

ANTONIO BRUNAZZO - ANGIOLA POLLASTRI

**PROPOSTA DI UNA NUOVA DISTRIBUZIONE:
LA LOGNORMALE GENERALIZZATA**

Estratto da:
SCRITTI IN ONORE DI FRANCESCO BRAMBILLA
Volume primo

Milano - Edizioni di «Bocconi Comunicazione» - 1986

PROPOSTA DI UNA NUOVA DISTRIBUZIONE: LA LOGNORMALE GENERALIZZATA *

INTRODUZIONE

In questo lavoro viene presentata una nuova variabile casuale (v.c.), la cui trasformata logaritmica è la cosiddetta v.c. normale generalizzata già ampiamente studiata in Italia¹. È noto infatti che

$$P(Y \leq y) = \int_{-\infty}^y \frac{c}{\theta_2} \exp \left\{ -\frac{1}{\alpha + 1} \left| \frac{t - \theta_1}{\theta_2} \right|^{\alpha + 1} \right\} dt \triangleq {}_rN(y | \theta_1, \theta_2, \alpha) \quad (0.1)$$

$$\theta_2 > 0, c = \frac{(\alpha + 1)^{\frac{\alpha}{\alpha + 1}}}{2\Gamma\left(\frac{1}{\alpha + 1}\right)}, \alpha \geq 0 \quad (0.2)$$

è la funzione di distribuzione della v.c. normale generalizzata.

Nel caso che una v.c. X sia legata alla v.c. Y dalla relazione

$$Y = \log X \quad (0.3)$$

diremo che la v.c. X ha una distribuzione lognormale generalizzata.

Pertanto la funzione di densità $f(x)$ della v.c. risulta

$$f(x) = \frac{c}{\theta_2 x} \exp \left\{ -\frac{1}{\alpha + 1} \left| \frac{\log x - \theta_1}{\theta_2} \right|^{\alpha + 1} \right\}, \quad x > 0 \quad (0.4)$$

e quindi la relativa funzione di ripartizione (che indicheremo con il simbolo ${}_rA(x | \theta_1, \theta_2, \alpha)$) sarà

$${}_rA(x | \theta_1, \theta_2, \alpha) \triangleq \int_0^x \frac{c}{\theta_2 t} \exp \left\{ -\frac{1}{\alpha + 1} \left| \frac{\log t - \theta_1}{\theta_2} \right|^{\alpha + 1} \right\} dt. \quad (0.5)$$

* I parr. 1, 2, 3 e 4 sono stati curati dal prof. Antonio Brunazzo. L'introduzione e il par. 5 nonché la predisposizione delle tabelle e dei grafici sono dovuti alla prof.ssa Angiola Pollastri. Il lavoro è stato eseguito con il contributo CNR n. 84.00616.10. Si ringrazia il professor Michele Zenga per aver suggerito agli autori l'argomento qui studiato e per gli utili consigli durante lo svolgimento del lavoro.

¹ Si veda S. VIANELLI (1963), G. LUNETTA (1963), A. MINEO (1980).

Come si può notare, detta funzione dipende dai tre parametri: θ_1 , θ_2 , α .

In questo lavoro si studieranno alcune caratteristiche della distribuzione ${}_{\alpha}A(x | \theta_1, \theta_2, \alpha)$, quali l'andamento della funzione di densità, la moda, la mediana, i momenti, la variabilità e l'asimmetria al variare dei parametri. Saranno dati inoltre (per alcune terne dei parametri $\theta_1, \theta_2, \alpha$) i grafici della $f(x) = {}_{\alpha}A'(x | \theta_1, \theta_2, \alpha)$.

Si è presentato infine un esempio di applicazione della distribuzione qui proposta al fine di descrivere la distribuzione del reddito in Italia ottenuto dall'indagine DOXA del 1948. Verranno utilizzati due metodi di stima che portano a stime di α molto simili.

Inoltre verrà mostrato come, per tale esempio, si ottiene un miglioramento considerevole nel grado di accostamento quando si utilizza la distribuzione lognormale generalizzata invece della classica distribuzione lognormale.

1. STUDIO DELLA FUNZIONE DI DENSITÀ DELLA VARIABILE CASUALE LOGNORMALE GENERALIZZATA

Premesso che il valore della mediana M_e della v.c. X è, $\forall \alpha \geq 0$

$$M_e = e^{\theta_1}, \quad (1.1)$$

distingueremo, per lo studio della funzione di densità della v.c. X

$$f(x) = \frac{c}{\theta_2 x} \exp \left\{ - \frac{1}{\alpha + 1} \left| \frac{\log x - \theta_1}{\theta_2} \right|^{\alpha + 1} \right\}, \quad x > 0 \quad (1.2)$$

i due casi: $\alpha = 0$, $\alpha > 0^2$.

a) $\alpha = 0$. La (2.1) diventa

$$f(x) = \frac{1}{2 \theta_2 x} \exp \left\{ - \left| \frac{\log x - \theta_1}{\theta_2} \right| \right\}, \quad x > 0. \quad (1.3)$$

Tale distribuzione è stata proposta anche da Uppuluri (1981) col nome di distribuzione log-Laplace.

Dalla (1.3) si deducono facilmente le caratteristiche analitiche.

Riassumendo abbiamo che:

per $\forall \theta_2 > 0$, nei relativi grafici (vedi fig. 1) il punto $(e^{\theta_1}, f(e^{\theta_1}))$ risulta angoloso con tangente sinistra distinta da tangente destra;

per $\theta_2 < 1$, la moda coincide con la mediana; mentre per $\theta_2 > 1$ la moda non esiste. Nel caso invece di $\theta_2 = 1$ qualsiasi punto di $[0, e^{\theta_1}]$ è punto modale.

Osserviamo, sempre per $\alpha = 0$, che i momenti μ_r della v.c. X esistono solamente se

$$r < \frac{1}{\theta_2} \quad (1.4)$$

cioè nel caso di $\theta_2 < 1$.

Infatti, dopo qualche calcolo

$$\mu_r = \int_0^{\infty} x^r f(x) dx = \frac{e^{r\theta_1}}{1 - (r\theta_2)^2}. \quad (1.5)$$

² Osserviamo che per $\alpha = 1$ — vedi fig. 4 — la (1.2) diventa la densità della v.c. lognormale.

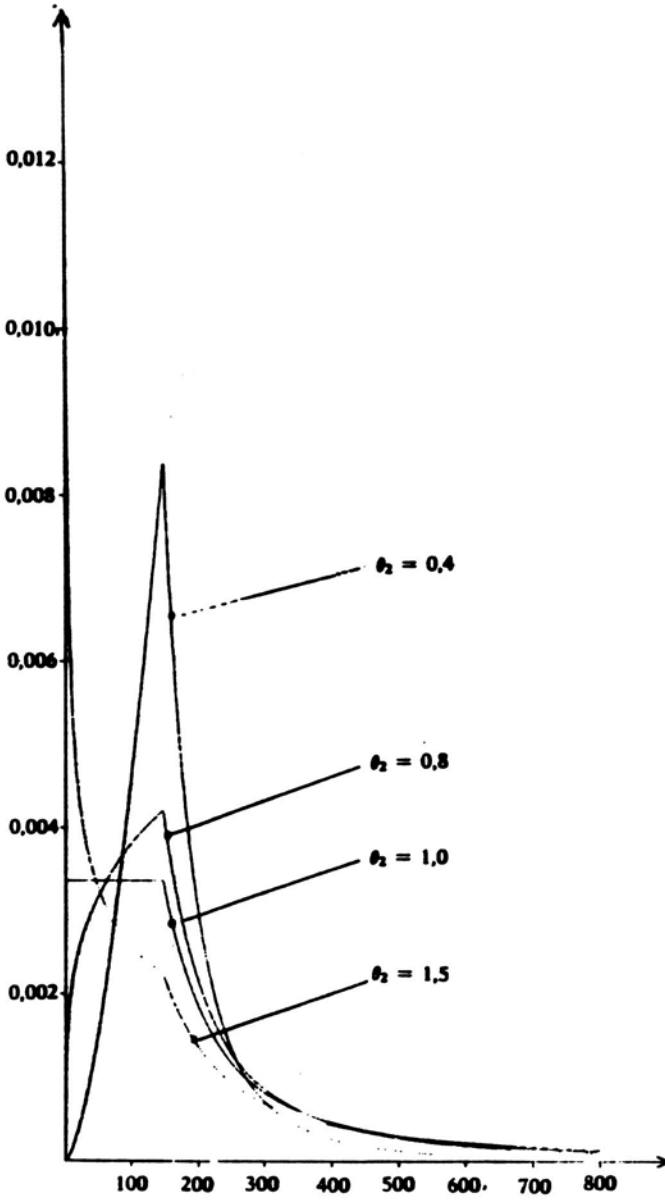


Fig. 1 - Grafici di $gA'(x | \theta_1, \theta_2, \alpha) = gA'(x | 5, \theta_2, 0)$

Segue che la media aritmetica μ risulta

$$\mu = \frac{e^{\theta_1}}{1 - \theta_2^2} \quad (1.6)$$

e (nel caso di $\theta_2 < 1/2$) lo scarto quadratico medio è uguale a

$$\sigma = \sqrt{\mu_2 - \mu^2} = \frac{e^{\theta_1} \theta_2}{1 - \theta_2^2} \sqrt{\frac{2 + \theta_2^2}{1 - 4\theta_2^2}} = \mu \theta_2 \sqrt{\frac{2 + \theta_2^2}{1 - 4\theta_2^2}}. \quad (1.7)$$

b) $\alpha > 0$.

Dalla (1.2) si ottiene $\lim_{x \rightarrow 0^+} f(x) = 0$, $\lim_{x \rightarrow \infty} f(x) = 0$.

$$\text{Per } 0 < x < e^{\theta_1}, \quad \lim_{x \rightarrow 0^+} f'(x) = 0, \quad \lim_{x \rightarrow e^{\theta_1}-} f'(x) = -\frac{c}{\theta_2 e^{2\theta_1}},$$

$$\text{per } x > e^{\theta_1}, \quad f'(x) < 0, \quad \lim_{x \rightarrow e^{\theta_1}+} f'(x) = -\frac{c}{\theta_2 e^{2\theta_1}},$$

ne segue quindi che $f(x)$ è ovunque derivabile nel suo dominio.

