000	0.0639	0.0623	0.1158	0.1097	1.5413	0.0173	1913	1981
	3.5	3.4991	0.0778	0.0787	0.0903	-0.8188	0.9980	0.2722	1897	1982
	4.5	4.4985	0.0897	0.0884	-0.0418	-0.0624	1.0427	0.2270	1907	1981
1000	0.5	0.5004	0.0142	0.0142	0.1186	-0.0505	0.7221	0.6742	1910	1983
	1.5	1.5023	0.0327	0.0329	0.0989	-0.0803	0.6845	0.7368	1904	1978
	2.5	2.5011	0.0452	0.0453	0.0188	-0.1174	1.3935	0.0411	1894	1987
	3.5	3.5011	0.0550	0.0543	0.0030	0.0254	1.1279	0.1570	1893	1980
	4.5	4.5019	0.0634	0.0631	0.0584	-0.1824	0.9330	0.3488	1910	1985
2000	0.5	0.5003	0.0101	0.0098	0.1667	0.1811	0.8608	0.4491	1916	1983
	1.5	1.5007	0.0231	0.0235	0.0001	-0.0665	0.4954	0.9669	1888	1982
	2.5	2.5022	0.0319	0.0318	-0.0462	-0.0011	1.2285	0.0978	1908	1982
	3.5	3.5019	0.0389	0.0398	0.0815	-0.0605	0.9541	0.3241	1895	1977
	4.5	4.4944	0.0448	0.0452	0.0432	-0.1146	1.6987	0.0062	1891	1987

Le entrate $m(\beta_n)$ ed $s(\beta_n)$ della tabella 1 corrispondono alla media aritmetica ed allo scarto quadratico medio dei valori di $\{\beta_n^i, i=1,2,\dots,L\}$ ottenuti su L repliche. Lo stesso vale

$$g_1 = \sum_{i=1}^n \left[\frac{\beta_n^i - b}{s(\beta_n)} \right]^3; \quad g_2 = \sum_{i=1}^n \left[\frac{\beta_n^i - b}{s(\beta_n)} \right]^4 - 3 \quad (4.8)$$

Le colonne K ed α afferiscono alla prova di ipotesi

$$\begin{cases} H_0 : F_n(\beta) = \phi(\beta) \\ H_1 : F_n(\beta) \neq \phi(\beta) \end{cases} \quad (4.9)$$

dove $F_n(\beta)$ è la distribuzione empirica (basata sulle L repliche) dello stimatore β_n e $\phi(\beta)$ la cumulata della gaussiana standardizzata. Più specificamente, nella colonna K si trovano i valori della statistica di Kolmogorov-Smirnov $K = \text{MAX}(K^+, K^-)$ con

$$K^+ = \sqrt{L} \max_{1 \leq i \leq L} \left\{ \frac{i}{L} - \Phi(B_i) \right\} \quad (4.10)$$

$$K^- = \sqrt{L} \max_{1 \leq i \leq L} \left\{ \Phi(B_i) - \frac{i-1}{L} \right\} \quad (4.11)$$

e è pari $B_i = \frac{\beta_n^i - b}{s(b)}$ che si trova in i -esima posizione nella graduatoria crescente di tali quantità. La colonna α riporta l'ampiezza del test (Cfr. Smirnov, 1948) associata al corrispondente valore di K

$$\alpha(K) = 1 - \frac{0.125\sqrt{2\pi}}{K} \sum_{i=1}^{\infty} e^{-\left[\frac{(2i-1)\pi}{K} \right]^2} \quad (4.12)$$

ad esempio, nella prima riga, si legge $K=0.8748$ e $\alpha=0.4285$; questo significa che qualora si rifiutasse H_0 , cioè si negasse che la distribuzione di b_n sia gaussiana, si sbaglierebbe nel 42.85% dei casi. L'analisi puntuale dello stimatore b_n non sempre è sufficiente a chiarirne il comportamento. In questo senso la costruzione di intervalli di confidenza risulta spesso più utile. La tabella 1 tiene conto di due elementi importanti degli intervalli di confidenza: Il livello effettivo di significatività degli intervalli asintotici e la lunghezza degli intervalli stessi. Su quest'ultima danno informazioni le colonne intestate $s(b)$ ed $s(b_n)$, mentre sul livello di significatività chiariscono le ultime due colonne; in esse è stato infatti annotato il numero di volte in cui il valore teorico di b è stato incluso nell'intervallo

$$\left[\beta - Z_{\frac{\alpha}{2}} s(\beta_n), \beta + Z_{\frac{\alpha}{2}} s(\beta_n) \right] \quad (4.13)$$

con $Z_{\alpha/2}$ pari a 1.960 e 2.576 corrispondenti ai livelli di significatività del 5% (attese 1900 inclusioni) e dell'1% (1980).

