Statistica Applicata Vol. 10, n. 1, 1998

DISTRIBUZIONI ASINTOTICHE E SIMULATE DI UNO STIMATORE DEL QUANTILE DI ORDINE p (*)

Filippo Domma

Dipartimento di Economia Politica, Università degli Studi della Calabria.

Riassunto

Nell'ambito degli studi sulla distribuzione personale del reddito, un'importante caratteristica è senz'altro il quantile di ordine p. Come è noto, il quantile campionario pur avendo distribuzione asintoticamente normale, presenta una varianza che dipende dalla funzione di densità valutata nel quantile di popolazione; quindi, essendo i parametri, generalmente, incogniti la varianza asintotica resta sconosciuta. Al fine di superare tale ostacolo, si utilizzano le proprietà asintotiche degli stimatori di massima verosimiglianza e il teorema d–Method. Così, sotto l'ipotesi che la distribuzione personale del reddito possa essere adattata dal modello di Dagum a tre parametri, è stata determinata la distribuzione asintotica dello stimatore del quantile di ordine p. Infine, tramite la simulazione dell'universo dei campioni, si sono verificate le proprietà asintotiche di tale stimatore per dimensioni campionarie finite.

1. INTRODUZIONE

In questa nota, sotto l'ipotesi che la distribuzione personale del reddito sia rappresentata da una variabile casuale (v.c.) X, continua e non–negativa, con funzione di densità (fd) $f(x;\theta)$ appartenente ad una famiglia di distribuzioni $P = \{f(.;\theta):\theta \in \Theta \subset \Re^r,\ r \geq 1\}$, dove Θ è lo spazio parametrico, studieremo la distribuzione asintotica di uno stimatore del quantile di ordine p. Quest'ultimo, indicato con ξ_p, ci fornisce il valore del reddito tale che la quota dei redditieri con redditi inferiorei o uguali a ξ_p sia esattamente p, con $p \in (0,1)$.

In letteratura [si veda, ad esempio, David (1970) pp. 201; Arnold *et al.* (1992), pp. 223], è noto che il p–esimo quantile campionario ha distribuzione asintoticamente normale con media asintotica pari al quantile di popolazione e varianza asintotica che dipende dalla funzione di densità valutata nel quantile di popolazione. Poiché quest'ultimo dipende dai parametri incogniti della popolazione la varianza asintotica resta sconosciuta.

(*) *Lavoro parzialmente finanziato dal C.N.R., contributo n. 96.01568.CT10.*

Una soluzione alternativa è quella di far riferimento alle proprietà asintotiche degli stimatori di massima verosimiglianza (m.v.); tale approccio è stato utilizzato in diversi lavori da Latorre (1987, 1988) per determinare la distribuzione asintotica di alcuni indici di concentrazione sotto diversi modelli per la distribuzione dei redditi, da cui vengono ripresi alcuni risultati. Lo stesso approccio risulta presente anche in Hosking e Wallis (1987) che forniscono la stima dei parametri e dei quantili della Pareto generalizzata, proposta da Pickands nel 1975, utilizzata per l'analisi degli eventi estremi, nonché in Grimshaw (1993) che della stessa distribuzione propone un algoritmo per il calcolo degli stimatori di m.v. dei parametri.

2. ALCUNE DEFINIZIONI

Sia X una v.c. continua e non negativa, con fd $f(x;\theta)$, $\in$ P e funzione di ripartizione (fr) $F(x;\theta)$, si dice che ξ_p (θ) è il quantile di ordine p se sussiste la relazione:

$$F(\xi_p;\theta) = p \qquad p \in (0,1) \tag{2.1}$$

Sotto l'ipotesi che la f.r. sia strettamente crescente la soluzione della (2.1) rispetto a ξ_p è unica e pari a

$$\xi_p(\theta) = F^{-1}(p) \tag{2.2}$$

Tra la vasta gamma di proposte presenti in letteratura riguardo i modelli adottati per la interpretazione della distribuzione dei redditi, negli ultimi anni particolare interesse ha suscitato il modello a tre parametri di Dagum (1977, 1980), il quale presenta una fd data da:

$$f(x;\theta) = \beta\lambda\delta x^{-(\delta+1)}(1+\lambda x^{-\delta})^{-(\beta+1)} \qquad (x>0) \tag{2.3}$$

