Marcello Chiodi: APPROSSIMAZIONI SADDLEPOINT ALLE DISTRIBUZIONI CAMPIONARIE DEGLI STIMATORI DI MASSIMA VEROSIMIGLIANZA DEI PARAMETRI DELLE CURVE NORMALI DI ORDINE P PER PICCOLI CAMPIONI.Atti della XXXVII Riunione Scientifica della SIS, San Remo 6-8 Aprile 1994, ed. CISU, vol.II, pagg. 139-146.

APPROSSIMAZIONI SADDLEPOINT ALLE DISTRIBUZIONI CAMPIONARIE DEGLI STIMATORI DI MASSIMA VEROSIMIGLIANZA DEI PARAMETRI DELLE CURVE NORMALI DI ORDINE P PER PICCOLI CAMPIONI.

Marcello Chiodi

Istituto di Statistica,
Facoltà di Economia e Commercio di Palermo.

1. Introduzione (1)

In questo lavoro vengono fornite delle approssimazioni, mediante determinazione del *saddlepoint*, alle distribuzioni campionarie degli stimatori di massima verosimiglianza dei parametri di posizione e di scala delle curve normali di ordine p.

Dopo un breve richiamo sulle curve normali di ordine p e sugli stimatori di massima verosimiglianza dei parametri di posizione e di scala, nei paragrafi successivi, dopo un cenno alla tecnica della determinazione del *saddlepoint* per gli *stimatori di tipo M*, classe nella quale rientrano gli stimatori di massima verosimiglianza, vengono esposte le peculiarità dell' applicazione al nostro caso; infine vengono riportati alcuni esempi di approssimazioni alle densità ed alle probabilità integrali dei suddetti stimatori che sono risultate più che soddisfacenti già per n=5 e in corrispondenza di un ampio range di valori di p. Nella presente comunicazione, per brevità, esporremo solo alcuni dei risultati ottenuti, senza entrare nel dettaglio degli aspetti formali (per alcuni dei quali si rimanda alla bibliografia) e delle tecniche di calcolo numerico impiegate che, insieme con altri sviluppi metodologici del presente approccio, verranno esposti in altri lavori.

2. Il modello delle curve normali di ordine p: stimatori dei parametri e loro distribuzioni campionarie asintotiche

Il modello delle curve normali di ordine p è costituito da una famiglia di curve simmetriche di densità

$$f(x)=[2p^{(1/p)}\Gamma(1+1/p)\sigma_p]^{-1}\exp[-|x-\mu|^p/(p\sigma_p^p)],\quad p>0,\ x\in\Re,$$

con μ e σ_p parametri di posizione e di scala, dato che si ha $E[X]=\mu$; $E[|X-\mu|^p]^{1/p}=\sigma_p$.

Queste curve sono state introdotte dal Subbotin (1923) e sono state impiegate, talvolta con diverse espressioni dei parametri, da numerosi autori (Box e Tiao, 1973; Vianelli, 1963; Mineo, 1986), poichè risultano utili quando si vuole descrivere il comportamento di errori accidentali con distribuzione anche diversa dalla legge di Gauss. Se p è noto, gli stimatori di massima verosimiglianza M_p e V_p di μ e σ_p^p sulla base di un campione casuale semplice di n osservazioni, per $p \geq 1$, sono forniti dalla soluzione del sistema di due equazioni

$$\sum_{i=1}^{n} |x_i - M_p|^{p-1} \mathrm{sign}(x_i - M_p) = 0 \; ; \qquad \sum_{i=1}^{n} [|x_i - M_p|^p - V_p] = 0, \tag{1}$$

La soluzione esiste ed è sempre unica se p>1. In generale per ottenere M_p occorre far ricorso a metodi iterativi (tranne che per p=1,2 e 4), mentre V_p è dato da: $V_p = \sum_{i=1}^{n} |x_i - M_p|^p / n$.

Gli stimatori di massima verosimiglianza risultano essere i migliori, almeno asintoticamente, ed anche per piccoli campioni sono superiori ad esempio agli stimatori robusti (Mineo, 1986). Per potere comunque fare inferenza su μ e σ_p^p, è necessario conoscere, anche approssimativamente, le distribuzioni campionarie di M_p e V_p al variare di p.

Quando μ è noto, V_p segue una distribuzione gamma con parametro di scala $\lambda = n/(p\sigma_p^p)$ e parametro di forma c=n/p, (Lunetta,1966), ed é corretto e sufficiente per σ_p^p (Cifarelli, Regazzini, 1974), mentre per valori di n finiti non esistono stimatori congiuntamente sufficienti per i parametri μ e σ_p^p, tranne ovviamente che per p=2; se invece μ non é noto e si utilizza M_p, la distribuzione di V_p non è più nota. In un precedente lavoro (Chiodi, 1988), abbiamo mostrato come una approssimazione soddisfacente a $E(V_p)$ sia data da: $E(V_p) = \sigma_p^p(1-p/(2n)) + o(n^{-1})$, solo per p<2n; ed inoltre, riaggiustando i parametri della distribuzione gamma in modo che abbia questa speranza matematica, avevamo ottenuto delle approssimazioni abbastanza buone già per n>10 e per valori di p non troppo lontani da 2; per n>20 l'approssimazione risultava comunque utilizzabile a fini pratici.

