

STUDI STATISTICI n. 1
IL MODELLO LINEARE:
INFERENZA E PREVISIONE

Pietro MULIERE e Marco SCARSINI

Gennaio 1982

UNIVERSITA' BOCCONI – Via Sarfatti, 25, 20136 MILANO

IL MODELLO LINEARE: INFERENZA E PREVISIONE*

Pietro MULIERE¹

Marco SCARSINI²

RIASSUNTO

In questo lavoro vengono messe a confronto diverse impostazioni del modello lineare. Viene in particolare esaminata l'impostazione bayesiana nei suoi aspetti ipotetici e previsivi. Sono considerati le distribuzioni finali, gli stimatori bayesiani e le distribuzioni predittive per modelli parametrici e non parametrici. Infine viene preso in esame un modello di analisi della varianza a due criteri con interazioni, e vengono effettuati confronti con altri modelli più semplici già noti.

* Lavoro eseguito nell'ambito del GNAFA-CNR

1 Istituto di Scienze Economiche e Statistiche, Università di Pavia

2 Istituto di Metodi Quantitativi, Università Bocconi (Milano)

I N D I C E

1. INTRODUZIONE E SOMMARIO	Pag. 1
2. ALCUNI CONCETTI DI INFERENZA BAYESIANA	" 3
3. L'IMPOSTAZIONE CLASSICA DEL MODELLO LINEARE ...	" 10
3.1. Definizione di modello lineare	" 10
3.2. Alcuni risultati	" 11
4. STIMATORI AMMISSIBILI	" 15
5. L'IMPOSTAZIONE BAYESIANA TRADIZIONALE DEL MODELLO LINEARE	" 20
5.1. Distribuzioni iniziali non informative ...	" 23
5.1.1. σ^2 nota	" 23
5.1.2. σ^2 non nota	" 23
5.2. Distribuzioni iniziali informative coniugate	" 24
5.2.1. σ^2 nota	" 24
5.2.2. σ^2 non nota	" 26
6. I MODELLI GERARCHICI	" 28
6.1. Modello a due stadi	" 29
6.2. Modello a tre stadi	" 29
6.3. Scambiabilità tra parametri	" 32
6.4. Scambiabilità tra equazioni	" 33
6.5. Osservazioni e confronti	" 35
6.6. Modello gerarchico con matrici di varianze e covarianze non note	" 36
7. L'IMPOSTAZIONE BAYESIANA NON PARAMETRICA DEL MODELLO LINEARE	" 40

7.1. Distribuzioni iniziali per modelli non parametrici	Pag. 41
7.2. Applicazione al modello lineare	" 43
8. PROBLEMI DI PREVISIONE NEL MODELLO LINEARE ...	" 47
8.1. Modello parametrico: varianza nota	" 49
8.2. Modello parametrico: varianza non nota ..	" 51
8.3. Modello non parametrico	" 52
9. ANALISI DELLA VARIANZA	" 56
BIBLIOGRAFIA	" 65

1. INTRODUZIONE E SOMMARIO

Uno dei principali problemi in discussione nella metodologia statistica è la possibilità di applicare il meccanismo bayesiano alle ricerche empiriche. Ritenere che i metodi bayesiani giocheranno un ruolo importante nel futuro dell'econometria ci ha spinto a studiare il loro impiego nell'analisi del modello lineare generale. La trattazione bayesiana di tale modello non è ancora entrata nell'uso corrente almeno nei testi di base di statistica ed è in questo senso che abbiamo ritenuto utile preparare una monografia che permettesse di accedere in modo piano a questa problematica.

Con il presente lavoro ci prefiggiamo di fornire un inquadramento generale all'analisi del modello lineare secondo l'impostazione bayesiana, mettendo a confronto diversi metodi ed evidenziando analogie e discordanze.

Nel paragrafo 2 illustreremo alcuni aspetti caratterizzanti l'inferenza (bayesiana). Per poter effettuare alcuni confronti presenteremo brevemente nel paragrafo 3 alcuni risultati dell'impostazione classica e nel paragrafo 4 qualche proposta tendente a superare alcune manchevolezze degli stimatori ottenuti con il metodo dei minimi quadrati.

Con il paragrafo 5 inizieremo l'esame dell'impostazione bayesiana del modello lineare. In particolare dopo esserci soffermati su alcuni risultati ormai tradizionali, nel paragrafo 6 presenteremo un modello gerarchico che per la sua flessibilità giudichiamo partico-

larmente interessante. In tale modello mostreremo una possibile applicazione dell'idea di scambiabilità ai parametri oltre che alle osservazioni.

Nel paragrafo 7 verrà considerata una generalizzazione dei risultati esaminati nei paragrafi precedenti in ambito non parametrico.

Il paragrafo 8 sarà dedicato ai problemi di carattere previsivo. Nel paragrafo 9 particolarizzeremo il modello lineare generale per l'analisi della varianza a uno e due criteri.

2. ALCUNI CONCETTI DI INFERENZA BAYESIANA

L'argomento di cui vogliamo occuparci in questo lavoro è in quadrabile, evidentemente, nel tema più generale del ragionamento per induzione. Il procedimento induttivo si propone di stabilire proposizioni aventi validità generale a partire dalla conoscenza di fatti particolari. Le proposizioni che si vengono a stabilire non potranno avere, in generale, valore di verità pieno. Non si potrà, quindi, dire che siano vere o false; saranno probabili. Affinchè la frase abbia significato compiuto occorrerà definire il termine probabilità. Nella concezione soggettiva, a cui faremo riferimento, la probabilità è la sintesi quantitativa di opinioni che un Soggetto coerentemente esprime al riguardo di fatti osservabili (de Finetti, 1937).

Consideriamo un vettore Y_1, \dots, Y_n di numeri aleatori (n.a.) ed interrogiamoci sull'atteggiamento da tenere ai fini del procedimento induttivo. Ovviamente per caratterizzare dal punto di vista probabilistico il vettore Y_1, Y_2, \dots, Y_n è necessario assegnare la sua funzione di ripartizione (f.r.). La f.r. congiunta dal vettore esprime le convinzioni del soggetto circa i n.a. ed in particolare segnala il tipo di dipendenza stocastica che intercorre tra questi. La f.r. pertanto dovrà essere conforme al modo in cui il Soggetto ritiene opportuno tener conto delle osservazioni ai fini del ragionamento induttivo.

Evidentemente i possibili legami di dipendenza ipotizzabili tra i n.a. sono infiniti ma tra questi alcuni, particolarmente sem-

plici, sono più frequentemente assunti, per esempio la scambiabilità
tà.

I n.a. (Y_1, \dots, Y_n) sono scambiabili se la f.r. di ogni
 $(Y_{i_1}, \dots, Y_{i_n})$ in cui i_1, \dots, i_n è una qualunque permutazione di
 $(1, \dots, n)$, coincide con quella di (Y_1, \dots, Y_n) , vale a dire:

$$\Pr(Y_{i_1} \leq y_1, \dots, Y_{i_n} \leq y_n) = \Pr(Y_1 \leq y_1, \dots, Y_n \leq y_n).$$

D'ora innanzi penseremo l'ennupla (Y_1, \dots, Y_n) come estratta
dalla successione $\{Y_n\}$ $n = 1, 2, \dots$ composta da una infinità numera
bile di n.a. che si riferiscono alle possibili osservazioni di un
determinato fenomeno.

Diremo che la successione $\{Y_n\}$ è scambiabile se ogni successione fi
nita da essa estratta soddisfa alla proprietà di scambiabilità.

Il significato di questa proprietà è che il Soggetto giudica
irrilevante ai fini della valutazione di probabilità l'ordine delle
realizzazioni sperimentali.

Da quanto detto emerge chiaramente che la scambiabilità non è
una proprietà intrinseca degli elementi $\{Y_n\}$ ma piuttosto uno schema
teorico che il Soggetto impone, in relazione ai propri convincimenti
ed alle proprie informazioni.

Vediamo ora di chiarire come l'ipotesi di scambiabilità si ri
levi di grande utilità pratica nell'affrontare il ragionamento indut
tivo. Per far ciò ci verrà in aiuto un fondamentale teorema di de Fi
netti.

Il teorema afferma che se $\{Y_n\}$ è una successione infinita di
n.a. ed F è la f.r. limite della f.r. empirica $\hat{F}_n(y)$ si ha:

$$\Pr(Y_1 \leq y_1, \dots, Y_k \leq y_k | F) = \prod_{i=1}^k F(y_i) \text{ q.c.} \quad (2.1)$$

con Y_1, Y_2, \dots, Y_k estratti da $\{Y_n\}$.

In altre parole, subordinatamente ad F , gli elementi di $\{Y_n\}$ sono indipendenti ed identicamente distribuiti secondo F .

Nella realtà F è aleatoria: $F \in \mathcal{F}$ dove \mathcal{F} indica la classe delle f.r. limite. Pertanto dalle (2.1) si ottiene:

$$\Pr\{Y_1 \leq y_1, \dots, Y_k \leq y_k\} = \int_{\mathcal{F}} \prod_{i=1}^k F(Y_i) d\mu(F) \quad (2.2)$$

μ è una misura di probabilità depositata sullo spazio \mathcal{F} e rappresenta il peso assegnato alle possibili conformazioni della legge F .