con $\theta' = (\beta,\lambda,\delta)$, $\beta > 0$, $\lambda > 0$ e $\delta > 0$. Il vincolo $\delta\beta > 1$ garantisce la unimodalità di detta densità, mentre l'esistenza del momento di ordine j richiede che $\delta > j$. La fr è data da:

$$F(x;\theta) = (1+\lambda x^{-\delta})^{-\beta} \tag{2.4}$$

dalla quale è immediato determinare il quantile di ordine p

$$\xi_p(\theta) = \lambda^{1/\delta}(p^{-1/\beta}-1)^{-1/\delta} \qquad p \in (0,1) \tag{2.5}$$

che sarà oggetto di stima.

3. METODOLOGIA STATISTICA

La metodologia statistica utilizzata per la determinazione della distribuzione asintotica degli stimatori dei quantili di popolazione, si basa sulle proprietà degli stimatori di m.v., alcune delle quali vengono qui riportate per comodità di lettura.

Sia X una v.c. con fd $f(x;\theta) \in P = \{f(.;\theta):\theta \in \Theta \subset \Re^r, r \geq 1\}$, di forma nota e vettore dei parametri incognito. Sia $X_1, X_2, \ldots, X_n$ un campione casuale (c.c.), le cui osservazioni sono v.c. i.i.d., lo stimatore di m.v. di θ è ottenuto massimizzando rispetto a θ la funzione di verosimiglianza (fv) data da:

$$L(\theta;\mathbf{x}) = \prod_{i=1}^{n} f(x_i;\theta) \tag{3.1}$$

rispetto a θ, dove $\mathbf{x}' = (x_1, x_2, \ldots, x_n)$ è il campione osservato. Quindi, $\hat{\theta}_n$ è lo stimatore di m.v. di θ se

$$L(\hat{\theta}_n;\mathbf{x}) = \sup_{\theta \in \Theta} L(\theta;\mathbf{x}) \tag{3.2}$$

Si dimostra [ad esempio: Cox–Hinkley (1974), pp. 295; Landenna–Marasini (1992), pp. 347] che, sotto le consuete condizioni di regolarità, lo stimatore di m.v. $\hat{\vartheta}_n$ di θ è tale che:

$$\sqrt{n}(\hat{\theta}_n - \theta) \underset{n\to\infty}{\overset{d}{\to}} N(\mathbf{0}, \Sigma_\theta) \tag{3.3}$$

cioè ha distribuzione asintotica normale multivariata con matrice varianze e covarianze data da $\Sigma_\theta = n[I(\theta)]^{-1}$ dove $I(\theta)$ è la mastrice di informazione attesa di Fisher circa θ definita da:

$$I(\theta) = -E\left\{\frac{\partial^2 \ln L}{\partial\theta_i \partial\theta_j}\right\} \quad i,j = 1,2,\ldots,r \tag{3.4}$$

Poiché siamo interessati alla distribuzione asintotica di una funzione del vettore dei parametri, si può far riferimento al "Principio d'Invarianza" degli stimatori di m.v. [Zehna, 1966], secondo il quale se $\tau = \tau(\theta)$ è una funzione reale del vettore dei parametri, allora lo stimatore della funzione τ di θ è dato da $\hat{\tau}_n = \tau(\hat{\theta}_n)$. Inoltre, dalla (3.3) ed ipotizzando che $\tau = \tau(\theta)$ sia una funzione con derivata prima continua in θ e $\tau'(\theta) \neq 0$, per il teorema δ–Method [Rao (1973), pp. 388; Landenna–Marasini (1992), pp. 60] la distribuzione asintotica dello stimatore $\hat{\tau}_n$ di τ, è.

$$\sqrt{n}\,\frac{\hat{\tau}_n - \tau}{\gamma(\hat{\tau}_n;\theta)} \underset{n\to\infty}{\overset{d}{\to}} N(0,1) \tag{3.5}$$

dove la varianza asintotica di $\hat{\tau}_n$ è data da:

$$v^2(\hat{\tau}_n;\theta) = \mathbf{t}'\,\Sigma_\theta \mathbf{t} \tag{3.6}$$

con

$$\mathbf{t}' = \left[\frac{\partial\tau(\theta)}{\partial\theta}\right]$$

È importante osservare che $\hat{\tau}_n = \tau(\hat{\theta}_n)$ risulta essere uno stimatore di m.v. se la funzione τ è invertibile mentre, se detta condizione non sussiste allora, come osservato da Berk (1967), non sembrerebbe giustificata l'attribuzione a $\hat{\tau}_n = \tau(\hat{\theta}_n)$ della denominazione di "stimatore di massima verosimiglianza", ma più correttamente quella di stimatore basato sul profilo di verosimiglianza [si veda, ad esempio, Barndorff–Nielsen e Cox (1994), pag. 89] restando in ogni caso valide le conclusioni tratte dalla (3.5).

