

IL MODELLO LINEARE NELL'APPROCCIO BAYESIANO  
NON PARAMETRICO

D.M.CIFARELLI , P.MULIERE, M.SCARSINI<sup>(\*)</sup>

1. INTRODUZIONE

Si supponga di avere a disposizione  $n$  osservazioni  $(x_{i1}, \dots, x_{ik}, y_i)$   $i=1, 2, \dots, n$  sul vettore aleatorio  $(k+1)$  dimensionale  $(X_1, X_2, \dots, X_k, Y)$  e di essere interessati a "spiegare" la variabile  $Y$  mediante  $X_1, \dots, X_k$ .

Un modello comunemente utilizzato a questo scopo è quello della regressione in cui si assume che le osservazioni soddisfino la relazione

$$\underline{y} = X\underline{\beta} + \underline{\varepsilon} \quad (1)$$

con  $\underline{y}$  vettore delle osservazioni sulla variabile  $Y$  del tipo  $n \times 1$ ,  $X$  matrice delle osservazioni sulle variabili  $(X_1, \dots, X_k)$  di tipo  $n \times k$ ,  $\underline{\beta}$  vettore dei parametri  $k \times 1$  ed  $\underline{\varepsilon}$  vettore aleatorio, non osservabile,  $n \times 1$ , con valore atteso nullo e matrice di varianze e covarianze  $\sigma^2 I_n$ ,  $\sigma^2 > 0$ .

(\*) Il coordinamento del testo del lavoro è di D.M.Cifarelli al quale sono dovuti altresì i paragrafi 1,2,4. A P.Muliere sono dovuti i paragrafi 3 e 6, mentre il paragrafo 5 è dovuto a M.Scarsini che lo ha tratto dalla sua tesi di laurea in Discipline Economiche e Sociali presso l'Università L.Bocconi di Milano.

I lavori contenuti in questo quaderno sono stati presentati nel ciclo di seminari sui Modelli Macroeconomici, organizzato dal gruppo di ricerca MONIF nell'A.A. 1979-80.

Il modello (1) è detto lineare, ma occorre sottolineare che tale qualifica non proviene dalla linearità di  $\underline{y}$  rispetto alla variabile  $X$  bensì dalla linearità riguardo ai parametri  $\beta_i$ . E' inoltre consuetudine assegnare denominazione differente al modello a seconda delle ipotesi che si adottano riguardo alle quantità che servono a definirlo. Si hanno così i modelli della regressione funzionale, condizionata, stocastica, con variabili affette da errore, etc.

L'approccio classico al problema della "stima" del parametro  $\underline{\beta}$  è quello dei minimi quadrati "arricchito" con risultati di natura distributiva nel caso di normalità del vettore  $\underline{\epsilon}$ .

C.Stein (1956), trattando il cosiddetto problema delle  $k$  medie

$$\underline{y} = \underline{\xi} + \underline{u} \quad (2)$$

in cui  $\underline{u}$  è un vettore aleatorio normale con valore atteso nullo e matrice di varianze e covarianze  $\sigma^2 I_k$ ,  $\underline{\xi}$  è un vettore di parametri  $k \times 1$  con  $E(\underline{y}|\underline{\xi}) = \underline{\xi}$ , mostrò che la stima dei minimi quadrati di  $\underline{\xi}$ , cioè  $\hat{\underline{\xi}} = \underline{y}$ , quando  $k \geq 3$  e la funzione di danno è quadratica, non è ammissibile, vale a dire esistono altri stimatori con funzione di rischio non maggiore per ogni  $\underline{\xi}$ . Tale risultato fu successivamente generalizzato da L.D.Brown (1966) al caso di una vasta classe di funzioni di danno. Poiché, trasformando opportunamente le variabili (ed il parametro  $\underline{\beta}$ ), il modello (1) può scriversi nella forma (2), il risultato di Stein si applica anche alla stima dei minimi quadrati di  $\underline{\beta}$ .

Si deve forse a risultati di tale natura la spinta degli statistici a tentare altri approcci ed in particolare quello

bayesiano. Ed infatti, lo stesso Stein, in un lavoro successivo (1962), dopo aver proposto uno stimatore ammissibile (ridge) si preoccupò di porre in luce come tale stimatore potesse essere interpretato come un'approssimazione di quello bayesiano nella consueta ipotesi di danno quadratico. D.V. Lindley (1962) fornì analogo interpretazione introducendo, in "nuce", quello che successivamente verrà detto "modello gerarchico". Una disamina abbastanza recente di questo aspetto del problema può rintracciarsi nel lavoro di A. Zellner e W. Vandaele (1974).

L'approccio bayesiano al modello lineare della regressione, sia per quanto riguarda i problemi di natura ipotetica (che riguardano cioè i parametri del modello statistico) sia per quanto concerne l'aspetto predittivo (che riguarda cioè la previsione di future altre osservazioni), procede, come di consueto, con l'assegnazione di una distribuzione iniziale sui parametri che compaiono nel modello statistico e quindi con la determinazione di quella finale per il tramite della formula di Bayes.

Senza eccezioni, il modello statistico impiegato nei problemi di regressione è quello normale (eventualmente con osservazioni correlate), cioè dati  $X$ ,  $\underline{\beta}$  e  $\sigma$ , si assume che  $\underline{y}$  posseda densità (ordinaria) data da

$$p(\underline{y}|X, \underline{\beta}, \sigma) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}(\underline{y} - X\underline{\beta})'(\underline{y} - X\underline{\beta})\right\}$$

$$\underline{y} \in \mathbb{R}^n, \quad \underline{\beta} \in \mathbb{R}^k, \quad \sigma \in \mathbb{R}_+$$

Le distribuzioni tradizionalmente utilizzate per i parametri  $\underline{\beta}$  e  $\sigma$  sono sia non informative improprie (H. Jeffreys (1961), L.J. Savage (1962)) sia di tipo informative e coniugate (G.C. Tiao e A. Zellner (1964), A. Zellner (1971)). Nel caso di distri

buzione iniziale non informativa à la Jeffreys, è noto che la distribuzione finale di  $\underline{\beta}$  è normale con valore atteso  $\hat{\underline{\beta}} = (X'X)^{-1}X'y$  e matrice di varianze e covarianze  $\sigma^2(X'X)^{-1}$ , nel caso della varianza  $\sigma^2$  nota e in assenza di multicollinearità, mentre è data da una distribuzione t multivariata con valore atteso  $\hat{\underline{\beta}}$  e matrice di varianze e covarianze

$$\frac{n-k}{n-k-2} S^2 (X'X)^{-1}, \quad (n-k)S^2 = (\underline{y} - X\hat{\underline{\beta}})'(\underline{y} - X\hat{\underline{\beta}})$$

quando la varianza non è nota.

A nostra conoscenza, la prima trattazione non tradizionale del modello lineare, sempre nell'ambito del modello statistico normale, è quella dovuta a D.V. Lindley (1971) il quale introdusse l'idea della scambiabilità dei parametri e di modello gerarchico, idea quest'ultima già utilizzata dallo stesso Lindley (1969) nel tentativo di giustificare, in ambito bayesiano, la stima dei minimi quadrati. Si deve infine a D.V. Lindley e A.F.M. Smith (1972) la trattazione generale del modello lineare con impostazione gerarchica.

Le stesse idee furono successivamente riprese da A.F.M. Smith (1973a), (1973b), e da D.V. Lindley (1974) per quanto attiene più strettamente alla problematica dell'analisi della varianza.

Come abbiamo accennato, caratteristica comune dei lavori riguardanti il modello regressivo è l'assunzione di normalità del modello statistico, cioè del meccanismo di generazione delle osservazioni ed, in subordine, di linearità della regressione.