L'utilità del teorema risiede nel fatto che in genere è più agevole esprimere le $F(Y_i)$ e la $\mu(F)$ piuttosto che direttamente la $\Pr\{Y_1 \leq y_1, \dots, Y_k \leq y_k\}$. Infatti con la (2.2) abbiamo sostituito alla legge di probabilità diretta della successione $\{Y_n\}$ due elementi:

- la legge di probabilità di $\{Y_n\}$ data un'ipotesi F .
- la probabilità dell'ipotesi F .

Grazie al fatto che, subordinatamente a ciascuna ipotesi gli elementi di $\{Y_n\}$ sono identicamente distribuiti ed indipendenti, sarà sufficiente assegnare la legge di probabilità subordinata del generico Y_n .

Sempre restando nell'ambito della scambiabilità, abbiamo visto che l'ingrediente fondamentale del ragionamento per induzione è la misura di probabilità $\mu(F)$ che, d'ora in poi designeremo con il termine di "iniziale".

Il teorema di Bayes consente di passare dalla distribuzione iniziale $\mu(F)$ a quella "finale" $\mu(F | (Y_1, \dots, Y_k) \in A)$ essendo A un evento che coinvolge i n.a. Y_1, \dots, Y_k :

$$d\mu(F | (Y_1, \dots, Y_k) \in A) = \frac{\Pr\{(Y_1, \dots, Y_k) \in A\} d\mu(F)}{\int_F \Pr\{(Y_1, \dots, Y_k) \in A\} d\mu(F)} \quad (2.3)$$

La (2.3) fornisce, come si può vedere, una risposta adeguata al problema dell'induzione. Infatti costituisce una valutazione (probabilistica) circa un'ipotesi generale F quando si conosca l'esito di un particolare esperimento $(Y_1, \dots, Y_k) \in A$.

Potrebbe apparire superfluo il ricorso alla scambiabilità visto che il teorema di Bayes ne prescinde; quello che ci ha spinto a farla intervenire è la concezione soggettiva della probabilità (valutazione soggettiva coerente circa il verificarsi di *fatti osservabili*). Sono, infatti, fatti osservabili Y_1, \dots, Y_k, \dots e la loro funzione \hat{F}_n .

Da quanto detto emerge che la (2.3) rappresenta effettivamente una probabilità, nonostante il fatto che F sia una legge generale. Ciò è possibile solo in quanto F è stata ottenuta come limite di f.r.

In alcune situazioni la classe di f.r. \hat{F} (se si vuole la classe di ipotesi) a priori ritenute possibili è costituita da tutte le distribuzioni aventi una determinata forma funzionale e caratterizzate da un parametro (scalare o vettoriale) $\theta \in \Theta$. In tal caso il modello statistico è detto *parametrico*. Quando non si verifica che \hat{F} contenga solo f.r. della stessa forma analitica il modello è detto *non parametrico*.

θ , in generale, non è noto per cui, bisognerà assegnargli una probabilità (iniziale).

Se le funzioni F e \bar{F} ammettono densità $f(\cdot|\theta)$, e se la probabilità iniziale è esprimibile in termini di una densità $h(\theta)$, allora il teorema di Bayes, risulta:

$$h(\theta|x) = \frac{h(\theta) p(x|\theta)}{\int_{\Theta} h(\theta) p(x|\theta) d\theta}$$

dove, in ipotesi di scambiabilità,

$$p(x|\theta) = \prod_{i=1}^n f(x_i|\theta),$$

con $x = (x_1, \dots, x_n)$.

Nel caso di modelli parametrici il teorema di de Finetti può essere scritto:

$$P\{y_1 \leq y_1, \dots, y_k \leq y_n\} = \int_{\mathbb{R}^n} \prod_{i=1}^k F(y_i|\theta) dH(\theta)$$

Abbiamo detto in precedenza che l'ipotesi di scambiabilità conduce a ritenere che le osservazioni agiscano simmetricamente sulle opinioni iniziali implicando l'irrilevanza dell'ordine secondo cui si presentano i risultati dell'esperienza. Si prospettano, tuttavia, piuttosto frequentemente situazioni in cui tale ruolo simmetrico delle osservazioni non risulta accettabile. Ciò condurrà a prendere in esame schemi diversi da quello della scambiabilità totale. Per esempio, nel caso in cui si abbia a che fare con dati raggruppati in classi omogenee, si può ritenere accettabile l'ipotesi di scambiabilità all'interno di ciascuna classe ma non per il complesso delle osservazioni. Si pensi alla distribuzione del reddito di una collettività ripartita per classi di età dei redditi oppure ad una successione di esperimenti fisi-

ci compiuti in diverse condizioni di temperatura, ecc.

Uno schema di questo tipo è detto della "scambiabilità parziale" (de Finetti 1938).

Si avranno in tal modo k successioni

$$\{y_{i,n}\} \quad n = 1, 2, \dots; \quad i = 1, 2, \dots, k.$$

di n.a. ciascuna delle quali è scambiabile.

Bisognerà ipotizzare a questo punto i legami di dipendenza tra le diverse successioni. La scelta di questi legami è assai ampia; può andare dal caso di indipendenza tra le k successioni a quello della scambiabilità totale delle osservazioni anche se provenienti da classi diverse.

Per la scambiabilità parziale valgono risultati analoghi a quelli visti per la scambiabilità totale.

In particolare vale una generalizzazione della (2.2).

Se abbiamo k successioni infinite parzialmente scambiabili

$$\{x_{i,n}\} \quad n = 1, 2, \quad i = 1, 2, \dots, k \text{ allora}$$

$$\Pr\{y_{11} \leq y_{11}', \dots, y_{1n_1} \leq y_{1n_1}', \dots, y_{k1} \leq y_{k1}', \dots, y_{kn_k} \leq y_{kn_k}'\}$$

$$= \int_{F^{(k)}} \prod_{i=1}^k \prod_{j=1}^{n_i} F_i(y_{ij}) \, d\mu(F_1, \dots, F_k)$$

Non ci soffermiamo sull'interpretazione di questo risultato che è analogo al caso di scambiabilità totale con le opportune varianti.

Riferimenti bibliografici

Generalità sul ragionamento induttivo: de Finetti (1959), Savage (1959).

L'impostazione soggettiva della probabilità: de Finetti (1937), (1970).

La scambiabilità: de Finetti (1931), (1937), (1938); Hewitt-Savage (1955), Kingman (1978), Lindley-Novick (1981).

L'inferenza statistica bayesiana: Lindley (1965), (1971); De Groot (1970); Box-Tiao (1973).

3. L'IMPOSTAZIONE CLASSICA DEL MODELLO LINEARE

3.1. Definizione di modello lineare

Un carattere (quantitativo) Y , detto anche variabile dipendente o endogena è considerato come parzialmente spiegato o determinato da un certo numero di altri caratteri, Z_1, Z_2, \dots, Z_k , detti variabili indipendenti o esplicative o esogene, tramite una relazione stocastica della forma

$$Y = \sum_{i=1}^k \beta_i h_i(z_i) + \varepsilon \quad (3.1)$$

nella quale $\beta_1, \beta_2, \dots, \beta_k$ sono k parametri non noti ed ε è una variabile aleatoria che dovrebbe tener conto dell'eventuale esistenza di altre variabili, che per vari motivi non possono essere identificate e di eventuali effetti accidentali.

In termini della t -esima osservazione, il modello (3.1) può allora scriversi

$$Y_t = \sum_{i=1}^k \beta_i x_{ti} + \varepsilon_t \quad t = 1, 2, \dots, n \quad (3.2)$$

dove $x_{ti} = h_i(Z_{ti})$.

Se si definisce la matrice

$$X \begin{matrix} (nxk) \end{matrix} = \begin{pmatrix} x_{11}, \dots, x_{1k} \\ x_{21}, \dots, x_{2k} \\ \vdots \\ x_{n1}, \dots, x_{nk} \end{pmatrix}$$

di tutte le osservazioni sulle k variabili indipendenti ed i vettori

$$\begin{array}{l} \underline{\beta} = (\beta_1, \dots, \beta_k)'; \quad \underline{Y} = (Y_1, \dots, Y_n)'; \quad \underline{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)', \\ (k \times 1) \qquad \qquad \qquad (n \times 1) \qquad \qquad \qquad (n \times 1) \end{array}$$

allora possiamo scrivere il sistema (3.2) in forma matriciale,

$$\underline{Y} = X\underline{\beta} + \underline{\varepsilon} \tag{3.3}$$

Naturalmente, se nel modello (3.2) compare un addendo costante, allora la matrice X avrà la prima colonna costituita da elementi eguali ad uno.

Il modello (3.2), (3.3) è detto *lineare* poichè è lineare la relazione che lega Y e i parametri $\beta_1, \beta_2, \dots, \beta_k$.