In altri termini, per n sufficientemente grande, lo stimatore $\hat{\tau}_n = \tau(\hat{\theta}_n)$ di $\tau = \tau(\theta)$ risulta essere approssimativamente normale con media $\tau(\theta)$ e varianza $\frac{1}{n}v^2(\hat{\tau}_n;\theta)$; quest'ultima consistentemente stimata dalla seguente $\frac{1}{n}v^2(\hat{\tau}_n;\hat{\theta}_n)$.

In definitiva, per quanto suddetto possiamo calcolare intervalli di confidenza e/o effettuare prove di ipotesi su $\tau(\theta)$ avendo a disposizione un c.c., i.i.d., estratto da $f(x;\theta)$.

4. DISTRIBUZIONE ASINTOTICA DI UNO STIMATORE DI ξ_p

La metodologia statistica appena illustrata verrà impiegata per determinare la distribuzione asintotica dello stimatore del quantile di ordine p dato dalla (2.5). Da quanto esposto nel precedente paragrafo è evidente che, per raggiungere tale obiettivo, è necessario determinare sia gli stimatori di m.v. del vettore parametrico θ che la matrice di informazione di Fisher. Per il modello di Dagum, questi elementi sono già noti in letteratura (si veda Latorre, 1988); in particolare, si evidenzia che la soluzione del sistema, composto dalle equazioni di verosimiglianza

$$\begin{cases} \dfrac{\partial l(\theta;\mathbf{x})}{\partial\beta} = \dfrac{n}{\beta} - \displaystyle\sum_{i=1}^{n}\left(1+\lambda x_i^{-\delta}\right) = 0 \\ \dfrac{\partial l(\theta;\mathbf{x})}{\partial\lambda} = \dfrac{n}{\lambda} - (\beta+1)\displaystyle\sum_{i=1}^{n}\dfrac{x_i^{-\delta}}{1+\lambda x_i^{-\delta}} = 0 \\ \dfrac{\partial\lambda(\theta;\mathbf{x})}{\partial\delta} = \dfrac{n}{\delta} - \displaystyle\sum_{i=1}^{n}\ln x_i + \lambda(\beta+1)\displaystyle\sum_{i=1}^{n}\dfrac{x_i^{-\delta}\ln x_i}{1+\lambda x_i^{-\delta}} = 0 \end{cases} \tag{4.1}$$

(Latorre, 1988) non è esplicita, quindi il vettore delle stime di m.v. $\hat{\theta}'_n = (\hat{\beta}_n, \hat{\lambda}_n, \hat{\delta}_n)$ può essere ottenuto solo per via numerica. Nel citato lavoro, vengono determinati gli elementi della matrice di informazione di Fisher (che qui non riportiamo per non appesantire il presente lavoro), l'inversa della quale ci fornisce la varianza asintotica dello stimatore di m.v. $\hat{\theta}'_n = (\hat{\beta}_n, \hat{\lambda}_n, \hat{\delta}_n)$. Sostituendo quest'ultimo nella (2.5) otteniamo lo stimatore del quantile di popolazione di ordine p, cioè

$$\hat{\xi}_{p,n} = \xi_p(\hat{\theta}_n) = \hat{\lambda}^{1/\hat{\delta}}\left(p^{-1/\hat{\beta}} - 1\right)^{-1/\hat{\delta}} \qquad p \in (0,1) \tag{4.2}$$