E' possibile prescindere da tali assunzioni? In altre parole, è possibile, senza impegnarsi eccessivamente sul modello sta-

tistico, procedere utilizzando ancora il meccanismo bayesiano? La risposta, come vedremo, è in qualche misura affermativa a patto di riuscire ad escogitare convenienti misure di probabilità (iniziali) sullo spazio delle funzioni di ripartizione.

## 2. IL PROCESSO DI DIRICHLET COME DISTRIBUZIONE INIZIALE SULLA CLASSE DELLE FUNZIONI DI RIPARTIZIONE

Per introdurre la problematica rifacciamoci per un momento al classico modello statistico parametrico

$$(Z, \mathcal{A}, P)$$

in cui  $Z$  è un insieme (spazio dei risultati),  $\mathcal{A}$  una  $\sigma$ -algebra di sottoinsiemi di  $Z$  e  $P$  una classe di misure di probabilità sullo spazio misurabile  $(Z, \mathcal{A})$  del tipo

$$P \equiv \{P_\theta, \theta \in \Theta\}$$

Quando si voglia impostare un problema inferenziale utilizzando il meccanismo bayesiano occorre, come è noto, pensare probabilizzato anche  $(\Theta, \mathcal{B}_\Theta)$  con una misura di probabilità  $Q$  (detta iniziale). In questo contesto la risposta globale bayesiana ai problemi inferenziali sotto forma ipotetica è rappresentata dalla distribuzione finale del parametro  $\theta \in \Theta$  ottenuta tramite la applicazione della formula di Bayes. La distribuzione iniziale  $Q$  su  $(\Theta, \mathcal{B}_\Theta)$  svolge direttamente l'ufficio di probabilizzare lo "spazio dei parametri" e indirettamente quello delle misure di probabilità  $P$  su  $(Z, \mathcal{A})$ . Se le misure di probabilità  $P_\theta$  non sono

analiticamente note, allora una versione molto semplice di modello "non parametrico" potrebbe essere

$$(\mathbb{R}^n, \mathcal{B}^n, F^n)$$

in cui  $F^n$  è la classe delle funzioni di ripartizione della forma  $\prod_{i=1}^n F(x_i)$ .

In analogia al caso parametrico ora si tratterà di probabilizzare la classe  $F = F^1$  delle funzioni di ripartizione da farsi però direttamente non potendo usufruire dell'intermediazione del parametro  $\theta$ . In altre parole occorrerà proporre dei processi aleatori che possano fungere da distribuzione iniziale sulla classe delle funzioni di ripartizione  $F \in \mathcal{F}$ .

Una svolta decisiva per la soluzione di questo difficile problema si ebbe nel 1973 allorché T.S. Ferguson (1973) introdusse una classe di processi aleatori i cui membri ben si adattano, da molti punti di vista come distribuzione iniziale su spazi misurabili astratti ed in particolare su  $(\mathbb{R}^k, \mathcal{B}^k)$ . Per l'applicazione che faremo nel paragrafo seguente ricorderemo ora la definizione ed alcune proprietà del processo proposto e studiato da T.S. Ferguson col nome di processo di Dirichlet.

DEFINIZIONE 1 (T.S. Ferguson (1973)). Sia  $\alpha$  una misura non negativa, finitamente additiva, finita su  $(X, \mathcal{A})$ . Si dice che  $(P(A), A \in \mathcal{A})$  è un processo di Dirichlet su  $(X, \mathcal{A})$  con parametro  $\alpha$ , se per ogni partizione misurabile di  $X$ ,  $(B_1, B_2, \dots, B_k)$ , il vettore aleatorio

$$(P(B_1), \dots, P(B_k))$$

possiede densità (ordinaria)  $(k-1)$  dimensionale data da

$$\frac{\Gamma(\alpha(X))}{\prod_{i=1}^k \Gamma(\alpha(B_i))} \left( \prod_{i=1}^{k-1} y_i^{\alpha(B_i)-1} \right) \left( 1 - \sum_{i=1}^{k-1} y_i \right)^{\alpha(B_k)-1}$$

su  $\{(y_1, \dots, y_{k-1}) : y_i \geq 0 \quad i=1, 2, \dots, k-1, \sum_{i=1}^{k-1} y_i \leq 1\}$ .

Un risultato importante riguardante il processo di Dirichlet è la sua proprietà di "chiusura" per precisare la quale occorre ricordare la definizione di campione estratto dal processo.

DEFINIZIONE 2 (T.S. Ferguson (1973)). Sia  $P$  un processo di Dirichlet su  $(X, \mathcal{A})$ . Diremo che il vettore  $(X_1, X_2, \dots, X_n)$  è un campione di ampiezza  $n \geq 1$  estratto da  $P$  se  $\forall m = 1, 2, \dots$  ed insiemi misurabili  $A_1, \dots, A_m, C_1, \dots, C_n$ ,

$$\begin{aligned} \Pr\{X_1 \in C_1, \dots, X_n \in C_n \mid P(A_1), \dots, P(A_m), P(C_1), \dots, P(C_n)\} = \\ = \prod_{j=1}^n P(C_j) \text{ q.c.} \end{aligned}$$

Il senso della Definizione 2 è che se è nota una realizzazione del processo  $P$ , allora  $(X_1, \dots, X_n)$  è un vettore di v.a. mutuamente indipendenti ed identicamente distribuite.

TEOREMA 1 (T.S. Ferguson (1973)). Se  $P$  è un processo di Dirichlet con parametro  $\alpha$  e se  $(X_1, \dots, X_n)$  è un campione estratto da  $P$ , allora il processo  $(P \mid X_1, \dots, X_n)$  è ancora di Dirichlet con parametro  $\beta = \alpha + \sum \delta_{X_i}$  essendo  $\delta_x(A) = 1$  se  $x \in A$ ,  $= 0$  se  $x \notin A$ .

Se  $X = \mathbb{R}^k$ ,  $A = B^k$  e  $F(\underline{x}) = P((-\infty, x_1], \dots, (-\infty, x_k])$ , essendo  $P$  un processo di Dirichlet con parametro  $\alpha$ , allora

$$(F(\underline{x}), \underline{x} \in \mathbb{R}^k)$$

è una f.r. aleatoria retta da un processo di Dirichlet "ordinato" con parametro  $\alpha$ , vale a dire il vettore

$$\{F(\underline{x}^{(1)}), F(\underline{x}^{(2)}), \dots, F(\underline{x}^{(p)}),\}$$

con  $\underline{x}^{(1)} < \underline{x}^{(2)} < \dots < \underline{x}^{(p)}$  vettori di  $\mathbb{R}^k$  e con la consueta interpretazione del segno di disuguaglianza, possiede densità data da:

$$\frac{\Gamma(\alpha(\mathbb{R}^k))}{\Gamma(\alpha(\underline{x}^{(1)}))\Gamma(\alpha(\underline{x}^{(2)})-\alpha(\underline{x}^{(1)}))\dots\Gamma(\alpha(\mathbb{R}^k)-\alpha(\underline{x}^{(p)}))} y_1^{\alpha(\underline{x}^{(1)})-1} \times$$

$$\times (y_2 - y_1)^{\alpha(\underline{x}^{(2)})-\alpha(\underline{x}^{(1)})-1} \dots (y_p - y_{p-1})^{\alpha(\underline{x}^{(p)})-\alpha(\underline{x}^{(p-1)})-1} \times$$

$$\times (1 - y_p)^{\alpha(\mathbb{R}^k)-\alpha(\underline{x}^{(p)})-1}$$

su  $0 \leq y_1 \leq y_2 \leq \dots \leq y_p \leq 1$  ed in cui abbiamo utilizzato la scrittura:

$$\alpha((-\infty, x_1^{(i)}], (-\infty, x_2^{(i)}], \dots, (-\infty, x_k^{(i)}]) = \alpha(\underline{x}^{(i)}) \quad i = 1, 2, \dots, p$$