3.2. Alcuni risultati

In questo numero presentiamo alcuni risultati dell'impostazione classica che ci permetteranno di effettuare, nel seguito, alcuni confronti:

Consideriamo il modello di Gauss-Markov.

$$\underline{Y} = X \underline{\beta} + \underline{\varepsilon} \tag{3.4}$$

in cui \underline{Y} è un vettore aleatorio (n x 1) di variabili osservabili; X è una matrice nota e non stocastica (n x k); $\underline{\beta}$ è un vettore (k x 1) di parametri non noti; $\underline{\varepsilon}$ è un vettore (n x 1) di errori aleatori (non osservabili).

Il problema principale posto dalla statistica classica connesso al modello (3.4) è quello della "stima" di $\underline{\beta}$. Supponiamo che siano soddisfatte alcune ipotesi "standard":

- (a) $E(\underline{\varepsilon}) = \underline{0}$
- (b) $E(\underline{\varepsilon} \underline{\varepsilon}') = \sigma^2 I_n, \sigma^2 > 0$
- (c) $\text{rango}(X) = k$

Le condizioni (a) e (b) sono di immediata percezione, la (c) è semplicemente la condizione di "indipendenza lineare", tra le colonne della matrice X. Essa è cioè la condizione di assenza di multicollinearità; vale a dire i valori osservati $(x_{11}, \dots, x_{1k}), \dots, (x_{n1}, \dots, x_{nk})$ in corrispondenza dei quali si osservano le Y_1, \dots, Y_n non devono disporsi su un iperpiano di dimensione inferiore a k.

Un procedimento assai noto per la stima di $\underline{\beta}$ e che assicura certe proprietà ottimali agli stimatori è quello dei minimi quadrati. Esso consiste nel determinare il vettore $\hat{\underline{\beta}}$ che minimizza la funzione

$$Q(\underline{\beta}) = (\underline{Y} - X \underline{\beta})' (\underline{Y} - X \underline{\beta})$$

E' agevole mostrare che:

$$\hat{\underline{\beta}} = (X'X)^{-1} X' \underline{Y} \quad (3.5)$$

La (3.5) mostra che lo stimatore dei minimi quadrati di $\underline{\beta}$ è una trasformazione lineare del vettore delle osservazioni \underline{Y} .

La matrice di varianze e covarianze di $\hat{\underline{\beta}}$ è data da:

$$\text{Cov}(\hat{\underline{\beta}}) = \sigma^2 (X'X)^{-1}$$

da cui

$$\text{Var}(\hat{\beta}_i) = \sigma^2 c_{ii}$$

essendo c_{ii} l'i-esimo elemento della diagonale principale di $(X'X)^{-1}$.

Per lo stimatore $\hat{\underline{\beta}}$, sotto le condizioni (a), (b), (c) e

$|x_{ti}| \leq L$ (*) valgono le seguenti proprietà:

(*) La condizione che x_{ti} sia uniformemente limitata non è strettamente necessaria e può sostituirsi con $\frac{X'X}{n} \rightarrow A$

- 1 - $\hat{\underline{\beta}}$ è nondistorto, cioè $E(\hat{\underline{\beta}}) = \underline{\beta}$
- 2 - $\hat{\underline{\beta}}$ è consistente, cioè $P \lim \hat{\underline{\beta}} = \underline{\beta}$
- 3 - $\hat{\underline{\beta}}$ è il miglior stimatore non distorto (B.L.U.E.) per $\underline{\beta}$;
cioè, se $\tilde{\underline{\beta}}$ è un qualunque altro stimatore non distorto
per $\underline{\beta}$, allora

$$\text{Cov}(\tilde{\underline{\beta}}) - \text{Cov}(\hat{\underline{\beta}}) = V \text{ (matrice semidefinita positiva)}$$

quindi

$$\text{Var}(\tilde{\beta}_i) \geq \text{Var}(\hat{\beta}_i) \quad i = 1, 2, \dots, k.$$

Poichè $\text{Cov}(\hat{\underline{\beta}})$ dipende dalla varianza di $\underline{\varepsilon}$, la quale in generale non è nota, occorrerà stimare anche quest'ultima. Uno stimatore non distorto e consistente di σ^2 è dato da:

$$\hat{\sigma}^2 = \frac{1}{n-k} (\underline{Y} - X\hat{\underline{\beta}})' (\underline{Y} - X\hat{\underline{\beta}})$$

Le ipotesi sul modello (3.4) possono venire rafforzate in senso parametrico, ipotizzando cioè di conoscere la distribuzione degli errori. L'ipotesi comunemente più adottata per $\underline{\varepsilon}$ è quella di normalità:

$$\underline{\varepsilon} \sim N_n(\underline{0}, \sigma^2 I_n)$$

che implica

$$\underline{Y} \sim N_n(X\underline{\beta}, \sigma^2 I_n)$$

Risulta allora immediato stabilire che lo stimatore di massima verosimiglianza di $\underline{\beta}$ è:

$$\hat{\underline{\beta}}_{ML} = (X'X)^{-1} X'\underline{Y}.$$

Come si vede tale stimatore coincide con quello dei minimi quadrati ed è il miglior stimatore non distorto per $\underline{\beta}$ (Minimum Variance Bound).

Giova sottolineare che sovente risulta inaccettabile l'ipotesi di omoschedasticità e non correlazione tra gli errori

$$(E (\underline{\varepsilon} \underline{\varepsilon}') = \sigma^2 I_n).$$

Qualora sia

$E (\underline{\varepsilon} \underline{\varepsilon}') = \Sigma$ (simmetrica, definita positiva, nota) lo stimatore

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

è detto "stimatore dei minimi quadrati generalizzati" o di Aitken e gode delle medesime proprietà di $\hat{\underline{\beta}}$.

Ovviamente

$$\text{Cov} (\hat{\underline{\beta}}_A) = (X' \Sigma^{-1} X)^{-1}.$$

Riferimenti bibliografici

Tra gli innumerevoli testi segnaliamo: Rao (1973), Johnston (1972), Goldberger (1964).

4. STIMATORI AMMISSIBILI

Le proprietà ricordate nel paragrafo precedente sono le proprietà più rilevanti che la statistica classica suole richiedere agli stimatori e sono quindi le proprietà che discriminano la "bontà" degli stimatori medesimi.

Con l'avvento della teoria delle decisioni l'attenzione si è spostata su altri aspetti ed altre proprietà.

In particolare si prende in considerazione una funzione di perdita $L(\hat{\theta}, \theta)$ che dipende dallo stimatore e dal valore incognito del parametro.

Essendo L una funzione aleatoria, come base per il confronto tra due stimatori, si considera il suo valore atteso che viene detto "rischio": $R_{\hat{\theta}}(\theta)$.

In generale uno stimatore $\hat{\theta}$ è detto "ammissibile" se non esiste un altro stimatore $\tilde{\theta}$ tale che

$$R_{\tilde{\theta}}(\theta) \leq R_{\hat{\theta}}(\theta) \quad \text{per ogni } \theta$$

e

$$R_{\tilde{\theta}}(\theta) < R_{\hat{\theta}}(\theta) \quad \text{per almeno un } \theta$$

Stein (1956) mostrò che nell'ipotesi di funzione di perdita quadratica

$$L(\hat{\theta}, \theta) = (\hat{\theta} - \theta)' (\hat{\theta} - \theta) \quad (4.1)$$

lo stimatore dei minimi quadrati e quello di massima verosimiglianza possono risultare non ammissibili.

Presentiamo in breve l'esempio di Stein. Sia

$$\underline{Y} \sim N(\underline{\xi}, I_p)$$

Lo stimatore di massima verosomiglianza per $\underline{\xi}$ è:

$$\tilde{\underline{\xi}} = \underline{Y}$$

Se impieghiamo una funzione di perdita del tipo (4.1) il rischio è $R_{\tilde{\underline{\xi}}}(\underline{\xi}) = p$. Quando $p \geq 3$, $\tilde{\underline{\xi}}$ non è uno stimatore ammissibile, esiste infatti almeno un altro stimatore avente rischio uniformemente minore.

Sia infatti:

$$\tilde{\underline{\xi}} = \left(1 - \frac{p-2}{\underline{Y}'\underline{Y}}\right) \underline{Y}$$

avremo

$$R_{\tilde{\underline{\xi}}}(\underline{\xi}) = E \{L(\tilde{\underline{\xi}}, \underline{\xi})\} = p - (p-2)^2 E \left(\frac{1}{p-2+2K} \right) < p$$

dove K è una variabile aleatoria con distribuzione di Poisson avente media $\frac{\underline{\xi}'\underline{\xi}}{2}$.

In sostanza Stein ha messo in evidenza come l'usuale stimatore per il vettore media per una normale multivariata non risulti soddisfacente e ne esistano altri preferibili che sono non lineari e distorti. Questi stimatori contraggono il vettore medio campionario verso lo zero.

Idee analoghe a quella di Stein hanno trovato sviluppo con la applicazione al modello (3.4). Hoerl e Kennard (1970 a) hanno introdotto l'idea di stimatore "ridge" che brevemente richiamiamo.