Per quanto detto nel paragrafo 3, per n sufficientemente grande, la distribuzione campionaria è approssimabile da:

$$\hat{\xi}_{p,n} \sim N\left(\xi_p; v^2\left(\hat{\xi}_{p,n}; \hat{\theta}_n\right)\right) \tag{4.3}$$

con

$$v^2\left(\hat{\xi}_{p,n}; \hat{\theta}_n\right) = \left[\frac{\partial \xi_p}{\partial \theta}\right]'_{\theta=\hat{\theta}} \left[\mathbf{I}\left(\hat{\theta}_n\right)\right]^{-1} \left[\frac{\partial \xi_p}{\partial \theta}\right]_{\theta=\hat{\theta}} \tag{4.4}$$

dove gli elementi del vettore gradiente sono dati da:

$$\left[\frac{\partial \xi_p}{\partial \beta}\right]_{\theta=\hat{\theta}} = \frac{-\ln(p)}{\hat{\delta}\hat{\beta}^2\left(1 - p^{1/\hat{\beta}}\right)} \hat{\xi}_{p,n} \tag{4.5}$$

$$\left[\frac{\partial \xi_p}{\partial \lambda}\right]_{\theta=\hat{\theta}} = \frac{1}{\hat{\delta}\hat{\lambda}} \hat{\xi}_{p,n} \tag{4.6}$$

$$\left[\frac{\partial \xi_p}{\partial \delta}\right]_{\theta=\hat{\theta}} = \frac{-1}{\hat{\delta}^2} \hat{\xi}_{p,n} \ln\left(\frac{\hat{\lambda} p^{1/\hat{\beta}}}{1 - p^{1/\hat{\beta}}}\right) \tag{4.7}$$

In conclusione, sulla base della (4.3) è possibile effettuare inferenza statistica sul quantile di popolazione di ordine p, sotto l'ipotesi che la distribuzione personale del reddito segua il modello di Dagum.

5. SIMULAZIONE

In questo paragrafo vengono esposti i risultati di una simulazione volta a verificare per quale dimensione campionaria finita i risultati asintotici, ottenuti per il modello di Dagum, siano ancora validi e se, tali dimensioni siano dell'ordine di grandezza pari a quelle delle indagini reali. A tal fine, si è utilizzata la metodologia della simulazione dell'universo dei campioni; si rinvia, ad esempio, a Latorre (1990) per una descrizione dettagliata.

Detto che l'analisi è stata effettuata sui quantili di ordine 0.1, 0.5 e 0.9, ma che verranno presentati solo i risultati relativi al quantile di ordine 0.5 poiché le conclusioni per i tre quantili studiati sono sostanzialmente coincidenti, si evidenzia che uno dei punti più delicati di tale metodologia è rappresentato dalla scelta dei valori teorici dei parametri β,λ,δ. Infatti, poiché i quantili sono funzioni di $\theta=(\beta,\lambda,\delta)$ allora la scelta dovrà ricadere su valori dei parametri tali che i corrispondenti valori dei quantili siano realistici.

Alla luce delle restrizioni sui parametri, $\beta\delta > 1$ per la unimodalità della (2.3) e $\delta > 2$ per l'esistenza del momento secondo, ed evitando inutili duplicazioni che non aggiungerebbero ulteriori informazioni all'analisi, si è deciso di scegliere i seguenti valori teorici dei parametri :

$$\beta = (0.66,0.99,1.29)\ ;\ \delta= (2.1,2.86,3.5)\ ;\ \lambda = (1000,3020,5000)$$

Fissata una combinazione dei parametri teorici, ovvero un elemento della famiglia P, la metodologia suddetta richiede l'estrazione di N campioni indipendenti di dimensione n dalla densità individuata. Per la scelta del numero N di campioni, si è tenuto conto di due esigenze contrapposte. Da un lato, N deve essere molto elevato in modo tale da essere rappresentativo dell'universo dei campioni che, nel caso di popolazioni continue, contiene una infinità continua di elementi. Dall'altro, si è tentato di contenere entro limiti accettabili i tempi di elaborazione, tenuto conto che questi crescono a dismisura in quanto, per ognuno degli N campioni, si deve far ricorso a metodi iterativi per la soluzione del sistema (4.1) che fornisce le stime di m.v. di θ. Sulla base di queste considerazioni, si è deciso di porre N=2000. Alle dimensioni campionarie n sono stati assegnati i seguenti valori crescenti : 500, 1000, 1500 e 2000. Si evidenzia che l'ultima dimensione campionaria è di gran lunga inferiore a quelle impiegate nelle indagini reali. Utilizzando la trasformazione integrale di probabilità (Mood *et al.*, pag. 202, 1974) e la *subroutine* RNUN della libreria IMSL (1987), la quale genera numeri casuali nell'intervallo (0,1), per ogni combinazione dei parametri teorici, è stata simulata l'estrazione di N campioni di dimensione n dall'elemento della famiglia P. Per ognuno di questi sono state determinate le stime di m.v. di θ e, conseguentemente, sulla base della (4.2), le stime dei quantili di ordine 0.1, 0.5 e 0.9. Il risultato finale delle simulazioni consiste in (3x3x3)x4=108 insiemi di 2000 osservazioni per ogni quantile considerato.

