



Department of Statistical Sciences
University of Padua
Italy

UNIVERSITÀ
DEGLI STUDI
DI PADOVA
DIPARTIMENTO
DI SCIENZE
STATISTICHE

Intervalli di confidenza non parametrici per i quantili con dati MAR

Gianfranco Adimari, Elisa Drago
Department of Statistical Sciences
University of Padua
Italy

Abstract: In questo lavoro si studia una metodologia per la costruzione di intervalli di confidenza non parametrici per i quantili di una variabile risposta Y , con dati *missing at random* (MAR). Tale metodologia prevede l'uso di una statistica ottenuta combinando opportunamente la funzione di verosimiglianza empirica per i quantili e uno stimatore della funzione di ripartizione di Y , proposto da Cheng e Chu (1996), basato su una procedura di imputazione non parametrica.

Keywords: Dati mancanti; Imputazione non parametrica; *Missing at random*; Verosimiglianza empirica.

Indice

1	Introduzione	1
2	Stima non parametrica della funzione di ripartizione e dei quantili con dati MAR.	2
2.1	Stima non parametrica della funzione di ripartizione e dei quantili .	2
2.2	Consistenza	3
2.3	Normalità asintotica.	4
2.4	Estensione al caso in cui $X \in \mathbb{R}^d$	6
3	Intervalli di confidenza per i quantili: l'approccio proposto	7
4	Alcuni risultati di simulazione	10
4.1	Descrizione dello studio di simulazione.	10
4.2	I risultati.	12
4.3	Analisi dei risultati e commenti.	12
5	Due esempi	14
6	Note conclusive	16

Department of Statistical Sciences
Via Cesare Battisti, 241
35121 Padova
Italy

tel: +39 049 8274168
fax: +39 049 8274170
<http://www.stat.unipd.it>

Corresponding author:
Gianfranco Adimari
tel: +39 049 827 4168
gianfranco.adimari@unipd.it
<http://www.stat.unipd.it/~adimari>

Intervalli di confidenza non parametrici per i quantili con dati MAR

Gianfranco Adimari, Elisa Drago
Department of Statistical Sciences
University of Padua
Italy

Abstract: In questo lavoro si studia una metodologia per la costruzione di intervalli di confidenza non parametrici per i quantili di una variabile risposta Y , con dati *missing at random* (MAR). Tale metodologia prevede l'uso di una statistica ottenuta combinando opportunamente la funzione di verosimiglianza empirica per i quantili e uno stimatore della funzione di ripartizione di Y , proposto da Cheng e Chu (1996), basato su una procedura di imputazione non parametrica.

Keywords: Dati mancanti; Imputazione non parametrica; *Missing at random*; Verosimiglianza empirica.

1 Introduzione

Nell'analisi statistica ci si scontra spesso con il problema dei dati mancanti. In molte situazioni, infatti, capita di non poter disporre di alcuni valori delle variabili rilevate, per ragioni che vanno dalla banale perdita delle informazioni raccolte, al fatto che alcune delle unità sondate si siano rifiutate di fornire questa o quella risposta. La *non risposta* è elemento che caratterizza le ricerche di mercato e le indagini socio-economiche; ma il problema dei dati mancanti si presenta anche negli studi medici o negli esperimenti scientifici e industriali così come nelle indagini sull'ambiente.

Con dati mancanti, le procedure inferenziali usuali non possono essere applicate direttamente (a meno che non si accetti, eventualmente, di rinunciare a parte dell'informazione disponibile) e c'è la necessità di ricorrere a tecniche specifiche. Diventano allora cruciali le ipotesi sul meccanismo che determina la perdita dei dati: le assunzioni su tale meccanismo consentono di scegliere la tecnica appropriata e di interpretare correttamente i risultati. La letteratura propone vari metodi per la trattazione dei dati mancanti e l'applicazione dell'uno o dell'altro dipende da diversi fattori, tra cui proprio le ipotesi sul meccanismo che li genera. Molte delle metodologie comunemente usate presuppongono che i dati mancanti siano *Missing at Random* (MAR).

Si consideri una variabile risposta univariata Y le cui determinazioni non sono tutte osservate (nel seguito diremo che Y è non completamente osservata), rilevata, su N unità, assieme ad un'altra variabile X di cui si osservano tutte le determinazioni (completamente osservata). Si supponga, quindi, che l'informazione campionaria sia

costituita dalle terne

$$(X_i, Y_i, \delta_i) \quad i = 1, \dots, N, \quad (1)$$

dove tutte le determinazioni X_i di X sono osservate e la variabile binaria δ_i vale 0 se la determinazione Y_i è mancante, altrimenti vale 1. Assumere che il meccanismo che determina la non completa osservabilità della variabile Y sia di tipo MAR, equivale a richiedere (in un contesto non parametrico) che δ e Y siano condizionatamente indipendenti dato X , ossia che

$$\text{pr} \{ \delta = 1 | Y, X \} = \text{pr} \{ \delta = 1 | X \} .$$

Operativamente, pertanto, l'ipotesi MAR è sensata quando sia ragionevole ritenere che la non osservabilità delle determinazioni della variabile Y possa dipendere solo dai livelli delle determinazioni della variabile X concomitanti.

Lo scopo di questo lavoro è quello di proporre una tecnica non parametrica per ottenere intervalli di confidenza per i quantili della distribuzione di una variabile risposta univariata Y (rilevata assieme ad X) quando sia plausibile l'ipotesi MAR per i dati mancanti. L'idea è quella di combinare la funzione log-rapporto di verosimiglianza empirica (si veda Owen, 2001, come riferimento generale) per i quantili con uno stimatore per la funzione di ripartizione di Y , per dati MAR, proposto da Cheng e Chu (1996) e basato sul metodo del nucleo. Ciò porta ad una versione modificata della statistica di verosimiglianza empirica per i quantili che, se opportunamente corretta, ha distribuzione limite χ_1^2 . Tale statistica può essere quindi usata, nella maniera nota, per costruire intervalli di confidenza per i quantili (o per risolvere problemi di verifica d'ipotesi) con dati mancanti.

Il Paragrafo 2 fornisce, in sintesi, gli elementi del lavoro di Cheng e Chu (1996) su cui si basa la tecnica proposta, presentata nel Paragrafo 3. Il Paragrafo 4 fornisce i risultati di uno studio di simulazione effettuato per verificare l'accuratezza degli intervalli di confidenza ottenuti seguendo tale tecnica, mentre nel Paragrafo 5 sono presentati due esempi di applicazione. Alcuni commenti finali sono riportati nel Paragrafo 6.

Nel seguito, si farà sempre riferimento al caso più semplice in cui anche la variabile X , rilevata assieme alla variabile d'interesse Y , è univariata. L'estensione al caso più realistico in cui $X \in \mathfrak{R}^d$, $d > 1$, non presenta particolari problemi. Alcune indicazioni al riguardo sono date nel Paragrafo 2.4.

2 Stima non parametrica della funzione di ripartizione e dei quantili con dati MAR

2.1 Stima non parametrica della funzione di ripartizione e dei quantili

Assumiamo che X e Y siano variabili casuali univariate continue e supponiamo che l'informazione campionaria sia costituita dalle terne di tipo (1). Supponiamo cioè che tutte le determinazioni X_i siano osservate e che δ_i assuma valore 1 se Y_i è osservata, altrimenti valga 0. Assumiamo, inoltre, che il meccanismo che determina la non osservabilità della variabile Y sia di tipo MAR e poniamo

$$p(X) = \text{pr} \{ \delta = 1 | X \} . \quad (2)$$

La tecnica proposta da Cheng e Chu (1996), per stimare la funzione di ripartizione (e i quantili) di Y , utilizza lo stimatore di Nadaraya e Watson (1964) per la funzione di regressione come strumento di base per una procedura di imputazione in ambito non parametrico. In effetti Cheng e Chu propongono di stimare la funzione di ripartizione di Y , $G(y) = \text{pr}\{Y \leq y\}$, combinando l'usuale funzione di ripartizione empirica e una stima della funzione di ripartizione condizionata $G(y|x) = \text{pr}\{Y \leq y|x\}$, calcolata per ogni valore reale y ed ogni X_i con $\delta_i = 0$, per il quale cioè la corrispondente risposta Y_i non è osservata. Operativamente, la stima della funzione di ripartizione condizionata $G(y|X_i)$ è definita come

$$\hat{G}(y|X_i) = \sum_{j \neq i} W_j(X_i) I(Y_j \leq y), \quad (3)$$

dove $I(\cdot)$ è la funzione indicatrice e

$$W_j(X_i) = \frac{K_b(X_j - X_i) \delta_j}{\sum_{j \neq i} K_b(X_j - X_i) \delta_j}, \quad (4)$$

con $K_b(X_j - X_i) = b^{-1}K((X_j - X_i)/b)$, è una sequenza di pesi. Qui $K(\cdot)$ è la funzione nucleo che integra ad 1 sull'asse reale e $b = b(N)$ è una successione di valori per il parametro di lisciamento che tende a 0 per $N \rightarrow \infty$. Lo stimatore per la funzione di ripartizione di Y , proposto da Cheng e Chu (1996), risulta essere, allora,

$$\hat{G}(y) = N^{-1} \sum_{i=1}^N \{\delta_i I(Y_i \leq y) + (1 - \delta_i) \hat{G}(y|X_i)\}. \quad (5)$$

Dallo stimatore per la funzione di ripartizione si può dedurre in maniera ovvia lo stimatore per un generico quantile della distribuzione di Y . In pratica, con riferimento al q -esimo quantile teorico θ_0 , definito come $\theta_0 = G^{-1}(q) = \inf\{y : G(y) \geq q\}$, per $0 < q < 1$, l'equivalente campionario risulta essere, evidentemente,

$$\hat{\theta} = \hat{G}^{-1}(q) = \inf\{y : \hat{G}(y) \geq q\},$$

con $\hat{G}(\cdot)$ definita nella (5). Sotto opportune condizioni, valgono le proprietà di consistenza (in senso forte) e di normalità asintotica per $\hat{G}(y)$ e $\hat{G}^{-1}(q)$.