Dato il modello (3.4) ed una funzione di perdita quadratica,

il rischio di $\hat{\beta} = (X'X)^{-1} X'Y$ è:

$$E\{L(\hat{\beta}, \beta)\} = E\{(\hat{\beta}, \beta)'(\hat{\beta}, \beta)\} = \sigma^2 \operatorname{tr} (X'X)^{-1} = \sigma^2 \sum_{i=1}^p \frac{1}{\lambda_i} \quad (4.2)$$

dove λ_i ($i = 1, 2, \dots, p$) sono gli autovalori di $(X'X)$. Se $(X'X)$ è espressa nella forma di matrice di correlazione la (4.2) è tanto minore quanto più $(X'X)$ è prossima alla forma di matrice identità. L'idea di Hoerl e Kennard è stata quindi quella di "distorcere" lo stimatore dei minimi quadrati in questa direzione, definendo un nuovo stimatore

$$\hat{\beta}_R = (X'X + kI)^{-1} X'Y \quad (4.3)$$

dove k è una costante positiva data. Il problema, ora, è quello di verificare quando lo stimatore ridge è migliore di quello dei minimi quadrati, vale a dire quando

$$R_{\hat{\beta}}(\beta) - R_{\hat{\beta}_R}(\beta) > 0 \quad \forall \beta \quad (4.4)$$

Un teorema di Hoerl e Kennard afferma che esiste sempre un $k > 0$ tale che valga la (4.4). In particolare se Λ è la matrice degli autovalori di $X'X$ e P è una trasformazione ortogonale tale che $P' \Lambda P = X'X$, e $(\alpha_1, \dots, \alpha_p)' = \underline{\alpha} = P \underline{\beta}$ allora la (4.4) vale se

$$0 < k < \frac{\sigma^2}{\max(\alpha_i^2)}$$

Come si vede la disuguaglianza su k dipende dal vettore dei parametri incogniti e questa appare una limitazione logica a tale impostazione del problema. In pratica, però, è sovente accettabile porre dei limiti superiori a $\underline{\beta}'\underline{\beta}$ e quindi stabilire i vincoli per k .

Ricordiamo, inoltre, che esiste una espressione più generale

degli stimatori ridge (Bibby (1974), Rao (1976)).

$$\underline{\hat{\beta}}_R = (X'X + G)^{-1} X' \underline{Y} \quad (4.5)$$

dove G è una matrice semidefinita positiva. Una diversa proposta di stimatore ammissibile per il modello (3.4) è stata avanzata da Mayer e Willke (1973), i quali hanno individuato la seguente classe di stimatori

$$\underline{C}_\lambda = \lambda \underline{\hat{\beta}} \quad \text{con } \lambda \in (0 + \infty)$$

dove $\underline{\hat{\beta}}$ è lo stimatore dei minimi quadrati. Gli stimatori \underline{C}_λ sono detti "shrunken" (contratti) e λ è detto fattore di contrazione (shrinkage factor).

\underline{C}_λ è ammissibile se

$$\frac{\underline{\beta}'\underline{\beta} - \sigma^2 (X'X)^{-1}}{\underline{\beta}'\underline{\beta} + \sigma^2 (X'X)^{-1}} < \lambda < 1$$

Lo scopo di questa rassegna altro non era se non quello di ricordare alcuni dei contributi più classici o più recenti nell'ambito della stima dei parametri in un modello lineare.

Pur dal poco che si è detto appare evidente la notevole disgregazione dei metodi e dei loro fini ultimi. Ciascuno di questi stimatori è meglio degli altri per qualche verso, ma diversa ne è l'ispirazione, e sempre - in fin dei conti - di carattere formale. Questi stimatori sono "buoni" (o meno) in quanto capaci (o meno) di soddisfare una qualche proprietà analitica, piuttosto che in quanto ragionevole conseguenza di un procedimento di logica induttiva coerentemente condotto.

Riferimenti bibliografici

Gli stimatori di Stein: Stein (1956), James-Stein (1961), Stein (1962), Lindley (1962), Zellner-Vandaele (1974).

Ammissibilità per funzioni di perdita non quadratiche: Brown (1966).

Gli stimatori Ridge: Hoerl-Kennard (1970a), (1970b) (1975), Bibby (1974), Goldstein-Smith (1974).

Gli stimatori shrunken: Mayer-Willke (1973)

Per una visione generale e possibili confronti: Rao (1975), Rao (1976).

5. L'IMPOSTAZIONE BAYESIANA TRADIZIONALE DEL MODELLO LINEARE.

L'impostazione bayesiana conduce a risultati diversi nella sostanza anche se spesso numericamente coincidenti con alcuni di quelli già visti. La diversità risiede nell'unicità del principio informatore: il teorema di Bayes.

Con questo procedimento ci si vincola solamente a criteri di coerenza logica e non alla ricerca di soluzioni che soddisfino a proprietà formali.

Si consideri il modello

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

con

\underline{Y} vettore (nx1) aleatorio di variabili osservabili

X matrice (nxk) non stocastica di rango K

$\underline{\beta}$ vettore (kx1) di parametri *aleatori*

$\underline{\varepsilon}$ vettore (nx1) di errori aleatori non osservabili.

In generale si assume che $\underline{\varepsilon}$ si distribuisca normalmente con $E(\underline{\varepsilon}) = \underline{0}$ e $E(\underline{\varepsilon} \underline{\varepsilon}') = \sigma^2 I_n$.

Quindi

$$\underline{Y}|X, \underline{\beta}, \sigma^2 \sim N(X \underline{\beta}, \sigma^2 I_n)$$

cioè

$$\begin{aligned} p(\underline{Y}|X, \underline{\beta}, \sigma^2) &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma^2} (\underline{Y} - X \underline{\beta})' (\underline{Y} - X \underline{\beta}) \right\} \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \left[(n-k) \sigma^2 + (\underline{\beta} - \hat{\underline{\beta}})' X'X (\underline{\beta} - \hat{\underline{\beta}}) \right] \right\} \quad (5.1) \end{aligned}$$

dove

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

$$s^2 = \frac{1}{n - K} (\underline{Y} - \underline{X} \hat{\underline{\beta}})' (\underline{Y} - \underline{X} \hat{\underline{\beta}}).$$

Per poter procedere all'inferenza sui parametri $\underline{\beta}$ e σ^2 è necessario assegnare loro una distribuzione iniziale. Sono stati proposti diversi metodi di assegnazione. I più comuni sono:

- a) il metodo delle distribuzioni iniziali non informative
- b) il metodo delle distribuzioni naturali coniugate.

L'idea connessa alle distribuzioni iniziali non informative è quella di rappresentare una situazione di ignoranza "completa". Jeffreys (1961) ha suggerito distribuzioni di questo tipo:

$$p(\theta) \propto |\text{Inf}_\theta|^{1/2}$$

dove Inf_θ è la matrice di informazione di Fisher del parametro θ .

In particolare se $-\infty < \theta < +\infty$ allora

$$p(\theta) \, d\theta \propto d\theta$$

e se $\theta > 0$ allora

$$p(\theta) \, d\theta \propto \frac{d\theta}{\theta}.$$

Qualora la situazione iniziale non sia di ignoranza completa e si ricerchi una distribuzione finale "trattabile" un utile strumento è dato dalle distribuzioni coniugate (Raiffa-Schlaifer, 1961).

Tali distribuzioni godono della seguente proprietà: se la distribuzione iniziale appartiene ad una certa classe di distribuzioni anche la distribuzione finale vi appartiene, cioè ha la stessa forma funzionale e differisce solo per il valore di alcuni parametri.

La formula di Bayes consente la determinazione della distribuzione finale:

$$p(\underline{\beta}, \sigma^2 | \underline{y}, x) = \frac{p(\underline{\beta}, \sigma^2) p(\underline{y} | x, \underline{\beta}, \sigma^2)}{\int p(\underline{\beta}, \sigma^2) p(\underline{y} | x, \underline{\beta}, \sigma^2) d\underline{\beta} d\sigma^2} \quad (5.2)$$

Integrando la (5.2) rispetto a σ^2 si ottiene la distribuzione finale del vettore $\underline{\beta}$.

La distribuzione finale esaurisce il problema dell'inferenza in termini bayesiani.

Esistono, tuttavia, situazioni in cui è utile ricercare una conveniente sintesi della distribuzione finale.

La determinazione di questa sintesi può avvenire mediante il ricorso alla teoria delle decisioni. In tale caso il procedimento conduce ad attribuire un valore al parametro in modo che risulti minimo il rischio associato a questo tipo di decisione. Il valore che realizza tale proprietà è detto "stimatore bayesiano".

Qualora si scelga una funzione di perdita quadratica lo stimatore bayesiano coincide con il valore atteso della distribuzione finale.

Consideriamo, ora, alcuni risultati relativi alla scelta di diverse distribuzioni iniziali.

5.1. Distribuzioni iniziali non informative

5.1.1. σ^2 nota

La distribuzione non informativa à la Jeffreys per il vettore $\underline{\beta}$ è

$$p(\underline{\beta}) d\underline{\beta} \propto d\underline{\beta}.$$

Applicando la (5.2) avremo

$$p(\underline{\beta} | \underline{y}, X) \propto p(\underline{\beta}) \cdot p(\underline{y} | X, \underline{\beta})$$

da cui per la (5.1)

$$p(\underline{\beta} | \underline{y}, X) \propto \exp \left\{ -\frac{1}{2\sigma^2} [(\underline{\beta} - \hat{\underline{\beta}})' X'X (\underline{\beta} - \hat{\underline{\beta}})] \right\} \quad (5.3)$$

La distribuzione finale di $\underline{\beta}$ è normale multivariata k -dimensionale con valore atteso

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

e matrice di varianze e covarianze $\sigma^2 (X'X)^{-1}$.