Inoltre

$$\text{per } 0 < x < e^{\theta_1}, \quad f'(x) \cong 0, \quad \text{se: } x \cong \exp\left\{\theta_1 - \theta_2 \frac{\alpha+1}{\alpha}\right\};$$

risulta perciò che la moda M_0 è

$$M_0 = \exp\left\{\theta_1 - \theta_2 \frac{\alpha+1}{\alpha}\right\} < M_e, \quad (1.8)$$

e la $f(x)$ risulta unimodale con asimmetria positiva.

Lo studio della concavità di $f(x)$ comporta qualche difficoltà analitica. Risulta infatti che la $f(x)$ possiede due o quattro flessi a seconda dei valori che il parametro θ_2 assume al variare del parametro α .

Distingueremo quindi i seguenti casi

b₁) $\alpha > 1$, $x > e^{\theta_1}$.

Dalla (1.2) si ottiene

$$f''(x) = \frac{f(x)}{x^2} \left\{ \left[\frac{1}{\theta_2^{\alpha+1}} (\log x - \theta_1)^\alpha + 1 \right]^2 + \right. \\ \left. + \left[\frac{1}{\theta_2^{\alpha+1}} (\log x - \theta_1)^\alpha + 1 \right] - \frac{\alpha}{\theta_2^{\alpha+1}} (\log x - \theta_1)^{\alpha-1} \right\}, \quad (1.9)$$

quindi

$$\text{segno } f''(x) = \text{segno } \{(\beta u^\alpha + 1)(\beta u^\alpha + 2) - \alpha \beta u^{\alpha-1}\} \quad (1.10)$$

dove

$$\beta = \frac{1}{\theta_2^{\alpha+1}}, \quad u = \log x - \theta_1. \quad (1.11)$$

Studiando le due funzioni

$$y_1(u) = (\beta u^\alpha + 1)(\beta u^\alpha + 2), \quad u > 0 \quad (1.12)$$

$$y_2(u) = \alpha \beta u^{\alpha-1}, \quad u > 0 \quad (1.13)$$

avremo

$$\text{segno } f''(x) = \text{segno } [y_1(u) - y_2(u)], \quad u > 0. \quad (1.14)$$

Dallo studio delle (1.12) (1.13) si deduce che i rispettivi grafici possono non incontrarsi affatto, essere tangenti in un punto, avere due sole intersezioni.

Supponendo che per $u = \lambda$ le due funzioni siano tangenti, questo valore λ dovrà soddisfare le due espressioni

$$\begin{aligned} y_1(\lambda) &= y_2(\lambda) \\ y_1'(\lambda) &= y_2'(\lambda) \end{aligned} \quad (1.15)$$

cioè

$$\begin{aligned} \beta^2 \lambda^{2\alpha} + 3\beta\lambda^\alpha + 2 &= \alpha\beta\lambda^{\alpha-1} \\ \alpha\beta\lambda^{\alpha-1}(2\beta\lambda^\alpha + 3) &= \alpha\beta(\alpha-1)\lambda^{\alpha-2}. \end{aligned} \quad (1.16)$$

Operando sulle (1.16) si trova che la relazione che lega β ad α , per la tangenza, è

$$\beta^* = \frac{3\sqrt{8\alpha^2 + 1} - 8\alpha - 1}{2[3\alpha - \sqrt{8\alpha^2 + 1}]^{\alpha+1}} \quad (1.17)$$

e dato che

$$\begin{aligned} \beta &= \frac{1}{\theta_2^{\alpha+1}} \\ \theta_2^* &= \frac{2^{\frac{1}{\alpha+1}} [3\alpha - \sqrt{8\alpha^2 + 1}]^{\frac{1}{\alpha+1}}}{[3\sqrt{8\alpha^2 + 1} - 8\alpha - 1]^{\frac{1}{\alpha+1}}} \end{aligned} \quad (1.18)$$

Il comportamento poi delle funzioni $y_1(u)$ e $y_2(u)$, al variare di β , porta ad affermare che:

se: $\theta_2 \geq \theta_2^*$ la $f''(x)$ si mantiene sempre positiva (per $\alpha > 1$, $x > e^{\theta_1}$) e quindi $f(x)$ non possiede nessun flesso dopo la mediana e la concavità sarà verso l'alto;

se: $\theta_2 \leq \theta_2^*$ la $f''(x)$ cambia tre volte di segno (per $\alpha > 1$, $x > e^{\theta_1}$) e quindi $f(x)$ possiede due flessi dopo la mediana.

$$b_2) \quad \alpha > 1, \quad 0 < x < e^{\theta_1}.$$

Dalla (1.2) si ottiene

$$f''(x) = \frac{f(x)}{x^2} \left\{ \left[\frac{1}{\theta_2^{\alpha+1}} (\theta_1 - \log x)^\alpha - 1 \right]^2 - \left[\frac{1}{\theta_2^{\alpha+1}} (\theta_1 - \log x)^\alpha - 1 \right] - \frac{\alpha}{\theta_2^{\alpha+1}} (\theta_1 - \log x)^{\alpha-1} \right\}, \quad (1.19)$$

e

$$\text{segno } f''(x) = \text{segno } \{(\beta z^\alpha - 1)(\beta z^\alpha - 2) - \alpha \beta z^{\alpha-1}\} \quad (1.20)$$

dove

$$\beta = \frac{1}{\theta_2^{\alpha+1}}, \quad z = \theta_1 - \log x.$$

Studiando le due funzioni

$$\begin{aligned} g_1(z) &= (\beta z^\alpha - 1)(\beta z^\alpha - 2), & z > 0 \\ g_2(z) &= \alpha \beta z^{\alpha-1}, & z > 0 \end{aligned} \quad (1.21)$$

sarà

$$\text{segno } f''(x) = \text{segno } \{g_1(z) - g_2(z)\}, \quad z > 0.$$

Dallo studio delle (1.21) risulta che la $f(x)$ possiede sempre due flessi per $0 < x < M_e$. Precisamente: un flesso è prima della moda e l'altro fra la moda e la mediana.

I casi b_1 e b_2 sono riassunti nella seguente tabella

$\alpha > 1$	$0 < x < M_0$	$M_0 < x < M_e$	$x > M_e$
$\theta_2 < \theta_2^*$	1 flesso per $f(x)$	1 flesso per $f(x)$	2 flessi per $f(x)$
$\theta_2 \geq \theta_2^*$	1 flesso per $f(x)$	1 flesso per $f(x)$	—

(1.22)
cioè, se $\alpha > 1$, la $f(x)$ possiede nel suo dominio:

- 1) un flesso prima della moda M_0 , per $\forall \theta_2 > 0$
- 2) un flesso fra la moda e la mediana M_e , per $\forall \theta_2 > 0$
- 3) altri due flessi, dopo la mediana, solo se $\theta_2 < \theta_2^*$.

I relativi grafici sono presentati dalle figg. 5 e 6.

Lo studio della concavità di $f(x)$ nel caso di $0 < \alpha < 1$ (operando come per $\alpha > 1$) è riassunto nella seguente tabella:

$0 < \alpha < 1$	$0 < x < M_0$	$M_0 < x < M_e$	$x > M_e$
$\theta_2 \leq \theta_2^{**}$	1 flesso per $f(x)$	—	1 flesso per $f(x)$
$\theta_2 > \theta_2^{**}$	1 flesso per $f(x)$	2 flessi per $f(x)$	1 flesso per $f(x)$

(1.23)

essendo

$$\theta_2^{**} = \frac{2^{\frac{1}{\alpha+1}} [\sqrt{8\alpha^2 + 1} - 3\alpha]}{[3\sqrt{8\alpha^2 + 1} - 8\alpha - 1]^{\frac{1}{\alpha+1}}} \quad (1.24)$$

I relativi grafici sono presentati dalle figg. 2 e 3.

Ricordiamo la proprietà, già detta, che la $f(x)$ per $\alpha > 0$ è ovunque derivabile nel suo dominio. Questo richiamo vien fatto per una lettura attenta dei grafici di $f(x)$ nel caso di $0 < \alpha < 1$. Infatti se $0 < \alpha < 1$ si ha l'impressione che i grafici di $f(x)$ possiedano un punto angoloso per $x = e^{\theta_1}$, cioè nel punto mediano. Per quanto è stato precedentemente affermato il detto punto non può essere angoloso, succede che non esiste $f'(e^{\theta_1})$ e che $\lim_{x \rightarrow e^{\theta_1}} f'(x) = -\infty$. Ne segue che in un intorno H di e^{θ_1} risulta elevata l'oscillazione $|f'(e^{\theta_1} - h) - f'(e^{\theta_1} + h)|$ e pertanto nelle vicinanze di $P(e^{\theta_1}, f(e^{\theta_1}))$ la curva si allontana (mantenendo la concavità verso il basso) rapidamente dalla tangente in P alla $f(x)$.

2. CALCOLO DEI MOMENTI μ_r DELLA V.C. X

È noto che, per la v.c. X

$$\mu_r = \int_0^{\infty} x^r f(x) dx. \quad (2.1)$$

Sostituendo in (2.1) la (1.2) si avrà

$$\mu_r = \int_0^{\infty} \frac{c}{\theta_2} x^{r-1} \exp \left\{ -\frac{1}{\alpha+1} \left| \frac{\log x - \theta_1}{\theta_2} \right|^{\alpha+1} \right\} dx. \quad (2.2)$$

L'integrale (2.2) converge per $r > 0$; posto quindi $u = \frac{\log x - \theta_1}{\theta_2}$, dopo qualche calcolo, si ottiene

$$\mu_r = 2c e^{\theta_1} \int_0^{\infty} \exp \left\{ -\frac{1}{\alpha+1} u^{\alpha+1} \right\} \coth(r\theta_2 u) du. \quad (2.3)$$

Essendo³

³ Cfr. I.S. GRADSHTEYN - I.M. RYZHIK (1980), p. ...