2.2 Consistenza

Supponiamo che la variabile X sia definita su un supporto compatto e che abbia funzione di densità $f(x)$. Poniamo

$$g(x) = f(x)p(x)$$

e indichiamo con

$$\sigma^2(y|X) = \text{var}(I[Y \leq y]|X) = G(y|X) - G^2(y|X)$$

la varianza condizionata. Assumiamo, inoltre, che $\text{pr}\{\delta = 1\} = E[p(X)] = p$ sia una costante strettamente compresa tra 0 e 1 e consideriamo le condizioni:

- 1) $g(x) \geq 2c > 0$, dove c è una costante fissata arbitrariamente e g ha derivata prima g' limitata all'interno del supporto compatto di X ;
- 2) il nucleo $K(\cdot)$ assume valori finiti ed è una funzione di densità simmetrica con supporto compatto (ad esempio $[-1, 1]$);
- 3) la sequenza di valori per il parametro di liscio b è tale che $b \rightarrow 0$ e $Nb^5 \simeq \log N$ quando $N \rightarrow \infty$;
- 4) la derivata rispetto a x , di $G(y|x)$, indicata con $G'(y|x)$, è limitata sia in x che in y .

Qui e nel seguito, il simbolo \simeq indica che due successioni di valori sono dello stesso ordine. In altri termini, se $a(N) \simeq b(N)$, allora $\lim_{N \rightarrow \infty} |a(N)/b(N)|$ è una costante finita e positiva. Vale il seguente teorema.

Teorema 1. (teorema 2.1 di Cheng e Chu, 1996) *Si supponga di disporre di un'osservazione campionaria di tipo (1), con dati MAR. Siano verificate le condizioni dalla 1) alla 4). Allora,*

$$\lim_{N \rightarrow \infty} \sup (Nb/\log N)^{1/2} \sup_{y \in \mathfrak{R}} |\hat{G}(y) - G(y)| \leq (\|K\|/c)^{1/2} + C/c, \quad (6)$$

dove $C \geq \sup_{x,y} (\|G'(y|x)\|) \sup_x (\|g'(x)\|) (\int_{-1}^1 u^2 K(u) du)$ e $\|K\| = \sup_u \|K(u)\|$.

La parte più stringente della condizione 1), e in particolare la richiesta di supporto compatto per X e la richiesta che la densità $f(x)$ sia limitata in basso (richiesta implicita nell'assunzione $g(x) \geq c > 0$), sono requisiti abbastanza usuali nella letteratura relativa alla regressione non parametrica. Come si vedrà più avanti, queste richieste possono essere indebolite quando lo scopo è quello di stabilire la distribuzione asintotica di $\hat{G}(y)$. Inoltre, dai risultati di uno studio di simulazione effettuato da Cheng e Chu (1996) per valutare il comportamento dello stimatore $\hat{G}(y)$, è emerso che la condizione di regolarità sulla funzione $g(x) = f(x)p(x)$ sembra non essere cruciale. Più specificatamente, cioè, il fatto che $p(x)$ presenti un piccolo numero di punti di discontinuità all'interno del supporto di X sembra non incidere in modo rilevante sulle buone proprietà di $\hat{G}(y)$.

Il Teorema 1 fornisce anche l'indicazione sulla velocità di convergenza uniforme di $\hat{G}(\cdot)$ a $G(\cdot)$. La condizione 3) è necessaria per ottenere questo tipo di informazione. Per ottenere il solo risultato sulla consistenza uniforme, è tuttavia sufficiente richiedere che $Nb/\log N \rightarrow \infty$ quando $N \rightarrow \infty$.

Il risultato sulla consistenza di $\hat{\theta}$ può essere ottenuto come conseguenza del Teorema 1 (si veda l'Osservazione 2.2 in Cheng e Chu, 1996).

2.3 Normalità asintotica.

Come già accennato, la parte più stringente della condizione 1), ossia il fatto che $g(x)$ sia una funzione strettamente maggiore di una costante positiva sul supporto compatto di X , è necessaria per ottenere il risultato sulla consistenza uniforme dato

nel Teorema 1. Tale condizione esclude di fatto possibili fluttuazioni radicali dei pesi $W_j(X_i)$. Senza entrare nei dettagli, consideriamo la quantità

$$\hat{g}(X_i) = \sum_{j \neq i} K_b(X_j - X_i) \delta_j / (N - 1)$$

e riscriviamo quindi i pesi $W_j(X_i)$ nel seguente modo

$$W_j(X_i) = K_b(X_j - X_i) \delta_j / ((N - 1)\hat{g}(X_i)) .$$

È facile convincersi che, quando $\hat{g}(X_i)$ assume valori molto piccoli, i pesi $W_j(X_i)$ possono portare a stimare in modo instabile $\hat{G}(y | X_i)$. Questo fenomeno si può verificare specialmente nelle aree con dati sparsi, ossia nelle regioni a bassa densità di osservazioni, quali possono essere quelle relative alle code della densità $f(x)$ o le aree in cui sono “presenti” molti dati mancanti. In pratica, cioè, soprattutto in queste regioni, proprio perchè le osservazioni sono “distanti” tra di loro, i pesi $W_j(X_i)$ potrebbero fluttuare, nel senso che potrebbero assumere valori anche molto diversi l’uno dall’altro. La parte più stringente della condizione 1) garantisce che

$$\hat{g}(X) \geq c + o(1)$$

con probabilità 1 quando $N \rightarrow \infty$ (si veda la relazione 2.9 di Cheng e Chu, 1996) ed esclude di fatto tali possibili fluttuazioni dei pesi $W_j(X_i)$, considerando gli insiemi $\{\hat{g}(X_i) < c\}$, per ogni costante $c > 0$ piccola fissata, eventi a probabilità nulla.

Quando lo scopo è ottenere la proprietà di consistenza in media quadratica, e in particolare stabilire la distribuzione asintotica di $\hat{G}(y)$, la condizione 1) può essere indebolita utilizzandone una versione meno stringente che ne conservi la richiesta delle caratteristiche di regolarità per $g(x)$, ma ammettendo che X possa essere definita su un supporto non compatto:

1*) $g(x)$ ha derivata prima limitata.

Ora, per evitare possibili fluttuazioni dei pesi W_j , non potendo contare sulla parte stringente della condizione 1) che, come già detto, “taglia” gli eventi $\{\hat{g}(X_i) < c\}$ rendendoli insiemi a probabilità nulla, è necessario ricorrere ad una versione “troncata” dello stimatore $\hat{G}(y | X_i)$, e quindi di $\hat{G}(y)$. Tale versione si ottiene “troncando” i pesi $W_j(X_i)$ secondo la relazione

$$W_j(X_i) I[\hat{g}(X_i) \geq c] , \tag{7}$$

dove $\hat{g}(\cdot)$ è definita in precedenza e $c = c(N) \simeq b^t$, con $0 < t \leq 1/4$, è una sequenza decrescente di costanti. Tale sequenza è chiamata sequenza di troncamento. Il “troncamento” degli eventi $\{\hat{g}(X_i) < c\}$ è quindi adesso direttamente espresso nello stimatore.

Introducendo le due ulteriori condizioni

3*) $\log N = O(Nb^2)$,

5) $E[\sigma^2(y | X) | p(X)] < \infty$ e $E[\{1 - p(X)\}\{G'(y | X)/g(X)\}^2] < \infty$,

il risultato sulla normalità asintotica della versione troncata di $\hat{G}(y)$ è dato dal seguente teorema.

Teorema 2. (teorema 3.1 di Cheng e Chu, 1996) *Si supponga di disporre di un'osservazione campionaria di tipo (1), con dati MAR. Siano verificate le condizioni 1*), 2), 3*), 4) e 5). Si supponga inoltre che $Nb^4 \rightarrow 0$ per $N \rightarrow \infty$ e che i pesi in $\hat{G}(y)$ risultino troncati secondo la (7). Allora, per ogni fissato y ,*

$$N^{1/2}[\hat{G}(y) - G(y)]$$

converge in distribuzione ad una variabile casuale normale di media 0 e varianza

$$\sigma^2(y) = E[\sigma^2(y|X)/p(X)] + E[G^2(y|X)] - G^2(y) . \quad (8)$$

Naturalmente, sotto le condizioni più forti date in precedenza, il risultato sulla distribuzione asintotica rimane valido anche per la versione non troncata di $\hat{G}(y)$. Inoltre, il risultato fornito dal Teorema 2 permette anche di scegliere in maniera ottimale la sequenza di valori per il parametro di lisciamento b . In particolare, la scelta ottimale per b può essere determinata minimizzando l'errore quadratico medio $E[\hat{G}(y) - G(y)]^2$. Ciò porta a stabilire (si veda Cheng e Chu, 1996, per i dettagli) che deve essere

$$b_{opt} \simeq N^{-2/5} .$$

Il risultato relativo alla normalità asintotica può essere esteso al caso dei quantili di Y . Vale infatti il seguente teorema.

Teorema 3. (teorema 3.2 di Cheng e Chu, 1996) *Siano verificate le condizioni del Teorema 2. Si supponga che G abbia densità $G^{(1)}(\cdot)$ positiva e continua in θ_0 , con $\theta_0 = G^{-1}(q)$ e $0 < q < 1$. Allora la quantità*

$$N^{1/2}(\hat{G}^{-1}(q) - \theta_0) ,$$

tende in distribuzione ad una variabile casuale normale che ha media nulla e varianza pari a $(\sigma(\theta_0)/G^{(1)}(\theta_0))^2$, dove $\sigma^2(\cdot)$ è definita dalla (8).

2.4 Estensione al caso in cui $X \in \mathfrak{R}^d$.

I risultati visti finora per X univariata, si estendono facilmente al caso in cui X è un vettore di dimensione d , con $d > 1$. In particolare, a titolo esemplificativo, di seguito vengono elencate le versioni modificate delle condizioni 1) – 4), indicate con 1') – 4'), che stanno alla base di tale estensione (si veda l'Osservazione 2.1 in Cheng e Chu, 1996). Più precisamente, quando X ha valori in \mathfrak{R}^d , si ha che $K_b(X_j - X_i) = b^{-d}K((X_j - X_i)/b)$ e le versioni modificate delle condizioni 1) – 4) sono:

- 1') $g(x) \geq 2c > 0$ e g ha derivate parziali limitate, entro il supporto compatto di X , fino all'ordine $k - 1$ dove k è pari e $2 \leq k \leq d$.