Inoltre, se  $\underline{X}_1, \dots, \underline{X}_n$  è un campione estratto da  $F$  si può dimostrare che

$$\Pr\{\underline{X}_i \leq \underline{x}\} = \frac{\alpha(\underline{x})}{\alpha(\mathbb{R}^k)} = F_0(\underline{x}) \quad i = 1, 2, \dots, n \quad (3)$$

ed in più la f.r. predittiva è data da

$$\Pr\{\underline{X}_{-n+1} \leq \underline{x} | \underline{X}_1 = \underline{x}_1, \dots, \underline{X}_n = \underline{x}_n\} =$$

$$= \frac{\alpha(\mathbb{R}^k)}{\alpha(\mathbb{R}^k) + n} F_0(\underline{x}) + \frac{n}{\alpha(\mathbb{R}^k) + n} \hat{F}_n(\underline{x}) \quad (4)$$

essendo  $\hat{F}_n(\underline{x})$  la f.r. empirica. La (3) fornisce il significato del parametro  $\alpha$  del processo facendolo apparire come la f.r. (opportunamente normalizzato) (iniziale) di  $\underline{X}_i$  (osservabile).

### 3. IL MODELLO STOCASTICO

Come abbiamo già detto nell'Introduzione, il modello comunemente utilizzato per lo studio delle relazioni tra la variabile  $Y$  e le variabili  $X_1, \dots, X_k$  è quello che assume

$$\underline{y} = X\underline{\beta} + \underline{\varepsilon}$$

La stima tradizionale (dei minimi quadrati) quando i disturbi si assumono non correlati e con la stessa varianza e  $X'X$  è non singolare, è data da  $(X'X)^{-1}X'\underline{y}$ . Come pure abbiamo detto un approccio alternativo è quello bayesiano di tipo parametrico. Entrambi i procedimenti presuppongono però l'esistenza di un vettore  $\underline{\beta}$  per cui

$$E(\underline{y} | X, \underline{\beta}) = X\underline{\beta}$$

Quando tale assunto non si ritiene accettabile si può procedere secondo una via indicata da M. Goldstein (1976) che ora

esamineremo nell'ambito del processo di Dirichlet.

Sia  $(X_1, \dots, X_k, Y)$  un vettore aleatorio con f.r.  $F(x_1, \dots, x_k, y)$  non nota ma retta da un processo di Dirichlet (ordinato) con parametro  $\alpha(x_1, \dots, x_k, y)$  e sia

$$\begin{pmatrix} x_{11} & x_{12} & \dots & x_{1k} & y_1 \\ x_{21} & x_{22} & \dots & x_{2k} & y_2 \\ \vdots & \vdots & & \vdots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nk} & y_n \end{pmatrix} \equiv (X; \underline{y})$$

un campione di ampiezza  $n$  estratto da  $F$ .

Se  $(X_{n+1,1}, \dots, X_{n+1,k}, Y_{n+1})$  indica una futura osservazione del processo, allora un modo naturale di affrontare il problema è quello di ricercare il vettore  $\underline{\beta}$  in modo che

$$\sum_{i=1}^k \beta_i X_{n+1,i}$$

riesca una buona previsione di  $Y_{n+1}$  nel senso di minimizzare

$$E\left\{ \left( Y_{n+1} - \sum_{i=1}^k \beta_i X_{n+1,i} \right)^2 \mid (X; \underline{y}) \right\}$$

in cui il valore atteso è calcolato, ovviamente, con la f.r. predittiva di  $(X_{n+1,1}, \dots, X_{n+1,k}, Y_{n+1})$ .

Se si indica con  $D$  la matrice di tipo  $k \times k$  con generico elemento

$$d_{ij} = E\{X_{n+1,i} X_{n+1,j} \mid (X; \underline{y})\} \quad i, j = 1, \dots, k$$

e con  $\underline{b}$  il vettore  $(k \times 1)$  con elementi

$$b_i = E(X_{n+1,i}, Y_{n+1} \mid (X; \underline{y})) \quad i = 1, 2, \dots, k$$

allora è noto che il vettore  $\underline{\beta}$  richiesto è dato da

$$\underline{\beta}^* = D^{-1} \underline{b}$$

Se ora utilizziamo l'espressione della f.r. predittiva per un processo di Dirichlet data dalla (4) del paragrafo precedente (in cui però al posto di  $k$  bisogna porre  $k+1$ ), l'espressione di  $d_{ij}$  è

$$d_{ij} = \frac{\alpha(\mathcal{R}^{k+1})}{\alpha(\mathcal{R}^{k+1}) + n} E_{F_0}(X_i X_j) + \frac{n}{\alpha(\mathcal{R}^{k+1}) + n} \left( \frac{1}{n} \sum_{t=1}^n x_{ti} x_{tj} \right)$$

in cui il valore atteso è calcolato con la f.r. iniziale

$$F_0(x_1, \dots, x_k, y) = \frac{\alpha(x_1, \dots, x_k, y)}{\alpha(\mathcal{R}^{k+1})}$$

Da ciò discende che la matrice  $D$  può scriversi come

$$D = pM + (1-p) \left( \frac{X'X}{n} \right)$$

essendo

$$p = \frac{\alpha(\mathcal{R}^{k+1})}{\alpha(\mathcal{R}^{k+1}) + n}$$

Similmente il vettore  $\underline{b}$  risulta

$$\underline{b} = p\underline{m} + (1-p) \left( \frac{X'Y}{n} \right)$$

essendo  $\underline{m}$  il vettore di elementi  $m_i = E_{F_0}(X_i Y)$ ;  $i = 1, 2, \dots, k$ .

Se il peso  $\alpha(\mathcal{R}^{k+1})$  assegnato alla distribuzione iniziale  $F_0$  è elevato (rispetto ad  $n$ ) ed in particolare se  $\alpha(\mathcal{R}^{k+1}) \rightarrow +\infty$ ,

allora la matrice  $D$  coincide con  $M$  ed il vettore  $\underline{b}$  con  $\underline{m}$ . In questo caso si avrebbe

$$\underline{\beta}^* = M^{-1} \underline{m} = \underline{\beta}_0 \quad (M \text{ non singolare})$$

vale a dire  $\underline{\beta}^*$  è determinato esclusivamente in base all'opinione iniziale formalizzata da  $F_0$ .

Viceversa se  $\alpha(\mathcal{R}^{k+1})$  è trascurabile, allora  $D$  coincide con  $X'X$  mentre  $p$  è prossimo a  $X'y$  sicchè si avrebbe approssimativamente

$$\underline{\beta}^* = (X'X)^{-1} X'y = \hat{\underline{\beta}}$$

Nel caso in cui  $0 < p < 1$ , si ha

$$\underline{\beta}^* = (pM + (1-p)\left(\frac{X'X}{n}\right))^{-1} (pM\underline{\beta}_0 + (1-p)\left(\frac{X'X}{n}\right)\hat{\underline{\beta}})$$

cioè una media ponderata di  $\underline{\beta}_0$  e  $\hat{\underline{\beta}}$ .

#### 4. UNA DISTRIBUZIONE INIZIALE SU UN VETTORE DI MISURE DI PROBABILITA' ALEATORIE

In questo paragrafo vogliamo presentare le definizioni e le principali proprietà di un processo proposto da D.M. Cifarelli e E. Regazzini (1978) che poi utilizzeremo per trattare il modello della regressione condizionata da un punto di vista non parametrico (in  $\mathcal{R}$ ).