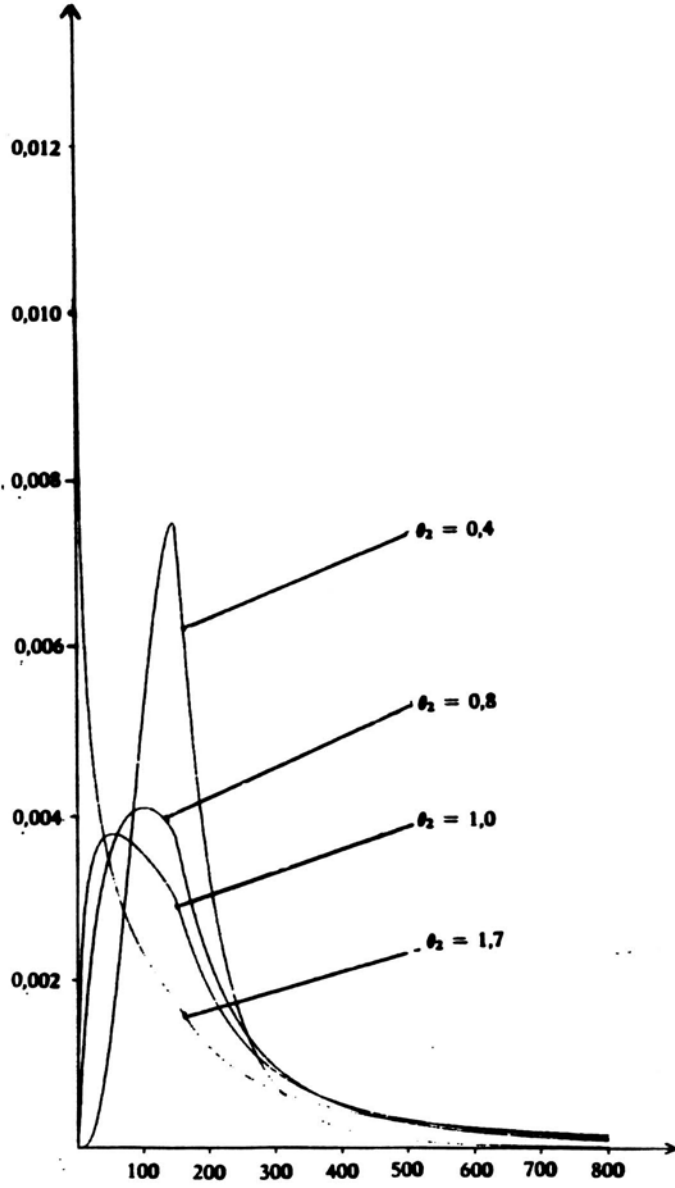


Fig. 2 - Grafici di $gA'(x | \theta_1, \theta_2, \alpha) = gA'(x | 5, \theta_2, 0.3)$; $\theta_2^* = 1,1357$

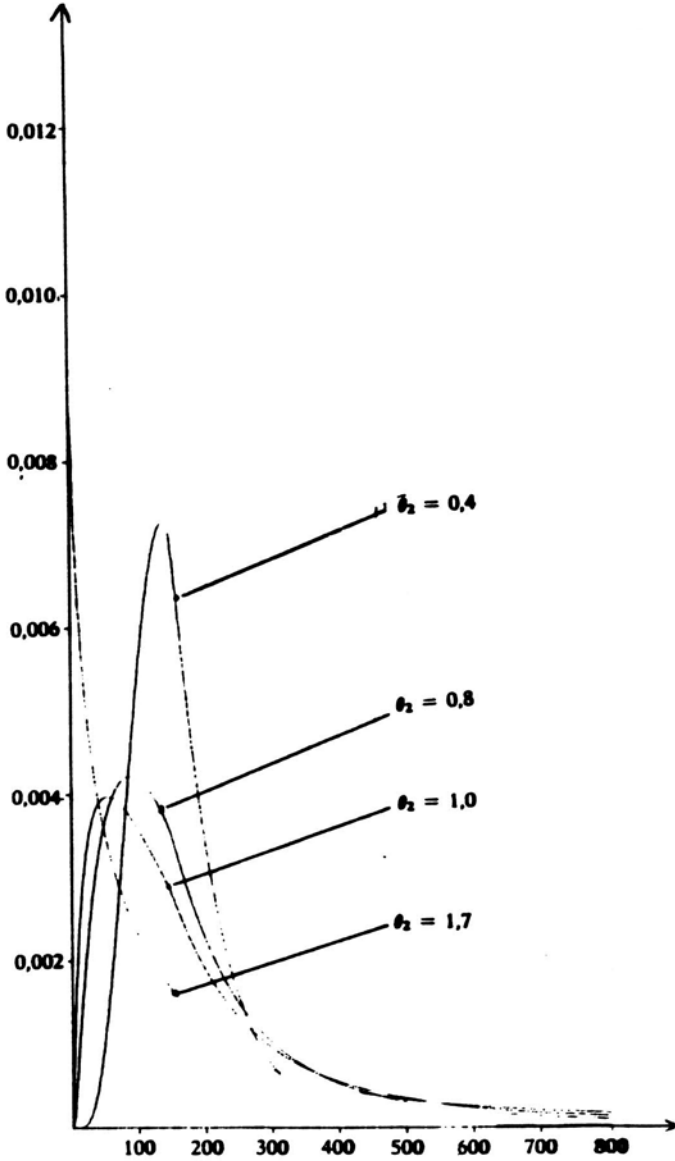


Fig. 3 - Grafici di $gA'(x | \theta_1, \theta_2, \alpha) = gA'(x | 5, \theta_2, 0.5)$; $\theta_1^* = 1,09112$

$$\operatorname{coh} z = \sum_{i=0}^{\infty} \frac{z^{2i}}{(2i)!} \quad (2.4)$$

$$\begin{aligned} \mu_r &= 2c e^{r\theta_1} \int_0^{\infty} \exp \left\{ -\frac{1}{\alpha+1} u^{\alpha+1} \right\} \left(\sum_{i=0}^{\infty} \frac{(r\theta_2 u)^{2i}}{(2i)!} \right) du = \\ &= 2c e^{r\theta_1} \sum_{i=0}^{\infty} \frac{(\theta_2 r)^{2i}}{(2i)!} \int_0^{\infty} u^{2i} \exp \left\{ -\frac{1}{\alpha+1} u^{\alpha+1} \right\} du, \end{aligned} \quad (2.5)$$

e dato che⁴

$$\int_0^{\infty} x^{v-1} e^{-ax^p} dx = \frac{1}{p} a^{-\frac{v}{p}} \Gamma \left(\frac{v}{p} \right); \quad v > 0, \quad a > 0, \quad p > 0 \quad (2.6)$$

si ottiene

$$\mu_r = \frac{e^{r\theta_1}}{\Gamma \left(\frac{1}{\alpha+1} \right)} \sum_{i=0}^{\infty} \frac{\theta_2^{2i} r^{2i} (\alpha+1)^{\frac{2i}{\alpha+1}}}{(2i)!} \Gamma \left(\frac{2i+1}{\alpha+1} \right), \quad (2.7)$$

che è l'espressione cercata.

Ne segue che la media μ risulta

$$\begin{aligned} \mu &= \mu_1 = 2c e^{\theta_1} \int_0^{\infty} \exp \left\{ -\frac{1}{\alpha+1} u^{\alpha+1} \right\} \operatorname{coh}(\theta_2 u) = \\ &= \frac{e^{\theta_1}}{\Gamma \left(\frac{1}{\alpha+1} \right)} \sum_{i=0}^{\infty} \frac{\theta_2^{2i} (\alpha+1)^{\frac{2i}{\alpha+1}}}{(2i)!} \Gamma \left(\frac{2i+1}{\alpha+1} \right). \end{aligned} \quad (2.8)$$

Dalla (2.7) si ottengono i seguenti casi particolari

i₁) $\alpha = 0, \quad r\theta_2 < 1.$

$$\mu_r = e^{r\theta_1} \sum_{i=0}^{\infty} \frac{\theta_2^{2i} r^{2i}}{(2i)!} \Gamma(2i+1) = e^{r\theta_1} \sum_{i=0}^{\infty} (\theta_2 r)^{2i} = \frac{e^{r\theta_1}}{1 - (r\theta_2)^2}.$$

espressione già ottenuta in (1.5).

i₂) $\alpha = 1$

$$\mu_r = \frac{e^{r\theta_1}}{\Gamma \left(\frac{1}{2} \right)} \sum_{i=0}^{\infty} \frac{\theta_2^{2i} r^{2i}}{(2i)!} 2^i \Gamma \left(\frac{2i+1}{2} \right) \quad (2.9)$$

Essendo⁵

⁴ Cfr. I.S. GRADSHTEYN - I.M. RYZHIK (1980), p. 342.

⁵ Cfr. I.S. GRADSHTEYN - I.M. RYZHIK (1980), p. 932.

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}, \quad \Gamma\left(i + \frac{1}{2}\right) = \frac{\sqrt{\pi}}{2^i} (2i-1)!!, \quad (2i-1)!! = \frac{(2i)!}{2^i i!}$$

la (2.9) diventa

$$\mu_r = e^{r\theta_1} \sum_{i=0}^{\infty} \frac{\theta_2^{2i} r^{2i}}{i! 2^i} = e^{r\theta_1} \sum_{i=0}^{\infty} \left(\frac{\theta_2^2 r^2}{2}\right)^i \frac{1}{i!} = e^{r\theta_1} e^{\frac{1}{2} r^2 \theta_2^2},$$

cioè

$$\mu_r = e^{r\theta_1 + \frac{1}{2} r^2 \theta_2^2} \quad (2.10)$$

che, come è noto⁶, rappresenta il momento di ordine r della v.c. lognormale.

3. VARIABILITÀ ED ASIMMETRIA

La tab. 1 presenta la variabilità e l'asimmetria della distribuzione lognormale generalizzata al variare dei parametri θ_2 ed α , relative ai grafici delle figg. 1, 2, 3, 4, 5, 6, per un fissato valore del parametro θ_1 .

Per la variabilità si sono calcolati, a mezzo dell'elaboratore elettronico, valori dei seguenti indici:

σ , cioè lo scarto quadratico medio utilizzando l'espressione

$$\sigma = \sqrt{\mu_2 - \mu^2}, \quad (3.1)$$

C.V. = $\frac{\sigma}{\mu}$, cioè il coefficiente di variazione,

S_{Me} , cioè lo scarto medio assoluto della mediana,

$$\frac{S_{Me}}{M_e}, \quad V = \frac{S_{Me}}{\mu} \quad (0 < V < 1).$$

Per il calcolo di S_{Me} abbiamo

$$\begin{aligned} S_{Me} &= \int_0^{\infty} |x - M_e| f(x) dx = \\ &= 2ce^{\theta_1} \int_0^{\infty} \exp\left\{-\frac{1}{\alpha+1} u^{\alpha+1}\right\} \operatorname{seh}(\theta_2 u) du. \end{aligned} \quad (3.2)$$

Utilizzando⁷

$$\operatorname{seh} z = \sum_{i=0}^{\infty} \frac{z^{2i+1}}{(2i+1)!} \quad (3.3)$$

⁶ Vedi, ad es. N.L. JOHNSON - S. KOTZ (1970) e J. AITCHISON - J.A.C. BROWN (1957).