5.1.2. σ^2 non nota.

La distribuzione iniziale non informativa à la Jeffreys assumendo $\underline{\beta}$ e σ^2 indipendenti è:

$$p(\underline{\beta}, \sigma^2) \propto \frac{1}{\sigma^2} \quad -\infty < \beta_i < +\infty \quad i = 1, 2, \dots, k \quad (5.4)$$

$$0 < \sigma < +\infty$$

Combinando la (5.1) e la (5.4) otteniamo la distribuzione finale congiunta:

$$p(\underline{\beta}, \sigma^2 | \underline{y}, X) \propto \sigma^{-n-2} \exp \left\{ -\frac{1}{2\sigma^2} [(n-k) s^2 + (\underline{\beta} - \hat{\underline{\beta}})' X'X (\underline{\beta} - \hat{\underline{\beta}})] \right\} \quad (5.5)$$

Integrando la (5.5) rispetto a σ^2 otteniamo la distribuzione finale marginale per $\underline{\beta}$:

$$p(\underline{\beta} | \underline{y}, X) = \int_0^{\infty} p(\underline{\beta}, \sigma^2 | \underline{y}, X) d\sigma^2$$

$$\propto \{ (n-k) s^2 + (\underline{\beta} - \hat{\underline{\beta}})' X'X (\underline{\beta} - \hat{\underline{\beta}}) \}^{-\frac{n}{2}}$$

che è una t di student k -variata con $(n-k)$ gradi di libertà, valore atteso $\hat{\underline{\beta}}$ e matrice di varianze e covarianze

$$\frac{n-k}{n-k-2} s^2 (X'X)^{-1}.$$

Analogamente integrando la (5.5) rispetto a $\underline{\beta}$ si ottiene la distribuzione finale marginale per σ^2 :

$$p(\sigma^2 | \underline{y}, X) \propto \frac{1}{\sigma^{n-k+2}} \exp \left\{ -\frac{(n-k)s^2}{2\sigma^2} \right\} \quad (5.6)$$

che è una gamma 1 inversa (si veda Raiffa-Schlaifer 1961, pag.227) di parametri $(n-k)/2$ e $(n-k)s^2/2$

Facciamo notare che dividendo la (5.5) per la (5.6) si ottiene $p(\underline{\beta} | \underline{y}, \sigma^2, X)$ che coincide con la (5.3)

5.2. Distribuzioni iniziali informative coniugate

5.2.1. σ^2 nota

La distribuzione iniziale per $\underline{\beta}$ coniugata al modello è

$$p(\underline{\beta}) \propto \exp \left\{ -\frac{1}{2} (\underline{\beta} - \underline{\beta}_0)' C_{\underline{\beta}}^{-1} (\underline{\beta} - \underline{\beta}_0) \right\}$$

cioè una normale multivariata con media $\underline{\beta}_0$ e matrice di varianze e covarianze C_β .

Data la (5.1) è immediato verificare che la distribuzione finale sarà ancora una normale con valore atteso

$$\underline{\beta}^* = \{(X'X)\sigma^{-2} + C_\beta^{-1}\}^{-1} \{\sigma^{-2} X'Y + C_\beta^{-1} \underline{\beta}_0\} \quad (5.7)$$

e matrice di varianze e covarianze

$$V = \{(X'X)\sigma^{-2} + C_\beta^{-1}\}^{-1}$$

Giova a questo punto fare due osservazioni:

a) nel caso in cui la matrice di varianze e covarianze di \underline{Y} non sia $\sigma^2 I_n$ bensì una matrice C_1 nota, la (5.7) diventa:

$$\underline{\beta}^* = \{(X' C_1^{-1} X + C_\beta^{-1})\}^{-1} \{X' C_1^{-1} Y + C_\beta^{-1} \underline{\beta}_0\}.$$

b) Ai risultati dati si può pervenire senza ricorrere all'ipotesi rango $(X) = k$. Questo significa che da un punto di vista bayesiano è possibile analizzare modelli con variabili esplicative collineari. Qualora l'ipotesi rango $(X) = k$ sia soddisfatta la (5.7) può scriversi:

$$\underline{\beta}^* = (X'X \sigma^{-2} + C_\beta^{-1})^{-1} (\sigma^{-2} (X'X) \hat{\underline{\beta}} + C_\beta^{-1} \underline{\beta}_0)$$

Si può notare allora che $\underline{\beta}^*$ è una media ponderata fra lo stimatore dei minimi quadrati $\hat{\underline{\beta}} = (X'X)^{-1} X'Y$ e l'opinione iniziale media rappresentata da $\underline{\beta}_0$, con pesi proporzionali all'inverso delle rispettive matrici di varianze e covarianze.

5.2.2. σ^2 non nota.

La distribuzione iniziale coniugata al modello per $\underline{\beta}$ e σ^2 è data da:

$$p(\underline{\beta}, \sigma^2) \propto \sigma^{-(m+k+2)} \exp\left\{-\frac{1}{2\sigma^2} [\gamma + (\underline{\beta} - \underline{\beta}_0)' C_{\beta}^{-1} (\underline{\beta} - \underline{\beta}_0)]\right\} \quad (5.8)$$

La (5.8) è data dal prodotto di una distribuzione normale multivariata per una gamma 1 inversa. Infatti $p(\underline{\beta}|\sigma^2)$ è una normale multivariata con media $\underline{\beta}_0$ e matrice di varianze e covarianze $\sigma^2 C_{\beta}$. La distribuzione marginale di σ^2 è una gamma 1 inversa:

$$p(\sigma^2) \propto \sigma^{-m-2} \exp\left\{-\frac{\gamma}{2\sigma^2}\right\}.$$

mentre la distribuzione marginale per $\underline{\beta}$ è una t multivariata con m gradi di variabilità:

$$p(\underline{\beta}) \propto [\gamma + (\underline{\beta} - \underline{\beta}_0)' C_{\beta}^{-1} (\underline{\beta} - \underline{\beta}_0)]^{-\frac{m+k}{2}}$$

combinando la (5.1) con la (5.8) otteniamo la distribuzione finale congiunta:

$$p(\underline{\beta}, \sigma^2 | \underline{y}, X) \propto \sigma^{-(n+m+k+2)} \exp\left\{-\frac{1}{2\sigma^2} [(n-k)s^2 + \gamma + (\underline{\beta} - \underline{\beta})' C_{\beta}^{-1} (\underline{\beta} - \underline{\beta}_0) + (\underline{\beta} - \hat{\underline{\beta}})' X'X (\underline{\beta} - \hat{\underline{\beta}})]\right\} \quad (5.9)$$

Integrando la (5.9) rispetto a σ^2 otteniamo la distribuzione finale marginale per $\underline{\beta}$:

$$p(\underline{\beta} | \underline{y}, X) \propto \left\{Z + (\underline{\beta} - \underline{\beta}^*)' (C_{\beta}^{-1} + X'X) (\underline{\beta} - \underline{\beta}^*)\right\}^{-\frac{n+m+k}{2}} \quad (5.10)$$

dove

$$\underline{\beta}^* = (C_{\beta}^{-1} + X'X)^{-1} (C_{\beta}^{-1} \underline{\beta}_0 + X'X \hat{\underline{\beta}}) \quad (5.11)$$

e

$$Z = (n-k)s^2 + \gamma + \underline{\beta}'_0 C_\beta^{-1} \underline{\beta}_0 + \underline{\beta}' X' X \hat{\underline{\beta}} - \underline{\beta}'^* (C_\beta^{-1} + X'X) \underline{\beta}^*$$

La (5.10) è una distribuzione t-multivariata con (m+n) gradi di libertà con valore atteso $\underline{\beta}^*$ e matrice di varianze e covarianze

$$(C_\beta^{-1} + X'X)^{-1} \frac{Z}{m+n-2}$$

La (5.11) rappresenta ancora una media ponderata tra l'opinione iniziale e lo stimatore dei minimi quadrati.

Ricordiamo che nulla cambia nella sostanza se il rango (X) < k.

La (5.11) si scriverà allora:

$$\underline{\beta}^* = (C_\beta^{-1} + X'X)^{-1} (C_\beta^{-1} \underline{\beta}_0 + X'Y)$$

Riferimenti bibliografici

Una rassegna dei principali risultati relativi all'impostazione bayesiana del modello lineare si trova in Zellner (1971). Si veda anche Maddala (1977), Box-Tiao (1973).

Distribuzioni iniziali: Jeffreys (1961), Raiffa-Schlaifer (1961), de Finetti-Savage (1962), Savage (1962), Edwards-Lindman-Savage (1963).

6. I MODELLI GERARCHICI

Sovente in econometria è avvertita la necessità di incorporare nell'analisi opinioni "a priori" riguardo ad alcuni coefficienti del modello lineare, opinioni che mal si esprimono attraverso le distribuzioni iniziali presentate nel paragrafo precedente. D'altra parte il ricorso a distribuzioni informative non coniugate farebbe sorgere grossi problemi computazionali.

