Working Paper Series, N. 16, dicembre 2007

Una nota sul comportamento asintotico della statistica di verosimiglianza empirica

Gianfranco Adimari, Annamaria Guolo
Department of Statistical Sciences
University of Padua
Italy

Abstract: In questo lavoro vengono presentati alcuni risultati teorici, di natura generale, sul comportamento asintotico delle statistiche di verosimiglianza empirica e di verosimiglianza empirica profilo generate da funzioni di stima. Tali risultati possono essere utili in diversi ambiti, in particolare per l'inferenza con dati dipendenti.

Keywords: Funzione di stima; Processo stazionario; Pseudo-verosimiglianza

Indice

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

Una nota sul comportamento asintotico della statistica di verosimiglianza empirica

Gianfranco Adimari, Annamaria Guolo
Department of Statistical Sciences
University of Padua
Italy

Abstract: In questo lavoro vengono presentati alcuni risultati teorici, di natura generale, sul comportamento asintotico delle statistiche di verosimiglianza empirica e di verosimiglianza empirica profilo generate da funzioni di stima. Tali risultati possono essere utili in diversi ambiti, in particolare per l'inferenza con dati dipendenti.

Keywords: Funzione di stima; Processo stazionario; Pseudo-verosimiglianza

1 Introduzione

Sia $\mathbf{Y} = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n\}$ una successione di variabili casuali (ciascuna di dimensione k) e $\theta_0 \in \Theta \subseteq \mathbb{R}^s$ un parametro ignoto da cui dipende la legge di $\mathbf{Y}$. Si supponga di disporre di un'adeguata funzione di stima per θ_0, del tipo

$$\overline{\eta}(\mathbf{Y};\theta) = \frac{1}{m}\sum_{j=1}^{m} \eta_j(\mathcal{Y}_j^{\ell};\theta), \tag{1}$$

con $\eta_1, \eta_2, \ldots, \eta_m$ funzioni reali da $\mathbb{R}^{k\ell} \times \Theta$ in $\mathbb{R}^s$, $m, \ell \in \{1, 2, \ldots, n\}$ e $\mathcal{Y}_j^{\ell}$ che indica, per ogni j, un qualche insieme di ℓ elementi della successione $\{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n\}$. La funzione di stima $\overline{\eta}(\mathbf{Y};\theta)$ è dunque molto generale, potendo le singole componenti $\eta_1, \eta_2, \ldots, \eta_m$ essere diverse tanto nella forma quanto nel modo di dipendere da $\mathbf{Y}$.

In tale contesto, una stima di θ_0 è tipicamente ottenuta come radice dell'equazione $\overline{\eta}(\mathbf{Y};\theta) = 0$. Inoltre, è spesso possibile associare a $\overline{\eta}(\mathbf{Y};\theta)$ una pseudo-verosimiglianza per l'inferenza su θ_0 ricorrendo alla funzione di verosimiglianza empirica (si veda Owen, 2001, come riferimento generale). In particolare, quando le variabili $\{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n\}$ sono indipendenti ed identicamente distribuite (i.i.d.), si ha, generalmente, $m = n$, $\ell = 1$, $\mathcal{Y}_j^{\ell} = \mathcal{Y}_j^{1} = \{\mathbf{y}_j\}$ e $\eta_1 = \eta_2 = \ldots = \eta_n = \eta$. La funzione log-rapporto di verosimiglianza empirica[1] per θ_0, generata da $\overline{\eta}(\mathbf{Y};\theta)$ e calcolata in un generico punto θ, ha allora espressione

$$l(\theta) = 2\sum_{j=1}^{n} \log\{1 + \lambda^{\top}\eta(\mathbf{y}_j;\theta)\}, \tag{2}$$

[1] Con questo termine si indicherà la funzione pari a meno due volte il logaritmo del rapporto di verosimiglianza empirica.

con $\lambda = \lambda(\theta)$ che soddisfa la relazione

$$\frac{1}{n}\sum_{j=1}^{n}\frac{\eta(\mathbf{y}_j;\theta)}{1+\lambda^\top\eta(\mathbf{y}_j;\theta)} = 0.$$

L'espressione (2) è corretta se il punto θ è tale che l'origine $\mathcal{O}$ è interna all'involucro convesso dei punti $\eta(\mathbf{y}_j;\theta)$, $j = 1,\ldots,n$; altrimenti bisogna porre $l(\theta) = +\infty$. Sotto deboli condizioni (Owen, 1990, Qin e Lawless, 1995), risulta che, quando $n \longrightarrow +\infty$,

$$l(\theta_0) \xrightarrow{d} \chi^2_s, \tag{3}$$

dove il simbolo χ^2_s indica la distribuzione chi-quadro con s gradi di libertà. Sotto ulteriori condizioni di regolarità, un risultato analogo si ottiene quando il parametro di interesse è una qualche funzione di θ, diciamo $\tau = h(\theta)$. In questo caso, per la funzione log-rapporto di verosimiglianza empirica profilo $l_P(\tau) = \inf_{\theta:h(\theta)=\tau} l(\theta)$, si ha

$$l_P(\tau_0) \xrightarrow{d} \chi^2_r, \tag{4}$$

dove r indica la dimensione di τ e $\tau_0 = h(\theta_0)$.