2') $K(\cdot)$ è la funzione nucleo di ordine k , ad esempio $K = \prod_{i=1}^d K_0$, dove la funzione nucleo simmetrica unidimensionale K_0 soddisfa

$$\int_{-1}^1 u^l K_0(u) du \begin{cases} = 1 & \text{se } l = 0 \\ = 0 & \text{se } 1 \leq l \leq k-1 \\ \neq 0 & \text{se } l = k \end{cases} .$$

3') $Nb^{d+2k} \simeq \log N$ se $N \rightarrow \infty$, con $d \leq 2k$.

4') $G(y|x)$ ha derivate parziali limitate rispetto ad x , all'interno del supporto di X , fino all'ordine $k-1$.

Sotto tali condizioni (e opportune varianti del tipo la 1*) e la 3*)) e la condizione 5), i risultati sulla convergenza uniforme e sulla normalità asintotica richiamati nei paragrafi precedenti valgono anche quando $X \in \mathfrak{R}^d$ ($d > 1$). Ciò, di fatto, determina l'immediata estensione al caso $X \in \mathfrak{R}^d$ della tecnica discussa in questo lavoro.

3 Intervalli di confidenza per i quantili: l'approccio proposto

Intervalli di confidenza per i quantili della variabile Y , quando completamente osservata, possono essere ottenuti via verosimiglianza empirica. Più precisamente, indicando con $\theta(\cdot)$ il funzionale che definisce il q -esimo quantile $\theta_0 = G^{-1}(q)$ di Y (per $0 < q < 1$) e assumendo che G sia strettamente crescente in un intorno di θ_0 , l'espressione della funzione log-rapporto di verosimiglianza empirica per $\theta(\cdot)$, basata su N determinazioni i.i.d. Y_1, Y_2, \dots, Y_N di Y e calcolata nel generico punto θ , risulta essere

$$l_E(\theta) = 2N \left\{ F_N(\theta) \log \frac{F_N(\theta)}{q} + \{1 - F_N(\theta)\} \log \frac{1 - F_N(\theta)}{1 - q} \right\} \quad (9)$$

(si veda Adimari, 1998). L'espressione (9), nella quale $F_N(\cdot)$ indica l'usuale funzione di ripartizione empirica ottenuta dal campione Y_1, Y_2, \dots, Y_N , è corretta se $\theta \in [Y_{(1)}, Y_{(n)}]$, ossia se il punto θ è compreso tra il più piccolo e il più grande dei valori osservati. Altrimenti, $l_E(\theta) = +\infty$. Inoltre,

$$l_E(\theta_0) \xrightarrow{d} \chi_1^2.$$

Quindi, per un fissato livello di copertura nominale γ , intervalli di confidenza per il q -esimo quantile di Y possono essere ottenuti come insiemi del tipo

$$\{\theta : l_E(\theta) \leq c_\gamma\},$$

con c_γ tale che $\text{pr}\{\chi_1^2 \geq c_\gamma\} = 1 - \gamma$.

Se lo scopo è ottenere intervalli di confidenza per i quantili, quando la variabile Y non è completamente osservata (ed è rilevata assieme ad X), si può pensare di agire modificando opportunamente la statistica $l_E(\theta)$. L'idea è quella di sostituire alla funzione di ripartizione empirica $F_N(\cdot)$ nella (9) lo stimatore $\hat{G}(\cdot)$ della funzione di ripartizione di Y proposto da Cheng e Chu (1996). In particolare, nel seguito

faremo riferimento alla versione troncata dello stimatore $\hat{G}(y)$, la cui espressione è fornita dalle equazioni (3), (4) e (7). L'approccio sembra plausibile in quanto lo stimatore $\hat{G}(\cdot)$ rappresenta un'estensione, al contesto di dati MAR, della funzione di ripartizione empirica $F_N(\cdot)$, della quale ripropone diverse proprietà e caratteristiche.

Quindi, la versione modificata della funzione log-rapporto di verosimiglianza empirica per il q -esimo quantile, $\theta_0 = G^{-1}(q)$, per $0 < q < 1$, quando la variabile Y non è completamente osservata, ha espressione

$$l_E^*(\theta) = 2N \left\{ \hat{G}(\theta) \log \frac{\hat{G}(\theta)}{q} + \{1 - \hat{G}(\theta)\} \log \frac{1 - \hat{G}(\theta)}{1 - q} \right\}. \quad (10)$$

Naturalmente per poter utilizzare la statistica l_E^* per costruire regioni di confidenza per θ_0 , è necessario conoscerne la distribuzione, almeno asintoticamente.

Teorema 4. *Sia G continua e strettamente crescente in un intorno del suo q -esimo quantile θ_0 ($0 < q < 1$). Siano verificate le condizioni del Teorema 2. Allora*

$$k(\theta_0)l_E^*(\theta_0) \xrightarrow{d} \chi_1^2,$$

con

$$k(\theta) = \frac{q(1-q)}{\sigma^2(\theta)},$$

dove $\sigma^2(\cdot)$ è definita dalla (8).

Dimostrazione. Poichè $\hat{G}(\theta_0)$ è uno stimatore consistente per $G(\theta_0) = q$ e, in base al Teorema 2, $\hat{G}(\theta_0) = q + O_p(N^{-1/2})$, attraverso lo sviluppo in serie di Taylor della funzione $\log(z)$ attorno a $z=1$,

$$\log(z) = z - 1 - (z - 1)^2/2 + o((z - 1)^2),$$

dalla (10) si ottiene, dopo alcuni passaggi,

$$l_E^*(\theta_0) = N\{\hat{G}(\theta_0) - q\}^2 \left\{ \frac{2q - 3q^2 - \hat{G}(\theta_0) + 2q\hat{G}(\theta_0)}{q^2(1-q)^2} \right\} + o_p(1).$$

Usando nuovamente la consistenza di $\hat{G}(\theta_0)$, si ha

$$l_E^*(\theta_0) = N \frac{\{\hat{G}(\theta_0) - q\}^2}{q(1-q)} + o_p(1). \quad (11)$$

Il risultato segue immediatamente dal Teorema 2, ossia dalla normalità asintotica di $N^{1/2}\{\hat{G}(\theta_0) - q\}$. \square

Il Teorema 4 stabilisce che, a meno del fattore di correzione $k(\theta_0)$, la statistica $l_E^*(\theta_0)$ ha distribuzione asintotica χ_1^2 . Inoltre, si può notare che $k(\theta_0)$ è un rapporto tra varianze e in particolare tra la varianza asintotica di $F_N(\theta_0)$, cioè dello stimatore della funzione di ripartizione di Y quando quest'ultima è completamente osservata, e

la varianza asintotica di $\hat{G}(\theta_0)$, cioè dello stimatore per $G(\theta_0)$ quando la variabile Y non è completamente osservata. Naturalmente, da un punto di vista pratico, ci sono diversi modi asintoticamente equivalenti per correggere la statistica $l_E^*(\theta_0)$ in modo da ricondurne la distribuzione asintotica a quella χ_1^2 . In particolare, una possibilità è quella di stimare $k(\theta_0)$ attraverso la quantità

$$\hat{k}(\theta_0) = \frac{\hat{G}(\theta_0)(1 - \hat{G}(\theta_0))}{\hat{\sigma}^2(\theta_0)},$$

dove

$$\hat{\sigma}^2(\theta_0) = \frac{1}{N} \sum_{i=1}^N \left[\frac{\hat{G}(\theta_0 | X_i) - \hat{G}^2(\theta_0 | X_i)}{\hat{p}(X_i)} + \hat{G}^2(\theta_0 | X_i) \right] - \hat{G}^2(\theta_0) \quad (12)$$

è semplicemente il corrispondente empirico della varianza asintotica $\sigma^2(\theta_0)$. Nella (12), $\hat{G}(\theta | X_i)$ è la versione troncata dello stimatore (3) della funzione di ripartizione condizionata, mentre

$$\hat{p}(x) = \frac{\sum_{j=1}^N K_b(X_j - x) \delta_j}{\sum_{j=1}^N K_b(X_j - x)}, \quad (13)$$

è uno stimatore consistente per $p(x)$ (si veda anche Cheng, 1994), ossia per la probabilità di osservare Y dato $X = x$. La sequenza di valori per il parametro di liscio b nella (13) non deve essere necessariamente la stessa usata per la stima di $G(\cdot | X_i)$ (e quindi di $G(\cdot)$).

Corollario 1. *Sotto le condizioni del Teorema 4,*

$$\hat{k}(\theta_0) l_E^*(\theta_0) \xrightarrow{d} \chi_1^2.$$

Una possibilità alternativa, che è poi quella considerata nel seguito, è quella di stimare $k(\theta_0)$ attraverso la quantità $\hat{k}(\hat{\theta})$. Questo approccio presenta il vantaggio di fornire un *pivot*

$$l_E^{**}(\theta) = \hat{k}(\hat{\theta}) l_E^*(\theta)$$

la cui forma è, di fatto, quella della funzione $l_E^*(\theta)$. È necessario sottolineare che, da un punto di vista tecnico, la sostituzione di θ_0 con $\hat{\theta}$ nella funzione $\hat{k}(\cdot)$ non è banale, visto che quest'ultima non è una funzione continua. Sebbene una dimostrazione rigorosa sia necessaria, è comunque ragionevole ritenere (e i risultati di simulazione riportati nelle pagine che seguono sembrano confermare queste congetture) che tale sostituzione non alteri la distribuzione asintotica della statistica almeno sotto le ipotesi qui considerate e che, quindi,

$$l_E^{**}(\theta_0) \xrightarrow{d} \chi_1^2.$$

In definitiva, sulla base delle argomentazioni appena sostenute, insiemi del tipo

$$\{\theta : l_E^{**}(\theta) \leq c_\gamma\},$$

con c_γ tale che $\text{pr}\{\chi_1^2 \geq c_\gamma\} = 1 - \gamma$, costituiscono delle regioni di confidenza per θ_0 di livello nominale γ quando la variabile Y non è completamente osservata. È

evidente che la funzione $l_E^*(\cdot)$ definita nella (10) è una funzione a gradini con salti nei valori di Y osservati, in quanto tale è lo stimatore della funzione di ripartizione di Y , $\hat{G}(\cdot)$. Di conseguenza $l_E^{**}(\theta)$ è una funzione a gradini con salti nei valori di Y osservati e quindi le regioni di confidenza costruite secondo l'approccio proposto sono in realtà intervalli del tipo $[Y_a, Y_b)$, dove Y_a e Y_b , $Y_a < Y_b$, sono due valori di Y osservati.