⁷ Cfr. I.S. GRADHSTEYN - I.M. RYZHIK (1980), pp. 33.

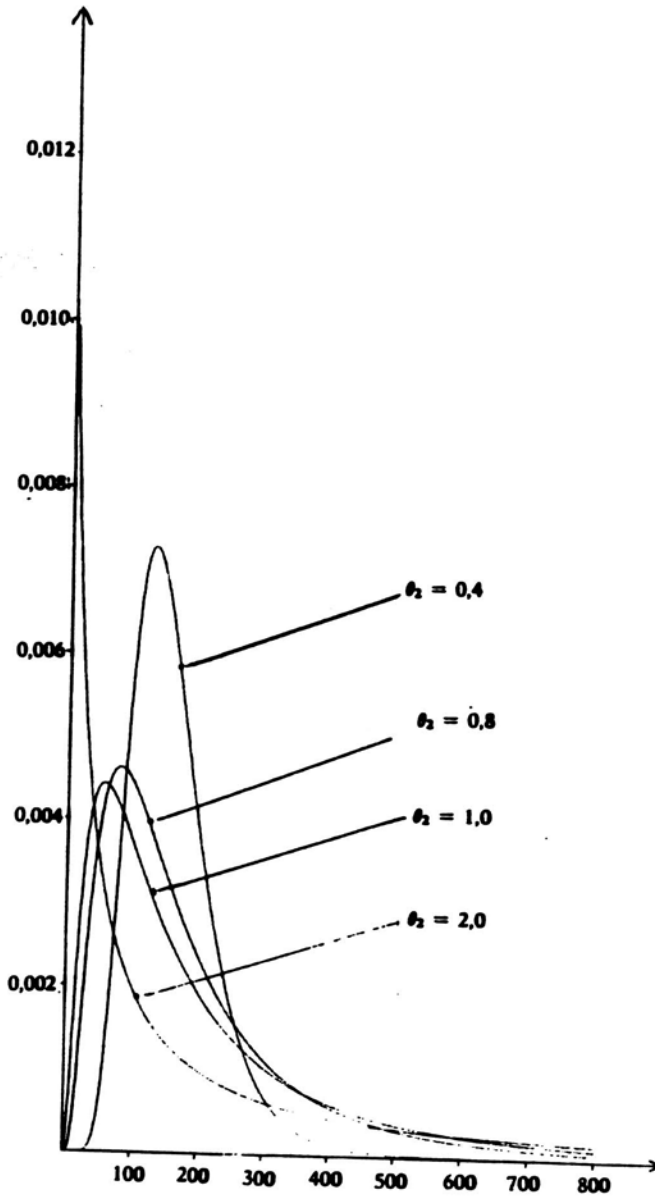


Fig. 4 - Grafici di $g\Lambda'(x | \theta_1, \theta_2, \alpha) = \Lambda'(x | 5, \theta_2, 1,0)$

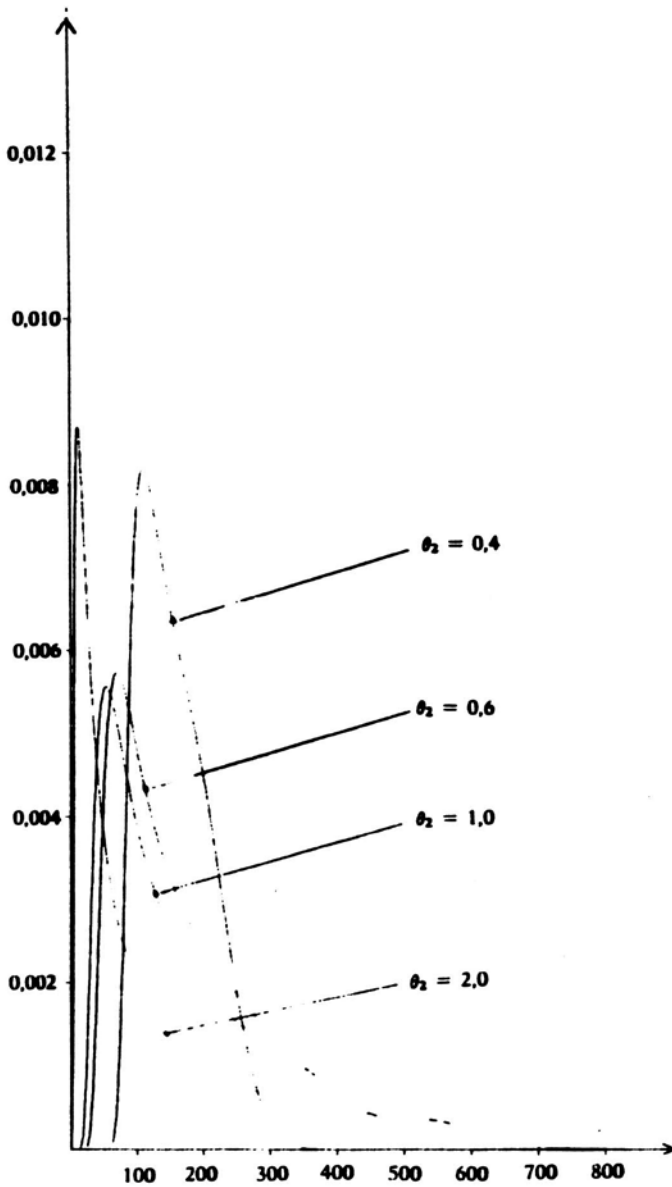


Fig. 5 - Grafici di $gA'(x | \theta_1, \theta_2, \alpha) = gA'(x | 5, \theta_2, 3,0)$; $\theta_1^2 = 0,60819$

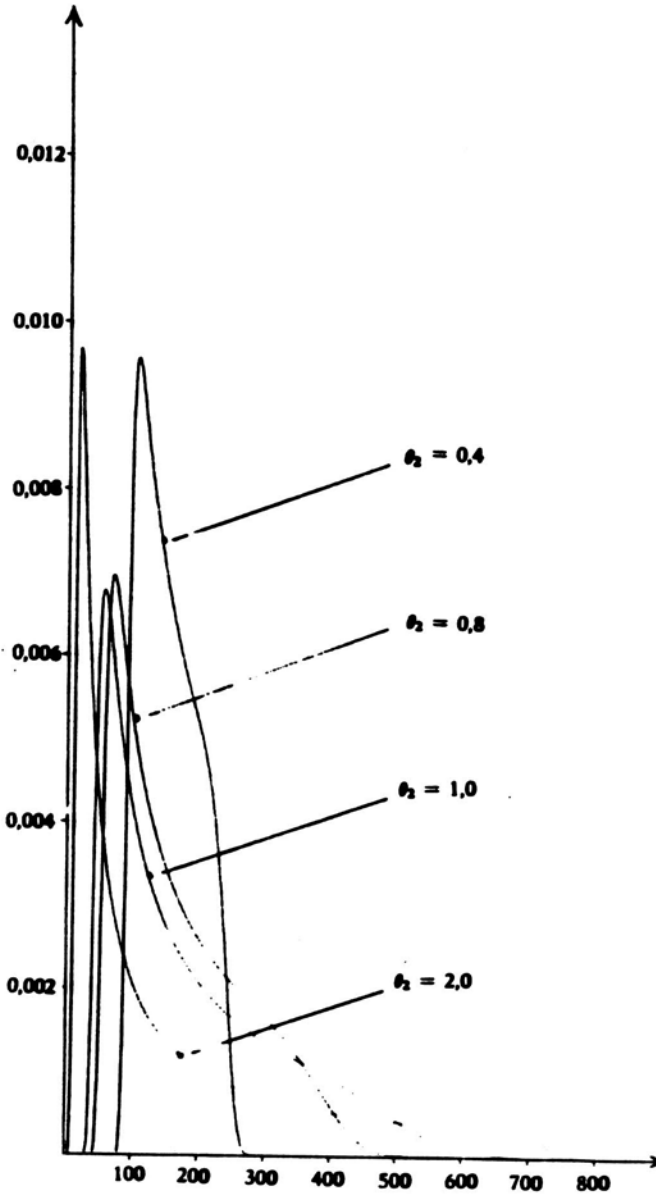


Fig. 6 - Grafici di $gA'(x | \theta_1, \theta_2, \alpha) = gA'(x | 5, \theta_2, 8.0)$; $\theta_1^2 = 1.29349$

e la (2.6), dopo qualche calcolo otteniamo

$$S_{M_0} = \frac{e^{\theta_1}}{\Gamma\left(\frac{1}{\alpha+1}\right)} \sum_{i=0}^{\infty} \frac{\theta_2^{2i+1} (\alpha+1)^{\frac{2i+1}{\alpha+1}}}{(2i+1)!} \Gamma\left(\frac{2i+2}{\alpha+1}\right). \quad (3.4)$$

Dalla tab. 1 risulta che tutti gli indici di variabilità hanno un andamento analogo, cioè:

1) nel caso di α costante, al crescere di θ_2 aumenta la variabilità (riducendosi, qualche volta anche notevolmente, il suo campo di variazione);

2) nel caso di θ_2 costante, al crescere di α diminuisce la variabilità (aumentando, qualche volta notevolmente, il suo campo di variazione).

Per quanto riguarda l'asimmetria oltre a calcolare i valori forniti dagli indici del Fisher e del Pearson, cioè

$$\gamma_1 = \frac{\bar{\mu}_3}{\sigma^3}, \quad s_k = \frac{\mu - M_0}{\sigma} \quad (3.5)$$

si è considerato anche l'indice⁸

$$a\ddagger = \frac{S_d - S_s}{S_d + S_s} \quad (3.6)$$

dove:

S_d (S_s) = scostamento semplice medio della mediana di tutti i valori X che non sono minori (superiori) della mediana.