I risultati confermano la distorsione positiva di b già osservata da molti autori; distorsione che si attenua, ma non scompare anche per campioni moderatamente grandi. La varianza campionaria di b_n è invece prossima alla soglia Cramer-Rao già a partire da un campione di ampiezza 250, segno sicuro dell'efficienza dello stimatore.

Gli intervalli di confidenza aderiscono piuttosto fedelmente al livello atteso di significatività già per campioni di 250 unità.

La tendenza alla normale della distribuzione empirica di b_n non appare accentuatissima se saggiata attraverso gli indici g_1 e g_2 . C'è infatti una certa asimmetria positiva (più marcata quando $b < 1$) che però decresce sia all'aumentare di b che all'aumentare di n .

Il grado di appuntimento della curva mostra un certo orientamento alla mesocurtosi anche se non sembra una tendenza dominante, almeno per le ampiezze campionarie considerate. Le realizzazioni empiriche dei due indici sulla forma vanno però valutate tenuto conto della difficoltà numeriche insite nel loro calcolo, difficoltà che la doppia precisione aritmetica può solo ridurre ma non eliminare del tutto.

La Statistica K non segnala forti discrepanze tra la distribuzione simulata e quella normale. Al livello di significatività del 5% l'ipotesi di normalità viene rifiutata in due casi ($b=2.5, n=500$), ($b=2.5, n=1000$); all'1% solo per ($b=4.5, n=2000$).

5. Stimatori di massima verosimiglianza, distribuzione ed intervalli asintotici per gli indici di concentrazione

McDonald e Jensen (1979) hanno realizzato il primo studio sul comportamento degli indici di concentrazione per campioni provenienti dalla densità Gamma. Lo studio presenta però due limitazioni: all'indice di Pietra e Gini viene affiancato l'indice di Theil che ha un diverso campo di variazione. Mancano inoltre considerazioni sull'effetto che ha, sugli indici, la distorsione (presente anche in campioni di ampiezza cospicua) dello stimatore del parametro di forma. Il presente paragrafo è incentrato proprio su questo tema.

Le relazioni (3.10)-(3.14) che legano gli indici di concentrazione al parametro di forma b sono tutte differenziabili rispetto allo stesso b . Questo è sufficiente per richiamare il principio di invarianza degli stimatori di massima verosimiglianza (cfr. ad esempio Mood, Graybill e Boes, 1974, pp. 284-285). Gli stimatori degli indici saranno quindi

$$R_n = \frac{\Gamma(2\beta_n + 1)}{\left[2^{\beta_n} \Gamma(\beta_n + 1) \right]^2} \quad (5.1)$$

$$P_n = \frac{e^{\beta_n [Ln(\beta_n) - 1]}}{\Gamma(\beta_n + 1)} \quad (5.2)$$

$$L_n = \frac{P(\beta_n) + IG(\beta_n, \beta_n)}{2} \quad (5.3)$$

$$G_n = P(\beta_n) IG(\beta_n, \beta_n) \quad (5.4)$$

$$Z_n = 1 - e^{-\frac{1}{\beta_n}} \quad (5.5)$$

le cui distribuzioni asintotiche sono, rispettivamente, le normali

$$N\left(R, \frac{R^2 4[\psi(2b+1) - \psi(b+1) - Ln(2)]^2}{s(b)}\right) \quad (5.6)$$

