STATISTICA, anno LXII, n. 1, 2002

SULLA DISTRIBUZIONE DI UNA BASE DI NORMA UNITARIA DEL COMPLEMENTO ORTOGONALE DI UN VETTORE GAUSSIANO: 11, CASO RIDIMENSIONALE ( )

A. Luati, P. Paruolo

1. INTRODUZIONE

In questa nota si deriva la distribuzione di una base del complemento ortogonale di un vettore aleatorio bidimensionale gaussiano. La stessa distribuzione si ottiene nel caso più generale di una distribuzione ellittica. L'interesse per questo problema è legato, fra l'altro, a problemi di inferenza in modelli di serie storiche con struttura fattoriale o con 'caratteristiche comuni' (common features) nel senso di Vahid e Engle (1993).

Nella presente nota si affronta il caso in cui la base del complemento ortogonale è normalizzata imponendone norma unitaria. In altri lavori (Paruolo, 1997) si è studiato lo stesso problema quando la normalizzazione è imposta mediante vincoli lineari affini.

La successione degli argomenti è così organizzata: nella sezione 2 si riportano alcuni riferimenti a modelli statistici che motivano il problema in oggetto; nella sezione 3 si definisce il problema e si enunciano le distribuzioni trovate; nella sezione 4 si riportano alcune rappresentazioni grafiche delle distribuzioni trovate; infine nella sezione 5 si riportano le dimostrazioni.

2. MOTIVAZIONE

Al solo scopo di motivare il presente lavoro, e senza alcuna pretesa di generalità, si consideri il seguente modello di regressione:

$$Z_t = \alpha U_t + \varepsilon_t \qquad t = 1, \ldots, T, \tag{1}$$

dove Z_t, α e ε_t sono vettori bidimensionali, U_t è scalare, ed ε_t è una successione di vettori aleatori i.i.d. a media nulla e con matrice di varianza e covarianza R. Si indichi con $\alpha_\perp$ una base qualsiasi del complemento ortogonale di $sp(\alpha)$, dove $sp(\cdot)$

(*) Questa nota raccoglie risultati originali della tesi di laurea di Alessandra Luati, di cui Paolo Paruolo è stato co-relatore.

indica lo spazio vettoriale generato dalle colonne dell'argomento. Anche $\alpha_\perp$ è pertanto, in questo caso, un vettore bidimensionale.

I1 modello (1) include numerosi casi particolari di modelli di interesse, ad esempio, nell'ambito dell'analisi di serie storiche. Infatti se $Z_t :=$ X, ,$U_t := \beta' X_{t-1}$, dove β è un vettore bidimensionale, la (1) corrisponde a un modello di "nested reduced rank autoregression" nel senso di Reinsel e Velu (1998), sotto opportune assunzioni di stazionarietà di X,. Alternativamente, nel modello (1) Z_t potrebbe rappresentare un vettore contenente il rendimento di due attività finanziarie e ti, potrebbe rappresentare un fattore comune di rischio; modelli fattoriali di questo tipo sono compatibili con le teorie del CAPM e di APT, si veda ad esempio Costa et al. (1997).

Se, invece, $Z_t :=$ AX, , ti, $:= \beta' X_{t-1}$, dove Δ indica l'operatore differenza prima, la (1) corrisponde a un modello a correzione dell'errore di un sistema co-integrato di ordine 1 nel senso di Johansen (1995), sotto opportune condizioni su α e β. In questo ambito si dimostra che $\alpha_\perp' \varepsilon_t$ rappresenta l'innovazione nei trend comuni del sistema.

I casi precedenti appartengono alla classe di processi con caratteristiche comuni fra le diverse variabili, o 'common features' nella terminologia di Vahid e Engle (1993). Si osservi che la combinazione lineare $\alpha_\perp' Z_t = \alpha_\perp' \varepsilon_t$ delle variabili osservabili Z_t non dipende da U, .Tale combinazione lineare è detta co-feature *vector*, in quanto elimina la dipendenza dal fattore comune, ed ha spesso un'interpretazione Eenomenica.