Nel nostro caso, dopo qualche calcolo, si ottiene

$$a\ddagger = \frac{\mu - M_0}{S_{M_0}}. \quad (3.7)$$

Si prova facilmente che, nel nostro caso, $0 < a\ddagger < 1$. Infatti, dato che per la v.c. X risulta $\mu > M_0$, segue $a\ddagger > 0$; mentre, essendo⁹

$$\frac{\mu - M_0}{S_{M_0}} = \operatorname{tg} \alpha, \quad 0 < \forall \alpha < \frac{\pi}{4} \quad (3.8)$$

sarà: $a\ddagger < 1$.

Dalla tab. 1 risulta che gli indici di asimmetria non hanno tutti lo stesso andamento, precisamente:

1) per α costante, al crescere di θ_2 aumentano i valori di γ_1 e $a\ddagger$ (riducendosi, anche notevolmente, il relativo campo di variazione).

2) Per θ_2 costante, al crescere di α diminuiscono i valori di γ_1 e $a\ddagger$ (aumentando, anche notevolmente, il relativo campo di variazione).

⁸ Cfr. G. LETI (1983), p. 451.

⁹ Cfr. A. BRUNAZZO (1983).

Tab. 1 - Indici di posizione, di variabilità e di asimmetria per le distribuzioni riportate nei grafici
(Mediana = 148.413)

α	θ_1	θ_2	Moda	μ	σ	C.V. = $\frac{\sigma}{\mu}$	S_{Mc}	$\frac{S_{Mc}}{M_e}$	$V = \frac{S_{Mc}}{\mu}$	$\gamma_1 = \frac{\bar{\mu}_3}{\sigma^3}$	af	S_k
0,0	5,0	0,4	148,413	176,682	0,17311E+03	0,980	70,672	0,476	0,400	(°)	0,400	0,163
0,0	5,0	0,8	148,413	412,259	(°)	(°)	329,789	2,222	0,800	(°)	0,800	(°)
0,0	5,0	1,0	148,413	(°)	(°)	(°)	(°)	(°)	(°)	(°)	(°)	(°)
0,0	5,0	1,7	148,413	(°)	(°)	(°)	(°)	(°)	(°)	(°)	(°)	(°)
0,3	5,0	0,4	145,640	167,344	0,95339E+02	0,570	59,346	0,400	0,355	0,39709E+01	0,319	0,228
0,3	5,0	0,8	101,470	249,934	0,61365E+03	2,455	165,934	1,118	0,664	0,22420E+04	0,612	0,242
0,3	5,0	1,0	54,598	353,673	0,32050E+04	9,062	278,258	1,875	0,787	0,36941E+09	0,738	0,093
0,3	5,0	1,7	0,007	5445,102	0,50942E+13	935559478,737	5389,923	36,317	0,990	0,87099E+87	0,983	0,000
0,5	5,0	0,4	139,212	164,475	0,81908E+02	0,498	55,483	0,374	0,337	0,24205E+01	0,289	0,308
0,5	5,0	0,8	88,943	227,479	0,32720E+03	1,438	142,617	0,961	0,627	0,19148E+02	0,554	0,423
0,5	5,0	1,0	54,598	294,840	0,75833E+03	2,573	218,714	1,474	0,742	0,15577E+03	0,669	0,317
0,5	5,0	1,7	1,091	1454,964	0,17241E+06	118,501	1399,555	9,430	0,962	0,34437E+11	0,934	0,008
1,0	5,0	0,4	126,469	160,774	0,66970E+02	0,417	49,976	0,337	0,311	0,13219E+01	0,247	0,512
1,0	5,0	0,8	78,257	204,384	0,19352E+03	0,947	117,784	0,794	0,576	0,36893E+01	0,475	0,652
1,0	5,0	1,0	54,598	244,692	0,32075E+03	1,311	167,049	1,126	0,683	0,61849E+01	0,576	0,882
1,0	5,0	2,0	2,718	1096,633	0,80285E+04	7,321	1046,735	7,053	0,954	0,41436E+03	0,906	0,136
3,0	5,0	0,4	110,529	156,599	0,51579E+02	0,329	42,296	0,285	0,270	0,59486E+00	0,194	0,893
3,0	5,0	0,8	70,622	183,163	0,12132E+03	0,662	92,546	0,624	0,505	0,12425E+01	0,375	0,928
3,0	5,0	1,0	54,598	205,196	0,17074E+03	0,832	123,710	0,834	0,603	0,16040E+01	0,459	0,882
3,0	5,0	2,0	11,943	480,124	0,84715E+03	1,764	428,021	2,884	0,891	0,41754E+01	0,775	0,553
8,0	5,0	0,4	103,885	154,574	0,44000E+02	0,285	37,291	0,251	0,241	0,38643E+00	0,165	1,152
8,0	5,0	0,8	68,170	174,035	0,97515E+02	0,560	79,411	0,535	0,456	0,77066E+00	0,323	1,086
8,0	5,0	1,0	54,598	189,633	0,13138E+03	0,693	104,004	0,701	0,548	0,96175E+00	0,396	1,028
8,0	5,0	2,0	16,760	358,595	0,46451E+03	1,295	302,989	2,042	0,845	0,19192E+01	0,694	0,736

3) Per α costante l'indice s_k non mantiene la monotonia al crescere di θ_2 , mentre per θ_2 costante assume valori crescenti al crescere di α .

Nel caso poi di $\alpha = 0$ e $\theta_2 < 1$, si ha

$$V = \frac{S_{M_0}}{\mu} = a_f^2 = \theta_2, \quad (3.9)$$

si ha cioè che il parametro θ_2 oltre che misurare la variabilità data dall'indice V , misura anche l'asimmetria data dall'indice a_f^2 .

4. OSSERVAZIONI CONCLUSIVE

È evidente che i parametri θ_1 , θ_2 , α concorrono a definire, in diversa misura, la funzione di densità $f(x)$ della v.c. X . Per esempio il parametro θ_1 , espresso nella forma e^{θ_1} risulta parametro di scala, mentre i due parametri θ_2 ed α dipendono fra loro e per essi si può dimostrare quanto segue

$$\lim_{\alpha \rightarrow \infty} M_0 = \lim_{\alpha \rightarrow \infty} \exp \left\{ \theta_1 - \theta_2 \frac{\alpha+1}{\alpha} \right\} = \exp \{ \theta_1 - \theta_2 \}, \quad \forall \theta_2 \geq 1. \quad (4.1)$$

In tal caso (cioè per α che diverge) la funzione di densità diventa

$$f(x) \begin{cases} = \frac{1}{2\theta_2 x}, & \exp \{ \theta_1 - \theta_2 \} \leq x \leq \exp \{ \theta_1 + \theta_2 \} \\ = 0 & , \text{altrove.} \end{cases}$$

Il suo grafico comprende quindi un arco di iperbole equilatera, con moda = $\exp \{ \theta_1 - \theta_2 \}$ e mediana $M_0 = e^{\theta_1} = \sqrt{\exp \{ \theta_1 - \theta_2 \} \cdot \exp \{ \theta_1 + \theta_2 \}}$.

Segue che, per qualsiasi valore di α finito e positivo, la rispettiva moda M_0 si trova fra 0 ed $\exp \{ \theta_1 - \theta_2 \}$ se $\theta_2 > 1$, mentre la stessa è fra $\exp \{ \theta_1 - \theta_2 \}$ ed M_0 se $\theta_2 < 1$.

Infine, per (4.1), al crescere di α la moda si avvicina al valore $\exp \{ \theta_1 - \theta_2 \}$.

Un altro legame fra θ_2 ed α lo abbiamo considerando la funzione caratteristica

$$\Theta_2 = \frac{2^{\frac{1}{\alpha+1}} \left| 3\alpha - \sqrt{8\alpha^2 + 1} \right|}{\left[3\sqrt{8\alpha^2 + 1} - 8\alpha - 1 \right]^{\frac{1}{\alpha+1}}}. \quad (4.2)$$

La (4.2) contiene le due espressioni (1.18), (1.24) rispettivamente per $\alpha > 1$ e $\alpha < 1$, che regolano il numero dei flessi della $f(x)$ al variare di θ_2 per ogni fissato $\alpha > 0$. Si può anzi notare che la (4.2) ha per dominio l'insieme di tutti i reali positivi esclusa l'unità. Nella tab. 2 sono stati riportati alcuni valori critici di θ_2^* e θ_2^{**} al variare di α .

5. APPLICAZIONI A PROBLEMI CONCRETI: PROBLEMI DI STIMA E STUDIO DELLA BONTÀ DI ADATTAMENTO

In questo paragrafo si individueranno innanzitutto due procedimenti per stimare i parametri della v.c. lognormale generalizzata. Quindi, sulla base delle stime ottenute su una distribuzione relativa ad una indagine effettuata dalla Doxa (Luzzato Fegiz (1956)) sulla distribuzione del reddito in Italia nel 1948, si individueranno di-

TAB. 2 - Valori di θ_1^* o di θ_2^* in corrispondenza di alcuni valori di α

α	θ_1^*	α	θ_2^*
0,01	1,0098	1,10	0,6108
0,05	1,0450	1,20	0,5684
0,10	1,0802	1,30	0,5437
0,15	1,1062	1,40	0,5287
0,20	1,1235	1,50	0,5200
0,25	1,1330	1,60	0,5156
0,30	1,1356	1,70	0,5143
0,35	1,1319	1,80	0,5153
0,40	1,1229	1,90	0,5182
0,45	1,1091	2,00	0,5225
0,50	1,0911	2,10	0,5280
0,55	1,0694	2,20	0,5345
0,60	1,0443	2,30	0,5418
0,65	1,0162	2,40	0,5498
0,70	0,9851	2,50	0,5584
0,75	0,9512	2,60	0,5676
0,80	0,9143	2,70	0,5772
0,85	0,8740	2,80	0,5872
0,90	0,8294	2,90	0,5975
0,95	0,7782	3,00	0,6082
0,99	0,7264	3,50	0,6654
1,00	non def.	4,00	0,7273

verse procedure per verificare se l'accostamento fra dati effettivi e dati stimati possa ritenersi soddisfacente.

5.1. Stima dei parametri

Se la distribuzione dei redditi nell'indagine in esame fosse una lognormale generalizzata allora la distribuzione del logaritmo degli stessi dovrebbe essere una normale generalizzata e quindi si potrebbero stimare i parametri α , θ_1 , θ_2 , attraverso le procedure utilizzate per la normale generalizzata. A tal fine si sono trasformati i limiti delle classi dei valori originari nei logaritmi naturali degli stessi in modo da poter verificare se questi ultimi sono distribuiti secondo una normale generalizzata.