$$N\left(P, \frac{P^2 [Ln(b) - \psi(b+1)]^2}{s(b)}\right) \quad (5.7)$$

$$N\left(L, \frac{L'^2}{s(b)}\right) \quad (5.8)$$

$$N\left(G, \frac{G'^2}{s(b)}\right) \quad (5.9)$$

$$N\left(Z, \frac{\left(\frac{1-Z}{b^2}\right)^2}{s(b)}\right) \quad (5.10)$$

dove, nelle (5.6)-(5.10), $s(b)=n\psi'(b)$. L'analisi degli stimatori degli indici di concentrazione, le cui *performance* sono riportate nelle tavole 2-6, può essere effettuata secondo lo schema già tracciato nel paragrafo precedente. Quando $b=0.5$ la distorsione nei valori medi è negativa per tutti gli indici e per tutte le ampiezze campionarie. Per i valori più grandi di b l'indice di Zenga è quello che si discosta (NB nella quarta cifra decimale) più spesso degli altri dal suo valore atteso; gli indici di Lee e di Giorgi sono quelli più aderenti ai valori teorici. La temuta sottostima della concentrazione si verifica quindi in modo evidente solo quando la densità è zeromodale. La distorsione si attenua comunque all'aumentare dell'ampiezza campionaria. Una conferma questa dei risultati di McDonald e Ransom (1981).

Le varianze campionarie coincidono con le soglie Cramer-Rao anche per ampiezze campionarie esigue; quindi, negli intervalli di confidenza, le varianze campionarie potranno sostituire quelle teoriche. Gli indici di asimmetria e di curtosi non sono mai troppo lontani (tranne forse per $b=0.5$) dall'ideale valore nullo atteso nell'ipotesi di normalità.

Anche la statistica K non segnala particolari anomalie: per tutti gli indici l'ipotesi di normalità viene respinta tre volte al 5% ed una sola volta all'1%. Gli stessi intervalli di confidenza asintotici corrispondono piuttosto fedelmente, anche per i campioni più piccoli, al livello fiduciario atteso. Inoltre, data l'ampiezza ridotta degli intervalli, variazioni nell'ordine della terza cifra decimale sono più da attribuire a fluttuazioni campionarie degli indici che a vere e proprie modifiche nella concentrazione

In conclusione si può dire che se il modello gamma descrivesse bene la densità dei redditi sarebbe sufficiente esaminare le entrate di un campione di 250 redditi (con dati non raggruppati) per trarre adeguate conclusioni inferenziali a riguardo delle misure di concentrazione (l'ampiezza dovrebbe però essere molto più grande (≥ 500) qualora ci fossero ragioni per ritenere che la densità sia zeromodale).

Tabella 2: Risultati delle limitazioni per l'indice di Gini.

n	R	m(R)	$\sigma(R)$	s(R)	g1	g2	K	α	5%	1%
250	0.6366	0.6464	0.0111	0.0113	0.0904	0.0283	0.6523	0.7884	1901	1975
	0.4244	0.4243	0.0078	0.0077	0.0241	0.0164	0.6861	0.7309	1909	1982
	0.3395	0.3396	0.0055	0.0055	0.0663	-0.2310	0.8918	0.4041	1911	1984
	0.2910	0.2910	0.0042	0.0043	0.0235	-0.1919	0.6749	0.7525	1988	1997
	0.2587	0.2587	0.0034	0.0034	0.0379	0.0734	0.8416	0.4781	1903	1981
500	0.6366	0.6365	0.0079	0.0078	-0.1682	-0.0032	1.0329	0.2363	1900	1976
	0.4244	0.4244	0.0055	0.0054	0.0834	-0.1473	0.8580	0.4532	1925	1982
	0.3395	0.3396	0.0039	0.0038	-0.0103	0.1057	1.4409	0.0314	1917	1979
	0.2910	0.2911	0.0030	0.0030	-0.0010	-0.0933	0.9666	0.3075	1896	1981
	0.2587	0.2587	0.0024	0.0024	0.1225	-0.0541	1.1625	0.1339	1916	1981
1000	0.6366	0.6365	0.0056	0.0055	-0.0462	-0.0834	0.6344	0.8157	1906	1983
	0.4244	0.4242	0.0039	0.0039	-0.0209	-0.0958	0.8304	0.4955	1906	1978
	0.3395	0.3394	0.0028	0.0028	0.0500	-0.1174	1.3717	0.0464	1892	1983
	0.2910	0.2910	0.0021	0.0021	0.0626	0.0346	1.1251	0.1589	1902	1980
	0.2587	0.2587	0.0017	0.0017	-0.0046	-0.1921	0.8820	0.4180	1909	1985
2000	0.6366	0.6365	0.0039	0.0038	0.1111	0.1489	0.8486	0.4674	1915	1981
	0.4244	0.4243	0.0027	0.0028	0.0566	-0.0564	0.5086	0.9588	1887	1981
	0.3395	0.3394	0.0019	0.0020	0.0976	0.0164	1.2025	0.1109	1895	1978
	0.2910	0.2909	0.0015	0.0015	-0.0358	-0.0759	0.8900	0.4066	1895	1978
	0.2587	0.2587	0.0012	0.0012	0.0432	-0.1147	1.6739	0.0073	1902	1986