DEFINIZIONE 3. Se subordinatamente alla realizzazione  $u_1, u_2, \dots, u_k$  dei numeri (o vettori aleatori  $U_1, \dots, U_k$  ciascuna

misura di probabilità aleatoria  $P_i$   $i=1, \dots, k$  costituisce un processo di Dirichlet con parametro  $\alpha_i(u_i, \cdot): \mathcal{R} \times \mathcal{B} \rightarrow [0, \infty)$ , diremo che il vettore  $(P_1, \dots, P_k)$  costituisce un processo mistura di prodotti di processi di Dirichlet se

$$\Pr\left\{ \bigwedge_{i=1}^k \bigwedge_{j=1}^{m_i-1} (P_i(B_{ij}) \leq y_{ij}) \right\} = \int_{\mathcal{R}^k} \prod_{i=1}^k \mathcal{D}(y_{i1}, \dots, y_{i, m_i-1} | \alpha(u_i, B_{i,1}) \dots \dots \alpha_i(u_i, B_{i, m_i})) d\varphi(\underline{u})$$

essendo  $B_{i,1}, \dots, B_{i, m_i}$   $i=1, \dots, k$  la generica partizione misurabile di  $\mathcal{R}$  e  $\varphi(\underline{u})$  la f.r. di  $\underline{u} = (u_1, \dots, u_k)$ .

DEFINIZIONE 4. Si dice che  $\underline{X}_i = (X_{i,1}, \dots, X_{i, n_i})$ ,  $i=1, 2, \dots, n$  è un campione di ampiezza  $(n_1, n_2, \dots, n_k)$  estratto dal processo  $(P_1, \dots, P_k)$  se, per ogni  $(n_1, \dots, n_k)$ ,  $(m_1, \dots, m_k)$  e per ogni coppia di classi di sottoinsiemi misurabili di  $\mathcal{R}$

$$\{C_{ij}\}_{j=1, \dots, n_i}; i=1, \dots, k), \quad (\{A_{iq}\}_{q=1, \dots, m_i}; i=1, \dots, k)$$

risulta

$$\Pr\left\{ \bigwedge_{i=1}^k \bigwedge_{j=1}^{n_i} (X_{ij} \in C_{ij}) \mid \underline{u}, P_i(A_{i,q}), P_i(C_{ij}); j=1, \dots, n_i; q=1, \dots, m_i \right\} = \prod_{i=1}^k \prod_{j=1}^{n_i} P_i(C_{ij}) \quad (\text{q.c.})$$

Il senso di questa definizione è immediato: subordinatamente ad  $\underline{u}$ , se è nota una realizzazione del processo  $(P_1, \dots, P_k)$  allora i n.a.  $\underline{X}_1, \underline{X}_2, \dots, \underline{X}_k$  sono mutuamente indipendenti e, allo interno di ogni classe, identicamente distribuiti.

Il processo della Definizione 3, al pari di quello di Dirichlet, gode della proprietà di riproducibilità, vale a dire, se  $\underline{X}_1, \dots, \underline{X}_k$  è un campione estratto dal processo  $(P_1, \dots, P_k)$ , allora  $(P_1, \dots, P_k | \underline{X}_1, \dots, \underline{X}_k)$  è ancora dello stesso tipo salvo che per i suoi parametri che risultano modificati e per la f.r.  $\varphi(\underline{u})$  sostituita da  $\varphi(\underline{u} | \underline{x}_1, \dots, \underline{x}_k)$ .

Precisamente, se  $\underline{X}_1, \dots, \underline{X}_k$  è un campione estratto da  $(P_1, \dots, P_k)$ , allora

$$\Pr\left\{ \bigwedge_{i=1}^k \bigwedge_{j=1}^{m_i-1} (P_i(B_{i,j}) \leq y_{ij}) \mid \underline{X}_1 = \underline{x}_1, \dots, \underline{X}_k = \underline{x}_k \right\} =$$

$$= \int_{\mathbb{R}^k} \prod_{i=1}^k \mathcal{D}(y_{i,1}, \dots, y_{i,m_i-1}) | \alpha_i(u_i, B_{i,1} + n_{i,1}), \dots$$

$$\dots, \alpha_i(u_i, B_{i,m_i} + n_{i,m_i}) d\varphi(\underline{u} | \underline{x}_1, \dots, \underline{x}_k)$$

in cui  $n_{ij} = n_i(B_{i,j})$ ;  $j=1, \dots, m_i$ ,  $i=1, \dots, k$  rappresenta il numero delle osservazioni di  $\underline{x}_i$  che assumono valori in  $B_{ij}$  mentre  $\varphi(\underline{u} | \underline{x}_1, \dots, \underline{x}_k)$  rappresenta la f.r. di  $\underline{u}$  subordinato a  $\underline{x}_1, \dots, \underline{x}_k$ .

Per precisare l'espressione di  $\varphi(\underline{u} | \underline{x}_1, \dots, \underline{x}_k)$  supponiamo di aver organizzato le osservazioni relative alla  $i$ -ma classe,  $i=1, \dots, k$ , nel modo seguente (si osservi che per una caratteristica del processo più osservazioni possono coincidere con probabilità positiva):

$$\begin{cases} x_{i,1}, x_{i,2}, \dots, x_{i,r_i} \\ n_{i,1}, n_{i,2}, \dots, n_{i,r_i} \end{cases} \quad \sum_{j=1}^{r_i} n_{i,j} = n_i$$

e di indicare con  $\mu_i$  una misura la quale coincide con quella di Lebesgue su  $\mathbb{R}$  salvo nei punti in cui  $\alpha_i(u_i, \cdot)$  presenta un atomo dove  $\mu_i$  concentra una massa unitaria.

Dopo aver posto  $\underline{\mu} = \prod_{i=1}^k \mu_i$  si ha

$$d_{\underline{\mu}} \varphi(\underline{u} | \underline{x}_1, \dots, \underline{x}_k) \propto \prod_{i=1}^k \frac{1}{(\alpha_i(u_i, \mathbb{R}))_{(n_i)}} \cdot \prod_{j=1}^{r_i} \alpha'(u_i, x_{ij}) \cdot (m(u_i, x_{ij})+1)_{(n_{ij}-1)} d_{\underline{\mu}} \varphi(\underline{u}) \quad (5)$$

in cui

$$a_{(n)} = a(a+1) \dots (a+n-1)$$

$\alpha'(u_i, \cdot)$  è la derivata secondo Radon-Nikodym di  $\alpha_i$  rispetto alla misura  $\mu_i$

$$m_i(u_i, x) = \begin{cases} \alpha'_i(u_i, x) & \text{se } x \text{ è atomo di } \alpha_i(u_i, \cdot) \\ 0 & \text{altrove.} \end{cases}$$

La f.r. predittiva di  $k$  future osservazioni (una per ogni classe), analogamente al caso del processo di Dirichlet, è data da

$$\begin{aligned} \Pr\{X_{1,n_1+1} \leq x_1, \dots, X_{k,n_k+1} \leq x_k \mid \underline{x}_1, \dots, \underline{x}_k\} = \\ = \int_{\mathbb{R}^k} \prod_{i=1}^k \left\{ \frac{\alpha_i(u_i, \mathbb{R})}{\alpha_i(u_i, \mathbb{R}) + n_i} \cdot \frac{\alpha_i(u_i, x_i)}{\alpha_i(u_i, \mathbb{R})} + \right. \\ \left. + \frac{n_i}{\alpha_i(u_i, \mathbb{R}) + n_i} \hat{F}_{i,n_i}(x_i) \right\} d_{\underline{\mu}} \varphi(\underline{u} | \underline{x}_1, \dots, \underline{x}_k) \quad (6) \end{aligned}$$