Inoltre, asintoticamente, M_p e V_p risultano indipendenti e distribuiti normalmente: l'approssimazione gaussiana per $f(V_p)$ è soddisfacente solo per grandissimi campioni, e anche per $f(M_p)$, in particolare quando 1<p<1.5, risulta accettabile solo per grandi campioni; per p prossimo ad uno l'approssimazione è utilizzabile a fini pratici solo per valori di n molto grandi (n>100).

I metodi fondati su sviluppi in serie di Edgeworth (Pfanzagl, 1973), risultano inapplicabili nel nostro caso per valori di p<2, dal momento che non sono soddisfatte le condizioni riguardanti le derivate di ordine superiore al secondo di $\log[L(\underline{x};\mu,\sigma_p^p)]$ rispetto a μ (Diananda, 1949). Inoltre l'accuratezza di tali espansioni è elevata verso il centro della distribuzione, mentre va peggiorando verso le code, che sono oltretutto la zona più interessante per effettuare test di significatività o per la costruzione di intervalli di confidenza.

3. Approssimazioni *saddlepoint* alle densitá degli stimatori di tipo M

Per i motivi sopra esposti, per approssimare le distribuzioni di M_p e V_p, si è impiegata la tecnica basata sulla determinazione del *saddlepoint* che, nella sua forma più semplice (Daniels, 1954), fornisce delle ottime approssimazioni alla distribuzione campionaria di una media aritmetica,

Marcello Chiodi: APPROSSIMAZIONI SADDLEPOINT ALLE DISTRIBUZIONI CAMPIONARIE DEGLI STIMATORI DI MASSIMA VEROSIMIGLIANZA DEI PARAMETRI DELLE CURVE NORMALI DI ORDINE P PER PICCOLI CAMPIONI.Atti della XXXVII Riunione Scientifica della SIS, San Remo 6-8 Aprile 1994, ed. CISU, vol.II, pagg. 139-146.

anche per piccolissimi campioni, essendo solitamente l'errore *relativo* di ordine $O(n^{-1})$ o $O(n^{-3/2})$, e risulta in generale sufficientemente accurata anche nelle code della distribuzione e per questo rientra nella categoria delle cosiddette *approssimazioni asintotiche per piccoli campioni*. Tale tecnica corrisponde ad usare la densità associata o coniugata, o la tecnica dell' *exponential tilting* (Barndorff-Nielsen e Cox, 1989).Nel caso che ci interessa, la coppia (M_p,V_p) è una funzione piú generale dei valori osservati che rientra nella categoria dei cosiddetti *stimatori multipli di tipo M*, definiti mediante un sistema di equazioni implicite

$$\sum_{i=1}^{n} \Psi(x_i,\mathbf{T}_n) = \mathbf{0},$$

ove $\mathbf{T}_n$ è uno stimatore di k componenti di un vettore $\mathbf{q}$ di k parametri, Ψ è un vettore di k equazioni indipendenti e $\mathbf{0}$ è il vettore nullo di k elementi. Gli stimatori di massima verosimiglianza, nel caso regolare, rientrano nella categoria degli stimatori di tipo M.

Per determinare la densità approssimata di $\mathbf{T}_n$ in un punto $\mathbf{t}$ (appartenente al dominio di $\mathbf{q}$), il punto di partenza è dato dalla determinazione, per ciascun valore di $\mathbf{t}$, del *saddlepoint* $\mathbf{a}(\mathbf{t})$ (vettore di k componenti), soluzione del sistema di k equazioni (Ronchetti, 1990; Field, 1982)

$$\int_{-\infty}^{\infty} \Psi(x,\mathbf{t})\exp\{\mathbf{a}^T(\mathbf{t})\Psi(x,\mathbf{t})\}f(x)dx = \mathbf{0} \qquad (2)$$

essendo $\int_{-\infty}^{\infty} \exp\{\mathbf{a}^T\Psi(x,\mathbf{t})\}f(x)dx$ la funzione generatrice dei momenti di Ψ ed $\mathbf{a}(\mathbf{t})$, determinato dal sistema (2), il punto che rende minimo, per ciascuno dei valori di $\mathbf{t}$, l'esponente della funzione integranda. La densità di T_n, $f_n(\mathbf{t})$, è quindi approssimata da

$$g_n(\mathbf{t})=[n/(2\pi)]^{k/2}C^{-n}(\mathbf{t})\ |\det B(\mathbf{t})|\ |\det \Sigma(\mathbf{t})|^{-1/2} \qquad (3)$$

nella quale

$$\begin{aligned} C^{-1}(\mathbf{t}) &= \int_{-\infty}^{\infty} \exp\{\mathbf{a}^T(\mathbf{t})\Psi(x,\mathbf{t})\}f(x)dx,\\ B(\mathbf{t}) &= C(\mathbf{t})\int_{-\infty}^{\infty} \{\partial\Psi(x,\mathbf{t})/\partial\mathbf{t}\}\exp\{\mathbf{a}^T(\mathbf{t})\Psi(x,\mathbf{t})\}f(x)dx, \qquad (4)\\ \Sigma(\mathbf{t}) &= C(\mathbf{t})\int_{-\infty}^{\infty} \Psi(x,\mathbf{t})\Psi^T(x,\mathbf{t})\exp\{\mathbf{a}^T(\mathbf{t})\Psi(x,\mathbf{t})\}f(x)dx. \end{aligned}$$