Osservazione:
le analisi che verranno svolte si basano sull'ipotesi che l'universo simulato sia abbastanza grande da essere rappresentativo dell'universo dei campioni. Inoltre, poiché non si sono riscontrate significative differenze al variare del parametro di

scala λ, verranno presentati solo i risultati relativi a combinazioni di parametri teorici con $\lambda = 3020$.

Indicando con $\hat{\xi}_{0.5}(\hat{\theta}_n)$ lo stimatore del quantile di ordine 0.5, dalla (4.3) si evince che asintoticamente si ha: $E\left\{\hat{\xi}_{0.5}(\hat{\theta}_n)\right\} = \xi_{0.5}(\theta)$. Sotto l'*osservazione* è sufficiente calcolare, per ogni combinazione di parametri, la media $M\left\{\hat{\xi}_{0.5}(\hat{\theta}_n)\right\}$ sull'universo finito degli N = 2000 campioni e verificare se questa è uguale al quantile teorico $\xi_{0.5}(\theta)$, cioè al quantile che si otterrebbe se fosse vera la combinazione di parametri teorici scelta.

Per ogni combinazione dei parametri teorici, dalle colonne (A) e (B) della tabella I che riportano, rispettivamente, la mediana teorica $\xi_{0.5}$ e la media $M\left\{\hat{\xi}_{0.5}(\hat{\theta}_n)\right\}$, si può constatare che a partire da dimensioni campionarie prossime a 500, le differenze tra le due quantità in esame sono significative dalla quarta (in alcuni casi dalla terza) cifra decimale; inoltre, tali differenze decrescono all'aumentare di n e, comunque, risultano inferiori all'1 per mille. Per avere indicazioni sulla precisione della media $M\left\{\hat{\xi}_{0.5}(\hat{\theta}_n)\right\}$, nella colonna (C) della stessa tabella, è stata riportata la varianza $V^2\left\{\hat{\xi}_{0.5}(\hat{\theta}_n)\right\}$ calcolata sulle 2000 repliche. È immediato verificare che detta varianza decresce rapidamente all'aumentare della dimensione campionaria.

Da quanto detto nei paragrafi precedenti, la varianza dello stimatore del quantile di ordine 0.5, $v^2\left[\hat{\xi}_{0.5};\theta\right]$, viene stimata sostituendo al vettore dei parametri incogniti θ le rispettive stime di m.v., ottenendo $v^2\left[\hat{\xi}_p;\hat{\theta}_n\right]$ il quale risulta essere asintoticamente non–distorto, cioè $E\left\{v^2\left[\hat{\xi}_{0.5};\hat{\theta}_n\right]\right\} = v^2\left[\hat{\xi}_{0.5};\theta\right]$ quando n tende ad infinito. Quest'ultima relazione, analogamente a quanto fatto in precedenza, può essere verificata calcolando la media $M\left[v^2\left\{\hat{\xi}_{0.5};\hat{\theta}_n\right\}\right]$ sugli N campioni dell'universo campionario finito e valutando se essa è uguale alla varianza teorica $v^2\left[\hat{\xi}_{0.5};\theta\right]$, cioè quella che si otterrebbe se fosse vera la combinazione dei parametri teorici di scelta. I risultati delle simulazioni sono sintetizzati nelle colonne (D) ed (E) della tabella I. Anche in tal caso si evidenzia che, a partire da dimensioni campionarie pari a 500, le differenze sono significative dalla terza cifra decimale, diminuendo rapidamente all'aumentare della dimensione campionaria. Per avere

Tab. I: Risultati delle simulazioni per il quantile di ordine 0.5. (λ = 3020).