4 Alcuni risultati di simulazione

4.1 Descrizione dello studio di simulazione.

Per verificare l'accuratezza nel finito degli intervalli di confidenza ottenuti seguendo l'approccio proposto, è stato effettuato uno studio di simulazione. In particolare si sono considerate cinque batterie di esperimenti, ognuna corrispondente ad una specifica scelta del quantile d'interesse θ_0 : mediana ($q = 0.5$), primo e terzo quartile ($q = 0.25, q = 0.75$), primo e nono decile ($q = 0.1, q = 0.9$). Per ogni batteria i dati sono stati generati da due diverse distribuzioni:

$$A) (X, Y) \sim \text{normale bivariata}((0, 0), (1, 1), \rho),$$

$$B) (X, Y) \sim t_3 \text{ bivariata}((0, 0), (1, 1), \rho),$$

scegliendo dei valori per il coefficiente di correlazione ρ pari a 0.8 e 0 nel caso A) e 0.5 nel caso B).

Per ogni esperimento, dalla distribuzione considerata sono stati generati 10000 campioni di dimensione N , prendendo per N i seguenti valori: 20, 30 quando $q = 0.25, 0.5, 0.75$; 60, 70 quando $q = 0.1, 0.9$. Sono inoltre stati utilizzati cinque diversi meccanismi per la determinazione dei dati mancanti, ciascuno relativo ad una particolare scelta della funzione $p(x)$. Più precisamente le funzioni $p(x)$ utilizzate sono:

$$1) p_1(x) = \begin{cases} 0.44 + 0.2|x| & \text{se } |x| \leq 2.76 \\ 1 & \text{altrimenti} \end{cases}$$

$$2) p_2(x) = \begin{cases} 0.9 - 0.2|x| & \text{se } |x| \leq 4.5 \\ 0.1 & \text{altrimenti} \end{cases}$$

$$3) p_3(x) = I(x \leq 0) + 0.5 I(x > 0)$$

$$4) p_4(x) = 0.6 \quad \forall x$$

$$5) p_5(x) = 0.75 \quad \forall x .$$

I tassi di dati mancanti corrispondenti, nel caso della distribuzione normale, sono rispettivamente:

$$1) 1 - E[p_1(X)] = 0.4,$$

- 2) $1 - E[p_2(X)] = 0.26,$
- 3) $1 - E[p_3(X)] = 0.25,$
- 4) $1 - E[p_4(X)] = 0.4,$
- 5) $1 - E[p_5(X)] = 0.25.$

In tutti i casi considerati, in ogni campione generato è stata calcolata la statistica $l_E^{**}(\theta_0)$, con θ_0 vero valore del quantile, e, in alcuni casi, l'ampiezza del relativo intervallo di confidenza con livello di copertura nominale del 95%. Evidentemente, il calcolo della statistica $l_E^{**}(\theta)$ impone la scelta della funzione nucleo $K(\cdot)$ sia per la stima $\hat{G}(\theta)$ di $G(\theta)$, sia per la stima $\hat{p}(x)$ (si veda la (13)) della funzione $p(x)$. Nello studio sono state utilizzate 3 funzioni nucleo $K(\cdot)$, tutte con supporto su $[-1,1]$:

- 1) Epanechnikov

$$K_1(u) = \begin{cases} 3/4(1 - u^2) & \text{se } |u| \leq 1 \\ 0 & \text{altrimenti,} \end{cases}$$

- 2) Uniforme

$$K_2(u) = \begin{cases} 1/2 & \text{se } |u| \leq 1 \\ 0 & \text{altrimenti,} \end{cases}$$

- 3) Normale troncata

$$K_3(u) = \begin{cases} \phi(u)/[1 - 2\Phi(-1)] & \text{se } |u| \leq 1 \\ 0 & \text{altrimenti,} \end{cases}$$

dove $\phi(u)$ e $\Phi(u)$ indicano rispettivamente la funzione di densità e la funzione di ripartizione della distribuzione normale standard. In particolare, fissato il quantile d'interesse, la numerosità campionaria e la funzione $p(x)$ per i dati mancanti, gli esperimenti di simulazione sono stati realizzati con riferimento ad una classe di 3 schemi, ognuno dei quali esprime una diversa combinazione tra la scelta della funzione nucleo utilizzata nella stima di $G(\cdot)$ e quella utilizzata nella stima di $p(x)$:

- Schema 1: usa la funzione nucleo Epanechnikov per la stima di $G(\cdot)$ e la funzione nucleo Epanechnikov per la stima di $p(x)$,
- Schema 2: usa la funzione nucleo Normale troncata per la stima di $G(\cdot)$ e la funzione nucleo Epanechnikov per la stima di $p(x)$,
- Schema 3: usa la funzione nucleo Uniforme per la stima di $G(\cdot)$ e la funzione nucleo Epanechnikov per la stima di $p(x)$.

Come si può osservare, si è deciso di utilizzare sempre il nucleo di Epanechnikov per la stima della funzione $p(x)$ e di alternare invece l'uso delle 3 funzioni nucleo per la stima della funzione di ripartizione.

Per quanto riguarda la sequenza di valori per il parametro di lisciamiento $b(N)$

relativa allo stimatore $\hat{G}(\cdot)$, si è scelto una sequenza del tipo $aN^{-2/5}$ e quindi ottimale, a livello teorico. Più precisamente si è posto $b(N) = 5\hat{\sigma}(X)N^{-2/5}$, dove $\hat{\sigma}(X)$ indica l'usuale stima campionaria dello scarto quadratico medio di X , in quanto è naturale che tale scelta dipenda, operativamente, dalla variabilità della variabile completamente osservata X . Per la stima della funzione $p(x)$, si è scelto $b(N) = 2.5\hat{\sigma}(X)N^{-2/5}$. Infine per la sequenza di troncamento $c = c(N)$, della (7), si è scelto di porre $c = 0.005b^{1/4}$. Indicazioni sulla scelta dei valori per la sequenza dei parametri di liscio nelle stime di $G(\cdot)$ e $p(x)$ e per la sequenza di troncamento $c(N)$ si sono ottenute da alcuni esperimenti di simulazione preliminari in cui si sono sondate possibili alternative (per esempio, per la scelta della sequenza $c(N)$ si sono sondate le alternative $0.0005b^{1/4}$, $0.0001b^{1/4}$ e $0.005b^{1/4}$).

4.2 I risultati.

Nelle pagine che seguono sono riportate 15 tabelle contenenti i risultati dello studio effettuato. La singola tabella è costruita per uno specifico quantile d'interesse e una determinata distribuzione da cui sono generati i dati e riporta i livelli di copertura reali degli intervalli di confidenza basati su $l_{F'}^{**}(\theta)$ corrispondenti ai livelli di copertura nominali (γ) 0.99, 0.95 e 0.9, per i differenti valori di N , i vari meccanismi per la determinazione dei dati mancanti (funzioni $p(x)$) e i tre schemi di inferenza considerati. Per ogni esperimento viene riportata anche l'ampiezza media dell'intervallo di confidenza di livello 0.95 relativo allo Schema 1, ossia relativo all'uso della statistica $l_{F'}^{**}(\theta)$ calcolata ricorrendo alla funzione nucleo Epanechnikov tanto per la stima di $G(\cdot)$ quanto per la stima di $p(x)$. La scelta di riportare tale informazione solo con riferimento allo schema inferenziale 1 è dettata da un'esigenza di chiarezza espositiva, visto che, comunque, i tre schemi considerati forniscono, in generale, dei risultati molto simili tra loro. Infine, solo nel caso di dati generati dalla distribuzione normale, sono riportati i tassi di dati mancanti associati alle varie funzioni $p(x)$ considerate.

4.3 Analisi dei risultati e commenti.

Esaminando i risultati di simulazione riportati nelle varie tabelle emerge che, nel caso del primo quartile ($q = 0.25$), gli intervalli di confidenza quando i dati hanno distribuzione normale bivariata con coefficiente di correlazione $\rho = 0.8$ (Tabella 1) sono abbastanza accurati già con $N = 20$ e con un tasso di dati mancanti basso. Aumentando la numerosità soltanto di dieci unità, ossia considerando il caso $N = 30$, i risultati sono decisamente più buoni per tutti e tre gli schemi selezionati anche con tassi di dati mancanti più elevati. Inoltre, passando da $N = 20$ a $N = 30$ l'ampiezza media degli intervalli di confidenza si riduce, come è naturale aspettarsi, in quanto si hanno più informazioni a disposizione (ricordiamo che nelle tabelle sono riportate solo le ampiezze medie degli intervalli con livello nominale 0.95 e relativi allo schema inferenziale 1). Sempre con distribuzione normale, quando la correlazione tra le variabili X e Y è nulla ($\rho = 0$, Tabella 6), si può notare che le ampiezze degli intervalli sono mediamente più grandi rispetto al caso di correlazione positiva. Anche in questo caso le ampiezze si riducono e i livelli di copertura si avvicinano

generalmente ai valori nominali quando si passa da $N = 20$ a $N = 30$. È ragionevole pensare che esista una certa relazione tra l'ampiezza media dell'intervallo di confidenza e la correlazione tra le variabili X e Y . Più le variabili X e Y sono correlate, più le ampiezze medie degli intervalli di confidenza sono piccole, in generale, perché di fatto aumenta l'efficienza della procedura inferenziale, nel senso che questa sfrutta la maggiore capacità informativa della variabile X sulla risposta Y . Infine, nel caso di distribuzione t_3 bivariata (Tabella 11), quando $N = 30$, gli intervalli di confidenza sono sufficientemente accurati con qualsiasi funzione per i dati mancanti considerata mentre, con una numerosità di 20, i livelli di copertura reali possono scostarsi significativamente dai valori nominali in particolare quando $\gamma = 0.99$.