Il primo problema per poter identificare il tipo di curva che meglio interpola i dati consiste nella stima del parametro α . Mineo (1980) suggerisce di stimare $r = \alpha + 1$ sulla base dell'indice di forma

$$\beta_p = \frac{\mu_{2p}'}{\mu_p'^2}$$

dove μ_p' e μ_{2p}' sono rispettivamente i momenti assoluti centrali di ordine p e $2p$. Per la normale generalizzata vale la relazione

$$\beta_p = \frac{\Gamma\left(2 + \frac{1}{p}\right) \Gamma\left(\frac{1}{p}\right)}{\left[\Gamma\left(1 + \frac{1}{p}\right)\right]^2} = p + 1$$

Al fine di stimare r conviene porre $p = r$. Allora risulta $\beta_r = r + 1$. Le stime di μ_1' e μ_2' sono pari a

$$\hat{\mu}_1' = \frac{1}{n} \sum |y_i - \hat{M}|^r f_i, \quad \hat{\mu}_2' = \frac{1}{n} \sum |y_i - \hat{M}|^{2r} f_i$$

essendo

\hat{M} la media aritmetica campionaria;

f_i la frequenza campionaria della classe i -ma;

y_i il valore centrale della classe i -ma dei logaritmi dei redditi.

Per trovare r occorre risolvere l'equazione

$$\hat{\mu}_2' - (r + 1) \hat{\mu}_1'^2 = 0.$$

A tal fine Mineo (1980) propone di usare un metodo di approssimazioni successive chiamato *Regula-falsi*.

Si parte da due valori di r entro cui si pensa sia compresa la radice di r . Mineo propone di basarsi sui valori di r_1 ed r_2 stimati rispettivamente partendo dall'indice di curtosi

$$\beta_2 = \frac{\mu_4'}{\mu_2'^2}$$

e dal rapporto fra lo scostamento semplice medio e lo scarto quadratico medio

$$I = \frac{\mu_1'}{\mu_2'}$$

Per il calcolo di r_1 e di r_2 ci si può servire delle tavole riportate in Vianelli - Mineo (1978) predisposte *ad hoc*.

Nel nostro esempio si arriva ai valori $r_1 = 1,3278$ e $r_2 = 1,4199$. Attraverso il metodo della *Regula-falsi* si giunge alla soluzione approssimata $r = 1,4476$.

Una volta dato r , le stime di θ_1 e θ_2 col metodo dei momenti, sono date rispettivamente da

$$\hat{\theta}_1 = \frac{1}{n} \sum y_i f_i, \quad \hat{\theta}_2 = \left[\frac{1}{n} \sum |y_i - \hat{\theta}_1|^r f_i \right]^{1/r}.$$

Per il logaritmo dei redditi dell'indagine Doxa tali stime risultano pari a

$$\hat{\theta}_1 = 6,1087 \quad \text{e} \quad \hat{\theta}_2 = 0,6120.$$

Se si effettua la trasformazione $T = \frac{Y - \theta_1}{\theta_2}$, la v.c. T ha f.d. data da

$$f(t) = \frac{1}{2r^{1/r} \Gamma\left(1 + \frac{1}{r}\right)} e^{-\frac{|t|^r}{r}}$$

Si sono allora trasformati i limiti delle classi dei logaritmi dei redditi (indicati con y_i) nel seguente modo

$$t_i = \frac{y_i - \hat{\theta}_1}{\hat{\theta}_2}$$

Per effettuare il calcolo delle probabilità teoriche di uno scarto compreso fra $-t_i$ e $+t_i$ ci si è serviti dello sviluppo in serie fornito da Mineo (1980)

$$\begin{aligned} P(-t_i < T < t_i) &= p(t_i) = \int_{-t_i}^{+t_i} f(z) dz = \\ &= \frac{K_i}{r^{1/r} \left(1 + \frac{1}{r}\right)} \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \frac{t_i^n}{r^n (1 + rn)}. \end{aligned} \quad (5.2)$$

Per giungere alle probabilità di ogni classe senza dover effettuare laboriose interpolazioni delle tavole e per ottenere una maggior precisione dei risultati, si è predisposto un programma sull'elaboratore elettronico del CINECA. La serie (5.2) è stata arrestata quando i termini risultavano inferiori a 10^{-11} . Dopo aver calcolato le $p(t_i)$ è stato semplice trovare le probabilità pf^* della distribuzione lognormale generalizzata da attribuire alle classi di reddito. Si sono poi calcolate le frequenze teoriche ff^* moltiplicando ogni probabilità teorica pf^* per la numerosità del campione.

Il programma di cui sopra, una volta fornitigli gli estremi delle classi dei dati originari del reddito (x_i), le frequenze f_i ed il valore di r , ha elaborato la tab. 3.

Si è calcolata la statistica

$$X^2 = \sum_{i=1}^K \frac{(f_i - ff^*)^2}{ff^*}$$

che è risultata pari a 191,718.

Con il programma di calcolo sopra descritto si può anche determinare quel valore di r che rende minima la statistica X^2 . Per questo secondo metodo di stima si è proceduto come segue: si sono fatti assumere ad r valori fra 1,1 e 3 con incrementi pari a 0,1 e calcolando per ogni r le frequenze teoriche, si è visto che il minimo di X^2 si raggiunge fra $r = 1,4$ ed $r = 1,5$.

Procedendo in tale intervallo con incrementi di r pari a 0,01 e continuando in modo analogo, si è trovato che il minimo di X^2 si raggiunge per $r = 1,412$. In corrispondenza di tale valore di r si ottiene $X^2 = 190,668$.

Come si vede, i valori di r ottenuti con i due metodi di stima non si discostano in modo eccessivo e nemmeno il valore di X^2 nella lognormale generalizzata con $r = 1,4476$ è molto diverso da quello della distribuzione con $r = 1,421$.

Una prima conclusione a cui si può giungere è che, almeno relativamente alla distribuzione esaminata, fra le stime di r ottenute con il primo procedimento e quelle ottenute con il secondo non vi sono sostanziali differenze «confermando» la validità pratica del metodo proposto da Mineo.

5.2. *Bontà di adattamento*

Si tratta ora di giudicare la bontà di adattamento della v.c. lognormale generalizzata alla distribuzione dei redditi della indagine DOXA del 1948. Per meglio valutare questo adattamento si è anche interpolato con una lognormale semplice per verificare se l'aggiunta di un parametro migliora sensibilmente l'accostamento o meno. Ponendo $r = 2$ nel programma di calcolo sopra descritto, si sono ottenute le informazioni relative all'interpolazione della distribuzione lognormale semplice ai dati effettivi della distribuzione dei redditi. Le informazioni ottenute vengono riportate nella tab. 4.

Per giudicare la bontà di adattamento si sono seguite diverse strade:

- osservazione visiva ed impiego di indici descrittivi;
- impiego del test di Kolmogorov-Smirnov;
- impiego degli intervalli di confidenza simultanei asintotici per le probabilità delle classi;
- impiego del test chi-quadrato con errori proporzionali.

a) Innanzitutto, osservando la tab. 3 e la tab. 4, si può notare che gli scarti fra probabilità effettive e teoriche nella distribuzione lognormale generalizzata cambiano di segno al più ogni due classi eccetto che nelle ultime cinque classi. Si tenga tuttavia presente che le ultime tre classi sono relative al 5,4% del totale delle frequenze.

TAB. 3 - *Adattamento della distribuzione lognormale generalizzata con $\alpha = 0,4476$ alla distribuzione del reddito. Indagine DOXA*

CLASSI DI REDDITO ANNUO (migl. L.)		f_i	$\frac{f_i}{n}$	$\frac{F_i}{n}$	p_i^*	$f_i^* = np_i^*$	$p_i^* - \frac{f_i}{n}$	$\frac{(f_i^* - f_i)^2}{f_i^*}$
70,0	130,0	305,0	0,02842	0,02842	0,04071	436,923	0,01229	39,83238
130,0	260,0	1704,0	0,15878	0,18720	0,15406	1653,416	-0,00471	1,54755
260,0	390,0	2479,0	0,23099	0,41819	0,20906	2243,594	-0,02193	24,69971
390,0	520,0	1906,0	0,17760	0,59579	0,19392	2081,174	0,01632	14,74446
520,0	650,0	1441,0	0,13427	0,73006	0,12773	1370,835	-0,00654	3,59135
650,0	780,0	845,0	0,07874	0,80880	0,08069	865,995	0,00196	0,50899
780,0	910,0	566,0	0,052742	0,86154	0,05262	564,687	-0,00012	0,00305
910,0	1040,0	357,0	0,03327	0,89480	0,03552	381,230	0,00226	1,54002
1040,0	1170,0	212,0	0,01975	0,91455	0,02475	265,622	0,00500	10,82482
1170,0	1300,0	269,0	0,02507	0,93962	0,01773	190,242	-0,00734	32,60488
1300,0	1625,0	229,0	0,02134	0,96096	0,02670	286,531	0,00536	11,55125
1625,0	1950,0	158,0	0,01472	0,97568	0,01375	147,581	-0,00097	0,73553
1950,0	2275,0	63,0	0,00587	0,98155	0,00774	83,075	0,00187	4,85095
2275,0	2600,0	62,0	0,00578	0,98733	0,00466	50,014	-0,00112	2,87230
2600,0	2925,0	21,0	0,00196	0,98928	0,00296	31,737	0,00100	3,63268
2925,0	3250,0	21,0	0,00196	0,99124	0,00196	21,010	0,00000	0,00001
3250,0	3575,0	16,0	0,00149	0,99273	0,00134	14,402	-0,00015	0,17731
3575,0	3900,0	20,0	0,00186	0,99460	0,00095	10,164	-0,00092	9,51752
3900,0	5200,0	25,0	0,00233	0,99693	0,00186	20,015	-0,00046	1,24172
5200,0	6500,0	16,0	0,00149	0,99842	0,00066	7,096	-0,00083	11,17460
6500,0	42324,0	17,0	0,00158	0,00000	0,00062	6,658	-0,00096	16,06692
		10732,0						

CHI QUADRATO 191,71800
 STAT. KS 0,01436
 STAT. SA 0,00439

Tab. 4 - Adattamento della distribuzione lognormale generalizzata alla distribuzione del reddito Indagine DOXA