Si indichi con 0! lo stimatore dei minimi quadrati di a in (1). Lo scarto 0! - α è distribuito (asintoticamente) come un vettore bidimensionale gaussiano sotto varie assunzioni su (U, ε_t')'. La distribuzione di $\hat{\alpha}_\perp$ può risultare di interesse, ad esempio per fare inferenza sul co-feature *vector* $\alpha_\perp$. Dato che $\alpha_\perp$ è una qualsiasi base del complemento ortogonale, occorre fissarne una particolare imponendo una normalizzazione. Paruolo (1997) ha studiato la distribuzione di una base del complemento ortogonale di uno spazio lineare euclideo di dimensione finita con normalizzazione di alcune delle componenti a valori noti (tipicamente 0 e 1); nella presente nota di considera il caso di vincolo a uno della norma di $\hat{\alpha}_\perp$. Quest'ultimo vincolo è ad esempio suggerito in Hillier (1990).

3. PROBLEMA E RISULTATI

Si consideri un vettore $x = (x_1, x_2)' \in \mathbb{R}^2$ distribuito secondo una legge gaussiana x - $N(0, \sigma^2 \Omega)$ con

$$\Omega = \begin{pmatrix} k & \rho\sqrt{k} \\ \rho\sqrt{k} & 1 \end{pmatrix},$$

dove p è il coefficiente di correlazione e k è il rapporto della varianza di x, rispetto a quella di x_2. Si indichi con $y = (y_1, y_2)'$ la base del complemento ortogonale di x di norma unitaria, $\|y\| = 1$, dove $\|\ \|$ indica la norma euclidea. È semplice verificare che $y_1 = x_2/\|x\|$, $y_2 = -x_1/\|x\|$, o, $y_1 = \sin 0$, $y_2 = -\cos\theta$, dove r e θ indicano le coordinate polari corrispondenti a x_1 e a x_2. Si osservi che il vincolo su y di norma unitaria rende la distribuzione di y singolare; in altri termini la distribuzione di y è unidimensionale ed è sufficiente la distribuzione di y_1 o di y_2 a caratterizzarla univocamente. Si osservi inoltre che il campo di variazione di y_1 e y_2 è l'intervallo $[-1, 1]$. Con queste premesse vale pertanto la seguente proposizione.

Proposizione 1. La funzione di densità di probabilità di y_1 è data da

$$f_{y_1}(u) = h(u, k, \rho) \qquad u \in [-1, 1] \tag{2}$$

dove

$$h(u,k,\rho) = \frac{\sqrt{k}}{\pi}\left(\frac{g_1(\rho)}{g_1(u)}\right)^{1/2} \frac{g(u,k)}{(g(u,k))^2 - 4\rho^2 g_1(u)\, g_2(u,k)}$$

e

$$g_1(u) = (1 - u^2) \qquad g_2(u, k) = ku^2 \qquad g(u, k) = g_1(u) + g_2(u, k)$$

Alternativamente, la funzione di densità di probabilità di y_2 è pari a $f_{y_2}(u) = h(u, 1/k, \rho)$, $u \in [-1, 11.$

Si osservi che la distribuzione in *(2)* non dipende da o'. Questa osservazione è sufficiente a garantire che la distribuzione *(2)* rimanga inalterata anche assumendo che la distribuzione di x sia ottenuta come mistura di scala di normali.

Corollario 2. La funzione di densità *(2)* rimane immutara se x ha distribuzione ellittica.

4. RAPPRESENTAZIONE GRAFICA

I grafici in fig. 1-2 rappresentano la densità $h(u, k, p)$ al variare di k e p. Si ricordi che per $k = 1$ e $p = 0$, θ ha distribuzione uniforme sul cerchio di raggio unitario; la corrispondente distribuzione $h(u, 1, 0) = \pi^{-1}(1 - u^2)^{-1/2}$ ha forma 'ad U'. Il grafico in fig. 1 mostra la densità $h(u, k, p)$ per k fisso ($k = 1$) al variare di ρ, $p = 0$, 0,8, 0,95, *0,995*. Al crescere del valore del coefficiente di correlazione lo scatter bidimensionale di x si concentra attorno alla bisettrice del primo e terzo quadrante; le ascisse corrispondenti del vettore base del complemento ortogonale y si addensano corrispondentemente attorno ai punti $\pm\sqrt{2}/2$, caratterizzanti l'intersezione della bisettrice del secondo e del quarto quadrante con il cerchio unitario.

Si consideri ora la densità $h(u, k, p)$ per p fisso al variare del rapporto delle varianze k. Il grafico in fig. 2 descrive l'andamento per $p = 0{,}5$ con $k = 1$, 10, *50*, *500*. All'aumentare della differenza fra le varianze, misurata dal rapporto k, lo scatter dei punti x si 'allunga' sempre più attorno all'asse delle ascisse. Corrispondentemente, la prima coordinata della base del complemento ortogonale y si addensa sempre più attorno a 0, valore dell'ascissa dell'intersezione dell'asse delle ordinate con il cerchio unitario.