## 5. IL MODELLO LINEARE DELLA REGRESSIONE CONDIZIONATA

Consideriamo il modello

$$\underline{y} = X\underline{\beta} + \underline{\varepsilon}$$

in cui  $\underline{y}$  è il vettore ( $n \times 1$ ) delle osservazioni sulla variabile dipendente  $Y$ ,  $X$  è una matrice di tipo ( $n \times p$ ) non aleatoria,  $\underline{\beta}$  il vettore dei parametri ( $p \times 1$ ) ed  $\underline{\varepsilon}$  un vettore aleatorio ( $n \times 1$ ). Supponiamo che i suddetti elementi siano partizionati nel modo seguente:

$$\underline{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_k \end{pmatrix} \text{ con } \underline{y}_i = \begin{pmatrix} y_{i1} \\ \vdots \\ y_{in_i} \end{pmatrix}, \quad i=1,2,\dots,k \quad (n_i \text{ eventualmente nullo})$$

$$\sum_{i=1}^k n_i = n, \quad k > p,$$

$$X = \begin{pmatrix} \dots & X_1 & \dots \\ \vdots & \vdots & \vdots \\ \dots & X_k & \dots \end{pmatrix} \text{ con } X_i = \underline{1}_{n_i} \underline{x}'_i = \begin{pmatrix} x'_{i1} \\ \vdots \\ x'_{in_i} \end{pmatrix}; \quad (n_i \times p), \quad \underline{x}'_i = (x_{i1}, \dots, x_{in_i})$$

$$\underline{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_k \end{pmatrix} \quad \underline{\varepsilon}'_i = (\varepsilon_{i,1}, \dots, \varepsilon_{i,n_i})$$

in modo che il sistema iniziale equivale a

$$\begin{aligned} y_1 &= \underline{1}_{n_1} \underline{x}'_1 \underline{\beta} + \varepsilon_1 \\ &\vdots \\ y_k &= \underline{1}_{n_k} \underline{x}'_k \underline{\beta} + \varepsilon_k. \end{aligned}$$

Supponiamo ora che, dati la matrice  $X$ , il vettore  $\underline{\beta}$  e le f.r.  $F_1, \dots, F_k$ , i vettori  $\underline{y}_1, \dots, \underline{y}_k$  siano mutuamente indipendenti con f.r. congiunta

$$\begin{aligned} \Pr\{\underline{y}_1 \leq \underline{\xi}_1, \dots, \underline{y}_k \leq \underline{\xi}_k | X, \underline{\beta}, F_1, \dots, F_k\} &= \\ &= \prod_{j=1}^{n_1} F_1(\xi_{1,j}) \cdot \prod_{j=1}^{n_2} F_2(\xi_{2,j}) \cdots \prod_{j=1}^{n_k} F_k(\xi_{k,j}). \end{aligned}$$

e pertanto, subordinatamente alla condizione detta, gli elementi di  $\underline{y}_i$  siano pure mutuamente indipendenti e somiglianti. Le f.r.  $F_1, \dots, F_k$  non siano note ma aleatorie e selezionate da un processo mistura di prodotti di processi di Dirichlet. Precisamente, dati  $\underline{x}_i, \underline{\beta}$  supponiamo che  $F_i$  sia retta da un processo di Dirichlet con parametro  $\alpha_i(\underline{x}_i, \underline{\beta}, \cdot)$  e che perciò il vettore  $F_1, F_2, \dots, F_k | X, \underline{\beta} \in \prod_{i=1}^k \mathcal{D}_i$ .

Indicata con  $\varphi(\underline{\beta})$  la f.r. del vettore  $\underline{\beta}$  si ha quindi

$$F_1, \dots, F_k | X \in \int \prod_{i=1}^k \mathcal{D}_i d\varphi(\underline{\beta})$$

Anche per effettuare successivamente qualche confronto supponiamo che

$$\alpha_i(\underline{x}_i, \underline{\beta}, \xi) = \alpha(\mathcal{R}) \phi\left(\frac{\xi - \underline{x}'_i \underline{\beta}}{\sigma_i}\right) \quad i = 1, \dots, k$$

in cui  $\phi$  è la f.r. normale standardizzata.

Prima di procedere ancora nella specificazione del modello osserviamo intanto che per le caratteristiche del processo utilizzato, la f.r. del generico vettore  $\underline{y}_i$ , dati  $\underline{x}_i$  e  $\underline{\beta}$ , risulta

$$\Pr\{y_i \leq \xi_i | x_i, \underline{\beta}\} = \prod_{j=i}^{n_i} \frac{\alpha(\mathcal{R}) \phi\left(\frac{\xi_i(j) - x_i' \underline{\beta}}{\sigma_i}\right)}{\alpha(\mathcal{R}) + j - 1}$$

in cui  $\xi_{i(1)} < \xi_{i(2)} < \dots < \xi_{i(n_i)}$ . Da ciò discende che  $(Y_{i1}, Y_{i2}, \dots, Y_{in_i})$  sono numeri aleatori scambiabili e non indipendenti (dati  $x_i$  e  $\underline{\beta}$ ) con f.r. marginali

$$\Pr\{y_{ij} \leq \xi | x_i, \underline{\beta}\} = \phi\left(\frac{\xi - x_i' \underline{\beta}}{\sigma_i}\right) \quad j = 1, 2, \dots, n_i$$

ed

$$E(y_{ij} | x_i, \underline{\beta}) = x_i' \underline{\beta}$$

$$\text{Var}(y_{ij} | x_i, \underline{\beta}) = \sigma_i^2$$

In più, dati  $X$  e  $\underline{\beta}$ , i vettori  $y_1, \dots, y_k$  sono mutuamente indipendenti (ma non somiglianti) e da ciò discende che, data la matrice  $X$ , i vettori sono parzialmente scambiabili.

Assumiamo ora che il vettore  $\underline{\beta}$  posseda distribuzione iniziale caratterizzata dalla densità (iniziale) normale

$$\varphi^{(p)}(\underline{\beta} | \beta_0, \sigma_\beta^2) \propto \exp\left\{-\frac{1}{2\sigma_\beta^2}(\underline{\beta} - \underline{1}\beta_0)'(\underline{\beta} - \underline{1}\beta_0)\right\}$$

con

$$E(\underline{\beta} | \beta_0, \sigma_\beta^2) = \underline{1}\beta_0, \quad \text{Cov}(\underline{\beta} | \beta_0, \sigma_\beta^2) = \sigma_\beta^2 I_p, \quad \sigma_\beta^2 > 0.$$