CLASSE DI REDDITO ANNUO (migl. L.)		f_i	$\frac{f_i}{n}$	$\frac{F_i}{n}$	p_i^*	$ff = np_i^*$	$p_i^* - \frac{f_i}{n}$	$\frac{(ff - f_i)^2}{ff}$
70,0	130,0	305,0	0,02842	0,02842	0,03795	407,238	0,00953	25,66704
130,0	260,0	1704,0	0,15878	0,18720	0,17864	1917,150	0,01986	23,69821
260,0	390,0	2479,0	0,23099	0,41819	0,20262	2174,500	-0,02837	42,63984
390,0	520,0	1906,0	0,17760	0,59579	0,16298	1749,075	-0,01462	14,07908
520,0	650,0	1441,0	0,13427	0,73006	0,11858	1272,595	-0,01569	22,28535
650,0	780,0	845,0	0,07874	0,80880	0,08369	898,195	0,00496	3,15047
780,0	910,0	566,0	0,05274	0,86154	0,05876	630,657	0,00602	6,62877
910,0	1040,0	357,0	0,03327	0,89480	0,04147	445,014	0,00820	17,40710
1040,0	1170,0	212,0	0,01975	0,91455	0,02953	316,919	0,00978	34,73448
1170,0	1300,0	269,0	0,02507	0,93962	0,02126	228,153	-0,00381	7,31290
1300,0	1625,0	229,0	0,02134	0,96096	0,03142	337,186	0,01008	34,71165
1625,0	1950,0	158,0	0,01472	0,97568	0,01514	162,503	0,00042	0,12480
1950,0	2275,0	63,0	0,00587	0,98155	0,00775	83,123	0,00188	4,87156
2275,0	2600,0	62,0	0,00578	0,98733	0,00417	44,741	-0,00161	6,65776
2600,0	2925,0	21,0	0,00196	0,98928	0,00234	25,153	0,00039	0,68562
2925,0	3250,0	21,0	0,00196	0,99124	0,00137	14,680	-0,00059	2,72127
3250,0	3575,0	16,0	0,00149	0,99273	0,00082	8,850	-0,00067	5,77667
3575,0	3900,0	20,0	0,00186	0,99460	0,00051	5,489	-0,00135	38,36014
3900,0	5200,0	25,0	0,00233	0,99693	0,00077	8,285	-0,00156	33,71967
5200,0	6500,0	16,0	0,00149	0,99842	0,00017	1,776	-0,00133	113,95050
6500,0	42324,0	17,0	0,00158	0,00000	0,00007	0,718	-0,00152	369,41732
		10732,0						

CHI QUADRATO 808,60020
 STAT. KS 0,02939
 STAT. SA 0,00677

Nel caso della lognormale gli scarti mantengono lo stesso segno persino per quattro classi contigue anche al centro della distribuzione e questo fa pensare che tale distribuzione interpoli molto bene i dati effettivi.

Si è poi calcolato lo scostamento medio assoluto fra probabilità teoriche e frequenze relative effettive, ossia

$$SA = \frac{1}{K} \sum_{i=1}^K \left| p_i^* - \frac{f_i}{n} \right|$$

Nel caso in esame è risultato che la media semplice degli scostamenti fra probabilità teoriche ed effettive è pari a 0,00439 nella lognormale generalizzata mentre è pari a 0,00677 nella lognormale semplice.

b) Si è calcolata la statistica del tipo di Kolmogorov-Smirnov

$$KS = \max | F_i - F_i^* |$$

dove F_i e F_i^* sono rispettivamente le frequenze cumulate effettive e teoriche calcolate in corrispondenza del limite superiore dell' i -ma classe.

Tale statistica di Kolmogorov-Smirnov è pari a 0,01436 per la distribuzione lognormale generalizzata e pari a 0,02939 nella lognormale. I valori critici del test di adattamento di Kolmogorov-Smirnov risultano, in questo esempio, pari a 0,0131 se

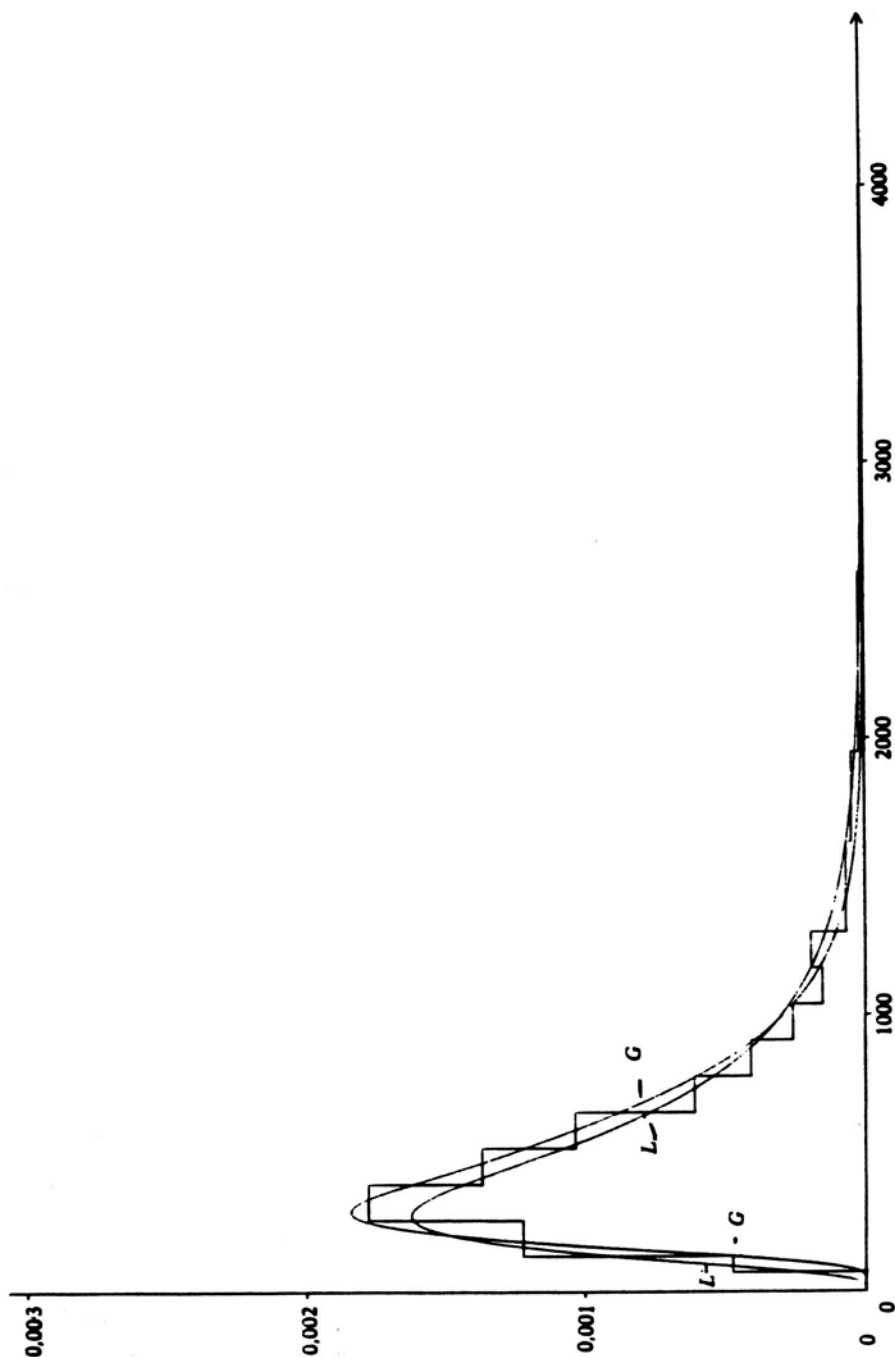


Fig. 7 - Istogramma di frequenze delle distribuzioni dei redditi dell'indagine Doxa, funzione di densità della v.c. lognormale generalizzata (G) con $\alpha = 0,4476$, funzione di densità della v.c. lognormale (L)

$\alpha = 0,05$ ed a $0,0157$ se $\alpha = 0,01$. Pertanto se si fissa $\alpha = 0,05$, si rifiuta che la distribuzione in esame sia lognormale generalizzata mentre, se si fissa $\alpha = 0,01$, si accetta che la distribuzione sia lognormale generalizzata. L'ipotesi di distribuzione lognormale viene invece rifiutata ad entrambi i livelli di significatività.

c) Ogni percettore di reddito può rientrare in una delle k classi di reddito e ciò può essere visto come un insieme di k eventi mutualmente escludentisi ed esaustivi. Pertanto la distribuzione del reddito in esame può essere vista come il risultato di n prove di una v.c. multinomiale avente probabilità p_1, p_2, \dots, p_k in cui si è ottenuto f_1 eventi nella prima classe, f_2 eventi nella seconda classe, ..., f_k eventi nella k -ma.

Si sono allora costruiti gli intervalli di confidenza simultanei con grado di fiducia pari ad $(1 - \alpha)$ per ogni probabilità p_i relativa alle k classi al fine di verificare se ogni intervallo contiene la probabilità teorica p_i della v.c. lognormale generalizzata. Se questo avviene si può accettare l'ipotesi che la distribuzione del reddito segua la suddetta legge con probabilità di errore di 1^a specie pari ad α .

A tal fine ci si è serviti dello studio di Zenga - Fedrizzi (1983) sulla distribuzione asintotica del massimo delle frequenze standardizzate assolute di una multinomiale. Secondo tali autori per n grande e per $k > 7$, l' $(1 - \alpha)$ -mo quantile della v.c.

$$Z = \max \left\{ \frac{|n_i - np_i|}{\sqrt{np_i(1 - p_i)}}; i = 1, 2, \dots, k \right\}$$

può esser ben approssimato da

$$z_{k, 1-\frac{\alpha}{k}}^{\pm} = \Phi^{-1} \left(\frac{1 + (1 - \alpha)^{1/k}}{2} \right)$$

dove $\Phi^{-1}(\cdot)$ è l'inverso della F.R. della v.c. normale standardizzata.

In base a ciò gli estremi degli intervalli di confidenza sono dati da

$$l_i = p_i \pm z_{k, 1-\frac{\alpha}{k}}^{\pm} \sqrt{\frac{p_i(1 - p_i)}{n}}$$

Poiché in (5.4) non si conosce purtroppo $p_i(1 - p_i)$, tale quantità viene sostituita da $\frac{f_i}{n} \left(1 - \frac{f_i}{n}\right)$.