n	δ	β	(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)
500	2.1	0.66	33.806254	33.814857	2.2636969	2.2657647	2.2637705	0.0604474	5.450	2.8150331
		0.99	45.108701	45.118275	2.8546523	2.9470048	2.9519940	0.0953064	4.4	3.7141736
		1.29	53.402592	53.355889	3.6965742	3.5113985	3.4930202	0.1385430	5.7	4.4977755
	2.86	0.66	13.264477	13.277878	0.1955484	0.1880648	0.1877249	0.0003144	5.1	0.2336556
		0.99	16.393311	16.393629	0.2029664	0.2098457	0.2099944	0.0003714	4.35	0.2644732
		1.29	18.556233			0.2285809				0.2927920
	3.5	0.66	8.2679574	8.2623751	0.0512911	0.0487887	0.0484889	0.0000177	5.65	0.0606161
		0.99	9.8300605	9.8234807	0.0530246	0.0503819	0.0500849	0.0000192	5.65	0.0634975
		1.29	10.877708	10.875908	0.0528439	0.0524485	0.0524606	0.0000210	5.9	0.0671816
1000	2.1	0.66	33.806254	33.795428	1.1455558	1.1328823	1.1293228	0.0076947	4.95	1.4075165
		0.99	45.108701	45.155508	1.5624303	1.4735024	1.4749430	0.0127097	5.8	1.8570881
		1.29	53.402592	53.366425	1.7309041	1.7556992	1.7552084	0.0167071	4.95	2.2488878
	2.86	0.66	13.264477	13.27154	0.0960264	0.0940324	0.0940478	0.0000403	5.75	0.1168278
		0.99	16.393311	16.4045577	0.1112566	0.1049228	0.1049142	0.0000487	5.8	0.1322366
		1.29	18.556233			0.1142905				0.1463960
	3.5	0.66	8.2679574	8.2637346	0.0242686	0.0243944	0.0243863	0.0000021	4.95	0.0303081
		0.99	9.8300605	9.8253167	0.0249144	0.0251909	0.0251835	0.0000023	5.2	0.0317487
		1.29	10.877708	10.874419	0.0261080	0.0262242	0.0261689	0.0000026	5.4	0.0335908
1500	2.1	0.66	33.806254	33.821134	0.7211002	0.7552549	0.7553774	0.0021090	4.25	0.9383443
		0.99	45.100334	45.100334	0.9682364	0.9823349	0.9804045	0.0034296	5.0	1.2380588
		1.29	53.402592	53.427664	1.1830546	1.1704661	1.1729612	0.0054732	4.6	1.4992585
	2.86	0.66	13.264477	13.266882	0.0611492	0.0626883	0.0626148	0.0000118	4.9	0.0778852
		0.99	16.393311	16.390321	0.0689329	0.0699486	0.0697971	0.0000132	5.0	0.0881577
		1.29	18.556233			0.0761937				0.0975973
	3.5	0.66	8.2679574	8.2702001	0.0165108	0.0162629	0.0162808	0.0000007	4.7	0.0202054
		0.99	9.8300605	9.8324902	0.0170019	0.0167939	0.0168149	0.0000008	5.2	0.0211658
		1.29	10.877708	10,881877	0,0164681	0,0174828	0,0174905	0,0000007	4,5	0,0223939
2000	2,1	066	33,806254	33,824992	0,5811242	0,5664412	0,5661113	0,0009895	5,2	0,7037583
		0,99	45,08701	45,127533	0,7563134	0,7367512	0,7373768	0,0015775	5,35	0,9285441
		1,29	53,402592	53,413464	0,8934840	0,8778496	0,8780452	0,0021002	5,75	1,124439
	2,86	0,66	13,264477	13,263471	0,0458095	0,0470162	0,0470360	0,0000047	4,85	0,0584139
		0,99	16,393311	16,397746	0,0538646	0,0524614	0,0524812	0,0000061	5,35	0,0661183
		1,29	18,556233			0,0571453				0,0731980
	3,5	0,66	8,2679574	8,2693008	0,0123118	0,0121972	0,012932	0,0000003	5,3	0,0151540
		0,99	9,8300605	9,8311531	0,0127831	0,0125955	0,0125922	0,0000003	5,55	0,0158744
		1,29	10,877708	10,875242	0,01265932	0,0131121	0,0130862	0,0000003	5,0	0,0167954