Tabella 3: Risultati delle simulazioni per l'indice di Pietra.

n	P	m(P)	$\sigma(P)$	s(P)	g1	g2	K	α	5%	1%
250	0.4839	0.4838	0.0100	0.0102	-0.0539	0.0228	0.6043	0.8585	1899	1975
	0.3084	0.3083	0.0060	0.0059	0.0355	0.0157	0.6843	0.7371	1909	1983
	0.2441	0.2441	0.0041	0.0041	0.0713	-0.2298	0.8911	0.4051	1911	1983
	0.2082	0.2082	0.0031	0.0031	0.0267	-0.1912	0.6747	0.7528	1897	1987
	0.1846	0.1846	0.0025	0.0025	0.0404	0.0740	1.6737	0.0073	1903	1981
500	0.4839	0.4839	0.0071	0.0071	-0.1435	-0.0129	1.0329	0.2363	1901	1973
	0.3084	0.3083	0.0042	0.0042	0.0909	-0.1453	0.8579	0.4533	1925	1983
	0.2441	0.2441	0.0029	0.0028	-0.0606	0.1064	1.4402	0.0315	1917	1979
	0.2082	0.2082	0.0022	0.0022	0.0014	-0.0931	0.9658	0.3084	1896	1981
	0.1846	0.1847	0.0018	0.0017	0.1242	-0.0537	1.1650	0.1324	1916	1982
1000	0.4839	0.4839	0.0050	0.0050	-0.0293	-0.0876	0.6331	0.8176	1904	1983
	0.3084	0.3082	0.0030	0.0030	-0.0155	-0.0960	0.8406	0.4797	1907	1978
	0.2441	0.2441	0.0021	0.0021	0.0527	-0.1170	1.3709	0.0466	1892	1983
	0.2082	0.2082	0.0016	0.0015	0.0643	0.0351	1.1250	0.1590	1902	1981
	0.1846	0.1846	0.0013	0.0012	-0.0035	-0.1921	0.8809	0.4196	1909	1985
2000	0.4839	0.4839	0.0035	0.0034	-0.0980	-0.1434	0.8480	0.4683	1914	1982
	0.3084	0.3083	0.0021	0.0022	0.0606	-0.0155	0.5085	0.9582	1888	1981
	0.2441	0.2440	0.0015	0.0014	0.0996	0.0173	1.2015	0.1114	1907	1982
	0.2082	0.2082	0.0011	0.0011	-0.0347	-0.0762	0.8885	0.4088	1896	1977
	0.1846	0.1846	0.0009	0.0009	-0.0024	-0.1286	1.6737	0.0073	1904	1986

Tabella 4: Risultati delle simulazioni per l'indice di Lee.