Nell'esempio in oggetto, ponendo $(1 - \alpha)$ pari al 90%, al 95% ed al 99% rimangono escluse, in ogni caso, sei probabilità teoriche dal rispettivo intervallo di confidenza.

Per la lognormale semplice, rimangono escluse nove probabilità ponendo $(1 - \alpha) = 0,99$, dieci probabilità ponendo il grado di fiducia pari al 95% ed undici probabilità ponendo il grado di fiducia pari al 90%.

Il risultato ottenuto, pur confermando un maggior adattamento della lognormale generalizzata rispetto alla lognormale semplice, non porta — anche con questa procedura — ad accettare l'ipotesi che la distribuzione dei redditi dell'indagine Doxa segua una lognormale generalizzata.

d) Ci si voglia ora servire del test chi-quadrato per verificare l'ipotesi H_0 che la distribuzione dei redditi in esame segua una v.c. lognormale generalizzata.

La statistica X^2 prende valore 191,718. Tale valore di X^2 porta a rifiutare H_0 ad uno qualsiasi dei consueti valori dell'errore di 1^a specie. La causa di ciò potrebbe essere attribuita al fatto che la numerosità del campione è elevata ed in questi casi,

essendo il test chi-quadrato consistente, la pur minima discrepanza dal modello ipotizzato porta a valori di X^2 elevati. Si è allora, pensato di costruire un test con errori proporzionali (Zenga (1978a), (1978b)) ossia tale che il rapporto fra la probabilità dell'errore di 2^a specie β e la probabilità dell'errore di 1^a specie α sia un valore ν prefissato. A tal fine si è fissata l'ipotesi alternativa H_1 in modo tale che la media quadratica ponderata η delle differenze assolute relative ξ fra probabilità p_i e probabilità p_i^* dell'ipotesi di distribuzione lognormale generalizzata sia almeno uguale a 0,05. In altre parole, sotto H_1 , dev'essere

$$\eta = \sqrt{\sum_{i=1}^k \xi_i^2 p_i} \geq 0,05$$

dove

$$\xi_i = \frac{|p_i - p_i^*|}{p_i}$$

Sotto H_0 la statistica X^2 tende a distribuirsi come una v.c. chi-quadrato con $(k-r-1)$ gradi di libertà (g.l.) dove r è il numero dei parametri stimati nella distribuzione ipotizzata.

Sotto H_1 tale statistica tende ad una v.c. chi-quadrato non centrale con lo stesso numero di g.l. e parametro di non centralità dato da $\lambda^2 = n \eta^2$ (Zenga (1978a), (1978b)).

Relativamente all'esempio in esame si è calcolato il rapporto

$$\frac{\beta(c)}{\alpha(c)} = \nu(c)$$

in funzione del valor critico c , essendo

$$\alpha(c) = P[X^2 \geq c/H_0]$$

(dove X^2 è una v.c. chi-quadrato con $k-r-1 = 21 - 3 - 1$ g.l.)

$$\beta(c) = \max P[X^2 < c/H_1]$$

(dove la v.c. X^2 è una v.c. chi-quadrato non centrale con 17 g.l. e parametro di non centralità $\lambda^2 = 191,718$).

Per il calcolo delle funzioni $\alpha(c)$ e $\beta(c)$ si è messo a punto un programma con il calcolatore utilizzando le subroutines della F.R. della v.c. chi-quadrato e della v.c. chi-quadrato non centrale.

Come si vede nella fig. 8, l'andamento di $\nu(c)$ è molto ripido. Infatti, per valori di ν compresi fra 0,4 e 7 (intervallo più che soddisfacente per tutte le esigenze pratiche) il valore critico c varia fra 25 e 31. Si ha cioè un andamento del rapporto $\nu(c)$ prossimo a quelli trovati asintoticamente da Zenga (1978b) per altri test statistici.

Essendo il valore sperimentale di X^2 di gran lunga più grande dei valori critici compresi fra 25 e 31, si osserva come anche questa particolare procedura porta al rifiuto dell'ipotesi H_0 . Si noti che con il calcolatore non si è riusciti a valutare il rapporto $\nu(c)$ in corrispondenza di $c = 191,718$ perché la subroutine del chi-quadrato non fornisce il valore della F.R. per valori così elevati.

Tuttavia, per la monotonia di $\nu(c)$, esso deve essere molto più elevato di 7. Se l'ipotesi H_1 prevede che le differenze assolute relative abbiano una media quadratica superiore al 10%, il valore di c varia fra 51 e 53 per valori di ν compresi fra 0,5

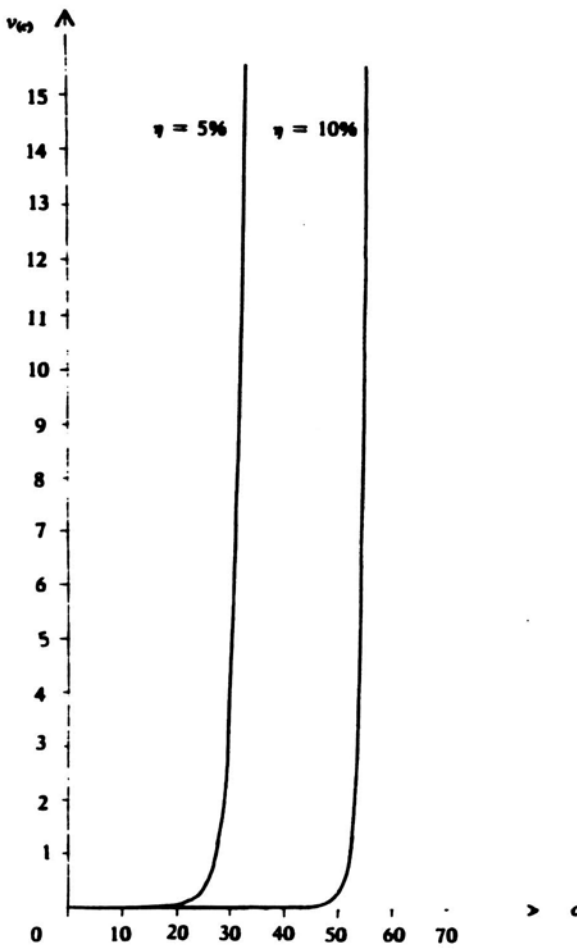


Fig. 8 - Valori del rapporto fra probabilità dell'errore di II specie e I specie in funzione del valore critico

e 8. Anche in questo caso il valore di X^2 sperimentale è considerevolmente più elevato di 53 e quindi, anche avendo aumentato la distanza minima di H_1 da H_0 , bisogna rifiutare l'ipotesi H_0 .

Ovviamente anche l'ipotesi H_0 che la distribuzione in esame è una lognormale verrebbe rifiutata.

La conclusione di quanto detto ai punti a), b), c) e d) è che, sebbene la distribuzione lognormale generalizzata si accosti decisamente meglio alla distribuzione dei redditi dell'indagine DOXA rispetto alla lognormale semplice, nemmeno essa può essere usata per sintetizzare tale andamento.

Trento, Università degli Studi

ANTONIO BRUNAZZO

ANGIOLA POLLASTRI

RIFERIMENTI BIBLIOGRAFICI

- ATCHEISON J. - BROWN J.A.C (1957), *The Lognormal Distribution*, London, Cambridge University Press.
- BRUNAZZO A. (1983), *Alcune considerazioni riguardanti lo scarto medio assoluto dalla media e dalla mediana*, Libera Università di Trento, «Quaderni di statistica e matematica applicata alle scienze economico-sociali», vol. VI, n. 1-2.
- CANAL L. (1981), *Sulla distribuzione normale generalizzata*, «Quaderni di statistica e matematica applicata alle scienze economico-sociali», vol. IV, n. 1-2, pp. 97-107.
- GRADSHTEYN I.S. - RYZHIK I.M. (1980), *Table of Integrals, Series and Products*, New York, Academic Press.
- JOHNSON N.L. - KOTZ S. (1970), *Continuous Univariate Distributions*, Boston, Houghton Mifflin Company.
- LETI G. (1983), *Statistica descrittiva*, Bologna, il Mulino.
- LUNETTA G. (1963), *Di una generalizzazione dello schema della curva normale*, «Annali della Facoltà di economia e commercio di Palermo», A. XVII, n. 2, pp. 237-244.
- (1967), *Le variabili normali di ordine*, in *Atti della XXV Riunione scientifica della Società Italiana di Statistica*, Bologna, pp. 249-258.
- LUZZATTO FEGEZ P. (1956), *Il volto sconosciuto dell'Italia*, vol. II, Milano, Giuffrè.
- MINEO A. (1980), *La stima dei parametri delle curve normali di ordine r per intervalli infiniti e di quelle di intervalli finiti*, in *Studi in onore di Paolo Fortunati*, vol. I, Bologna, CLEUP.
- UPPULURI V.R.R. (1981), *Some Properties of the Log-Laplace Distributions*, in C. TAILLER *et al.* (eds), *Statistical Distributions in Scientific Work*, vol. IV, pp. 105-110, Dordrecht, Reidel.
- VIANELLI S. (1963), *La misura della variabilità condizionata in uno schema generale delle curve normali di frequenza*, «Statistica», a. XXIII, n. 4, ottobre dicembre, pp. 447-474.
- - MINEO A. (1978), *Prontuari delle probabilità integrali di ordine r* , Palermo, Edigrafia Sud Europa.
- ZENGA M. (1975), *I tests con errori proporzionali*, «Statistica», a. XXXV, n. 4, ottobre-dicembre, pp. 731-750.
- (1977), *Soglie critiche asintotiche di alcuni tests con errori proporzionali*, «Saggi di statistica metodologica ed applicata», Contributi del Centro interdipartimentale di studi statistici dell'Università degli Studi della Calabria, vol. I, pp. 111-123.
- (1978), *Suggerimenti del test Chi-Quadrato*, «Quaderni di statistica e matematica applicata alle scienze economico-sociali», vol. I, n. 1, pp. 5-16.
- (1978), *La distanza delle alternative nelle verifiche di ipotesi riguardanti le distribuzioni multinomiali*, «Quaderni di statistica e matematica applicata alle scienze economico-sociali», vol. I, n. 2, pp. 3-23.
- - FEDRIZZI M. (1983), *The Asymptotic Distributions of the Maximum Absolute Standardized Frequencies of the Multinomial Distribution*, «Metron», Vol. XLI, n. 3-4, pp. 31-48.