I risultati (3) e (4) rappresentano un'estensione del teorema di Wilks all'ambito non parametrico (o semi-parametrico) e permettono di utilizzare le statistiche di verosimiglianza empirica e verosimiglianza empirica profilo per costruire regioni di confidenza o effettuare verifiche d'ipotesi, nella maniera in cui si utilizzerebbero le corrispondenti versioni parametriche.

Owen (1991), Kolazcyk (1994), Kitamura (1997), Monti (1997), Chuang e Chan (2002) e Nordmann, Sibbertsen e Lahiri (2007), tra gli altri, studiano estensioni dei risultati (3) e (4) in situazioni più complicate, in cui la struttura della successione $\mathbf{Y} = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n\}$ non è quella di variabili i.i.d.. Tali estensioni si riferiscono, in particolare, a problemi inferenziali in modelli lineari, modelli lineari generalizzati e modelli per serie temporali.

In questo lavoro, specificatamente nel Paragrafo 2, vengono presentati alcuni risultati teorici, di natura generale, sul comportamento asintotico delle statistiche di verosimiglianza empirica e di verosimiglianza empirica profilo generate da funzioni di stima di tipo (1). Tali risultati possono essere utili in diversi ambiti, in particolare per l'inferenza con dati dipendenti. Alcuni esempi del loro uso sono riportati nel Paragrafo 3.

2 Risultati principali

Sia

$$l(\theta) = 2\sum_{j=1}^{m}\log\{1+\lambda^\top\eta_j(\mathcal{Y}^\ell_j;\theta)\}, \tag{5}$$

con $\lambda = \lambda(\theta)$ che soddisfa la relazione

$$\frac{1}{m}\sum_{j=1}^{m}\frac{\eta_j(\mathcal{Y}^\ell_j;\theta)}{1+\lambda^\top\eta_j(\mathcal{Y}^\ell_j;\theta)} = 0, \tag{6}$$

la funzione log-rapporto di verosimiglianza empirica per θ_0, generata dalla funzione di stima (1). Siano

$$\overline{\Omega}(\theta) = \frac{1}{m}\sum_{j=1}^{m} \eta_j(\mathcal{Y}_j^\ell;\theta)\eta_j^\top(\mathcal{Y}_j^\ell;\theta) \quad \text{e} \quad \overline{\nu}(\theta) = \max_{1\leq j\leq m} ||\eta_j(\mathcal{Y}_j^\ell;\theta)||,$$

e si indichi con $\mathcal{I}(\theta)$ l'involucro convesso dei punti $\eta_1(\mathcal{Y}_1^\ell;\theta), \eta_2(\mathcal{Y}_2^\ell;\theta), \ldots, \eta_m(\mathcal{Y}_m^\ell;\theta)$. Nel seguito si userà la notazione concisa $\overline{\eta}(\theta)$ per $\overline{\eta}(\mathbf{Y};\theta)$ e si assumerà θ_0 essere un punto interno a Θ. Vale il seguente risultato.

Teorema 1. *Siano a_m e b_m due successione reali tali che $b_m \to b > 0$ (potendo essere $b = +\infty$), $a_m \to +\infty$ e $b_m/a_m \to 0$ quando $n \to +\infty$. Si assuma che, quando $n \to +\infty$, $m/n \to \kappa > 0$ (costante finita) e*

- *A0)* $\Pr\{\mathcal{O} \in \mathcal{I}(\theta_0)\} \to 1$*;*
- *A1)* $a_m\overline{\eta}(\theta_0)$ *sia asintoticamente normale, con media zero e matrice di covarianze* $\Omega(\theta_0)$ *finita e non singolare;*
- *A2)* $b_m\overline{\Omega}(\theta_0) = \Omega^*(\theta_0) + o_p(1)$, *con* $\Omega^*(\theta_0)$ *matrice finita e non singolare;*
- *A3)* $\overline{\nu}(\theta_0) = o_p(a_m/b_m)$.

Allora, sotto le condizioni A0-A3, si ha

$$\frac{a_m^2}{mb_m} l(\theta_0) = a_m^2\overline{\eta}^\top(\theta_0)\Omega^{*-1}(\theta_0)\overline{\eta}(\theta_0) + o_p(1),$$

quando $n \to +\infty$.