($C^{-1}(\mathbf{t})$ è uno scalare, $B(\mathbf{t})$ e $\Sigma(\mathbf{t})$ sono matrici $k\times k$). Si puó mostrare che

$$f_n(\mathbf{t})=g_n(\mathbf{t})[1+O(n^{-1})].$$

In generale comunque la rinormalizzazione di $g_n(t)$ con l'area trovata eventualmente mediante quadratura numerica migliora in modo sensibile l'approssimazione.

Inoltre, nel caso univariato (k=1), si può ottenere un' approssimazione diretta a $F_n(t)$, funzione di ripartizione di t (Lugannani e Rice, 1980; Daniels, 1987)

$$F_n(t)=G_n(t)[1+O(n^{-1})]$$

$$\text{e}\quad 1-G_n(t)=1-\Phi\{[2n\log(C(t))]^{1/2}\}+[n/(2\pi)]^{-1/2}C^{-n}(t)\{1/(\alpha(t)|\Sigma|)-1/[2\log(C(t))]^{1/2}\} \qquad (5)$$

con $\Phi\{.\}$ funzione di ripartizione di una variabile normale standardizzata. Nel caso di stima simultanea di piú parametri, non conosciamo sviluppi in serie troncati di analoga semplicitá, per cui in generale occorrerá integrare numericamente la (3), oppure ricorrere ad approssimazioni di Laplace delle probabilità integrali (Daniels e Young, 1991).

Marcello Chiodi: APPROSSIMAZIONI SADDLEPOINT ALLE DISTRIBUZIONI CAMPIONARIE DEGLI STIMATORI DI MASSIMA VEROSIMIGLIANZA DEI PARAMETRI DELLE CURVE NORMALI DI ORDINE P PER PICCOLI CAMPIONI.Atti della XXXVII Riunione Scientifica della SIS, San Remo 6-8 Aprile 1994, ed. CISU, vol.II, pagg. 139-146.

4. Approssimazioni alla distribuzione campionaria di M_p

Sulla base di quanto esposto finora, abbiamo applicato la tecnica della ricerca del *saddlepoint* inizialmente al problema dell'approssimazione della distribuzione campionaria del solo stimatore M_p, per un campione estratto da una distribuzione normale di ordine p standardizzata, ossia con $\mu=0$ e $\sigma_p^p=1$ (questa semplificazione non altera la generalità dei risultati qui esposti, poichè è immediato ricavare le distribuzioni di M_p e V_p quando X non è standardizzata, utilizzando $(M_p-\mu)/\sigma_p$ e V_p/σ_p^p).

M_p è uno stimatore di tipo M definito dalla relazione implicita

$$\sum_{i=1}^{n} \psi(x_i,T_n)=0, \text{ con } \psi(x_i,M_p)=|x_i - M_p|^{p-1}\text{sign}(x_i - M_p).$$

Pertanto, considerando p noto, occorre trovare il *saddlepoint* $\alpha(m_p)$, per ciascuno dei valori m_p per i quali si vuole approssimare la distribuzione di M_p, risolvendo l' equazione implicita

$$\int_{-\infty}^{+\infty}|x-m_p|^{p-1}\text{sign}(x-m_p)\exp\{\alpha(m_p)|x-m_p|^{p-1}\text{sign}(x-m_p)\}A_p\exp[-|x|^p/p]dx=0 \qquad (6)$$

ove $A_p=[2p^{1/p}\Gamma(1+1/p)]^{-1}$ è una costante di normalizzazione, ininfluente per la determinazione di $\alpha(m_p)$. Trovato $\alpha(m_p)$ si possono calcolare le altre quantità definite dalle relazioni (4) e sostituendo queste quantità nella (3) si ottiene $g_n(m_p)$, approssimazione alla densità della distribuzione campionaria dello stimatore di massima verosimiglianza M_p. Sostituendo invece nella (5) si ha direttamente un'approssimazione della funzione di ripartizione di M_p. E' da notare che l'integrale che fornisce $B(m_p)$ può essere riespresso, integrando per parti, in modo da eliminare la discontinuità della funzione integranda per p<2 e $x=m_p$, rendendone così più agevole l'integrazione numerica.

Per p=2 M_2 è la media aritmetica, e il metodo fornisce l'espressione esatta di $f_n(m_2)$, (Daniels, 1954) con $\alpha(m_2)=m_2$; per p=1 (distribuzione di Laplace) e con n dispari, M_p è la mediana, la cui distribuzione campionaria è in questo caso nota, e coincide con l'approssimazione *saddlepoint* a meno del fattore di normalizzazione e si ha $\alpha(m_1)=\log(2\exp(m_1)-1)$.