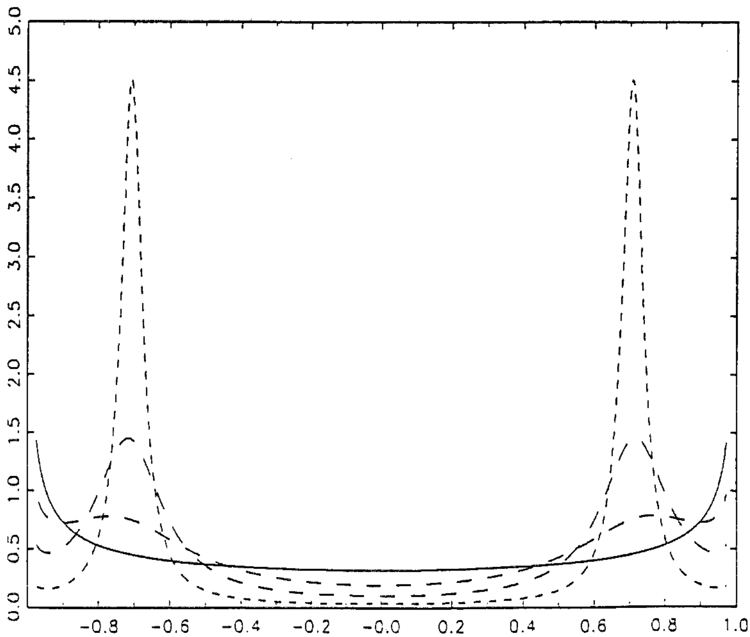

Figura 1 - Densità di probabilità $h(u, k, p)$ per $k = 1$ e $p = 0$, 0,8, 0,95, 0,995.

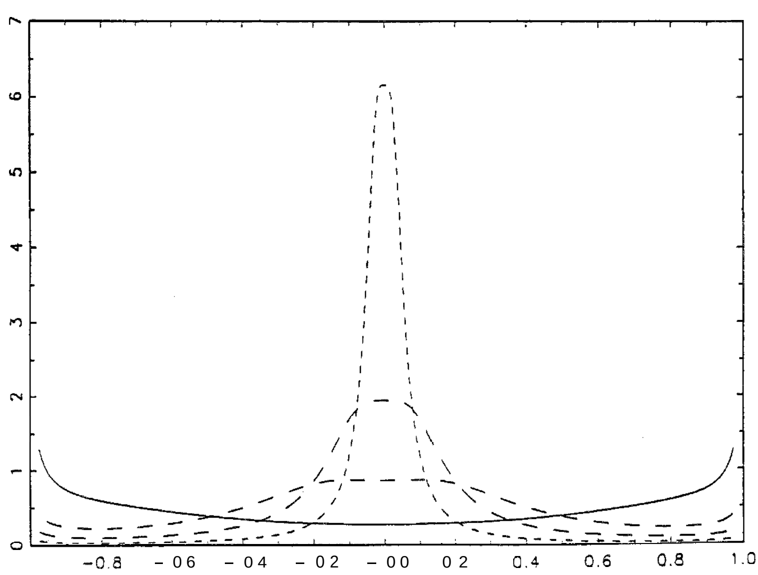

Figura 2 - Densità di probabilità $h(u, k, p)$ per $p = 0,5$ e $k = 1$, 10, 50, 500

5. DIMOSTRAZIONI

La dimostrazione della proposizione 1 è preceduta dal seguente lemma, che riporta un ben noto risultato relativo alla distribuzione gaussiana.

Lemma 3. La funzione di densità di probabilità di *0* è data da

$$f_\theta(v) = (2\pi)^{-1}(kg_1(\rho))^{1/2}(1 - 2\rho\sqrt{k}\sin v\cos v + (k - 1)\sin^2 v)^{-1} \tag{3}$$

per $\theta \in [0, 2\pi)$.