Quando $q = 0.75$, ossia l'interesse è rivolto al terzo quartile, nel caso di distribuzione normale, sia con $\rho = 0.8$ che con $\rho = 0$ (Tabelle 3 e 8) i risultati ottenuti sono accettabili quando $N = 20$ e migliorano quando aumenta la numerosità campionaria ($N = 30$). I risultati peggiori si hanno quando $\rho = 0.8$ e i dati mancanti sono determinati dalla funzione $p_3(x)$. Ciò è probabilmente dovuto alla forma della funzione $p_3(x)$ che è definita in modo che i valori mancanti nella risposta siano relativi a valori di X positivi. Più precisamente, se $x < 0$, $p_3(x) = \text{pr}\{\delta = 1 | x\} = 1$, ossia a determinazioni negative di X corrispondono determinazioni di Y osservate con probabilità 1; se $x > 0$, $p_3(x) = 0.5$ e quindi operativamente Y_i è mancante con una probabilità del 50% quando $X_i > 0$. Siccome la funzione $p_3(x)$ è definita in modo che solo se $X_i > 0$ la risposta Y_i corrispondente può essere mancante, poiché il terzo quartile corrisponde al valore 0.67449 (distribuzione normale) è intuitivo pensare di avere risultati di simulazione peggiori quando c'è forte correlazione positiva tra le variabili, in quanto la regione dove giace il quartile può risultare "più ricca" di dati mancanti. Evidentemente, questo fenomeno si attenua o scompare quando $\rho = 0$, cioè quando X e Y sono incorrelate (Tabella 8), e si ripropone quando la distribuzione dei dati è t_3 bivariata ($q = 0.5$, Tabella 12).

Se il quantile d'interesse è la mediana ($q = 0.5$), quando la distribuzione considerata è normale con $\rho = 0.8$ (Tabella 2), i risultati sono già molto buoni con $N = 20$ per qualsiasi funzione per i dati mancanti $p(x)$, quindi anche con alti tassi di dati mancanti. Quando $\rho = 0$ (Tabella 7) si osserva una generale riduzione dell'ampiezza media degli intervalli. Infine, anche con dati generati dalla distribuzione t_3 bivariata (Tabella 12), le simulazioni effettuate forniscono dei risultati molto buoni già con $N = 20$ per tutti i casi considerati, anche se leggermente peggiori rispetto a quelli ottenuti nel caso di distribuzione normale.

Analizzando i risultati di simulazione relativi al nono decile ($q = 0.9$), sia nel caso di distribuzione normale che di distribuzione t_3 bivariata (Tabelle 5, 10 e 15) si nota che i livelli di copertura reali sono generalmente buoni quando $N = 70$, mentre per $N = 60$ la discrepanza tra livello reale e nominale può essere in alcuni casi rilevante quando $\gamma = 0.99$. Inoltre, come per il terzo quartile, anche in questo caso, quando la distribuzione è t_3 bivariata o normale con $\rho = 0.8$, i risultati peggiori si hanno in corrispondenza del meccanismo per i dati mancanti determinato dalla funzione $p_3(x)$.

Considerazioni analoghe a quelle fatte per il caso del nono decile valgono nel caso del primo decile ($q = 0.1$, Tabelle 4, 9 e 14). In questo caso i risultati peggiori si hanno quando il meccanismo per i dati mancanti è determinato dalla funzione

$p_2(x)$.

In definitiva, dall'analisi dei risultati ci sembra di poter concludere che:

- gli intervalli di confidenza basati su l_E^{**} sono sufficientemente accurati anche per campioni relativamente piccoli ($N = 20$ per la mediana, $N = 30$ per primo e terzo quartile, $N = 70$ per primo e nono decile), almeno per tassi di dati mancanti non superiori al 50%. Naturalmente, trattandosi di un approccio non parametrico, la numerosità campionaria richiesta per un'inferenza sufficientemente accurata cresce quanto più il percentile d'interesse è estremo;
- per N fissato, l'accuratezza degli intervalli si riduce all'aumentare del tasso di dati mancanti;
- l'ampiezza media degli intervalli tendenzialmente si riduce all'aumentare della numerosità campionaria e, per N fissato, quanto più è grande la correlazione tra X e Y ;
- i risultati ottenuti sembrano sufficientemente stabili rispetto alle possibili scelte della funzione nucleo da utilizzare nella stima della funzione di ripartizione $G(\cdot)$;
- a conferma di quanto ottenuto da Cheng e Chu (1996), la presenza di pochi punti di discontinuità nelle funzioni $p(x)$ considerate sembra non incidere in modo evidente sui risultati di simulazione.

5 Due esempi

In questo paragrafo sono riportati due esempi di applicazione, per illustrare operativamente l'uso del metodo proposto. Sui due insiemi di dati considerati è stata calcolata la versione modificata della funzione log-rapporto di verosimiglianza empirica, $l_E^{**}(\theta)$, considerando come parametro d'interesse la mediana della variabile risposta Y , in un caso, e la mediana, il terzo quartile e il primo decile nell'altro caso.

Per quanto riguarda la scelta delle sequenze di valori per i parametri di liscio e di troncamento coinvolti nel calcolo di $l_E^{**}(\theta)$, sono stati utilizzati i valori considerati nello studio di simulazione descritto nel Paragrafo 4.1; inoltre si è deciso di utilizzare la funzione nucleo Epanechnikov sia per la stima della funzione $p(x)$ che per la stima di $G(\cdot)$.

Il primo esempio riportato (si veda Schafer, 1997, pag. 195) si riferisce ai dati relativi ad un campione di 28 individui colpiti da attacco cardiaco, sui quali sono state rilevate le variabili X e Y , di cui X è completamente osservata mentre Y presenta circa il 30% (9 valori) di dati mancanti, e che indicano, rispettivamente, il livello di colesterolo nel sangue registrato dopo 4 e dopo 14 giorni dall'attacco.

La Figura 1 riporta il profilo della funzione $l_E^{**}(\theta)$ per il livello mediano di colesterolo dopo 14 giorni dall'attacco. La linea orizzontale individua l'intervallo di confidenza per θ_0 al 95%.

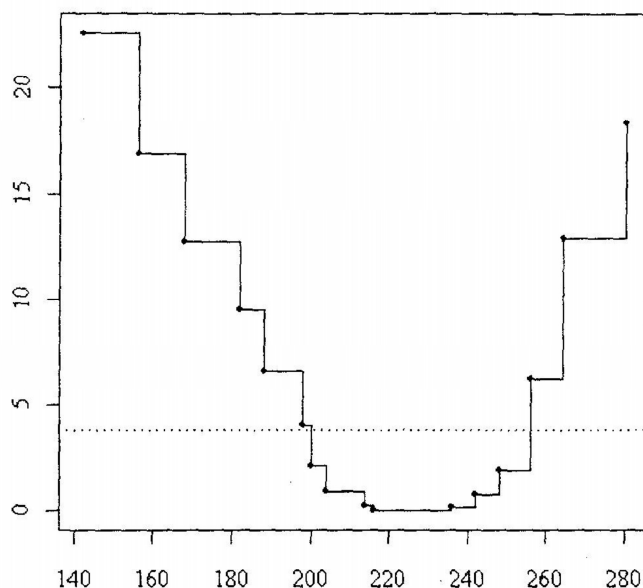


Figura 1: profilo della funzione $l_E^{**}(\theta)$ per la mediana della distribuzione del livello di colesterolo dopo 14 giorni dall'attacco cardiaco.

Il secondo esempio si riferisce a 88 dati (si veda Brinkman, 1981) relativi ad un esperimento condotto per rilevare l'ammontare di NO_x (monossido di azoto NO e di biossido di azoto NO_2), rappresentato dalla variabile Y , emesso da un motore che è stato fatto funzionare utilizzando come carburante l'etanolo. Assieme alla variabile risposta Y è rilevata anche la variabile X che rappresenta il rapporto di equivalenza con cui funziona il motore, ossia una misura di quanto è ricca la miscela di aria ed etanolo. Entrambe le variabili rilevate sono state completamente osservate. I dati mancanti per la variabile Y sono quindi stati determinati artificialmente con un meccanismo $p(x)$ così definito:

$$p(x) = I(x \leq 1) + 0.6 I(x > 1) .$$

Si sono ottenuti 17 valori non osservati per la variabile Y che corrispondono ad un tasso di dati mancanti del 20% circa.

Come nell'altro esempio, le Figure 2, 3 e 4 riportano l'andamento della funzione $l_E^{**}(\theta)$, rispettivamente per la mediana, il terzo quartile e il primo decile della distribuzione dell'ammontare di NO_x . Anche in questi casi la linea orizzontale individua l'intervallo di confidenza per θ_0 al 95%.

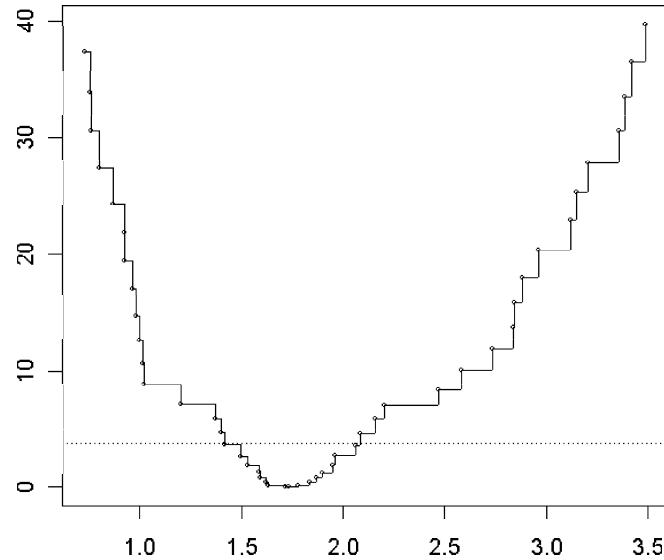


Figura 2: profilo della funzione $l_E^{**}(\theta)$ per la mediana della distribuzione dell'ammontare di NO_x emesso.

6 Note conclusive

In questo lavoro si studia una metodologia per la costruzione di intervalli di confidenza non parametrici per i quantili di una variabile risposta Y con dati MAR. Tale metodologia prevede l'uso di una statistica ottenuta combinando opportunamente la funzione di verosimiglianza empirica per i quantili e uno stimatore della funzione di ripartizione di Y , con dati MAR, proposto da Cheng e Chu (1996) e basato sul metodo del nucleo.