Per la risoluzione della (6) abbiamo impiegato il metodo di Newton, che converge sempre ed in 4-5 iterazioni abbiamo ottenuto una soluzione esatta alla quinta cifra decimale, partendo da un opportuno valore iniziale (una prima scelta puó essere data interpolando rispetto a p fra i valori esatti di $\alpha(m_p)$ per p=1 e p=2); in corrispondenza dell'iterazione finale la derivata risulta proporzionale a $\Sigma(m_p)$. Per il calcolo degli integrali che compaiono nella funzione implicita che definisce $\alpha(m_p)$, e nelle espressioni che definiscono $C(m_p)$, $B(m_p)$ e $\Sigma(m_p)$), per ottenere degli algoritmi soddisfacenti per i diversi valori di p e m_p, si é suddiviso opportunamente il campo di integrazione, in modo da isolare gli eventuali punti di discontinuitá delle derivate delle funzioni integrande, e su ciascuno di essi si é proceduto ad integrare numericamente con metodi adattivi.

Dal momento che in generale per le approssimazioni di tipo *saddlepoint* non sono disponibili espressioni approssimate del massimo dell' errore commesso, é stato necessario confrontare le nostre approssimazioni, per un' ampia gamma di valori di n, p e di m_p, con i valori ottenuti simulando, per ciascun valore di n e p, l' estrazione di N=1.000.000 di campioni di ampiezza n da una distribuzione normale standardizzata di ordine p (Chiodi,1986), e calcolando poi la distribuzione campionaria empirica dei valori di M_p calcolati su ciascun campione, distribuzione che é stata usata come base per i nostri confronti (il numero N=1.000.000 di campioni generati ci é sembrato sufficiente, dal momento che i risultati ottenuti per simulazione hanno un errore quadratico medio proporzionale a $N^{-1/2}$=0,001). Si è considerato sempre m_p>0, per la simmetria di $f_n(m_p)$.

Nella figura 1 sono riportati, per 6 valori di p, (p=1,1;1,3;1,5;1,7;2,5;3,5) e per n=5, tre serie di valori di Prob(M_p>m_p) in corrispondenza dei valori di m_p; le tre curve sono: i valori ottenuti per simulazione, i valori ottenuti dall'approssimazione *saddlepoint* fornita dalla relazione (5) e i valori forniti dalla consueta approssimazione normale; si sono rappresentate nei grafici solo le code delle distribuzioni (0,75<m_p<1,5). Come può vedersi, sebbene n sia molto piccolo, l'approssimazione proposta risulta tanto buona da risultare, nei grafici riportati, quasi indistinguibile dai valori ottenuti per simulazione, per tutti i valori di p considerati.

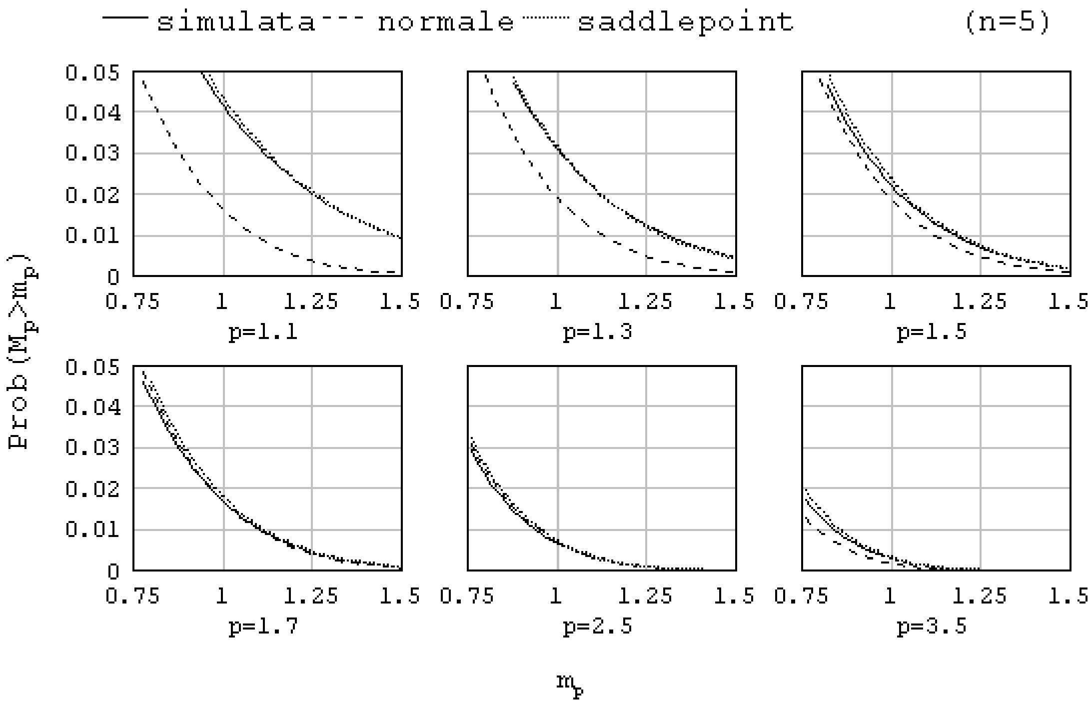

Fig. 1. Confronto fra i valori di Prob(M_p>m_p) ottenuti mediante simulazione, mediante approssimazione *saddlepoint* e con l'approssimazione *normale* per 6 valori di p e per n=5