Il problema della stima di  $\underline{\beta}$  consiste ovviamente nella ricerca della distribuzione finale di  $\underline{\beta}$ . La (5) del paragrafo precedente, poichè  $\alpha$  è assunta assolutamente continua, si semplifi

ca notevolmente nella

$$\begin{aligned} \varphi^{(p)}(\underline{\beta} | \beta_0, \sigma_\beta^2, y_1, \dots, y_k, x_1, \dots, x_k) &\propto \\ &\propto \prod_{i=1}^k \prod_{j=1}^{r_i} \phi\left(\frac{y_{ij} - x_i' \underline{\beta}}{\sigma_i}\right) \varphi^{(p)}(\underline{\beta} | \beta_0, \sigma_\beta^2) \\ &\propto \exp\left\{-\frac{1}{2}[(\tilde{y} - \tilde{X}\underline{\beta})' \tilde{\Sigma}^{-1}(\tilde{y} - \tilde{X}\underline{\beta}) + (\underline{\beta} - \underline{1}\beta_0)'(\underline{\beta} - \underline{1}\beta_0)\sigma_\beta^{-2}]\right\} \end{aligned}$$

in cui  $r_i$  indica il numero delle osservazioni distinte in  $(Y_{i1}, \dots, Y_{in_i})$ ,  $\tilde{y}$  è il vettore  $y$  dei soli elementi distinti e similmente la matrice  $\tilde{X}$ .  $\tilde{\Sigma}$  è invece data dalla matrice partizionata

$$\tilde{\Sigma} = \begin{pmatrix} \Sigma_1 & & \\ & \ddots & \\ & & \Sigma_k \end{pmatrix} \quad \text{di tipo } r \times r, \quad r = \sum_{i=1}^k r_i$$

e

$$\Sigma_i = \sigma_i^2 I_{r_i}, \quad i = 1, \dots, k$$

Qualche calcolo permette allora di stabilire che la distribuzione finale di  $\underline{\beta}$  è normale con valore atteso

$$\begin{aligned} \underline{b} &= E(\underline{\beta} | \beta_0, \sigma_\beta^2, y_1, \dots, y_k, x_1, \dots, x_k) = \\ &= (\tilde{X}' \tilde{\Sigma}^{-1} \tilde{X} + \sigma_\beta^{-2} I_p)^{-1} (\tilde{X}' \tilde{\Sigma}^{-1} \tilde{X} \tilde{y} + \sigma_\beta^{-2} \underline{1}\beta_0) \end{aligned}$$

essendo

$$\hat{\underline{\beta}} = (\tilde{X}' \tilde{\Sigma}^{-1} \tilde{X})^{-1} \tilde{X}' \tilde{\Sigma}^{-1} \tilde{y}$$

e matrice di varianze e covarianze

$$V = \text{Cov}(\underline{\beta} | \beta_0, \sigma_\beta^2, y_1, \dots, y_k, x_1, \dots, x_k) = (\hat{X}' \Sigma^{-1} \hat{X} + \sigma_\beta^{-2} I_p)^{-1}.$$

Anche in questo caso pertanto la stima bayesiana di  $\underline{\beta}$ ,  $\underline{b}$  (perdita quadratica) appare come media ponderata della stima dei minimi quadrati  $\hat{\underline{\beta}}$  e del valore atteso  $\underline{1}\beta_0$  della distribuzione iniziale. Se poi  $\sigma_\beta^{-2} \rightarrow 0$ , ovvero la distribuzione iniziale di  $\underline{\beta}$  è quella non informativa à la Jeffreys, allora

$$\underline{b} = \hat{\underline{\beta}} \quad e \quad V = (\hat{X}' \Sigma^{-1} \hat{X})^{-1}.$$

Aricchiamo ora il modello ipotizzando che anche  $\beta_0$  sia in certo con densità pure normale data da

$$\varphi_1(\beta_0 | \mu, \sigma_c^2) \propto \exp\left\{-\frac{1}{2\sigma_c^2}(\beta_0 - \mu)^2\right\} \quad \sigma_c^2 > 0.$$

Con l'aggiunta di questo stadio, naturalmente, il vettore  $\underline{\beta}$  è ora scambiabile con densità

$$\varphi_1^{(p)}(\underline{\beta} | \sigma_\beta^2, \sigma_c^2, \mu) \propto \int_{-\infty}^{+\infty} \exp\left\{-\frac{1}{2}\left[\frac{1}{\sigma_\beta^2}(\underline{\beta} - \underline{1}\beta_0)'(\underline{\beta} - \underline{1}\beta_0) + \frac{1}{\sigma_c^2}(\beta_0 - \mu)^2\right]\right\} d\beta_0.$$

Calcoli simili ai precedenti mostrano che la densità finale di  $\underline{\beta}$  è ancora normale e precisata dal valore atteso

$$\begin{aligned} \underline{b}_1 &= [\hat{X}' \Sigma^{-1} \hat{X} + \sigma_\beta^{-2} (I - \frac{\sigma_c^2}{p\sigma_c^2 + \sigma_\beta^2} J)]^{-1} [\hat{X}' \Sigma^{-1} \hat{Y} + \sigma_\beta^{-2} (I - \frac{\sigma_c^2}{p\sigma_c^2 + \sigma_\beta^2} J) \underline{1}\mu] \\ &= [I + (\hat{X}' \Sigma^{-1} \hat{X})^{-1} \sigma_\beta^{-2} (I - \frac{\sigma_c^2}{p\sigma_c^2 + \sigma_\beta^2} J)]^{-1} [\hat{\underline{\beta}} + (\hat{X}' \Sigma^{-1} \hat{X})^{-1} \times \\ &\quad \times \sigma_\beta^{-2} (I - \frac{\sigma_c^2}{p\sigma_c^2 + \sigma_\beta^2} J) \underline{1}\mu] \end{aligned}$$

e matrice di varianze e covarianze

$$V_1 = [(\hat{X}' \Sigma^{-1} \hat{X}) + \sigma_\beta^{-2} (I - \frac{\sigma_c^2}{p\sigma_c^2 + \sigma_\beta^2} J)]^{-1}$$

in cui

$$J = \underline{1}_p \cdot \underline{1}'_p$$

Se poi v'è incertezza anche su  $\mu$  e la sua densità è assunta non informativa (impropria) allora i parametri della distribuzione finale (ancora normale) sono dati da:

$$\begin{aligned} \underline{b}_0 &= [I + (\hat{X}' \Sigma^{-1} \hat{X})^{-1} \sigma_\beta^{-2} (I - \frac{1}{p} J)]^{-1} \hat{\underline{\beta}} \\ V_0 &= [\hat{X}' \Sigma^{-1} \hat{X} + \sigma_\beta^{-2} (I - \frac{1}{p} J)]^{-1}, \end{aligned}$$

avendo utilizzato negli sviluppi un risultato matriciale (C.R. Rao (1973), pag. 33, es. 2.9).

Particolarizzando convenientemente la matrice  $\hat{X}$  è immediato constatare l'uguaglianza di tali espressioni con quelle fornite da D. Michele Cifarelli (1979) relative al modello dell'analisi della varianza.

Riguardo a tali stime conviene fare alcune osservazioni.

In primo luogo, tali stime sono, in generale, differenti da quelle ottenute, nelle medesime condizioni per quanto riguarda le distribuzioni iniziali dei parametri, da D.V. Lindley e A.F.M. Smith (1972) nel caso in cui le f.r.  $F_1, \dots, F_k$  sono certe e normali in quanto nell'espressione, ad esempio, di  $\underline{b}_0$  non compaiono le ampiezze campionarie  $n_1, \dots, n_k$  bensì  $r_1, \dots, r_k$ . In secondo luogo, solo apparentemente tale stima non dipende da  $\alpha(\mathcal{R})$ , il peso inizialmente assegnato alle varie distribuzioni normali. In effetti  $\alpha(\mathcal{R})$  influenza il numero delle osservazioni distinte in ogni classe nel senso che se  $\alpha(\mathcal{R}) \rightarrow +\infty$ , allora, come si può dedurre da un risultato di R.M. Korwar e M. Hollander (1973),  $r_i \rightarrow n_i$  quasi certamente, ed in questo caso le nostre stime coincidono con quelle degli Autori citati. Ma ciò è per altri versi abbastanza naturale poichè in questa circostanza il processo da noi utilizzato per probabilizzare  $(F_1, \dots, F_k)$  degenera proprio sulle distribuzioni normali ed il modello non parametrico viene a confondersi con quello parametrico utilizzato dai due Autori precedenti.

In tutta la discussione precedente abbiamo assunto che le varianze  $\sigma_i^2$  e  $\sigma_\beta^2$  fossero note.