I risultati di simulazione ottenuti sono incoraggianti e mettono in luce come la metodologia proposta sia adeguata in questo contesto e possa consentire di ottenere degli intervalli di confidenza sufficientemente accurati anche in campioni di dimensione relativamente piccola. La tecnica proposta è inoltre semplice da usare. Da un punto di vista operativo, forse, le perplessità maggiori possono nascere dalla sostanziale arbitrarietà della scelta delle sequenze di valori per i parametri di liscio e di troncamento necessarie per le stime di $G(\cdot)$ e di $p(x)$. Si tratta di un problema comune alle procedure inferenziali basate sul metodo del nucleo; comunque, i valori utilizzati nello studio di simulazione (Paragrafo 4.1) sembrano fornire buoni risultati in generale e possono essere usati come valori di riferimento.

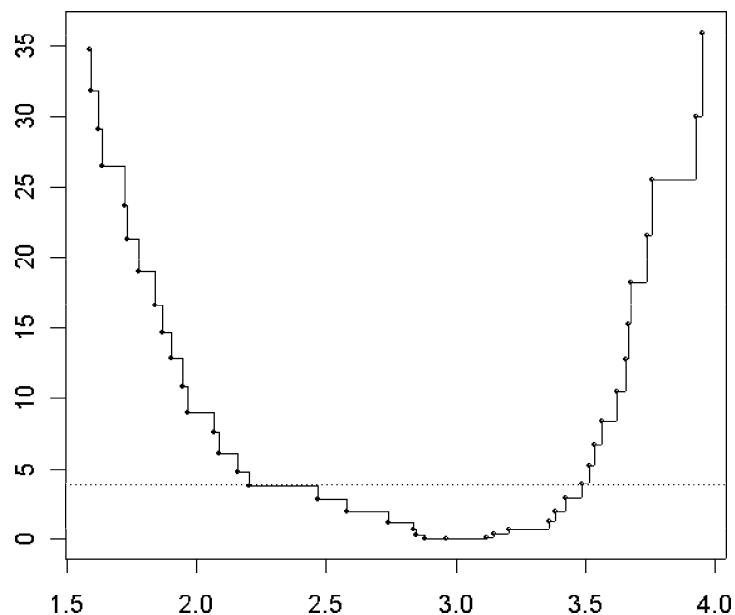


Figura 3: profilo della funzione $l_E^{**}(\theta)$ per il terzo quartile della distribuzione dell'ammontare di NO_x emesso.

Riferimenti bibliografici

Adimari G. (1998). An empirical likelihood statistic for quantiles. *J. Statist. Comput. Simul.*, 60, pp. 85-95.

Brinkman N.D. (1981). Ethanol fuel — a single cylinder engine study of efficiency and exhaust emissions. *SAE Transactions*, 90, No. 810345, pp. 1410-1427.

Cheng P.E. (1994). Nonparametric estimation of mean functionals with data missing at random. *J. Amer. Statist. Assoc.* 89, pp. 81-87.

Cheng P.E., Chu C.K. (1996). Kernel estimation of distribution functions and quantiles with missing data. *Statistica Sinica*, 6, pp. 63-78.

Nadaraya E.A. (1964). On estimating regression. *Theory Probab. Appl.* 9, pp. 141-142.

Owen, A.B. (2001). *Empirical likelihood*. Chapman and Hall, London.

Schafer J.L. (1997). *Analysis of incomplete multivariate data*. Monographs on Sta-

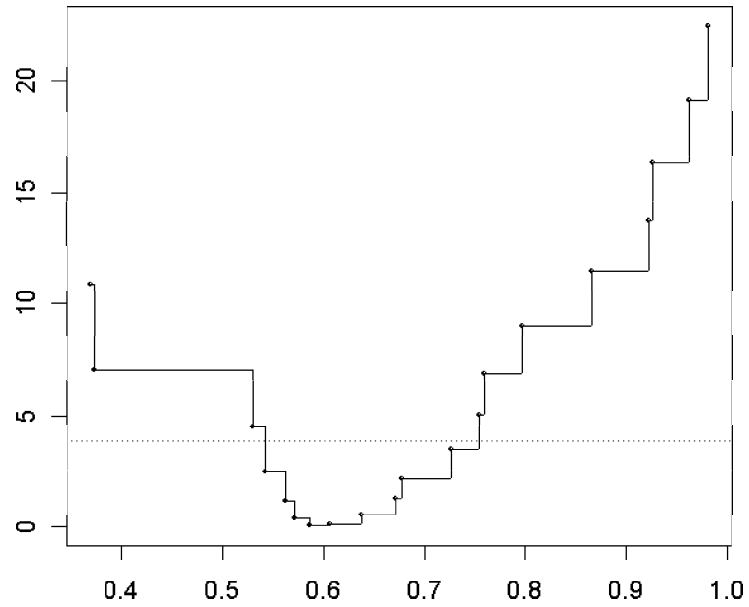


Figura 4: profilo della funzione $l_E^{**}(\theta)$ per il primo decile della distribuzione dell'ammontare di NO_x emesso.

tistics and Applied Probability, 72. Chapman & Hall, London.

Distribuzione $N((0,0), (1,1), 0.8)$			$N = 20$			$N = 30$		
			γ			γ		
$p(x)$	$1 - E[p(X)]$	schema	0.990	0.950	0.900	0.990	0.950	0.900
$p_1(x)$	0.4	1	0.965	0.936	0.889	0.989	0.953	0.904
		2	0.965	0.936	0.892	0.989	0.954	0.905
		3	0.964	0.937	0.892	0.989	0.954	0.905
			ampiezza = 1.677			ampiezza = 1.361		
$p_2(x)$	0.26	1	0.972	0.933	0.881	0.982	0.933	0.877
		2	0.973	0.936	0.884	0.982	0.934	0.880
		3	0.973	0.936	0.884	0.982	0.934	0.879
			1.340			ampiezza = 1.072		
$p_3(x)$	0.25	1	0.984	0.942	0.891	0.987	0.945	0.895
		2	0.983	0.939	0.885	0.986	0.942	0.892
		3	0.982	0.937	0.883	0.985	0.940	0.889
			ampiezza = 1.358			ampiezza = 1.079		
$p_4(x)$	0.4	1	0.958	0.935	0.887	0.986	0.946	0.894
		2	0.959	0.936	0.891	0.986	0.946	0.895
		3	0.959	0.936	0.891	0.986	0.946	0.895
			ampiezza = 1.546			ampiezza = 1.261		
$p_5(x)$	0.25	1	0.998	0.941	0.889	0.986	0.942	0.886
		2	0.980	0.941	0.888	0.987	0.942	0.886
		3	0.980	0.941	0.888	0.986	0.941	0.886
			ampiezza = 1.397			ampiezza = 1.110		

Tabella 1. Risultati di simulazione per il primo quartile ($q = 0.25$).

Distribuzione $N((0,0), (1,1), 0.8)$			$N = 20$			$N = 30$		
			γ			γ		
$p(x)$	$1 - E[p(X)]$	schema	0.990	0.950	0.900	0.990	0.950	0.900
$p_1(x)$	0.4	1	0.991	0.952	0.904	0.991	0.956	0.903
		2	0.992	0.954	0.906	0.991	0.957	0.907
		3	0.992	0.955	0.907	0.991	0.956	0.908
			ampiezza = 1.621			ampiezza = 1.258		
$p_2(x)$	0.26	1	0.990	0.950	0.897	0.992	0.955	0.908
		2	0.991	0.951	0.901	0.992	0.955	0.910
		3	0.991	0.950	0.899	0.992	0.954	0.910
			ampiezza = 1.244			ampiezza = 0.998		
$p_3(x)$	0.25	1	0.987	0.948	0.893	0.989	0.951	0.905
		2	0.986	0.943	0.886	0.989	0.947	0.899
		3	0.986	0.940	0.881	0.988	0.945	0.885
			ampiezza = 1.394			ampiezza = 1.096		
$p_4(x)$	0.4	1	0.990	0.954	0.908	0.989	0.952	0.903
		2	0.990	0.954	0.909	0.990	0.954	0.904
		3	0.990	0.954	0.909	0.990	0.954	0.904
			ampiezza = 1.483			ampiezza = 1.158		
$p_5(x)$	0.25	1	0.989	0.949	0.899	0.990	0.947	0.897
		2	0.989	0.949	0.901	0.991	0.948	0.899
		3	0.989	0.948	0.900	0.991	0.948	0.899
			ampiezza = 1.303			ampiezza = 1.033		

Tabella 2. Risultati di simulazione per la mediana ($q = 0.5$).

Distribuzione $N((0,0), (1,1), 0.8)$			$N = 20$			$N = 30$		
			γ			γ		
$p(x)$	$1 - E[p(X)]$	schema	0.990	0.950	0.900	0.990	0.950	0.900
$p_1(x)$	0.4	1	0.966	0.943	0.897	0.988	0.958	0.911
		2	0.966	0.944	0.900	0.989	0.958	0.915
		3	0.966	0.944	0.900	0.989	0.959	0.914
			ampiezza = 1.751			ampiezza = 1.402		
$p_2(x)$	0.26	1	0.973	0.944	0.901	0.988	0.955	0.909
		2	0.973	0.946	0.901	0.989	0.954	0.911
		3	0.973	0.947	0.900	0.989	0.954	0.911
			ampiezza = 1.438			ampiezza = 1.176		
$p_3(x)$	0.25	1	0.943	0.924	0.888	0.982	0.948	0.906
		2	0.943	0.926	0.889	0.983	0.949	0.905
		3	0.943	0.926	0.889	0.982	0.948	0.903
			ampiezza = 1.607			ampiezza = 1.384		
$p_4(x)$	0.4	1	0.958	0.938	0.899	0.987	0.954	0.908
		2	0.958	0.939	0.902	0.987	0.955	0.909
		3	0.957	0.939	0.902	0.987	0.955	0.908
			ampiezza = 1.608			ampiezza = 1.328		
$p_5(x)$	0.25	1	0.983	0.955	0.906	0.991	0.954	0.905
		2	0.983	0.957	0.906	0.991	0.956	0.908
		3	0.983	0.957	0.906	0.991	0.956	0.908
			ampiezza = 1.532			ampiezza = 1.197		

Tabella 3. Risultati di simulazione per il terzo quartile ($q = 0.75$).