Le approssimazioni risultano in ogni caso tanto migliori delle corrispondenti approssimazioni "*normali*", quanto più ci si avvicina a valori di p prossimi ad uno. L'approssimazione normale risulta in generale carente nella coda della distribuzione, almeno in termini di errore relativo.

5. Approssimazioni alle distribuzioni congiunte di M_p e V_p

Le relazioni (4) possono essere utilmente impiegate per trovare delle approssimazioni alle distribuzioni campionarie congiunte di M_p e V_p, stimatori di massima verosimiglianza dei parametri

μ e σ_p^p, rispettivamente. Va ricordato che nel caso in cui μ è noto, l'approssimazione *saddlepoint* alla distribuzione di V_p risulta esatta a meno del fattore di normalizzazione, poichè V_p è la media aritmetica di n variabili gamma i.i.d. e la distribuzione della media aritmetica di variabili gamma viene esattamente riprodotta dall' approssimazione *saddlepoint* a meno di un fattore di normalizzazione (Daniels, 1954). Dal momento che inoltre le approssimazioni alla distribuzione di M_p, supposto σ_p^p noto, sono risultate più che soddisfacenti anche per valori di n molto piccoli, ci è sembrato giustificato cercare di approssimare ancora mediante tecnica *saddlepoint* la distribuzione congiunta dei due stimatori, sebbene i problemi numerici e computazionali siano chiaramente amplificati rispetto al precedente problema, riguardante solo M_p.

Come è facile vedere dalla (1), la coppia (M_p,V_p) può essere vista come uno stimatore M di un parametro bidimensionale: infatti M_p e V_p vengono forniti dalle due equazioni implicite

$$\sum_{i=1}^{n} \psi_j(x_i,M_p,V_p) = 0, \quad (j=1,2)$$

con $\psi_1(x_i,M_p,V_p)=|x_i - M_p|^{p-1}\mathrm{sign}(x_i-M_p)$ e $\psi_2(x_i,M_p,V_p)=V_p-|x_i-M_p|^p$.

Per ottenere dalla (3) l' approssimazione alla densità di (M_p,V_p), occorre risolvere per ciascuna coppia (m_p,v_p) il sistema (2) di due equazioni nelle due incognite $\alpha_1(m_p,v_p)$ e $\alpha_2(m_p,v_p)$ e quindi calcolare $C(m_p,v_p)$ e gli elementi delle matrici $B(m_p,v_p)$ e $\Sigma(m_p,v_p)$ dalle relazioni (4), le cui formule esplicite, insieme con alcune semplificazioni dei calcoli, non vengono riportate per brevità. I metodi numerici impiegati per la determinazione di $\alpha_1(m_p,v_p)$ e $\alpha_2(m_p,v_p)$ per ciascuna coppia (m_p,v_p), e per la quadratura degli integrali, sono nella sostanza simili a quelli accennati nel paragrafo precedente per la distribuzione di M_p.

In effetti, per calcolare le densitá congiunte approssimate di (M_p,V_p), abbiamo sostituito nella (3) al fattore di normalizzazione $2\pi/n$ il fattore di normalizzazione *esatto* che si ottiene quando $m_p=0$, ossia $C_1= \exp(n/p)[2\pi/np)]^{1/2}\Gamma[n-1)/p]/(n/p)^{(n-1)/p}$

Tab. 1. Valori di $C_1/(2\pi/n)$ e C_2 in corrispondenza di alcuni valori di p ed n

	$C_1/(2\pi/n)$		C_2	
p	n=5	n=10	n=5	n=10
1.1	1,261	1,116	0,982	0,983
1.3	1,248	1,110	0,974	0,984
1.5	1,239	1,106	0,972	0,992
1.7	1,234	1,104	0,974	0,994
2.5	1,229	1,101	1,028	0,998
3.5	1,238	1,105	1,066	1,015

Infatti, quando $m_p=0$, si ha: $\alpha_1=0$ e $\alpha_2=[1/v_p-1]/p)$ e l'approssimazione *saddlepoint* alla distribuzione parziale di V_p risulta in questo caso proporzionale alla densitá di una distribuzione gamma, e dunque C_1 é calcolabile esattamente. Abbiamo impiegato C_1 come fattore di normalizzazione per tutti i valori di m_p e v_p, perchè, sebbene esatto solo se $m_p=0$, è risultato senz'altro migliore del fattore $2\pi/n$. Nella tabella 1 sono riportati, per alcuni valori di p e per n=5, i valori del rapporto $C_1/(2\pi/n)$, ed i valori di C_2, costante di normalizzazione ottenuta integrando poi numericamente la $f(m_p,v_p)$ cosí corretta. Come può vedersi, l' impiego di C_1 risulta indispensabile,