Supponiamo ora che  $\sigma_i^2 = \sigma^2 \quad \forall i = 1, 2, \dots, k$  non sia nota come pure  $\sigma_\beta^2$ .

Se si assegna come distribuzione iniziale di  $(\underline{\beta}, \sigma)$  (trascurando per semplicità i parametri condizionati)

$$p_1(\underline{\beta}, \sigma) \propto \sigma^{-p} \exp \left\{ -\frac{1}{2\sigma^2} (\underline{\beta} - \underline{1}\beta_0)' (\underline{\beta} - \underline{1}\beta_0) \right\} \cdot \sigma^{-h-1} \exp \left\{ -\frac{h}{2\sigma^2} \right\}$$

e con densità di  $(\beta_0, \sigma_\beta)$

$$p_2(\beta_0, \sigma_\beta) \propto \sigma_\beta^{-k-1} \exp \left\{ -\frac{k}{2\sigma_\beta^2} \right\}$$

allora si può mostrare che la densità finale di  $\underline{\beta}$  è data da

$$(h + rs^2 + (\underline{\beta} - \hat{\underline{\beta}})' \hat{\underline{X}}' \hat{\underline{X}} (\underline{\beta} - \hat{\underline{\beta}}))^{-(r+h)/2} \cdot (k + (\underline{\beta} - \underline{1}\bar{\beta})' (\underline{\beta} - \underline{1}\bar{\beta}))^{-(p+k-1)/2}$$

in cui

$$\hat{\underline{\beta}} = (\hat{\underline{X}}' \hat{\underline{X}})^{-1} \hat{\underline{X}}' \underline{y}, \quad \bar{\beta} = \frac{1}{p} \sum_{i=1}^p \beta_i, \quad s^2 = (\underline{y} - \hat{\underline{X}}\hat{\underline{\beta}})' (\underline{y} - \hat{\underline{X}}\hat{\underline{\beta}})^{-1}.$$

Un'analisi delle distribuzioni di tale natura (dette double t) può rintracciarsi nel lavoro di G.C. Tiao e A. Zellner (1964) e più in generale in quello di J.H. Dréze (1976).

## 6. LA DISTRIBUZIONE PREDITTIVA

In molte occasioni, come ad esempio in problemi di controllo, date le osservazioni  $y_1, \dots, y_k$  si desidera inferire su future osservazioni supposte generate dallo stesso modello. Come è noto il mezzo idoneo a questo scopo è la cosiddetta distribuzione predittiva, la quale, tra l'altro, svolge un ruolo unificante nella soluzione di molti problemi di analisi multivariata impostati dal punto di vista bayesiano e, di per sé, costituisce uno strumento che si è rivelato utile per nuove analisi di concetti tradizionali della Statistica.

Senza entrare in eccessivi dettagli, in questo paragrafo vogliamo determinare tale distribuzione predittiva, in presenza delle stesse ipotesi utilizzate nel numero precedente.

Per semplicità indichiamo con  $Y_1, \dots, Y_k$  i risultati di  $k$  future osservazioni da eseguirsi su ognuna delle  $k$  classi (una osservazione per classe). Essa, per definizione verrà caratterizzata dalla f.r.

$$\psi(y_1, \dots, y_k) = \Pr\{Y_1 \leq y_1, \dots, Y_k \leq y_k | \underline{y}_1, \dots, \underline{y}_k, \underline{x}_1, \dots, \underline{x}_k\}.$$

La (6) del paragrafo 4, nel nostro caso diviene

$$\psi(y_1 \dots y_k) = \int_{\mathbb{R}^p} \prod_{i=1}^k \left\{ \frac{\alpha(\mathcal{R})}{\alpha(\mathcal{R}) + n_i} \phi \left( \frac{y_i - \underline{x}_i' \underline{\beta}}{\sigma_i} \right) + \frac{n_i}{\alpha(\mathcal{R}) + n_i} \hat{F}_{i, n_i}(y_i) \right\} \varphi^{(p)}(\underline{\beta} | \cdot) d\underline{\beta}$$

in cui  $\hat{F}_{i, n_i}$  è la funzione di ripartizione empirica determinata con le osservazioni della  $i$ -ma classe,  $y_{i1}, \dots, y_{in_i}$  e  $\varphi^{(p)}$  è una delle densità finali di  $\underline{\beta}$  di cui abbiamo detto in precedenza.

Per semplicità di riferimento supponiamo che

$$\varphi^{(p)}(\underline{\beta} | \cdot) = (2\pi)^{-p/2} |V|^{-1/2} \exp \left\{ -\frac{1}{2} (\underline{\beta} - \underline{b})' V^{-1} (\underline{\beta} - \underline{b}) \right\}$$

con

$$\underline{b} = (\hat{X}' \Sigma^{-1} \hat{X} + \sigma_{\beta}^{-2} I)^{-1} (\hat{X}' \Sigma^{-1} \hat{X} \underline{\beta}_0 + \sigma_{\beta}^{-2} \underline{\beta}_0)$$

$$V = (\hat{X}' \Sigma^{-1} \hat{X} + \sigma_{\beta}^{-2} I)^{-1}$$

Consideriamo ora la f.r. predittiva della generica margine  $Y_i$  la quale è naturalmente data da

$$\psi_i(y) = \frac{\alpha(\mathcal{R})}{\alpha(\mathcal{R}) + n_i} \int_{\mathbb{R}^p} \phi \left( \frac{y - \underline{x}_i' \underline{\beta}}{\sigma_i} \right) \varphi^{(p)}(\underline{\beta} | \cdot) d\underline{\beta} + \frac{n_i}{\alpha(\mathcal{R}) + n_i} \hat{F}_{i, n_i}(y)$$

e quindi, integrando,

$$\psi_i(y) = \frac{\alpha(\mathcal{R})}{\alpha(\mathcal{R}) + n_i} \phi \left( \frac{y - \underline{x}_i' \underline{b}}{\sigma_i^*} \right) + \frac{n_i}{\alpha(\mathcal{R}) + n_i} \hat{F}_{i, n_i}(y)$$

in cui

$$\begin{aligned} \sigma_i^{*2} &= \sigma_i^2 (1 - \underline{x}_i' (\underline{x}_i \underline{x}_i' + \sigma_i^2 V^{-1})^{-1} \underline{x}_i)^{-1} \\ &= \sigma_i^2 + \underline{x}_i' V \underline{x}_i \end{aligned}$$

con  $\underline{b}$  e  $V$  specificati in precedenza.

Come si rileva tale f.r. costituisce una mistura di due f.r.:

- quella normale con valore atteso  $\underline{x}_i' \underline{b}$  e varianza  $\sigma_i^{*2}$
- quella empirica delle sole osservazioni che si riferiscono al  $i$ -mo gruppo.

Inoltre

- i) essa dipende da tutte le osservazioni  $\underline{y}_1, \dots, \underline{y}_k$  e non solo da quelle della classe cui la f.r. si riferisce;
- ii) la f.r. normale è influenzata da tutti i dati a disposizione tramite i suoi parametri  $\underline{b}$  e  $\sigma_i^*$ , mentre le osservazioni della  $i$ -ma classe determinano anche (da sole) la f.r. empirica;
- iii) la f.r.  $\psi_i$  concentra masse di probabilità su tutte (e sole) le osservazioni (distinte) della  $i$ -ma classe e ciò permette di apprezzare in maggior misura l'influenza esercitata dalle osservazioni delle varie classi;

$$\text{iv) } \psi_i(y) \rightarrow \phi\left(\frac{y - x_i' b}{\sigma_i^*}\right) \quad \text{se } \alpha(\mathcal{R}) \rightarrow +\infty$$

$$\rightarrow \hat{F}_{i, n_i}(y) \quad \text{se } \alpha(\mathcal{R}) \rightarrow 0_+$$

le quali sono di agevole interpretazione appena si pensi al significato attribuito al parametro  $\alpha(\mathcal{R})$ .