Dimostrazione. Evidentemente, n e m sono dello stesso ordine. La condizione A0 implica che la funzione log-rapporto di verosimiglianza empirica calcolata in $\theta = \theta_0$ è finita con probabilità tendente a 1 quando $n \to +\infty$ ed ha espressione data dalla (5). Sia $\lambda_0 = \lambda(\theta_0)$ la soluzione dell'equazione (6) quando $\theta = \theta_0$ e sia $\rho_0 = \rho(\theta_0) = \|\lambda_0\|$. Pertanto, $\lambda_0 = \rho_0\xi$ con $\xi = \rho_0^{-1}\lambda_0$ e $\|\xi\| = 1$. Si può mostrare che

$$\rho_0 \leq \frac{|u^\top\overline{\eta}(\theta_0)|}{\xi^\top\overline{\Omega}(\theta_0)\xi - \overline{\nu}(\theta_0)|u^\top\overline{\eta}(\theta_0)|}, \tag{7}$$

dove $u = (1, \ldots, 1)^\top$. Sia ω il più piccolo autovalore di $\Omega^*(\theta_0)$. Allora, dalla (7) e nelle ipotesi fatte, si ha

$$\rho_0 \leq \frac{O_p(b_m/a_m)}{\xi^\top\Omega^*(\theta_0)\xi + o_p(1)} \ .$$

Perciò $\lambda_0 = O_p(b_m/a_m)$ dato che $\xi^\top\Omega^*(\theta_0)\xi \geq \omega > 0$.

Si osservi anche che, dall'ipotesi A3, $\max_j |\lambda_0^\top\eta_j(\mathcal{Y}_j^\ell;\theta_0)| = O_p(b_m/a_m)o_p(a_m/b_m) = o_p(1)$. Quindi si può usare lo sviluppo

$$\{1 + \lambda_0^\top\eta_j(\mathcal{Y}_j^\ell;\theta_0)\}^{-1} = 1 - \lambda_0^\top\eta_j(\mathcal{Y}_j^\ell;\theta_0) + \frac{\{\lambda_0^\top\eta_j(\mathcal{Y}_j^\ell;\theta_0)\}^2}{1 + \lambda_0^\top\eta_j(\mathcal{Y}_j^\ell;\theta_0)}$$

nell'equazione (6) per ottenere

$$\begin{aligned} 0 &= \overline{\eta}(\theta_0) - \left\{ \frac{1}{m} \sum_{j=1}^{m} \eta_j(\mathcal{Y}_j^\ell;\theta_0)\eta_j^\top(\mathcal{Y}_j^\ell;\theta_0) \right\} \lambda_0 + \frac{1}{m} \sum_{i=1}^{m} \frac{\{\lambda_0^\top \eta_j(\mathcal{Y}_j^\ell;\theta_0)\}^2}{1+\lambda_0^\top \eta_j(\mathcal{Y}_j^\ell;\theta_0)} \eta_j(\mathcal{Y}_j^\ell;\theta_0) \\ &= \overline{\eta}(\theta_0) - \overline{\Omega}(\theta_0)\lambda_0 + resto. \end{aligned}$$

L'ipotesi A2 implica che $\frac{1}{m}\sum_{j=1}^{m} \|\eta_j(\mathcal{Y}_j^\ell;\theta_0)\|^2 = O_p(1/b_m)$. Quindi il *resto* ha norma limitata da

$$\begin{aligned} \frac{1}{m}\sum_j \|\eta_j(\mathcal{Y}_j^\ell;\theta_0)\|^3 \|\lambda_0\|^2 |1+\lambda_0^\top \eta_j(\mathcal{Y}_j^\ell;\theta_0)|^{-1} &= o_p(a_m/b_m) O_p(b_m^2/a_m^2) O_p(1/b_m) \\ &= o_p(1/a_m), \end{aligned}$$

avendo usato anche l'ipotesi A3 e il fatto che $\lambda_0 = O_p(b_m/a_m)$. Ne segue che,

$$\lambda_0 = b_n \Omega^{*-1}(\theta_0)\overline{\eta}(\theta_0) + o_p(b_m/a_m). \tag{8}$$

In modo analogo, usando lo sviluppo in serie di McLaurin di $\log(1+x)$ nel membro di destra della (5), si ottiene

$$l(\theta) = 2\sum_{j=1}^{m} \lambda_0^\top \eta_j(\mathcal{Y}_j^\ell;\theta_0) - \lambda_0^\top \left\{ \sum_{j=1}^{m} \eta_j(\mathcal{Y}_j^\ell;\theta_0)\eta_j^\top(\mathcal{Y}_j^\ell;\theta_0) \right\} \lambda_0 + \frac{2}{3}\sum_{j=1}^{m} \frac{\{\lambda_0^\top \eta_j(\mathcal{Y}_j^\ell;\theta_0)\}^3}{(1+\beta_j)^3}, \tag{9}$$

dove $|\beta_j| < |\lambda_0^\top \eta_j(\mathcal{Y}_j^\ell;\theta_0)|$. Si ricava facilmente che $\sum_{j=1}^{m}\{\lambda_0^\top \eta_j(\mathcal{Y}_j^\ell;\theta_0)\}^3/(1+\beta_j)^3 = o_p(mb_m/a_m^2)$. Quindi, usando la (8), il fatto che $\overline{\eta}(\theta_0) = O_p(1/a_m)$ e l'ipotesi A3, dalla (9) si ottiene

$$\frac{a_m^2}{mb_m} l(\theta_0) = a_m^2 \overline{\eta}^\top(\theta_0)\Omega^{*-1}(\theta_0)\overline{\eta}(\theta_0) + o_p(1).$$

□

Come diretta conseguenza del Teorema 1, si ha il seguente corollario.