n	L	m(M)	$\sigma(L)$	s(L)	g1	g2	K	α	5%	1%
250	0.5833	0.5832	0.0074	0.0074	-0.0509	-0.0467	0.4099	0.9961	1899	1981
	0.4584	0.4582	0.0042	0.0040	0.0176	-0.0896	1.3320	0.0575	1919	1987
	0.4141	0.4141	0.0028	0.0027	0.0614	0.0114	0.8961	0.3981	1911	1988
	0.3897	0.3896	0.0021	0.0021	0.0512	-0.2232	0.9937	0.2768	1917	1991
	0.3738	0.3738	0.0017	0.0017	0.0602	-0.2432	0.9177	0.3687	1895	1985
500	0.5833	0.5833	0.0052	0.0051	-0.0891	-0.0107	0.9483	0.3296	1914	1977
	0.4584	0.4584	0.0030	0.0029	0.0349	-0.0235	0.4762	0.9772	1898	1979
	0.4141	0.4141	0.0020	0.0020	0.1470	-0.0445	0.8894	0.4076	1891	1985
	0.3897	0.3897	0.0015	0.0015	-0.0146	0.0441	0.8327	0.495	1887	1977
	0.3737	0.3737	0.0012	0.0012	-0.0191	-0.0139	1.3270	0.0591	1886	1976
1000	0.5833	0.5833	0.0037	0.0037	-0.0008	-0.1484	0.5657	0.9062	1905	1982
	0.4584	0.4582	0.0021	0.0020	0.0572	-0.0758	1.8775	0.0017	1912	1985
	0.4141	0.4141	0.0014	0.0014	0.0576	0.1173	0.6782	0.7472	1892	1978
	0.3897	0.3897	0.0011	0.0011	0.0323	-0.0177	0.8893	0.4076	1886	1975
	0.3737	0.3737	0.0009	0.0008	-0.0046	-0.0107	0.8539	0.4594	1911	1987
2000	0.5833	0.5832	0.0026	0.0026	-0.1266	0.2023	0.9374	0.3433	1909	1980
	0.4584	0.4584	0.0015	0.0015	0.1129	0.0187	0.8139	0.5217	1890	1980
	0.4141	0.4141	0.0010	0.0010	-0.0096	-0.2254	1.5533	0.016	1905	1984
	0.3897	0.3897	0.0008	0.0008	-0.0013	0.0048	1.2152	0.1043	1913	1978
	0.3737	0.3736	0.0006	0.0006	-0.0182	0.0461	1.5901	0.0127	1896	1973

Tabella 5: Risultati delle simulazioni per l'indice di Giorgi.

n	G	m(G)	$\sigma(G)$	s(G)	g1	g2	K	α	5%	1%
250	0.3304	0.3303	0.0091	0.0091	-0.0230	-0.0464	0.3305	0.9996	1898	1980
	0.1876	0.1875	0.0044	0.0042	0.0333	-0.0856	1.3189	0.0617	1919	1987
	0.1426	0.1426	0.0028	0.0027	0.0731	0.0160	0.9188	0.3673	1911	1988
	0.1189	0.1189	0.0020	0.0020	0.0594	-0.2229	0.9874	0.2837	1916	1990
	0.1039	0.1039	0.0016	0.0016	0.0671	-0.2423	0.9084	0.3812	1895	1985
500	0.3304	0.3304	0.0064	0.0064	-0.0694	-0.0068	0.9483	0.3296	1914	1977
	0.1876	0.1876	0.0031	0.0031	0.0469	-0.0197	0.4797	0.9754	1896	1980
	0.1426	0.1426	0.0020	0.0020	0.1534	-0.0411	0.8876	0.4101	1891	1986
	0.1189	0.1189	0.0014	0.0014	-0.0146	-0.0077	0.8371	0.4892	1886	1975
	0.1039	0.1039	0.0011	0.0011	-0.0139	-0.1277	1.3336	0.0570	1888	1977
1000	0.3304	0.3304	0.0046	0.0046	-0.0008	-0.1484	0.5712	0.8996	1904	1983
	0.1876	0.1874	0.0022	0.0021	0.0572	-0.0758	1.8767	0.0017	1911	1986
	0.1426	0.1426	0.0020	0.0020	0.1554	-0.0411	0.8876	0.4101	1891	1986
	0.1189	0.1189	0.0010	0.0010	0.0372	-0.0160	0.8848	0.4140	1886	1974
	0.1039	0.1039	0.0008	0.0008	-0.0009	-0.0108	0.8504	0.4648	1910	1987
2000	0.3304	0.3303	0.0032	0.0032	-0.1158	0.1962	0.9369	0.3438	1910	1980
	0.1876	0.1876	0.0016	0.0016	0.1192	0.0239	0.8129	0.5232	1890	1980
	0.1426	0.1425	0.0010	0.0010	-0.0058	-0.2250	1.5528	0.0161	1905	1983
	0.1189	0.1189	0.0007	0.0007	0.0018	0.0052	1.2122	0.1058	1913	1978
	0.1039	0.1039	0.0006	0.0006	-0.0154	0.0464	1.5885	0.0129	1896	1973

Tabella 6: Risultati delle simulazioni per l'indice di Zenga.