Marcello Chiodi: APPROSSIMAZIONI SADDLEPOINT ALLE DISTRIBUZIONI CAMPIONARIE DEGLI STIMATORI DI MASSIMA VEROSIMIGLIANZA DEI PARAMETRI DELLE CURVE NORMALI DI ORDINE P PER PICCOLI CAMPIONI.Atti della XXXVII Riunione Scientifica della SIS, San Remo 6-8 Aprile 1994, ed. CISU, vol.II, pagg. 139-146.

almeno per piccoli campioni, mentre il miglioramento ulteriore apportato dall' integrazione numerica di $f(m_p,v_p)$ non è della stessa entità, specie considerando la mole di calcoli necessaria per ottenerla.

Nella figura 2 é riportata, solo per p=1,1 ed n=5, la densitá simulata $f(m_p,v_p)$ e quella ottenuta mediante l'approssimazione proposta, soltanto per $m_p>0$. Come puó vedersi, le due superfici sono molto simili, anche lontano dal centro della distribuzione.

Nella figura (3) sono riportati i valori delle densità simulate e di quelle approssimate di V_p, ottenute integrando numericamente rispetto a m_p l'approssimazione *saddlepoint* a $f(m_p,v_p)$, per 6 diversi valori di p e per n=5: anche in questo caso le approssimazioni risultano soddisfacenti per tutti i valori di p considerati. Per ottenere i valori approssimati delle densità marginali di V_p, abbiamo impiegato una formula di quadratura gaussiana a 8 punti basata sui polinomi ortogonali di Laguerre, insieme con alcune trasformazioni delle funzioni integrande, che hanno portato a delle approssimazioni numeriche soddisfacenti anche rispetto a formule di quadratura a 16 punti.

Le approssimazioni risultano complessivamente soddisfacenti. Va aggiunto che, seguendo un approccio proposto da Daniels e Young (1991), e integrando mediante una particolare approssimazione di Laplace rispetto a m_p l'approssimazione *saddlepoint* a $f(m_p,v_p)$, si ottiene una densità marginale di V_p proporzionale ad una v.c. gamma con parametri $\lambda=n/p$ e $c=n/p-1/2$, risultato che coincide con quello da noi trovato per altra via nel nostro lavoro (Chiodi, 1988) citato nel secondo paragrafo. Questa approssimazione é pure riportata nella figura 3.

Marcello Chiodi: APPROSSIMAZIONI SADDLEPOINT ALLE DISTRIBUZIONI CAMPIONARIE DEGLI STIMATORI DI MASSIMA VEROSIMIGLIANZA DEI PARAMETRI DELLE CURVE NORMALI DI ORDINE P PER PICCOLI CAMPIONI.Atti della XXXVII Riunione Scientifica della SIS, San Remo 6-8 Aprile 1994, ed. CISU, vol.II, pagg. 139-146.

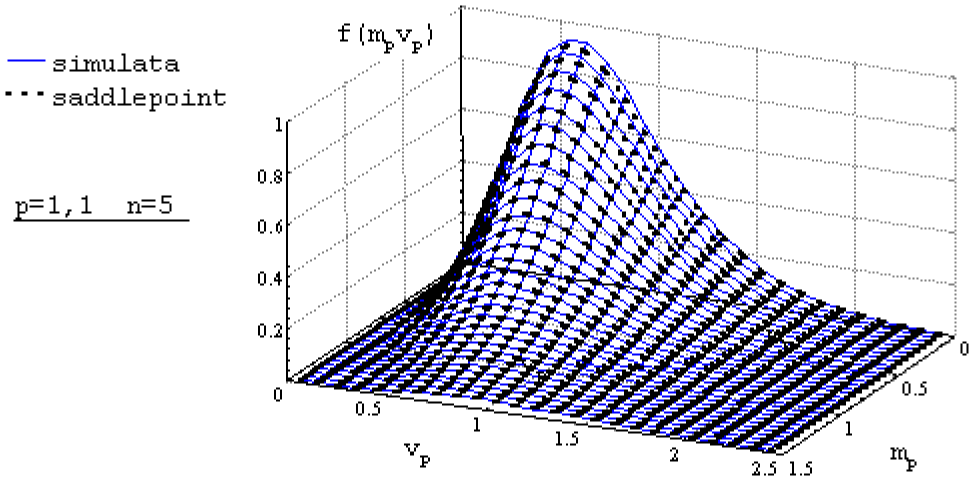

Fig. 2. Confronto,per p=1,1 ed n=5, fra la densitá simulata $f(m_p,v_p)$ e quella ottenuta mediante l'approssimazione *saddlepoint*