(A) Mediana teorica $\xi_{0.5}(\theta)$; (B) Media di $\hat{\xi}_{0.5}(\hat{\theta}_n)$ sulle 2000 repliche; (C) Varianza di $\hat{\xi}_{0.5}(\hat{\theta}_n)$ sulle 2000 repliche; (D) Varianza asintotica teorica $v^2[\hat{\xi}_{0.5};\theta]$; (E) Media di $v^2[\hat{\xi}_{0.5};\hat{\theta}_n]$ sulle 2000 repliche; (F) Varianza di $v^2[\hat{\xi}_{0.5};\hat{\theta}_n]$ sulle 2000 repliche; (G) % di intervalli di confidenza non–centrati; (H) Varianza asintotica della mediana campionaria $\{4n[f(\xi_{0.5})]^2\}^{-1}$.

un'idea della rilevanza numerica di detta distorsione, si osserva che la massima distorsione relativa percentuale è pari al 6 per mille (n = 500, $\delta = 3.5$ e $\beta = 0.66$).

Per avere delle indicazioni sulla precisione di tali stimatori, è necessario valutare l'ordine di grandezza delle medie degli stimatori delle varianze degli stimatori del quantile di ordine 0.5, tramite un indice di dispersione, ad esempio la varianza $V^2\left\{v^2\left[\hat{\xi}_{0.5};\hat{\theta}_n\right]\right\}$. Tale valutazione può essere effettuata osservando la colonna (F) dalla quale si evince che, per ogni combinazione di parametri, l'ordine di grandezza diminuisce rapidamente all'aumentare della dimensione campionaria.

Conferme sull'adattamento alla distribuzione normale degli stimatori in esame, si sono ottenute effettuando il test di Lilliefors [Conover (1971), pp. 302–302]. Infatti, per tutti i quantili considerati e per ogni combinazione di parametri teorici, si è sempre accettata l'ipotesi di normalità dello stimatore in questione.

L'analisi fin qui svolta ci porta a concludere che le proprietà asintotiche di non–distorsione, efficienza e normalità, evidenziate nei paragrafi precedenti, circa lo stimatore del quantile di ordine p per il modello di Dagum, possono ritenersi valide anche per le dimensioni campionarie finite a partire da n = 500.

Poiché la sola analisi puntuale non chiarisce fino in fondo il comportamento dello stimatore considerato, sono stati costruiti degli intervalli di confidenza per il quantile di ordine 0.5. Fissata una combinazione dei parametri teorici, per ognuno dei N = 2000 campioni di dimensone n, tramite la soluzione del sistema (4.1), è possibile ottenere una stima dei parametri β, λ, δe, quindi, una stima del quantile di ordine 0.5, cioè $\hat{\xi}_{0.5} = \hat{\xi}_{0.5}(\hat{\theta}_n)$, e della relativa varianza $v^2\left\{\hat{\xi}_{0.5};\hat{\theta}_n\right\}$. Così, ipotizzando la normalità dello stimatore $\hat{\xi}_{0.5}$, possiamo calcolare un intervallo di confidenza all'α% per $\xi_{0.5}$, cioè:

$$\xi_{0.5} = \hat{\xi}_{0.5} \pm z_{\alpha/2}\sqrt{v^2\left(\hat{\xi}_{0.5};\hat{\theta}_n\right)}$$

A tal punto è sufficiente verificare che non più dell'α% degli intervalli non–contengono il quantile teorico $\xi_{0.5}(\theta)$, essendo quest'ultimo noto in quanto θ è fissato. Anche in questo caso, si illustrano i risultati relativi al quantile di ordine 0.5, ma le conclusioni sono le medesime per i quantili di ordine 0.1 e 0.9. Nella colonna (G) della tabella I si riporta la percentuale di intervalli non–centrati. Risulta immediato constatare che, a partire da dimensioni campionarie pari a 500, la percentuale di intervalli non–centrati risulta minore del 5%, salvo casi particolari in cui tale percentuale arriva al massimo al 6%.

Infine, nella colonna (F) è stata riportata, per ogni dimensione campionaria e per ogni combinazione dei parametri scelti, la varianza asintotica della mediana campionaria, cioè $\{4n[f(\xi_{0.5})]^2\}^{-1}$. Questa risulta essere sempre maggiore della varianza asintotica $v^2\left\{\hat{\xi}_{0.5};\theta\right\}$, evidenziando una maggiore efficienza dello stimatore $\hat{\xi}_{0.5}(\hat{\theta}_n)$ considerato in questo lavoro.