Corollario 1. *(A) Sotto le condizioni del Teorema 1, se* $\Omega(\theta_0) = \Omega^*(\theta_0)$,

$$\frac{a_m^2}{mb_m} l(\theta_0) \xrightarrow{d} \chi_s^2$$

quando $n \to +\infty$. *(B) Inoltre, se* $a_m^2/(mb_m) \to 1$,

$$l(\theta_0) \xrightarrow{d} \chi_s^2.$$

Il Teorema 1 e il Corollario 1 mostrano che il comportamento asintotico della statistica di verosimiglianza empirica $l(\theta_0)$ dipende dal comportamento della funzione di stima $\overline{\eta}(\theta_0)$ e dal comportamento della successione di matrici $\overline{\Omega}(\theta_0)$. In particolare, il risultato standard di distribuzione asintotica χ^2 per $l(\theta_0)$ è disatteso se $\overline{\Omega}(\theta_0)$ non converge alla matrice di covarianze asintotica di $\overline{\eta}(\theta_0)$ e/o se esistono incongruenze tra gli ordini di convergenza di $\overline{\eta}(\theta_0)$ e $\overline{\Omega}(\theta_0)$.

In molti casi, le condizioni del Corollario 1(B) sono verificate con $a_m = m^{1/2}$ e $b_m = 1 \ \forall \ m$. Comunque, anche quando il risultato asintotico standard non è conseguibile direttamente per $l(\theta_0)$, è spesso possibile aggiustare opportunamente la funzione log-rapporto di verosimiglianza empirica, in modo che l'approssimazione χ^2 sia ancora valida. La cosa è particolarmente interessante (e tipicamente agevole) nel caso scalare (cioè quando la dimensione di θ è 1), in cui vale il seguente corollario.

Corollario 2. *(A) Sia $s = 1$ e siano verificate le condizioni del Teorema 1. Allora,*

$$\frac{a_m^2}{mb_m}\frac{\Omega^*(\theta_0)}{\Omega(\theta_0)}l(\theta_0) \xrightarrow{d} \chi_1^2$$

quando $n \to +\infty$. (B) Inoltre, se $a_m^2/(mb_m) \to 1$,

$$\frac{\Omega^*(\theta_0)}{\Omega(\theta_0)}l(\theta_0) \xrightarrow{d} \chi_1^2.$$

In base a tale risultato, evidentemente, l'aggiustamento della funzione log-rapporto di verosimiglianza empirica $l(\theta)$ che rende lecita la calibrazione χ^2 consiste nel recuperare, ove possibile, una stima consistente di $\frac{\Omega^*(\theta_0)}{\Omega(\theta_0)}$ (o di $\frac{a_m^2}{mb_m}\frac{\Omega^*(\theta_0)}{\Omega(\theta_0)}$ nel caso in cui a_m e/o b_m dovessero dipendere da entità ignote) e di moltiplicare $l(\theta)$ per tale stima.

Sia ora $l_P(\tau)$ la funzione log-rapporto di verosimiglianza empirica profilo derivata da $l(\theta)$ e riferita ad un parametro $\tau_0 = h(\theta_0)$, con $h(\cdot)$ funzione a valori in $\mathbb{R}^r$ la quale ammette il seguente sviluppo di Taylor

$$h(\theta) = \tau_0 + V_0(\theta - \theta_0) + o(||\theta - \theta_0||).$$

Qui V_0 indica la matrice Jacobiana della funzione $h(\cdot)$, calcolata in $\theta = \theta_0$. Per $l_P(\tau)$ vale allora il seguente risultato.