n	Z	m(Z)	$\sigma(Z)$	s(Z)	g1	g2	K	α	5%	1%
250	0.8647	0.8640	0.0154	0.0156	-0.2434	-0.0918	0.8556	0.4568	1894	1976
	0.4866	0.4865	0.0149	0.0148	0.0354	0.0115	0.6845	0.7368	1907	1982
	0.3297	0.3298	0.0097	0.0096	0.0938	-0.2259	0.8885	0.4088	1913	1984
	0.2485	0.2485	0.0067	0.0068	0.0513	-0.1866	0.7239	0.6711	1895	1984
	0.1993	0.1993	0.0050	0.0049	0.0692	0.0810	0.8001	0.5439	1900	1978
500	0.8647	0.8644	0.0109	0.0109	-0.2724	0.0538	1.0329	0.2363	1901	1978
	0.4866	0.4865	0.0105	0.0104	0.0907	-0.1469	0.8579	0.4533	1924	1983
	0.3297	0.3298	0.0068	0.0067	0.0907	0.1086	1.4383	0.0319	1916	1979
	0.2485	0.2487	0.0048	0.0198	-0.0924	0.0132	0.9593	0.3162	1897	1980
	0.1993	0.1994	0.0035	0.0035	0.1432	-0.0483	1.1933	0.1159	1916	1982
1000	0.8647	0.8644	0.0077	0.0077	-0.1180	-0.0572	0.7207	0.6765	1907	1982
	0.4866	0.4862	0.0075	0.0075	-0.0156	-0.0968	0.8408	0.4794	1907	1978
	0.3297	0.3296	0.0048	0.0049	0.0629	-0.1159	1.3676	0.0474	1892	1983
	0.2485	0.2485	0.0034	0.0033	0.0779	0.0388	1.1244	0.1594	1902	1981
	0.1993	0.1992	0.0025	0.0025	0.0915	-0.1930	0.8690	0.4369	1909	1985
2000	0.8647	0.8645	0.0054	0.0054	-0.1664	0.1772	0.8634	0.4451	1913	1983
	0.4866	0.4865	0.0053	0.0054	0.0605	-0.0556	0.5084	0.9583	1888	1981
	0.3297	0.3295	0.0034	0.0034	0.1074	0.0206	1.1976	0.1135	1906	1982
	0.2485	0.2484	0.0024	0.0024	-0.0253	-0.0785	0.8774	0.4246	1896	1977
	0.1933	0.1933	0.0018	0.0018	0.0701	-0.1311	1.6676	0.0076	1904	1986

Bibliografia

- Aitchison J., Brown J.A.C. (1954). On Criteria for Descriptions of Income Distribution, *Metroeconomica*, 6, 88-107.
- Ammon D.(1898). Die Gesellschaftsordnung und ihre naturlichen grundlagen, Jena.
- Amoroso L. (1924-25).Ricerche intorno alla curva dei redditi. *Annali di matematica pura e applicata*, Serie IV,21,123-159,
- Arnold B.C., Press S.J. (1983). Baysian Inference for Pareto Populations, *Journal of Econometrics*, 21, 287-306.
- Atkinson A.C., Pierce M.C. (1976). The Computer Generation of Beta, Gamma and Normal Random Variable, *Journal of the Royal Statistical Society, A*, 139, 431-438.
- Benini R. (1905). I diagrammi a scala logaritmica (a proposito della graduazione per valore delle successioni ereditarie in Italia, Francia e Inghilterra, *Giornale degli Economisti*, 222-237.
- Brush G.S. (1961). On the Optimum Distribution of Income, *Trabajos de Estadistica*, 12, 154-170.
- Cantelli F.P. (1957). Sur les difficultés qui se présentent quand on se fonde sur des considérations de probabilités pour en déduire des lois de fréquence, *Economie Appliquée*, 10, 151-160.
- Choi S.C. Wette R. (1969). Maximum Likelihood Estimation of the Parameters of the Gamma Distributions and Their Bias, *Technometrics*, 11, 683-690.
- Cohen C.A. (1975). Estimation in the Three-parameter Gamma Distribution, *Bulletin of the International Statistical Institute*, 3, 167-170.
- Dagum C. (1980). The Generation and Distribution of Income, the Lorenz Curve and Gini Ratio, *Économie Appliquée*, 33, 327-367
- D'Addario R. (1932). Intorno alla curva dei redditi di Amoroso, *Rivista Italiana di Statistica, Economia, Finanza*, 4, 89-108.
- D'Addario R. (1949). Ricerche sulla curva dei redditi, *Giornale degli Economisti ed Annali di Economia*, 8, 91-115.
- Èltetö.O., Frygies E. (1968). New Income Inequality Measures as Efficient Tools for Causal Analysis and Planning, *Econometrica*, 36, 383-396.
- Giorgi G.M. (1981). Note su alcuni indici di concentrazione, *NoteEconomiche*, 3, 112-119
- Glaser R.E. (1967).The Ratio of the Geometric Mean to the Arithmetic Mean for a Random Sample from a Gamma Distribution, *Journal of the American Statistical Association*, 71, 480-481.
- Greenwood J.A., Durand D. (1960).Aids for Fitting the Gamma Distributionby Maximum Likekihood, *Technometrics*, 2, 55-65.
- Griffiths P., Hill I.D.(Eds.). (1985). Applied Statistics algorithms,Ellis Horwood Limited, Chichester, GB.
- Kloek T., Van Dijk H.K. (1977). Further Results on Efficient Estimation of Income Distribution Parameters, *Écomie Appliquée*, 30, 439-459.
- Kloek T., Van Dijk H.K. (1978). Efficient Estimation of Income Distribution Parameters, *Journal of Econometrics*, 8, 61-74.
- Lee P.S. (1979).A New Measurement of Income Inequality as an Economic and Social Index, *The American Economist*, 23, 49-52.
- March L. (1898). Quelques exemples de la distribution de salaires, *Journal de la Société Statistique de Paris*, 193-206,241-248.