Distribuzione $N((0,0), (1,1), 0.8)$			$N = 60$			$N = 70$		
			γ			γ		
$p(x)$	$1 - E[p(X)]$	schema	0.990	0.950	0.900	0.990	0.950	0.900
$p_1(x)$	0.4	1	0.984	0.945	0.894	0.988	0.948	0.987
		2	0.984	0.945	0.893	0.988	0.947	0.895
		3	0.984	0.944	0.892	0.988	0.947	0.895
		ampiezza = 1.124			ampiezza = 1.030			
$p_2(x)$	0.26	1	0.974	0.912	0.857	0.977	0.922	0.863
		2	0.974	0.912	0.855	0.977	0.921	0.860
		3	0.973	0.911	0.854	0.977	0.920	0.857
		ampiezza = 0.970			ampiezza = 0.905			
$p_3(x)$	0.25	1	0.985	0.933	0.881	0.987	0.941	0.880
		2	0.983	0.934	0.880	0.986	0.936	0.871
		3	0.983	0.933	0.880	0.986	0.935	0.870
		ampiezza = 0.854			ampiezza = 0.816			
$p_4(x)$	0.4	1	0.971	0.933	0.880	0.981	0.936	0.881
		2	0.971	0.936	0.882	0.982	0.938	0.882
		3	0.971	0.935	0.880	0.981	0.938	0.881
		ampiezza = 1.103			ampiezza = 1.014			
$p_5(x)$	0.25	1	0.983	0.932	0.881	0.984	0.934	0.876
		2	0.983	0.931	0.880	0.984	0.934	0.875
		3	0.983	0.930	0.879	0.984	0.933	0.874
		ampiezza = 0.974			ampiezza = 0.891			

Tabella 4. Risultati di simulazione per il primo decile ($q = 0.1$).

Distribuzione $N((0,0), (1,1), 0.8)$			$N = 60$			$N = 70$		
			γ			γ		
$p(x)$	$1 - E[p(X)]$	schema	0.990	0.950	0.900	0.990	0.950	0.900
$p_1(x)$	0.4	1	0.982	0.947	0.900	0.991	0.954	0.905
		2	0.982	0.950	0.904	0.991	0.956	0.908
		3	0.982	0.951	0.904	0.991	0.957	0.908
		ampiezza = 1.191			ampiezza = 1.083			
$p_2(x)$	0.26	1	0.977	0.944	0.908	0.986	0.958	0.914
		2	0.978	0.946	0.910	0.987	0.960	0.916
		3	0.978	0.946	0.911	0.987	0.961	0.917
		ampiezza = 1.125			ampiezza = 1.074			
$p_3(x)$	0.25	1	0.952	0.936	0.894	0.973	0.947	0.901
		2	0.953	0.939	0.898	0.973	0.949	0.907
		3	0.953	0.940	0.899	0.974	0.950	0.908
		ampiezza = 1.253			ampiezza = 1.189			
$p_4(x)$	0.4	1	0.971	0.949	0.901	0.985	0.953	0.905
		2	0.972	0.951	0.904	0.985	0.955	0.909
		3	0.972	0.952	0.905	0.985	0.956	0.909
		ampiezza = 1.203			ampiezza = 1.106			
$p_5(x)$	0.25	1	0.987	0.950	0.907	0.990	0.954	0.906
		2	0.987	0.951	0.912	0.991	0.957	0.908
		3	0.987	0.952	0.913	0.992	0.957	0.909
		ampiezza = 1.088			ampiezza = 1.024			

Tabella 5. Risultati di simulazione per il nono decile ($q = 0.9$).

Distribuzione $N((0,0), (1,1), 0)$			$N = 20$			$N = 30$		
			γ			γ		
$p(x)$	$1 - E[p(X)]$	schema	0.990	0.950	0.900	0.990	0.950	0.900
$p_1(x)$	0.4	1	0.955	0.929	0.884	0.988	0.946	0.897
		2	0.955	0.930	0.889	0.988	0.948	0.898
		3	0.955	0.931	0.889	0.988	0.948	0.899
		ampiezza = 1.648			ampiezza = 1.431			
$p_2(x)$	0.26	1	0.976	0.945	0.888	0.989	0.948	0.896
		2	0.977	0.947	0.892	0.989	0.949	0.899
		3	0.977	0.946	0.893	0.989	0.949	0.899
		ampiezza = 1.564			ampiezza = 1.247			
$p_3(x)$	0.25	1	0.975	0.944	0.893	0.990	0.952	0.901
		2	0.976	0.945	0.896	0.991	0.954	0.907
		3	0.976	0.945	0.896	0.991	0.955	0.908
		ampiezza = 1.591			ampiezza = 1.296			
$p_4(x)$	0.4	1	0.956	0.931	0.887	0.986	0.944	0.894
		2	0.956	0.933	0.890	0.987	0.946	0.898
		3	0.956	0.933	0.890	0.987	0.946	0.898
		ampiezza = 1.643			ampiezza = 1.397			
$p_5(x)$	0.25	1	0.980	0.943	0.889	0.988	0.948	0.894
		2	0.980	0.945	0.892	0.988	0.948	0.896
		3	0.980	0.945	0.893	0.988	0.949	0.896
		ampiezza = 1.540			ampiezza = 1.211			

Tabella 6. Risultati di simulazione per il primo quartile ($q = 0.25$).

Distribuzione $N((0,0), (1,1), 0)$			$N = 20$			$N = 30$		
			γ			γ		
$p(x)$	$1 - E[p(X)]$	schema	0.990	0.950	0.900	0.990	0.950	0.900
$p_1(x)$	0.4	1	0.989	0.948	0.899	0.989	0.950	0.900
		2	0.990	0.950	0.903	0.990	0.952	0.903
		3	0.990	0.951	0.903	0.990	0.952	0.903
			ampiezza = 1.611			ampiezza = 1.274		
$p_2(x)$	0.26	1	0.989	0.944	0.894	0.989	0.948	0.897
		2	0.991	0.946	0.896	0.991	0.952	0.901
		3	0.990	0.947	0.896	0.991	0.952	0.902
			ampiezza = 1.396			ampiezza = 1.120		
$p_3(x)$	0.25	1	0.989	0.948	0.897	0.990	0.949	0.901
		2	0.990	0.952	0.902	0.991	0.953	0.909
		3	0.991	0.953	0.903	0.991	0.954	0.910
			ampiezza = 1.455			ampiezza = 1.160		
$p_4(x)$	0.4	1	0.987	0.944	0.896	0.989	0.947	0.896
		2	0.988	0.945	0.899	0.990	0.949	0.901
		3	0.988	0.946	0.900	0.990	0.950	0.901
			ampiezza = 1.583			ampiezza = 1.257		
$p_5(x)$	0.25	1	0.988	0.946	0.896	0.992	0.953	0.904
		2	0.989	0.946	0.897	0.992	0.954	0.906
		3	0.989	0.945	0.898	0.992	0.954	0.906
			ampiezza = 1.359			ampiezza = 1.092		

Tabella 7. Risultati di simulazione per la mediana ($q = 0.5$).

Distribuzione $N((0,0), (1,1), 0)$			$N = 20$			$N = 30$		
			γ			γ		
$p(x)$	$1 - E[p(X)]$	schema	0.990	0.950	0.900	0.990	0.950	0.900
$p_1(x)$	0.4	1	0.952	0.930	0.885	0.988	0.948	0.893
		2	0.953	0.932	0.890	0.989	0.951	0.897
		3	0.953	0.932	0.891	0.989	0.951	0.898
			ampiezza = 1.640			ampiezza = 1.418		
$p_2(x)$	0.26	1	0.977	0.939	0.886	0.986	0.942	0.889
		2	0.977	0.942	0.890	0.987	0.944	0.891
		3	0.976	0.943	0.890	0.987	0.944	0.892
			ampiezza = 1.550			ampiezza = 1.247		
$p_3(x)$	0.25	1	0.980	0.943	0.895	0.987	0.947	0.896
		2	0.981	0.948	0.900	0.989	0.950	0.902
		3	0.981	0.948	0.900	0.989	0.951	0.903
			ampiezza = 1.569			ampiezza = 1.291		
$p_4(x)$	0.4	1	0.956	0.929	0.877	0.986	0.949	0.901
		2	0.957	0.931	0.882	0.986	0.951	0.904
		3	0.957	0.932	0.882	0.986	0.951	0.904
			ampiezza = 1.626			ampiezza = 1.412		
$p_5(x)$	0.25	1	0.981	0.944	0.889	0.989	0.947	0.897
		2	0.981	0.946	0.891	0.989	0.949	0.899
		3	0.981	0.946	0.891	0.989	0.948	0.898
			ampiezza = 1.534			ampiezza = 1.210		

Tabella 8. Risultati di simulazione per il terzo quartile ($q = 0.75$).

Distribuzione $N((0,0), (1,1), 0)$			$N = 60$			$N = 70$		
			γ			γ		
$p(x)$	$1 - E[p(X)]$	schema	0.990	0.950	0.900	0.990	0.950	0.900
$p_1(x)$	0.4	1	0.969	0.946	0.888	0.982	0.945	0.893
		2	0.947	0.929	0.893	0.983	0.947	0.894
		3	0.947	0.929	0.893	0.983	0.947	0.895
			ampiezza = 1.262			ampiezza = 1.190		
$p_2(x)$	0.26	1	0.984	0.938	0.897	0.989	0.946	0.889
		2	0.984	0.941	0.899	0.990	0.947	0.892
		3	0.984	0.942	0.899	0.990	0.947	0.982
			ampiezza = 1.122			ampiezza = 1.029		
$p_3(x)$	0.25	1	0.985	0.948	0.898	0.989	0.947	0.903
		2	0.985	0.948	0.901	0.990	0.951	0.907
		3	0.985	0.949	0.902	0.990	0.952	0.907
			ampiezza = 1.165			ampiezza = 1.067		
$p_4(x)$	0.4	1	0.969	0.946	0.886	0.988	0.951	0.900
		2	0.969	0.946	0.887	0.987	0.954	0.903
		3	0.969	0.946	0.889	0.988	0.954	0.903
			ampiezza = 1.276			ampiezza = 1.177		
$p_5(x)$	0.25	1	0.986	0.936	0.892	0.989	0.944	0.887
		2	0.986	0.937	0.893	0.989	0.944	0.888
		3	0.986	0.936	0.892	0.989	0.944	0.889
			ampiezza = 1.081			ampiezza = 0.997		

Tabella 9. Risultati di simulazione per il primo decile ($q = 0.1$).