E' agevole constatare che la distribuzione predittiva bidimensionale è data da:

$$\psi_{i,j}(y_i, y_j) = \int_{\mathcal{R}^p} \left[ \frac{(\alpha(\mathcal{R}))^2}{(\alpha(\mathcal{R}) + n_i)(\alpha(\mathcal{R}) + n_j)} \phi\left(\frac{y_i - x_i' \beta}{\sigma_i^*}\right) \phi\left(\frac{y_j - x_j' \beta}{\sigma_j^*}\right) + \right. \\ \left. + \frac{\alpha(\mathcal{R}) n_j}{(\alpha(\mathcal{R}) + n_i)(\alpha(\mathcal{R}) + n_j)} \phi\left(\frac{y_i - x_i' \beta}{\sigma_i^*}\right) \hat{F}_{j, n_j}(y_j) + \right. \\ \left. + \frac{\alpha(\mathcal{R}) n_i}{(\alpha(\mathcal{R}) + n_i)(\alpha(\mathcal{R}) + n_j)} \phi\left(\frac{y_j - x_j' \beta}{\sigma_j^*}\right) \hat{F}_{i, n_i}(y_i) + \right. \\ \left. + \frac{n_i n_j}{(\alpha(\mathcal{R}) + n_i)(\alpha(\mathcal{R}) + n_j)} \hat{F}_{i, n_i}(y_i) \cdot \hat{F}_{j, n_j}(y_j) \right] \varphi^{(p)}(\underline{\beta} | \cdot) d\underline{\beta}$$

e quindi, integrando

$$\psi_{i,j}(y_i, y_j) = \frac{(\alpha(\mathcal{R}))^2}{(\alpha(\mathcal{R}) + n_i)(\alpha(\mathcal{R}) + n_j)} N_{ij}(X_{ij}' b, \Sigma_{ij}^*) + \\ + \frac{\alpha(\mathcal{R}) n_j}{(\alpha(\mathcal{R}) + n_i)(\alpha(\mathcal{R}) + n_j)} \phi\left(\frac{y_i - x_i' b}{\sigma_i^*}\right) \hat{F}_{j, n_j}(y_j) + \\ + \frac{\alpha(\mathcal{R}) n_i}{(\alpha(\mathcal{R}) + n_i)(\alpha(\mathcal{R}) + n_j)} \phi\left(\frac{y_j - x_j' b}{\sigma_j^*}\right) \hat{F}_{i, n_i}(y_i) + \\ + \frac{n_i n_j}{(\alpha(\mathcal{R}) + n_i)(\alpha(\mathcal{R}) + n_j)} \hat{F}_{i, n_i}(y_i) \cdot \hat{F}_{j, n_j}(y_j)$$

in cui

$N_{ij}$  indica la funzione di ripartizione della normale doppia in  $y_i, y_j$  con valore atteso  $X_{ij}' b$  e matrice di varianza e covarianza  $\Sigma_{ij}^*$ .

$X_{ij}'$  è una matrice  $(2 \times p)$  ottenuta sovrapponendo i vettori  $\underline{x}_i'$ ,  $\underline{x}_j'$  e  $\Sigma_{ij}^* = \Sigma_{ij} + X_{ij} V X_{ij}'$  con  $\Sigma_{ij} = \text{diag}(\sigma_i^2, \sigma_j^2)$  e  $V$  specificato in precedenza.

Si constata immediatamente un comportamento di  $\psi_{ij}$  analogo a quello di  $\psi_i$ .

## BIBLIOGRAFIA

- [1] L.D. BROWN (1966), *On the admissibility of invariant estimators of one or more location parameters*. Ann. Math. Stat., 37, 1087-1136.
- [2] D.M. CIFARELLI, E. REGAZZINI (1978), *Problemi statistici non parametrici in condizioni di scambiabilità parziale. Impiego di medie associative*. Ist. di Mat. Fin., Univ. di Torino.
- [3] D.M. CIFARELLI (1979), *Impostazione bayesiana di un problema di analisi della varianza con approccio non parametrico*. Istituto di Matematica Finanziaria, n. 17, Univ. di Torino.
- [4] J.H. DREZE (1976), *Bayesian Regression Analysis using Poly-T Densities*. In *New Developments in the Applications of Bayesian Methods*. (A. Aykaç e C. Brumat eds) North-Holland Publishing Co., Amsterdam, pp. 153-184.
- [5] T.S. FERGUSON (1973), *A Bayesian analysis of some non parametric problems*. Ann. Stat., 1, 209-230.
- [6] M. GOLDSTEIN (1976), *Bayesian analysis of regression problems*. Biometrika, 63, 1, 51-58.
- [7] H. JEFFREYS (1961), *Theory of Probability*. (3 ed.), Oxford, Clarendon Press.
- [8] R.M. KORWAR, M. HOLLANDER (1973), *Contributions to the theory of Dirichlet processes*. Ann. Prob., 1, 705-711.
- [9] D.V. LINDLEY (1962), *Discussion on Prof. Stein paper*. J.R.S.S., B, 24, 285-287.
- [10] D.V. LINDLEY (1969), *Bayesian least squares*. Bull. Inst. Int. Stat., 43, 152-153.
- [11] D.V. LINDLEY (1971), *The estimation of many parameters*. In *Foundation of Statistical Inference* (V.P. Godambe e D.A. Sprott eds.), pp. 435-455, Holt, Rinehart e Winston, Toronto.
- [12] D.V. LINDLEY, A.F.M. SMITH (1972), *Bayes estimates for the linear model*. J.R.S.S., B; 34, 1-41.
- [13] D.V. LINDLEY (1974), *A bayesian solution for two-way analysis of variance*. In *Progress in Statistics*, pp. 475-496, J. Gani, K. Sarkadi, I. Vincze (eds.), North Holland.
- [14] C.R. RAO (1973), *Linear statistical inference and its applications*. 2<sup>nd</sup> ed. Wiley, New York.
- [15] L.J. SAVAGE (1962), *Bayesian Statistics*. In *Decision and Information Processes*. Mac Millan, New York.
- [16] A.F.M. SMITH (1973a), *A general Bayesian linear model*. J.R.S.S., B, 35, 67-75.
- [17] A.F.M. SMITH (1973b), *Bayes estimates in one-way and two-way models*. Biometrika, 60, 2, 319-329.
- [18] C. STEIN (1956), *Inadmissibility of the usual estimator for the mean of a multivariate normal distribution*. Proc. 3<sup>rd</sup> Berkeley Simp. on Math. Stat. and Prob., I, 197-206, Berkeley, Univ. of Cal. Press.

- [19] C. STEIN (1962), *Confidence sets for the mean of a multivariate normal distribution*. Journal of the Royal Statistical Society, Serie B, 24, pp. 265-296.
- [20] G.C. TIAO, A. ZELLNER (1964), *Bayes' theorem and the use of prior knowledge in regression analysis*. Biometrika, 51, pp. 219-230.
- [21] A. ZELLNER (1971), *An Introduction to Bayesian Inference in Econometrics*. Wiley, New York.
- [22] A. ZELLNER, W. VANDAELE (1974), *Bayes-Stein estimators for  $k$  means regression and simultaneous equation model*. Studies in bayesian Econometrics and Statistics, pp. 627-653, S.E. Fienberg and A. Zellner (eds.), North-Holland.