8. CONCLUSIONI

In questa nota, utilizzando le proprietà asintotiche degli stimatori di m.v. ed il teorema δ–Method, si è determinata la distribuzione asintotica degli stimatori di uno stimatore del quantile di ordine p, sotto l'ipotesi che la distribuzione personale del reddito possa essere adattata dal modello di Dagum a tre parametri.

Le simulazioni effettuate ci portano a concludere che in realtà i risultati asintotici circa lo stimatore studiato, sotto l'ipotesi che la distribuzione dei redditi possa essere adattata dal modello di Dagum a tre parametri, possono essere utilizzati in contesti reali con dimensioni campionarie finite a partire da 500.

RIFERIMENTI BIBLIOGRAFICI

ARNOLD B.C., BALAKRISHNAN N., NAGARAJA H.N., (1992), *A first Course in Order Statistics*. J. Wiley & Sons, Inc., New York.

BARNDORFF–NIELSEN O.E., COX D.R., (1994), *Inference and Asymptotics,* Chapman & Hall, London.

BERK R.H., (1967), "Review of Zehna", *Math. Rev.,* 33, n.1922, pag. 342–343.

CONOVER W.J., (1971), *Practical nonparametric statistics*. J. Wiley & Sons Inc., New York.

COX D.R., HINKLEY D.V., (1974), *Theoretical Statistics,* Chapman and Hall, London.

DAGUM C., (1977), "A new model of personal income distribution: specification and estimation". *Economie Appliquée,* XXX, 3, pp. 413–437.

DAGUM C., (1980), "The generation and distribution of income, the Lorenz curve and the Gini ratio". *Economie Appliquée,* XXXIII, 2, pp. 327–367.

DAVID H.A., (1970), *Order Statistics,* J. Wiley & Sons, New York.

GRIMSHAW S.D., (1993) "Computing maximum likelihood estimates for generalized Pareto distribution". *Technometrics,* 35, 2, pp. 185–191.

HOSKING J.R.M., WALLIS J.R., (1987) "Parameter and quantile estimation for the Pareto distribution". *Technometrics,* 29, 3, pp. 339–349.

IMSL, (1987), *Fortran Subroutines for statistical analysis*. Houston.

LANDENNA G., MARASINI D., (1992), *La teoria della Stima Puntuale,* Cacucci editore, Bari.

LATORRE G., (1987), "Distribuzioni campionarie asintotiche di indici di concentrazione: approccio parametrico". *Statistica,* XLVII, n.4, pp.573–587.

LATORRE G., (1988), "Proprietà campionarie del modello di Dagum per la distribuzione dei redditi". *Statistica,* XLVIII, 1, pp.15–27.

LATORRE G., (1990), "Asymptotic distributions of indices of concentration: empirical verification and applications". In C. Dagum, M. Zenga (Eds.):*Income and Wealth Distribution, Inequality and Poverty.* Springer–Verlag, Berlin, 1990.

MOOD A.M., GRAYBILL F.A., BOES D.C., (1974), *Introduction to the Theory of Statistics.* Third Edition. McGraw–Hill.

RAO C.R., (1973), *Linear Statistical Inference and Its Applications.* J. Wiley & Sons, New York.

ZEHNA P.W., (1966), "Invariance of maximum likelihood estimators" . *Ann. Math. Statist.,* 37, pag.744.

ASYMPTOTIC AND SIMULATED DISTRIBUTIONS OF AN ESTIMATOR OF THE p–ORDER QUANTILE

Summary

In the field of personal income distribution studies, an important characteristic is the p–order quantile of the distribution. Nevertheless, the sample quantile, even if it has an asymptotic normal distribution, shows a variance which depends on the probability density function evaluated on the population quantile; since the parameters are not known, the asymptotic variance remains unknown. To solve this problem, we have used the asymptotic maximum likelihood properties in conjunction with the δ–Method. In this way, assuming that the personal income distribution can be fitted by a Dagum model, we have found the asymptotic distribution of the p–order quantile estimator. Finally, simulating the sample space, we have verified the asymptotic properties of the p–order quantile estimator for finite sample sizes.