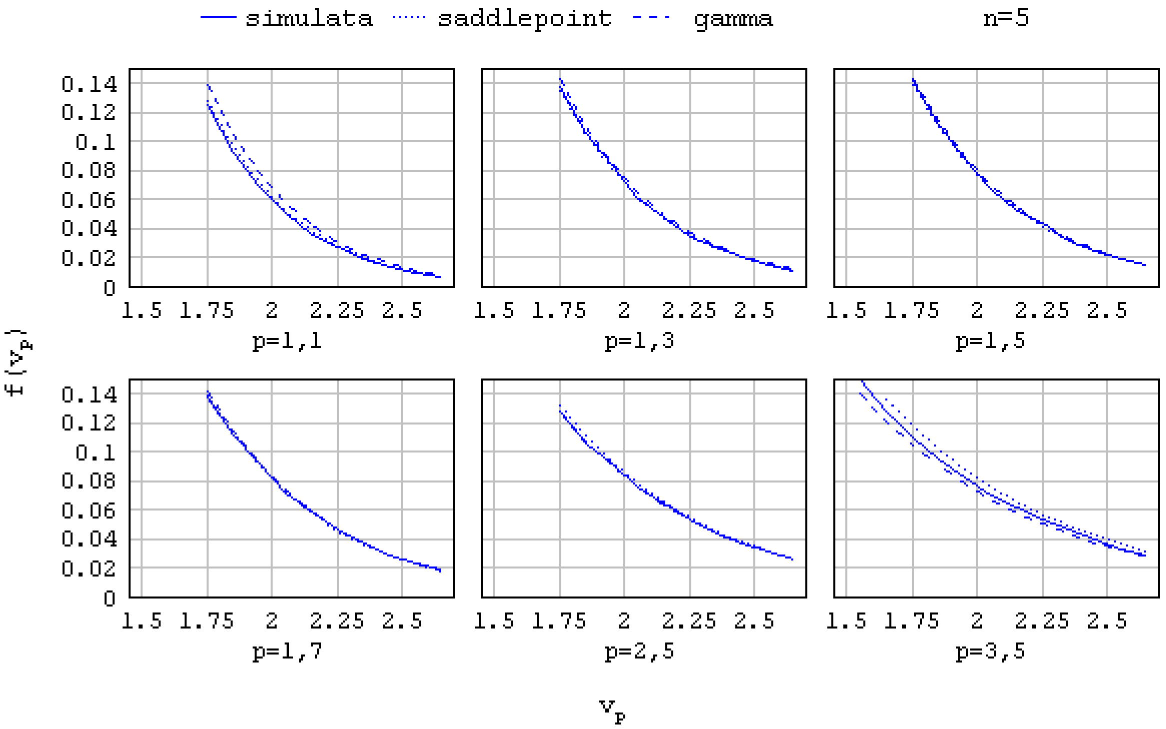

Fig. 3. Confronto fra le code di $f(v_p)$ calcolate mediante simulazione, integrazione numerica di $f(m_p,v_p)$ e distribuzione gamma

NOTE

(1) Studio compiuto con fondi C.N.R. per la ricerca dal titolo: "Distribuzioni di campionamento e test per la verifica di ipotesi su costanti statistiche di ordine p" (resp. Prof.A.Mineo).

RIFERIMENTI BIBLIOGRAFICI

BARNDORFF-NIELSEN O.E., COX D.R. (1989), "Asymptotic Techniques for Use in Statistics". Chapman and Hall Ed., London.

BOX G.E.P., TIAO G.C. (1973), "Bayesian inference in statistical analysis". Addison-Wesley Ed.; Reading, Massachusetts.

CHIODI M. (1986), "Procedures for generating pseudo-random numbers from a normal distribution of order p(p>1)". Rivista di Statistica Applicata, Milano, 19, p. 7-26.

CHIODI M. (1988),"Sulle distribuzioni di campionamento delle stime di massima verosimiglianza dei parametri delle curve normali di ordine p". Pubbl. a cura dell'Istituto di Statistica di Palermo il 21-12-1988 ai sensi dell' Art.1 D.L. 31-8-1945,n.660.

CIFARELLI D.M., REGAZZINI E. (1974), "Sulla caratterizzazione di una famiglia di distribuzioni basata sulla efficienza di una funzione campionaria". Rend. Sem. Mat. Univers. Politecn. Torino, 33, p. 299-311.

DANIELS H.E. (1954),"Saddlepoint approximations in statistics". Ann. Math. Stat.,25, p. 631-650.

DANIELS H.E. (1987), "Tail probability approximations". Int. Statist. Rev., 55, p. 37-48.

DANIELS H.E., YOUNG G.A. (1991), "Saddlepoint approximations for the studentized mean, with an application to the bootstrap". Biometrika, 78, p. 169-179.

DIANANDA P.H. (1949), "Note on some proprieties of maximum likelihood estimates". Proceedings of the Cambridge Philosophical Society, 45, p. 536-544.

FIELD C.A., (1982), "Small sample asymptotic expansion for multivariate M-estimates". Ann. Statist., 10, p. 672-689.

LUGANNANI R., RICE S. (1980), "Saddlepoint approximation for the distribution of the sum of independent random variables". Adv. Appl. Prob.,12, p. 475-490.