Teorema 2. *Sia $\mathcal{D}$ un opportuno intorno di θ_0. Si assuma che, quando $n \to +\infty$, $m/n \to \kappa > 0$ (costante finita) e*

- *B0)* $\Pr\{\mathcal{O} \in \mathcal{I}(\theta), \ \ \forall \ \theta \in \mathcal{D}$ *contestualmente*$\} \to 1$;
- *B1)* $m^{1/2}\overline{\eta}(\theta_0)$ *sia asintoticamente normale con media zero e matrice di covarianze $\Omega(\theta_0)$ finita e non singolare.*

Si assuma inoltre che, per ogni sequenza di punti casuali $\tilde{\theta} = \tilde{\theta}(\mathbf{Y})$ tale che $\|\tilde{\theta} - \theta_0\| = O(m^{-1/2})$ quasi certamente (q.c.),

- *B2)* $\overline{\eta}(\tilde{\theta}) = \overline{\eta}(\theta_0) + \{J(\theta_0) + o_p(1)\}(\tilde{\theta} - \theta_0)$, *per qualche matrice $J(\theta_0)$ finita e non singolare;*
- *B3)* $\overline{\Omega}(\tilde{\theta}) = \Omega(\theta_0) + o_p(1)$;
- *B4)* $\overline{\nu}(\tilde{\theta}) = o_p(m^{1/2})$.

Allora, sotto le condizioni B0-B4, $l_P(\tau_0) = \inf_{\theta:h(\theta)=\tau_0} l(\theta)$ *ha distribuzione asintotica* χ^2_r*, almeno quando l'estremo inferiore è ottenuto in una opportuna sequenze di sfere, diciamo* $\mathcal{C}_m = \{\theta : \|\theta - \theta_0\| \leq Cm^{-1/2}\}$*, di raggio di ordine* $m^{-1/2}$ *e centro in* θ_0*.*

Il Teorema 2 assicura, dunque, l'esistenza di una successione di soluzioni locali del problema di minimizzazione coinvolto nel calcolo di $l_P(\tau_0)$, successione che ben si comporta asintoticamente. La dimostrazione del teorema si basa su argomentazioni analoghe a quelle usate in Adimari e Preo (2007) (dimostrazione del Teorema 1) ed è, pertanto, omessa.

3 Alcune applicazioni

Diversi risultati presenti in letteratura sulla distribuzione asintotica della statistica di vorosimiglianza empirica e della statistica di verosimiglianza empirica profilo possono essere ottenuti applicando i teoremi e i corollari del paragrafo precedente. Nel seguito vengono presentate alcune esemplificazioni relative a problemi di inferenza in modelli per serie temporali.

3.1 Verosimiglianza empirica per modelli autoregressivi stazionari

Sia dato un processo $\{y_t\}$ autoregressivo stazionario di ordine p,

$$y_t = \theta_{01} y_{t-1} + \theta_{02} y_{t-2} + \ldots + \theta_{0p} y_{t-p} + \varepsilon_t,$$

dove $\{\varepsilon_t\}$ rappresenta una successione di martingale differenza rispetto ad una successione crescente di σ-algebre $\mathcal{F}_t$. Ciò significa che ε_t è $\mathcal{F}_t$ misurabile e che $E(\varepsilon_t|\mathcal{F}_{t-1}) = 0$ q.c. per ogni t. Si supponga che $E\left[\varepsilon_t^2|\mathcal{F}_{t-1}\right] = \sigma^2 < \infty$ q.c., per ogni t, e $\sup_{t\geq p} E\left[|\varepsilon_t|^{2+\alpha}|\mathcal{F}_{t-1}\right] < \infty$, per qualche $\alpha > 0$.

Data l'osservazione campionaria $\{y_1, y_2, \ldots, y_n\}$, la stima ai minimi quadrati $\hat{\theta}$ di $\theta_0 = (\theta_{01}, \ldots, \theta_{0p})^\top$ è il valore che minimizza in $\theta = (\theta_1, \ldots, \theta_p)^\top$ la quantità

$$-\frac{1}{2}\sum_{t=p+1}^{n} (y_t - E\left[y_t|\mathcal{F}_{t-1}\right])^2 = -\frac{1}{2}\sum_{t=p+1}^{n} (y_t - \theta^\top \mathbf{y}_{t-1})^2,$$

dove con $\mathbf{y}_{t-1}$ si indica il vettore, di dimensione p, $(y_{t-1}, \ldots, y_{t-p})^\top$. Risulta che

$$\hat{\theta} = \sum_{t=p+1}^{n} (\mathbf{y}_{t-1}\mathbf{y}_{t-1}^\top)^{-1} \sum_{t=p+1}^{n} \mathbf{y}_{t-1} y_t,$$

e la funzione di stima che definisce $\hat{\theta}$ è

$$\frac{1}{n-p}\sum_{t=p+1}^{n} (y_t - \theta^\top \mathbf{y}_{t-1})\mathbf{y}_{t-1}. \tag{10}$$

Tale funzione di stima può essere utilizzata per ottenere, via verosimiglianza empirica, una funzione di pseudo-verosimiglianza per θ_0 (si veda Chuang e Chan, 2002).