Una interessante mediazione delle diverse esigenze è stata illustrata da Lindley e Smith in svariati articoli.

L'idea è la seguente: sempre nell'ambito del modello normale (1° stadio) si assegnano distribuzioni normali ai coefficienti (2° stadio); queste distribuzioni sono specificate a meno di parametri aleatori a cui ancora una volta si assegna una distribuzione normale (3° stadio) e così via.

Con questo procedimento si possono esprimere opinioni di questo tipo: "Ritengo probabile che all'interno di una certa equazione i regressori X_2 e X_3 abbiano il medesimo peso quindi β_2 e β_3 siano dello stesso ordine di grandezza, ma non so indicare quale questo debba essere".

Il secondo stadio esprimerà allora la somiglianza tra β_2 e β_3 il terzo esprimerà l'incertezza circa il valore di β_2 e β_3 .

In questo numero illustreremo l'impostazione in generale per poi particolarizzarla ad alcuni casi che ci sembrano interessanti: scambiabilità tra parametri e scambiabilità tra equazioni.

6.1. Modello a due stadi

Dato il modello $\underline{Y} = X_1 \underline{\beta}_1 + \underline{\varepsilon}$ sia \underline{Y} , dato $\underline{\beta}_1$, distribuito normalmente con valore atteso $X_1 \underline{\beta}_1$ e matrice di varianze e covarianze C_1 . Indicheremo

$$\underline{Y} | \underline{\beta}_1 \sim N (X_1 \underline{\beta}_1, C_1) \quad (1^\circ \text{ stadio})$$

Assumiamo poi:

$$\underline{\beta}_1 | \underline{\beta}_2 \sim N (X_2 \underline{\beta}_2, C_2) \quad (2^\circ \text{ stadio})$$

D'ora innanzi assumeremo che tutte le matrici di varianze e covarianze C_i siano definite positive e di dimensioni appropriate.

Con $X_1, C_1, X_2, C_2, \underline{\beta}_2$ noti si ottiene facilmente:

$$\underline{\beta}_1 | \underline{Y} \sim N (V_0 \underline{b}_0, V_0)$$

dove

$$V_0 = (C_2^{-1} + X_1' C_1^{-1} X_1)^{-1}$$

$$\underline{b}_0 = (X_1' C_1^{-1} \underline{Y} + C_2^{-1} X_2 \underline{\beta}_2).$$

6.2. Modello a tre stadi

Sia

$$\underline{Y} | \underline{\beta}_1 \sim N (X_1 \underline{\beta}_1, C_1) \quad (1^\circ \text{ stadio})$$

$$\underline{\beta}_1 | \underline{\beta}_2 \sim N (X_2 \underline{\beta}_2, C_2) \quad (2^\circ \text{ stadio})$$

$$\underline{\beta}_2 | \underline{\beta}_3 \sim N (X_3 \underline{\beta}_3, C_3) \quad (3^\circ \text{ stadio})$$

con $X_i, C_i, i = 1, 2, 3$ note.

Se moltiplichiamo $p(\underline{\beta}_1 | \underline{\beta}_2)$ per $p(\underline{\beta}_2 | \underline{\beta}_3)$ ed integriamo rispetto a $\underline{\beta}_2$ otteniamo che:

$$\underline{\beta}_1 | \underline{\beta}_3 \sim N(x_2 x_3 \underline{\beta}_3, C_2 + x_2 C_3 x_2') \quad (6.1)$$

L'applicazione del teorema di Bayes conduce a:

$$\underline{\beta}_1 | \underline{Y} \sim N(V_1 \underline{b}_1, V_1) \quad (6.2)$$

con

$$V_1 = \{x_1' C_1^{-1} x_1 + (C_2 + x_2 C_3 x_2')^{-1}\}^{-1} \quad (6.3)$$

$$\underline{b}_1 = \{x_1' C_1^{-1} \underline{Y} + (C_2 + x_2 C_3 x_2')^{-1} x_2 x_3 \underline{\beta}_3\} \quad (6.4)$$

Osservazioni:

Si noti che la (6.1) è ricavabile sulla base delle seguenti semplici considerazioni:

$$\underline{\beta}_1 = x_2 \underline{\beta}_2 + \underline{u}_1, \quad \underline{u}_1 \sim N(\underline{0}, C_2)$$

$$\underline{\beta}_2 = x_3 \underline{\beta}_3 + \underline{u}_2, \quad \underline{u}_2 \sim N(\underline{0}, C_3)$$

quindi

$$\underline{\beta}_1 = x_2 x_3 \underline{\beta}_3 + x_2 \underline{u}_2 + \underline{u}_1$$

Poichè \underline{u}_1 e \underline{u}_2 sono indipendenti, $x_2 \underline{u}_2 + \underline{u}_1$ è una funzione lineare di variabili normali indipendenti e pertanto

$$x_2 \underline{u}_2 + \underline{u}_1 \sim N(\underline{0}, x_2 C_3 x_2' + C_2),$$

da cui segue la (6.1).

Applicando il teorema di Bayes

$$p(\underline{\beta}_1 | \underline{Y}) \propto p(\underline{\beta}_1) p(\underline{Y} | \underline{\beta}_1)$$

si deriva la (6.2)

La struttura del valore atteso della (6.2) è ancora quella di

una media ponderata tra le opinioni iniziali e lo stimatore dei minimi quadrati

Se fossimo interessati alla distribuzione di $\underline{\beta}_2 \mid \underline{Y}$ potremmo procedere in modo analogo a quello visto nell'osservazione ottenendo

$$\underline{Y} \mid \underline{\beta}_2 \sim N(\underline{x}_1 \underline{x}_2 \underline{\beta}_2 ; C_1 + \underline{x}_1 C_2 \underline{x}_1')$$

Poichè

$$\underline{\beta}_2 \mid \underline{\beta}_3 \sim N(\underline{x}_3 \underline{\beta}_3 , C_3)$$

applicando il teorema di Bayes si ottiene:

$$\underline{\beta}_2 \mid \underline{Y} \sim N(\underline{v}_2 \underline{b}_2 , \underline{v}_2)$$

dove

$$\underline{v}_2 = \{ \underline{x}_2' \underline{x}_1' (C_1 + \underline{x}_1 C_2 \underline{x}_2')^{-1} \underline{x}_1 \underline{x}_2 + C_3^{-1} \}^{-1} \quad (6.5)$$

$$\underline{b}_2 = \{ \underline{x}_2' \underline{x}_1' (C_1 + \underline{x}_1 C_2 \underline{x}_1')^{-1} \underline{Y} + C_3^{-1} \underline{x}_3 \underline{\beta}_3 \} \quad (6.6)$$

Il procedimento come è ovvio potrebbe proseguire un numero arbitrario di volte. E' vero però che le informazioni che possono essere incorporate in successivi stadi vanno riducendosi sempre più (si veda Goel- De Groot (1981)). proprio per questa ragione è spesso conveniente dare una distribuzione diffusa già a $\underline{\beta}_2$. Tale condizione è esprimibile assumendo $C_3^{-1} = 0$. Riscrivendo pertanto la (6.3) e la (6.4) in termini di C_3^{-1} e sostituendo si ottiene

$$\underline{\beta}_1 \mid \underline{Y} \sim N(\hat{\underline{v}}_1 \hat{\underline{b}}_1 , \hat{\underline{v}}_1)$$

dove

$$\hat{\underline{v}}_1 = \{ \underline{x}_1' C_1^{-1} \underline{x}_1 + C_2^{-1} - C_2^{-1} \underline{x}_2 (\underline{x}_2' C_2^{-1} \underline{x}_2)^{-1} \underline{x}_2' C_2^{-1} \}^{-1}$$

$$\hat{\underline{b}}_1 = \underline{x}_1' C_1^{-1} \underline{Y} .$$

Se si effettua lo stesso procedimento sulla (6.5) e (6.6)

si ha:

$$\underline{\beta}_2 | \underline{Y} \sim N(\hat{\underline{v}}_2 \hat{\underline{b}}_2, \hat{\underline{v}}_2)$$

con

$$\hat{\underline{v}}_2 = \{x_2' [(x_1' c_1^{-1} x_1)^{-1} + c_2]^{-1} x_2\}^{-1}$$

$$\hat{\underline{b}}_2 = \{x_2' x_1' (c_1 + x_1 c_2 x_1')^{-1} \underline{Y}\}.$$

6.3. Scambiabilità tra parametri

In molte occasioni il ricercatore si trova nelle condizioni di assegnare lo stesso peso ai regressori, vale a dire suppone che i coefficienti siano scambiabili. Per quanto detto allora nel paragrafo 2 si potrà assegnare agli elementi del vettore β_1 una distribuzione tale per cui questi siano indipendenti condizionatamente ad un parametro aleatorio β_2 (medie dei parametri β_1).