Distribuzione $N((0,0), (1,1), 0)$			$N = 60$			$N = 70$		
			γ			γ		
$p(x)$	$1 - E[p(X)]$	schema	0.990	0.950	0.900	0.990	0.950	0.900
$p_1(x)$	0.4	1	0.969	0.940	0.882	0.983	0.945	0.894
		2	0.969	0.942	0.885	0.983	0.946	0.897
		3	0.969	0.942	0.886	0.983	0.946	0.897
			ampiezza = 1.259			ampiezza = 1.176		
$p_2(x)$	0.26	1	0.985	0.939	0.895	0.987	0.948	0.890
		2	0.985	0.940	0.898	0.988	0.950	0.892
		3	0.985	0.941	0.898	0.988	0.950	0.892
			ampiezza = 1.138			ampiezza = 1.083		
$p_3(x)$	0.25	1	0.983	0.939	0.884	0.986	0.941	0.889
		2	0.984	0.942	0.891	0.987	0.944	0.892
		3	0.985	0.944	0.892	0.987	0.945	0.893
			ampiezza = 1.146			ampiezza = 1.068		
$p_4(x)$	0.4	1	0.965	0.940	0.883	0.980	0.942	0.893
		2	0.966	0.942	0.885	0.980	0.945	0.894
		3	0.966	0.942	0.884	0.980	0.945	0.893
			ampiezza = 1.255			ampiezza = 1.167		
$p_5(x)$	0.25	1	0.983	0.935	0.890	0.986	0.937	0.880
		2	0.983	0.936	0.891	0.986	0.939	0.882
		3	0.983	0.937	0.889	0.986	0.939	0.882
			ampiezza = 1.076			ampiezza = 1.005		

Tabella 10. Risultati di simulazione per il nono decile ($q = 0.9$).

Distribuzione $t_3((0,0), (1,1), 0.5)$		$N = 20$			$N = 30$		
		γ			γ		
$p(x)$	schema	0.990	0.950	0.900	0.990	0.950	0.900
$p_1(x)$	1	0.974	0.945	0.899	0.988	0.943	0.887
	2	0.974	0.946	0.900	0.988	0.944	0.890
	3	0.974	0.945	0.900	0.988	0.944	0.889
		ampiezza = 2.818			ampiezza = 1.997		
$p_2(x)$	1	0.963	0.930	0.874	0.983	0.935	0.875
	2	0.964	0.934	0.877	0.984	0.938	0.880
	3	0.964	0.935	0.877	0.984	0.939	0.881
		ampiezza = 2.367			ampiezza = 1.649		
$p_3(x)$	1	0.983	0.935	0.889	0.985	0.939	0.888
	2	0.982	0.934	0.884	0.985	0.935	0.880
	3	0.981	0.930	0.879	0.984	0.932	0.876
		ampiezza = 2.386			ampiezza = 1.640		
$p_4(x)$	1	0.959	0.935	0.890	0.985	0.941	0.891
	2	0.959	0.937	0.892	0.986	0.942	0.893
	3	0.959	0.938	0.892	0.987	0.943	0.894
		ampiezza = 2.617			ampiezza = 1.952		
$p_5(x)$	1	0.979	0.939	0.884	0.985	0.937	0.879
	2	0.980	0.941	0.886	0.985	0.938	0.881
	3	0.980	0.940	0.886	0.985	0.938	0.882
		ampiezza = 2.282			ampiezza = 1.575		

Tabella 11. Risultati di simulazione per il primo quartile ($q = 0.25$).

Distribuzione $t_3((0,0), (1,1), 0.5)$		$N = 20$			$N = 30$		
		γ			γ		
		0.990	0.950	0.900	0.990	0.950	0.900
$p(x)$	schema						
$p_1(x)$	1	0.990	0.952	0.902	0.990	0.951	0.900
	2	0.990	0.953	0.904	0.991	0.952	0.901
	3	0.990	0.953	0.903	0.991	0.952	0.900
			ampiezza = 1.930			ampiezza = 1.448	
$p_2(x)$	1	0.991	0.955	0.905	0.991	0.952	0.903
	2	0.991	0.955	0.907	0.991	0.952	0.906
	3	0.991	0.955	0.907	0.991	0.952	0.905
			ampiezza = 1.644			ampiezza = 1.243	
$p_3(x)$	1	0.986	0.938	0.884	0.985	0.932	0.879
	2	0.984	0.934	0.876	0.983	0.925	0.867
	3	0.983	0.930	0.873	0.981	0.921	0.860
			ampiezza = 1.665			ampiezza = 1.250	
$p_4(x)$	1	0.984	0.934	0.881	0.987	0.936	0.876
	2	0.983	0.928	0.874	0.985	0.929	0.865
	3	0.981	0.926	0.870	0.984	0.925	0.859
			ampiezza = 1.663			ampiezza = 1.263	
$p_5(x)$	1	0.987	0.943	0.891	0.990	0.949	0.899
	2	0.987	0.943	0.892	0.991	0.952	0.900
	3	0.987	0.943	0.892	0.991	0.952	0.900
			ampiezza = 1.552			ampiezza = 1.196	

Tabella 12. Risultati di simulazione per la mediana ($q = 0.5$).

Distribuzione $t_3((0,0), (1,1), 0.5)$		$N = 20$			$N = 30$		
		γ			γ		
		0.990	0.950	0.900	0.990	0.950	0.900
$p(x)$	schema						
$p_1(x)$	1	0.971	0.942	0.892	0.987	0.945	0.893
	2	0.971	0.945	0.893	0.987	0.946	0.896
	3	0.971	0.944	0.892	0.987	0.946	0.896
			ampiezza = 2.840			ampiezza = 2.033	
$p_2(x)$	1	0.964	0.938	0.902	0.989	0.959	0.913
	2	0.964	0.940	0.904	0.990	0.959	0.915
	3	0.964	0.940	0.903	0.990	0.960	0.916
			ampiezza = 2.438			ampiezza = 2.174	
$p_3(x)$	1	0.948	0.919	0.876	0.975	0.933	0.881
	2	0.948	0.919	0.875	0.975	0.931	0.879
	3	0.948	0.918	0.873	0.975	0.956	0.875
			ampiezza = 2.463			ampiezza = 1.986	
$p_4(x)$	1	0.957	0.932	0.889	0.985	0.947	0.898
	2	0.958	0.935	0.890	0.986	0.947	0.898
	3	0.958	0.934	0.889	0.986	0.947	0.897
			ampiezza = 2.605			ampiezza = 2.191	
$p_5(x)$	1	0.973	0.940	0.888	0.987	0.946	0.887
	2	0.973	0.942	0.889	0.987	0.946	0.890
	3	0.973	0.941	0.889	0.987	0.946	0.889
			ampiezza = 1.507			ampiezza = 1.193	

Tabella 13. Risultati di simulazione per il terzo quartile ($q = 0.75$).

Distribuzione $t_3((0,0), (1,1), 0.5)$		$N = 60$			$N = 70$		
		γ			γ		
		0.990	0.950	0.900	0.990	0.950	0.900
$p(x)$	schema						
$p_1(x)$	1	0.985	0.946	0.893	0.989	0.944	0.894
	2	0.984	0.947	0.895	0.989	0.944	0.893
	3	0.984	0.946	0.894	0.989	0.944	0.892
			ampiezza = 2.746			ampiezza = 2.315	
$p_2(x)$	1	0.968	0.917	0.863	0.977	0.919	0.862
	2	0.968	0.917	0.866	0.978	0.920	0.862
	3	0.968	0.916	0.865	0.978	0.920	0.863
			ampiezza = 2.405			ampiezza = 2.062	
$p_3(x)$	1	0.988	0.943	0.889	0.985	0.939	0.883
	2	0.987	0.942	0.887	0.984	0.938	0.880
	3	0.986	0.940	0.885	0.984	0.937	0.881
			ampiezza = 2.132			ampiezza = 1.796	
$p_4(x)$	1	0.973	0.938	0.881	0.982	0.936	0.879
	2	0.973	0.940	0.882	0.982	0.936	0.880
	3	0.973	0.941	0.883	0.982	0.936	0.881
			ampiezza = 2.861			ampiezza = 2.389	
$p_5(x)$	1	0.985	0.929	0.873	0.983	0.940	0.883
	2	0.985	0.930	0.874	0.983	0.939	0.883
	3	0.985	0.930	0.874	0.984	0.940	0.883
			ampiezza = 2.118			ampiezza = 1.900	

Tabella 14. Risultati di simulazione per il primo decile ($q = 0.1$).

Distribuzione $t_3((0,0), (1,1), 0.5)$		$N = 60$			$N = 70$		
		γ			γ		
$p(x)$	schema	0.990	0.950	0.900	0.990	0.950	0.900
$p_1(x)$	1	0.982	0.940	0.886	0.987	0.939	0.883
	2	0.982	0.942	0.889	0.988	0.942	0.886
	3	0.982	0.942	0.889	0.988	0.942	0.887
		ampiezza = 2.704			ampiezza = 2.300		
$p_2(x)$	1	0.969	0.940	0.906	0.980	0.951	0.908
	2	0.969	0.944	0.912	0.981	0.955	0.917
	3	0.969	0.945	0.913	0.981	0.956	0.917
		ampiezza = 3.121			ampiezza = 3.193		
$p_3(x)$	1	0.959	0.930	0.887	0.978	0.944	0.896
	2	0.959	0.932	0.890	0.978	0.948	0.900
	3	0.959	0.932	0.889	0.978	0.947	0.898
		ampiezza = 2.840			ampiezza = 2.721		
$p_4(x)$	1	0.970	0.936	0.886	0.982	0.941	0.892
	2	0.971	0.936	0.888	0.982	0.944	0.894
	3	0.971	0.937	0.888	0.983	0.944	0.894
		ampiezza = 2.999			ampiezza = 2.831		
$p_5(x)$	1	0.984	0.940	0.890	0.986	0.945	0.890
	2	0.985	0.940	0.893	0.987	0.947	0.892
	3	0.985	0.940	0.882	0.987	0.947	0.983
		ampiezza = 2.497			ampiezza = 2.368		

Tabella 15. Risultati di simulazione per il nono decile ($q = 0.9$).

Working Paper Series
Department of Statistical Sciences, University of Padua

You may order paper copies of the working papers by emailing wp@stat.unipd.it

Most of the working papers can also be found at the following url: <http://wp.stat.unipd.it>