Posto $\psi_t(\theta) = (y_t - \theta^\top \mathbf{y}_{t-1})\mathbf{y}_{t-1}$ e $\overline{\psi}(\theta) = \frac{1}{n-p}\sum_{t=p+1}^{n} \psi_t(\theta)$, si ha che $\psi_t(\theta_0) = \varepsilon_t \mathbf{y}_{t-1}$ cosicché $\{\psi_t(\theta_0)\}$ forma una successione di martingale differenza e la funzione $\overline{\psi}(\theta_0)$ è, a sua volta, una martingala. Da un'applicazione della legge dei grandi numeri e del teorema limite centrale per martingale (si veda, per esempio, Hall e Heyde, 1980) si ottiene allora che $1/(n-p)\sum_{t=p+1}^{n} \mathbf{y}_{t-1}\mathbf{y}_{t-1}^\top \xrightarrow{p} \Gamma$, con Γ matrice non singolare di dimensione $p \times p$ costituita dalla autocovarianze di $\{y_t\}$, e

$$\sqrt{n-p}\;\overline{\psi}(\theta_0) \xrightarrow{d} N(0, \sigma^2\Gamma),$$

quando $n \to +\infty$. Inoltre, si può mostrare che, quando $n \to +\infty$, la successione di matrici $\frac{1}{n-p}\sum_{t=p+1}^{n} \psi_t(\theta_0)\psi_t(\theta_0)^\top = \frac{1}{n-p}\sum_{t=p+1}^{n} \varepsilon_t^2 \mathbf{y}_{t-1}\mathbf{y}_{t-1}^\top$ ha lo stesso comportamento della successione $\frac{\sigma^2}{n-p}\sum_{t=p+1}^{n} \mathbf{y}_{t-1}\mathbf{y}_{t-1}^\top$, e che, come conseguenza del Lemma 5(ii) di Chuang e Chan (2002), $\max_{p+1 \le t \le n} ||\psi_t(\theta_0)|| = o(n^{1/2})$ q.c..

Evidentemente, la funzione di stima (10) può essere scritta nella forma (1) ponendo $m = n - p$, $j = t - p$, $\ell = p + 1$, $\mathcal{Y}_j^\ell = \mathcal{Y}_j^{p+1} = \{y_{j+p}, y_{j+p-1}, y_{j+p-2} \dots, y_j\}$ e $\eta_j(\mathcal{Y}_j^\ell; \theta) = \psi_{j+p}(\theta)$. Inoltre, si ha $s = p$ e $k = \kappa = 1$. Quindi, per quanto detto sopra, il risultato di Chuang e Chan (2002, Lemma 1(ii)) sulla distribuzione asintotica della statistica di verosimiglianza empirica per il parametro θ_0 generata dalla (10) può essere ottenuto da un'applicazione del Corollario 1(B), essendo, in particolare, in questo caso $a_m = m^{1/2}$ e $b_m = 1\ \forall\ m$.

3.2 Verosimiglianza empirica per la media di processi stazionari

Si supponga che l'osservazione campionaria $\{y_1, y_2, \dots, y_n\}$ rappresenti la realizzazione di un processo $\{y_t\}$ strettamente stazionario, con media θ_0 e funzione di densità spettrale integrabile $g(\omega)$, $|\omega| < \pi$, tale che

$$\lim_{\omega \to 0} g(\omega)|\omega|^{2d}/u_d = 1.$$

Qui, d è un opportuno reale appartenente all'intervallo $(-1/2, 1/2)$ e $u_d > 0$ è un'opportuna costante. A seconda che sia $d = 0$, $d > 0$ o $d < 0$, il processo è detto, rispettivamente, debolmente dipendente (o a memoria corta), fortemente dipendente (o a memoria lunga) o antipersistente.

Si consideri la funzione di stima

$$\frac{1}{m}\sum_{j=1}^{m}(M_j - \theta), \tag{11}$$

dove $m = n - \ell + 1$ e $M_j = \sum_{i=j}^{j+\ell-1} y_i/\ell$ è la media campionaria calcolata sul blocco $\mathcal{B}_j^\ell = \{y_j, y_{j+1}, \dots, y_{j+\ell-1}\}$ costituito da ℓ variabili, con $1 \le \ell \le n$. Tale funzione di stima definisce, come stimatore per θ_0, la media delle medie mobili, definisce cioè lo stimatore $\hat{\theta} = \frac{1}{m}\sum_{j=1}^{m} M_j$. Tipicamente, a causa della sovrapposizione dei blocchi $\mathcal{B}_j^\ell$, $\hat{\theta}$ è diverso dalla media campionaria $\overline{y} = (1/n)\sum_{i=1}^{n} y_i$. Comunque, $\hat{\theta}$ e $\overline{y}$ sono asintoticamente equivalenti sotto opportune condizioni (si veda Nordmann, Sibbertsen e Lahiri, 2007).