Si osservi che anche tale distribuzione non dipende da σ^2

Dim. lemma 3. La densità bivariata di r e *0* è data da

$$f_{r,\theta}(r, v) = (2\pi)^{-1}\sigma^{-2}(kg_1(\rho))^{-1/2} r\exp\left(-\frac{1}{2}r^2\sigma^{-2}(kg_1(\rho))^{-1}a\right)$$

dove lo jacobiano della traformazione da x a (r, O) è pari a r e $a = 1 - 2\rho\sqrt{k}\sin v$ $\cos v + (k - 1)\sin^2 v$. La distribuzione marginale di *0* si ottiene integrando $f_{r,\theta}(r, v)$ rispetto a r; l'integrale è del tipo $c_1\int_0^\infty r\exp\left(-\frac{1}{2}c_2r^2\right)dr = c_1/c_2$, dove $c_1 = (2\pi)^{-1}$ $\sigma^{-2}(kg_1(\rho))^{-1/2}$ e $c_2 = \sigma^{-2}(kg_1(\rho))^{-1}a$. Dividendo c_1 per c_2 si ottiene la (3). □

Si dimostra ora la proposizione 1.

Dim. proposizione 1. Sia $1\{\cdot\}$ la funzione indicatrice e si divida il supporto della distribuzione di *0* nei sottointervalli $A_1 = \{\pi/2 \le \theta \le 3\pi/2\}$, $B = \{3\pi/2 < O < 2\pi\}$, $C = 0 \le 0 < \pi/2$. Si osservi che $0 = \arcsin(y_1)$ è una funzione biunivoca negli intervalli disgiunti A_1 e $A_2 = B \cup C$. Pertanto

$$f_{y_1}(u) = (g_1(u))^{-1/2}\sum_{i=1}^{2} f_\theta(\arcsin(u))1\{\arcsin(u) \in A_i\},$$

dove $(g_1(y_1))^{-1/2}$ è lo jacobiano della trasformazione da *0* a y_1. Sostituendo e semplificando si ottiene la *(2)*; analogamente si procede per y_2. □

Si dimostra infine che il risultato vale anche per distribuzioni ellittiche.

Dim. corollario 2. Se x ha distribuzione ellittica, ha rappresentazione $x = \sigma z$ con $z \sim N(0, \Omega)$ e σ è una qualsiasi variabile aleatoria positiva (Fang *et al., 1990)*. La distribuzione *(2)* è la distribuzione di y_1 condizionata a σ. Essendo la *(2)* indipendente da σ, essa coincide anche con la distribuzione marginale di y_1. □

Dipartimento di Scienze statistiche "P. Fortunati"
Università di Bologna
ALESSANDRA LUATI

Dipartimento di Economia
Università dell'Insubria, Varese
PAOLO PARUOLO

RIFERIMENTI HIHLIOGKAFICI

M. COSTA, A. GARDINI, P. PARUOLO (1997), A *reduced rank regression approach to tests of asset pricing,* "Oxford Bulletin of Economics and Statistics", 59, pp. 163-181.

K.T. FANG, S. KOTZ, K.W. NG (1990), *Symmetric multivariate and related distrihutions,* Chapman and Hall.

G. HILLIER (1990), *On the normalization of structural equations: propertiers of direction estimators,* "Econometrica", 58, pp. 1181-1194.

S. JOHANSEN (1995), *Likelihood based inference in cointegrated vector autoregressie models,* Advanced texts in econometrics, Oxford Universitv Press.

P. PARUOLO (1997), *Asymptotic inference on the moving average impact matrix in cointegrated I(1) VAR systems,* "Econometric Theory", 13, pp. 79-118.

G.C. REINSEL, R.P. VELU (1998), *Multivariate reduced rank regression, theory and applications,* Lecture notes in Statistics 136, Springer Verlag, New York.

F. VAHID, R. ENGLE (1993), *Common trends and common cycles,* "Journal of Applied Econometrics", 8, pp. 341-360.

RIASSUNTO

Sulla distribuzione di una base di norma unitaria del complemento ortogonale di un vettore gaussiano: il caso bidimensionale

Questa nota illustra la distribuzione della base di norma unitaria del complemento ortogonale di un vettore aleatorio bidimensionale gaussiano. Si mostra inoltre che si ottiene la stessa distribuzione nel caso più generale di una distribuzione ellittica.

SUMMARY

The distribution of a unit norm basis of the orthogonal complement of a bidimensional Gaussian vector

The present note shows the distribution of a unit norm basis of the orthogonal complement of a two-dimensional Gaussian random vector. It also shows that the same distrihution can be obtained in the more general case of an elliptic distribution.