LUNETTA G. (1966): "Di alcune distribuzioni deducibili da una generalizzazione dello schema della curva normale". Annali della Facoltà di Economia e Commercio di Palermo, 20, 1, p. 117-143.

MINEO A. (1986), "La migliore combinazione delle osservazioni per la stima dei parametri di intensità e di dispersione". Studi in onore di Silvio Vianelli: Istituto di Statistica, Palermo;1, p.517-548.

PFANZAGL J. (1973), "Asymptotic expansions related to minimum contrast estimators". Ann. Statist., 1, p. 993-1026. Correzioni (1974) 2, p.1357-1358.

RONCHETTI E. (1990), "Small sample asymptotics and bootstrap". Quaderni di Statistica e Econometria, Centro di Specializzazione e Ricerche Economiche Agrarie per il Mezzogiorno, Portici (NA), 12, p. 19-38.

SUBBOTIN M.T. (1923), "On the law of frequency errors". Matematicheskii Sbornik., Mosca; 31, 2, p.296-300.

VIANELLI S. (1963), "La misura della variabilità condizionata in uno schema generale delle curve normali di frequenza". Statistica,23, p.447-474.

Marcello Chiodi: APPROSSIMAZIONI SADDLEPOINT ALLE DISTRIBUZIONI CAMPIONARIE DEGLI STIMATORI DI MASSIMA VEROSIMIGLIANZA DEI PARAMETRI DELLE CURVE NORMALI DI ORDINE P PER PICCOLI CAMPIONI.Atti della XXXVII Riunione Scientifica della SIS, San Remo 6-8 Aprile 1994, ed. CISU, vol.II, pagg. 139-146.

SADDLEPOINT APPROXIMATIONS TO THE SAMPLING DISTRIBUTIONS OF THE MAXIMUM LIKELIHOOD ESTIMATORS OF THE PARAMETERS OF THE NORMAL DISTRIBUTIONS OF ORDER P.

Marcello Chiodi

Istituto di Statistica,
Facoltà di Economia e Commercio di Palermo.

Summary

The family of normal distributions of order p, has density

$$f(x)=[2p^{(1/p)}\Gamma(1+1/p)\sigma_p]^{-1}\exp[-|x-\mu|^p/(p\sigma_p^p)], \qquad p>0,\ x\in\Re,$$

and is a class of symmetrical curves useful for the description of a wide class of distributions of random errors (Mineo, 1986). The ML estimators M_p and V_p of μ and σ_p^p are defined by the equations (for p>1)

$$\sum_{i=1}^{n} |x_i - M_p|^{p-1} \operatorname{sign}(x_i - M_p)=0 \quad \text{and} \quad \sum_{i=1}^{n} [|x_i - M_p|^p - V_p]=0,$$

the first of which must be solved numerically, in the general case, in order to find M_p. The sampling distributions of M_p and V_p are not known, for small values of n, except for some special cases; so, since the ML estimators belong to the class of the M-estimators, we used *saddlepoint* approximations (Daniels, 1954; Ronchetti, 1990).

These approximations have been compared with sampling distributions obtained by simulations performed on 1.000.000 samples for each value of p and n.

The approximations to $f(M_p)$ (as well to the tail probabilities) and to the joint density of M_p and V_p, after a special kind of renormalization which avoided the need for numerical integration, turned out to be very good even for small values of n (n=5) and for a wide range of values of p expecially in the tails of the distribution, which are crucial in order to make inference about the parameters μ and σ_p^p. For the marginal density of V_p, numerical integration with respect to M_p has been performed, with very good results. A laplacian integration gave also fairly good results, leading to a gamma distribution, a result similar to an approximation found in an our previous work (Chiodi, 1988).

Some key words:Normal distributions of order p, ML estimators, saddlepoint approximations

Essential references:

CHIODI M. (1988),"Sulle distribuzioni di campionamento delle stime di massima verosimiglianza dei parametri delle curve normali di ordine p". Istituto di Statistica di Palermo.

DANIELS H.E. (1954),"Saddlepoint approximations in statistics". Ann. Math. Stat.,25, p. 631-650.

MINEO A. (1986), "La migliore combinazione delle osservazioni per la stima dei parametri di intensità e di dispersione". Studi in onore di Silvio Vianelli, Istituto di Statistica, Palermo; vol.1, p.517-548

Marcello Chiodi: APPROSSIMAZIONI SADDLEPOINT ALLE DISTRIBUZIONI CAMPIONARIE DEGLI STIMATORI DI MASSIMA VEROSIMIGLIANZA DEI PARAMETRI DELLE CURVE NORMALI DI ORDINE P PER PICCOLI CAMPIONI.Atti della XXXVII Riunione Scientifica della SIS, San Remo 6-8 Aprile 1994, ed. CISU, vol.II, pagg. 139-146.

RONCHETTI E. (1990), "Small sample asymptotics and bootstrap". Quaderni di Statistica e Econometria, Centro di Specializzazione e Ricerche Economiche Agrarie per il Mezzogiorno, Portici (NA), 12, p. 19-38.