Anche in questo caso, la funzione di stima (11) può essere combinata con la funzione di verosimiglianza empirica per ottenere una pseudo-verosimiglianza per l'inferenza sulla media θ_0 del processo $\{y_t\}$. Evidentemente, la funzione di stima (11) è del tipo (1), con $\mathcal{Y}_j^\ell = \mathcal{B}_j^\ell$ e $\eta_j(\mathcal{Y}_j^\ell;\theta) = M_j - \theta$. Inoltre, sotto le condizioni A.1, A.2 e A.3 di Nordmann, Sibbertsen e Lahiri (2007) (si veda, in particolare la loro Proposizione 2), sono verificati gli assunti del Corollario 1(A) del paragrafo precedente, con $b_m = \ell^{1-2d}$, $a_m = \sqrt{m^{1-2d}}$ e $s = \kappa = 1$. Quindi, il risultato del Teorema 2 di Nordmann, Sibbertsen e Lahiri (2007) sulla distribuzione asintotica della statistica di verosimiglianza empirica per la media del processo $\{y_t\}$, generata dalla (11), segue da un'applicazione del Corollario 1(A).

3.3 Verosimiglianza empirica e inferenza nel dominio delle frequenze

Sia $\{y_t\}$ un processo stazionario lineare definito come $y_t = \sum_{v=0}^{\infty} \gamma_v \varepsilon_{t-v}$, con $\gamma_0 = 1$, $\sum_{v=0}^{\infty} \gamma_v^2 < \infty$ e dove $\{\varepsilon_t\}$ è una sequenza di variabili casuali i.i.d., tali che $E[\varepsilon_t] = 0$, $E[\varepsilon_t^2] = \sigma^2 > 0$ e $E[\varepsilon_t^4] < +\infty$. Sia $g(\omega,\theta)$ la funzione di densità spettrale del processo, che si assume dipendere da un parametro $\theta = (\theta_1, \ldots, \theta_{p-1}, \sigma^2)^\top$ di dimensione p, a valori in uno spazio Θ. Si assuma che sia $g(\omega,\theta) \geq c > 0$, per $\omega \in [-\pi, \pi]$ e per ogni valore di θ. Si assuma, inoltre, che la funzione $g(\omega,\theta)$ sia derivabile tre volte in θ, con derivate parziali continue in ω e maggiorate (in modulo) da opportune funzioni di θ, continue in un qualche intorno del vero valore, diciamo θ_0, del parametro.

Sia $y_1, \ldots, y_n$ una realizzazione del processo $\{y_t\}$ e sia $\overline{y}$ la media delle n osservazioni campionarie. Le ordinate del periodogramma, relative alle frequenze $\omega_j = 2\pi j/n$, per $j = 1, \ldots, n-1$, sono date da

$$I(\omega_j) = \frac{1}{2\pi n}\left[\left\{\sum_{t=1}^{n}(y_t - \overline{y})sin(\omega_j t)\right\}^2 + \left\{\sum_{t=1}^{n}(y_t - \overline{y})cos(\omega_j t)\right\}^2\right].$$

Data la simmetria di $I(\cdot)$, per cui $I(\pi + a) = I(\pi - a)$, si può focalizzare l'attenzione sulle sole frequenze ω_j, $j = 1, \ldots, m$, con $m = (n-1)/2$.

Uno stimatore consistente $\hat{\theta}$ per θ_0 è lo stimatore di Whittle (1953), che è definito da una funzione di stima di tipo (1), con $\ell = n$, $\mathcal{Y}_j^\ell = \mathcal{Y}_j^n = \{y_1, y_2, \ldots, y_n\}$ per ogni j e

$$\eta_j(\mathcal{Y}_j^n;\theta) = \left\{\frac{I(\omega_j)}{g(\omega_j,\theta)} - 1\right\}\frac{\partial \log g(\omega_j,\theta)}{\partial\theta}, \quad j = 1, \ldots, m. \tag{12}$$

Una pseudo-verosimiglianza per θ_0 può dunque essere ottenuta combinando la funzione di verosimiglianza empirica e la funzione di stima $\overline{\eta}(\theta)$ la cui generica componente è data dalla (12) (si veda Monti, 1997).

È noto (Dzhaparidze, 1986, §2.6) che $\sqrt{m}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, W^{-1})$, quando $n \to +\infty$, con

$$W = \frac{1}{4\pi}\int_{-\pi}^{\pi} \frac{\partial \log g(\omega,\theta)}{\partial\theta}\left(\frac{\partial \log g(\omega,\theta)}{\partial\theta}\right)^\top\Bigg|_{\theta=\theta_0} d\omega.$$