- McDonald J.B., Butler R.J. (1986). Some Generalized Mixture Distributions, Brigham Young University, Provo, Utah
- McDonald J.B., Jensen B.C. (1979). An Analysis of some Properties of Alternative measures of Income Inequality Based on the Gamma Distribution Function, *Journal of the American Statistical Association*, 74, 856-860.
- McDonald J.B., Ransom M.R. (1981). An Analysis of the Bounds for the Gini Coefficient, *Journal of Econometrics*, 17, 177-188.
- Mood M.A., Graybill F.A., Boes D.C. (1974). Introduction to the theory of statistics, 3rd edition, McGraw-Hill Kogakusha Ltd, Tokio.
- Ord J.K., Patil G.P., Taillie C. (1981). The Choice of a Distribution to Describe Personal Incomes, in Statistical Distributions in Scientific Work, Vol.6, 193-201, C. Taillie et al.(eds.).
- Piegorsch W.W. (1983). Regularity Conditions, Asymptotics, and the Exponential Class, American Statistical Association. Proceedings of the section on statistical education, 126-129.
- van Praag B., Hagenars A., van Eck W. (1983). The influence of classification and observation errors on the measurement of income inequality, *Econometrica*, 51, 1093-1108.
- Salem A.B., Mount T.D. (1974). A convenient descriptive model of income distribution: the Gamma density, *Econometrica*, 42, 1115- 1127.
- Shenton L.R., Bowman (1969). Maximum Likelihood Estimators Moment for the 2-parameters Gamma Distributions, *Sankhya*, 31, 379-396.
- Shenton L.R., Bowman (1972). Further remarks on maximum likelihood estimators for the Gamma distributions, *Technometrics*, 14, 725-733.
- Singh S.K., Maddala G.S. (1976). A function for size distribution of incomes, *Econometrica*, 44, 963-970
- Smirnov N. (1948). Tables for estimating the goodness of fit of empirical distributions, *The Annals of Mathematical Statistics*, 19, 279-281.
- Stacy C.A., Mihram G.A. (1965). Parameter Estimation for a Generalized Gamma Distribution, *Technometrics*, 7, 349-358.
- Taguchi T. (1980). On an Interpretation and an Estimation of Shape Parameters of the Generalized Gamma Distribution, *Metron*, 38, 27-40
- Taillie C. (1981). Lorenz Ordering Within the Generalized Gamma Family of Income Distributions, in Statistical Distributions in Scientific Work, Vol.6, 181-192, C. Taillie et al.(eds.).
- Vinci F. (1921). Nuovi contributi allo studio della distribuzione dei redditi, *Giornale degli Economisti*.
- Zenga M. (1985). Un secondo indice di concentrazione basato sui rapporti tra quantili di reddito e quantili di popolazione, *Rivista di statistica Applicata*, 18, 143-154.