Allora se si assume che i coefficienti del vettore $\underline{\beta}_1 = \underline{\beta}$ siano scambiabili e $C_1 = \sigma_A^2 I_n$ il modello a tre stadi può scriversi:

$$\underline{Y} | \underline{\beta} \sim N(x \underline{\beta}, \sigma_A^2 I_n)$$

$$\underline{\beta} | \mu_\beta \sim N(\mu_\beta \frac{1}{p}, \sigma_\beta^2 I_p)$$

$$\mu_\beta \sim N(\mu_0, \sigma_C^2)$$

Ricordando la (6.1) avremo:

$$\begin{aligned} \underline{\beta} | \mu_0 &\sim N\left(\frac{1}{p} \mu_0, \sigma_\beta^2 I + \frac{1}{p} \sigma_C^2 \frac{1}{p}'\right) \\ &\sim N\left(\frac{1}{p} \mu_0, \sigma_\beta^2 I_p + \sigma_C^2 J_p\right) \end{aligned}$$

dove J_p è una matrice $p \times p$ costituita da elementi unitari.

La distribuzione finale è

$$\underline{\beta} | \underline{Y} \sim N(\underline{V}_* \underline{b}_*, \underline{V}_*)$$

con

$$\underline{V}_* = \left\{ \sigma_A^{-2} (X'X) + \sigma_\beta^{-2} \left(I_p - \frac{\sigma_C^2}{p\sigma_C^2 + \sigma_\beta^2} J_p \right) \right\}^{-1}$$
$$\underline{b}_* = \left\{ \sigma_A^{-2} X'Y + \sigma_\beta^{-2} \left(I_p - \frac{\sigma_C^2}{p\sigma_C^2 + \sigma_\beta^2} J_p \right) \underline{1} \mu_0 \right\}$$

Se la distribuzione di μ_β è diffusa, cioè $\sigma_C^{-2} = 0$ avremo

$$\underline{\beta} | \underline{Y} \sim N(\hat{\underline{V}}_* \hat{\underline{b}}_*, \hat{\underline{V}}_*)$$

in cui

$$\hat{\underline{V}}_* = \left\{ \sigma_A^{-2} X'X + \sigma_\beta^{-2} \left(I_p - \frac{1}{p} J_p \right) \right\}^{-1}$$
$$\hat{\underline{b}}_* = \sigma_A^{-2} X'Y.$$

6.4. Scambiabilità tra equazioni

Una diversa applicazione dell'idea di scambiabilità è la seguente: consideriamo un modello costituito da m equazioni, ciascuna delle quali faccia dipendere linearmente un vettore di osservazioni \underline{y}_i da una matrice di regressori X_i . Supponiamo che il numero dei regressori sia uguale per ciascuna equazione.

In simboli

$$\underline{y}_i | \underline{\beta}_i \sim N(X_i \underline{\beta}_i, \sigma_i^2 I_{n_i})$$

per $i = 1, 2, \dots, m$.

Consideriamo scambiabili i vettori dei coefficienti delle m equazioni

$$\underline{\beta}_i \mid \underline{\xi} \sim N(\underline{\xi}, \Sigma)$$

$$\underline{\xi} \sim \text{diffusa.}$$

La distribuzione finale sarà:

$$\underline{\beta}_i \mid \underline{y} \sim N(v_i \underline{b}_i, v_i)$$

con

$$v_i = (\sigma_i^{-2} \underline{x}_i' \underline{x}_i + \Sigma^{-1})^{-1}$$

$$\underline{b}_i = (\sigma_i^{-2} \underline{x}_i' \underline{y} + \Sigma^{-1} \underline{\beta}^*)$$

dove

$$\underline{\beta}^* = \frac{m}{\sum_{j=1}^m} W_j \hat{\underline{\beta}}_j$$

essendo

$$W_i = \left[\sum_{j=1}^m (\underline{x}_j' \underline{x}_j \sigma_j^{-2} + \Sigma^{-1})^{-1} \underline{x}_j' \underline{x}_j \sigma_j^{-2} \right]^{-1} (\underline{x}_i' \underline{x}_i \sigma_i^{-2} + \Sigma^{-1})^{-1} \underline{x}_i' \underline{x}_i \sigma_i^{-2}$$

Come è facile verificare anche la media di quest'ultima distribuzione finale (stimatore bayesiano con funzione di perdita quadratica) è una media ponderata tra lo stimatore dei minimi quadrati della singola equazione e la media $\underline{\beta}^*$ degli m vettori $\underline{\beta}_i$. Infatti:

$$\underline{\beta}^* = \frac{1}{m} \sum_{j=1}^m (v_j \underline{b}_j)$$

6.5. Osservazioni e Confronti

Nei numeri precedenti abbiamo già fatto notare che lo scopo dell'impostazione bayesiana non è la ricerca di stimatori che soddisfino a proprietà formali; ciò nonostante crediamo che valga la pena di effettuare qualche confronto.

Assumendo una funzione di perdita quadratica abbiamo visto che lo stimatore coincide con il valore atteso della distribuzione finale. Sotto questa ipotesi gli stimatori ottenuti con il procedimento gerarchico sono "ammissibili" quando le matrici C_i sono non singolari. Lo sono anche nel caso in cui la matrice di varianze e covarianze considerata nell'ultimo stadio sia singolare (ipotesi di distribuzione diffusa) a condizione che questa abbia dimensione non superiore a due (si veda Cohen (1966)).

Il principio delle ammissibilità porta a dire che qualunque metodo di decisione statistico è bayesiano (sia pur non intenzionalmente) oppure può essere sostituito da un metodo bayesiano che gli sia incondizionatamente preferibile.

Ritornando al modello lineare è immediato verificare che trasformando opportunamente le variabili del modello (3.4), il risultato di Stein si applica anche alla stima dei minimi quadrati di β . Lo stesso Stein (1962) dopo aver proposto uno stimatore ammissibile si preoccupò di porre in luce come tale stimatore potesse essere interpretato come una approssimazione di quello bayesiano nell'ipotesi di funzione di perdita quadratica. Analoga interpretazione è stata fornita da Lindley (1962).

Facciamo inoltre notare che lo stimatore $V_* \underline{b}_*$ è del tipo ridge generalizzato; può essere infatti scritto:

$$\hat{V}_* \hat{b}_* = \left\{ X'X + \frac{\sigma^2}{\sigma_B^2} (I_p - \frac{1}{p} J_p) \right\}^{-1} X'Y ;$$

la matrice G della (4.6) è in questo caso:

$$G = \left\{ \frac{\sigma^2}{\sigma_B^2} (I_p - \frac{1}{p} J_p) \right\}$$

ed è singolare.

Lo stimatore ridge (4.4) corrisponde ad uno stimatore bayesiano ottenuto con una distribuzione iniziale avente media nulla.

Infatti se

$$Y|\beta \sim N(X\beta, \sigma^2 I)$$

$$\beta \sim N(\underline{0}, \sigma_B^2 I)$$

si ottiene

$$E(\beta|Y) = (X'X + k I)^{-1} X'Y$$

$$\text{con } k = \frac{\sigma^2}{\sigma_B^2}$$

6.6. Modello gerarchico con matrici di varianze e covarianze non note.

Finora abbiamo assunto che le matrici di varianze e covarianze nei diversi stadi fossero note. Succede spesso che questa ipotesi non sia accettabile. In tal caso è necessario assegnare una distribuzione iniziale anche a tali componenti. Le distribuzioni iniziali più utilizzate per matrici di varianze e covarianze sono del tipo Wishart.

Consideriamo il modello:

$$\underline{y} | \underline{\beta}_1, C_1 \sim N(\underline{x}_1 \underline{\beta}_1, C_1)$$

$$\underline{\beta}_1 | \underline{\beta}_2, C_2 \sim N(\underline{x}_2 \underline{\beta}_2, C_2)$$

Supponiamo che C_1 abbia una distribuzione Wishart di parametri p ed R e che C_1 e $\underline{\beta}_1$ siano tra loro indipendenti, inoltre C_2 e $\underline{\beta}_2$ siano noti.

La distribuzione iniziale su $\underline{\beta}_1$ e C_1 è:

$$p(\underline{\beta}_1, C_1 | \underline{\beta}_2, C_2, R, p) \propto$$

$$\exp\left\{-\frac{1}{2}(\underline{\beta}_1 - \underline{x}_2 \underline{\beta}_2)' C_2^{-1} (\underline{\beta}_1 - \underline{x}_2 \underline{\beta}_2)\right\} \cdot |C_1|^{-\frac{1}{2}(p+n+1)} \exp\left\{-\frac{1}{2} \text{tr } C_1^{-1} R\right\}$$

Applicando il teorema di Bayes otteniamo la seguente distribuzione finale

$$p(\underline{\beta}_1, C_1 | \underline{y}) \propto \exp\left\{-\frac{1}{2} [(\underline{y} - \underline{x}_1 \underline{\beta}_1)' C_1^{-1} (\underline{y} - \underline{x}_1 \underline{\beta}_1)] - \frac{1}{2} [(\underline{\beta}_1 - \underline{x}_2 \underline{\beta}_2)' C_2^{-1} (\underline{\beta}_1 - \underline{x}_2 \underline{\beta}_2)] - \frac{1}{2} \text{tr } C_1^{-1} R\right\} \cdot |C_1|^{-\frac{1}{2}(p+n+2)}$$

Integrando rispetto a C_1 si ottiene la distribuzione finale di $\underline{\beta}_1$.

$$p(\underline{\beta}_1 | \underline{y}) \propto \exp\left\{-\frac{1}{2} (\underline{\beta}_1 - \underline{x}_2 \underline{\beta}_2)' C_2^{-1} (\underline{\beta}_1 - \underline{x}_2 \underline{\beta}_2)\right\} \cdot \left|R + (\underline{y} - \underline{x}_1 \underline{\beta}_1) (\underline{y} - \underline{x}_1 \underline{\beta}_1)'\right|^{-\frac{p+1}{2}}$$

Il nucleo di questa distribuzione è dato dal prodotto dei nuclei di una normale e di una t di Student. Distribuzioni di questo tipo sono dette double- t (Dréze(1976)).