D'altro canto, uno sviluppo di Taylor di $\overline{\eta}(\theta)$ attorno a θ_0 porta all'espressione

$$\overline{\eta}(\tilde{\theta}) = \overline{\eta}(\theta_0) + \overline{\eta}^{'}(\theta_0)(\tilde{\theta} - \theta_0) + \overline{\eta}^{''}(\theta^*)(\tilde{\theta} - \theta_0)^2,$$

nella quale $||\tilde{\theta} - \theta_0|| = O_p(m^{-1/2})$, $||\theta^* - \theta_0|| \leq ||\tilde{\theta} - \theta_0||$ e i simboli $'$ e $''$ a esponente indicano le operazioni di derivazione (una e due volte) rispetto a θ. Con argomentazioni analoghe a quelle usate in Monti (1997), si può mostrare che, quando $n \to +\infty$, $\overline{\eta}'(\theta_0) \xrightarrow{p} -W$ e che $\overline{\eta}''(\theta^*)(\tilde{\theta} - \theta_0)^2 = o_p(m^{-1/2})$. Di conseguenza, si ha che

$$\overline{\eta}(\tilde{\theta}) = \overline{\eta}(\theta_0) - (W + o_p(1))(\tilde{\theta} - \theta_0),$$

per ogni sequenza casuale $\tilde{\theta} = \tilde{\theta}(\mathbf{Y})$ tale che $||\tilde{\theta} - \theta_0|| = O(m^{-1/2})$ q.c., e che

$$\sqrt{m}\overline{\eta}(\theta_0) \xrightarrow{d} N(0, W).$$

In maniera analoga, si mostra che per una successione casuale $\tilde{\theta} = \tilde{\theta}(\mathbf{Y})$, tale che $||\tilde{\theta} - \theta_0|| = O(m^{-1/2})$ q.c., risulta $\overline{\Omega}(\tilde{\theta}) = \frac{1}{m}\sum_{j=1}^{m} \eta_j(\mathcal{Y}_j^n; \tilde{\theta})\eta_j^\top(\mathcal{Y}_j^n; \tilde{\theta}) = W + o_p(1)$. Infine, si ha

$$\overline{\nu}(\tilde{\theta}) = \max_{1 \leq j \leq m} ||\eta_j(\mathcal{Y}_j^n; \tilde{\theta})|| = \max_{1 \leq j \leq m} \left| \frac{I(\omega_j)}{g(\omega_j, \tilde{\theta})} - 1 \right| \left|\left| \frac{g'(\omega_j, \tilde{\theta})}{g(\omega_j, \tilde{\theta})} \right|\right|$$

e risulta

$$\left| \frac{I(\omega_j)}{g(\omega_j, \tilde{\theta})} - 1 \right| \leq \left| \frac{I(\omega_j)}{g(\omega_j, \tilde{\theta})} \right| + 1 \leq \left| \frac{\sum_{t=1}^{n}(y_t - \overline{y})^2}{n\pi c} \right| + 1$$

e $||g'(\omega_j, \tilde{\theta})/g(\omega_j, \tilde{\theta})|| \leq ||g'(\omega_j, \tilde{\theta})/c|| = O_p(1)$. Pertanto, purchè sia verificata la condizione B0, il ricorso alla calibrazione χ^2 per la funzione log-rapporto di verosimiglianza empirica profilo generata dalla funzione di stima $\overline{\eta}(\theta)$ con generica componente data dalla (12) è giustificato dal Teorema 2 del paragrafo precedente.

Riferimenti bibliografici

Adimari, G., Preo, P. (2007). Robust confidence intervals for log-location-scale models with right censored data, *J. Stat. Plann. Infer.*, **137**, 2832-2849.

Chuang, C.S., Chan, N.H. (2002). Empirical likelihood for autoregressive models, with applications to unstable time series, *Stat. Sinica*, **12**, 387-407.

Dzhaparidze, K.O. (1986), *Parameter estimation and hypothesis testing in spectral analysis in stationary time series.* Springer and Verlag, New York.

Hall, P., Heide, C.C. (1980), *Martingale limit theory and its application.* Academic Press, London.

Kitamura, Y. (1997). Empirical likelihood methods with weakly dependent processes, *Ann. Statist.*, **25**, 2084-2102.

Kolaczyk, E.D. (1994), Empirical likelihood for generalized linear models, *Stat. Sinica*, **4**, 199-218.

Monti, A.C. (1997). Empirical likelihood confidence regions in time series models, *Biometrika*, **84**, 395-405.

Norman, D.J., Sibbertsen, P., Lahiri, S.N. (2007). Empirical likelihood confidence intervals for the mean of a long-range dependent process, *J. Time Series Analysis*, **28**, 576-599.

Owen, A.B. (1990), Empirical likelihood ratio confidence regions, *Ann. Statist.*, **18**, 90-120.

Owen, A.B. (1991), Empirical likelihood for linear models, *Ann. Statist.*, **19**, 1725-1747.

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

Qin, J., Lawless, J. (1995), Estimating equations, empirical likelihood and constraints on parameters, *Canad. J. Statist.*, **23**, 145-159.

Whittle, P. (1953), Estimation and information in stationary time series, *Archiv. Math.*, **2**, 423-434.

Acknowledgements

Gli autori desiderano ringraziare Luisa Bisaglia e Silvano Bordignon per alcuni utili suggerimenti, che hanno reso più semplice la stesura dell'articolo.

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