Tali distribuzioni non sono facili da trattare infatti una double- t è, in generale, bimodale ed inoltre non è possibile calcolarne per via analitica, nè la costante di normalizzazione, nè i

momenti. E' necessario ricorrere a metodi di analisi numerica (si veda Richard e Tompa, (1980)).

Consideriamo, ora, in particolare il caso dei coefficienti scambiabili:

$$\underline{y} | \underline{\beta}, \sigma_A^2 \sim N(\underline{x}\underline{\beta}, \sigma_A^2 I_n)$$

$$\underline{\beta} | \mu_\beta, \sigma_\beta^2 \sim N(\underline{1} \mu_\beta, \sigma_\beta^2 I_p)$$

$$\mu_\beta \sim \text{diffusa}$$

$$\sigma_A^2 \sim I \text{ Ga}(\nu/2, \nu\lambda/2).$$

$$\sigma_\beta^2 \sim I \text{ Ga}(\nu_\beta/2, \nu_\beta\lambda_\beta/2).$$

dove I Ga indica la distribuzione Gamma 1 inversa.

La distribuzione iniziale sar :

$$p(\underline{\beta}, \sigma_A^2, \mu_\beta, \sigma_\beta^2) \propto \exp\left\{-\frac{1}{2\sigma_\beta^2} [(\underline{\beta}-\underline{1}\mu_\beta)'(\underline{\beta}-\underline{1}\mu_\beta) + \nu_\beta\lambda_\beta]\right\} \cdot \sigma_\beta^{-\nu} \sigma_A^{-\nu-2} \exp\left\{-\frac{\lambda}{2\sigma_A^2}\right\} \quad (6.7)$$

Integrando la (6.7) rispetto a μ_β e σ_β^2 otteniamo:

$$p(\underline{\beta}, \sigma_A^2) \propto [\nu_\beta \lambda_\beta + \underline{\beta}' (I_p - \frac{1}{p} J_p) \underline{\beta}]^{-\frac{\nu_\beta+p}{1}} \cdot \sigma_A^{-\nu-2} \exp\left\{-\frac{\nu\lambda}{2\sigma_A^2}\right\}$$

pertanto la distribuzione finale per $\underline{\beta}$ sar :

$$p(\underline{\beta} | \underline{y}) \propto [\nu_\beta \lambda_\beta + \underline{\beta}' (I_p - \frac{1}{p} J_p) \underline{\beta}]^{-\frac{\nu_\beta+p}{2} - 1} \cdot [\nu\lambda + (\underline{y}-\underline{x}\underline{\beta})'(\underline{y}-\underline{x}\underline{\beta})]^{-\frac{\nu+n}{2}} \quad (6.8)$$

La (6.8) rappresenta ancora una distribuzione double-t.

Un metodo per approssimare lo stimatore bayesiano nel caso di

campioni di ampiezza elevata è stato proposto da (Tiao-Zellner, (1964)).

Riferimenti bibliografici

Il modello gerarchico: Lindley-Smith (1972), Smith (1973a), Lindley (1971), Leonard (1975), Goel-De Groot (1981).

Le distribuzioni poly-t: Tiao-Zellner (1964), Dréze (1976), Richard-Tompa (1980).

7. L'IMPOSTAZIONE BAYESIANA NON PARAMETRICA DEL MODELLO LINEARE

7.1. Distribuzioni iniziali per modelli non parametrici.

Una peculiarità dell'impostazione bayesiana, come abbiamo visto, è la necessità di assegnare una misura di probabilità iniziale ai parametri del modello statistico. Può darsi però il caso che non sia possibile esprimere il modello statistico in forma parametrica, cioè restringere l'attenzione ad una determinata famiglia di misure di probabilità.

In tali situazioni occorrerà depositare una misura di probabilità sullo spazio delle funzioni di ripartizione. Il primo efficace tentativo di individuare una conveniente misura di probabilità da depositare sullo spazio delle f.r. è venuto da Ferguson(1973).

L'estensione dei risultati relativi al processo di Dirichlet e il livello formale richiesto rendono impossibile un'esposizione dettagliata nell'economia di questo lavoro.

Il lettore interessato è rimandato agli articoli originali.

Consideriamo una misura α , non nulla, finita, non negativa e finitamente additiva su $(\mathcal{R}, \mathcal{B})$ dove \mathcal{R} è l'asse reale e \mathcal{B} è la σ -algebra di Borel. La misura di probabilità aleatoria P su $(\mathcal{R}, \mathcal{B})$ è un processo di Dirichlet su $(\mathcal{R}, \mathcal{B})$ con parametro α , se, per ogni $k = 1, 2, \dots$, e per ogni partizione misurabile di \mathcal{R} :

$\{B_1, B_2, \dots, B_k\}$, la distribuzione del vettore aleatorio

$\{P(B_1), P(B_2), \dots, P(B_k)\}$ è di Dirichlet con parametri $(\alpha(B_1), \dots, \alpha(B_k))$. Scriveremo $P \in \mathcal{D}(\alpha)$.

Dati k n.a. Z_1, \dots, Z_k , mutuamente indipendenti e tali che $Z_i \sim \text{Ga}(\alpha_i, 1)$, $\alpha_i \geq 0 \forall i$ e $\alpha_i > 0$ per qualche $i = 1, 2, \dots, k$, allora i n.a. Y_1, \dots, Y_k si distribuiscono secondo una Dirichlet di parametri $(\alpha_1, \dots, \alpha_k)$ se

$$Y_i = \frac{Z_i}{\sum_{j=1}^k Z_j} \quad \text{per } i = 1, 2, \dots, k.$$

Tale definizione di distribuzione di Dirichlet è più generale di quella comunemente usata (Wilks, 1962) poichè consente che qual-
che α_i sia uguale a zero.

Il parametro α rappresenta l'opinione iniziale media circa la conformazione della misura aleatoria P . Si ha infatti

- a) se $A \in \mathcal{B}$: $\alpha(A) = 0$ implica $P(A) = 0$ q.c.
- b) $E(P(A)) = \frac{\alpha(A)}{\alpha(\mathcal{R})}$

$\alpha(\mathcal{R})$ rappresenta il peso che si assegna a questa nostra opinione iniziale. Questi brevi cenni mettono in evidenza come il processo di Dirichlet si possa fruttuosamente utilizzare come distribuzione iniziale.

In questo senso è utile ricordare la proprietà di "chiusura" del processo. Infatti:

Se $P \in \mathcal{D}(\alpha)$ e (X_1, X_2, \dots, X_n) è un campione estratto da P , allora il processo $(P|X_1, \dots, X_n)$ è ancora un processo di Dirichlet con parametro $\beta = \alpha + \sum_{i=1}^n \delta_{X_i}$, essendo δ_x la misura che assegna massa unitaria al punto x .

$$\alpha(P(A) \mid X_1, X_2, \dots, X_n) = \frac{\alpha(A) + \sum_{i=1}^n \delta_{X_i}(A)}{\alpha(\mathbb{R}) + n}$$

E' utile per i nostri scopi ricordare una generalizzazione del pro
cesso di Dirichlet dovuta a Antoniak (1974).

Si dice che P è una mistura di processi di Dirichlet se
 $P|u \in \mathcal{D}(\alpha|u)$

dove u è un n.a. distribuito secondo una f.r. H. Allora

$$P \in \int_{\mathbb{R}} \mathcal{D}(\alpha|u) dH(u)$$

Il processo mistura di processi di Dirichlet ha rilevanza so
prattutto per l'estensione che è possibile farne quando le osser
vazioni sono solo parzialmente scambiabili. Nei casi di parziale
scambiabilità infatti è improponibile l'uso del processo di Dirichlet a reggere la distribuzione delle osservazioni poichè queste sono disomogenee, se provenienti da classi diverse. Neppure è pos
sibile scegliere un prodotto di k processi di Dirichlet a meno che le classi non siano tra loro indipendenti. L'interdipendenza tra le diverse classi può essere introdotta indicizzando il parametro di ciascuno dei k processi ad un parametro aleatorio u. Misturando il prodotto dei k processi di Dirichlet si ottiene un processo conforme all'ipotesi di scambiabilità parziale, detto ap
punto "Processo Mistura di Prodotti di Processi di Dirichlet" (PMPPD) (si veda Cifarelli e Regazzini (1978)).

In simboli:

$$P \in \int_U \prod_{i=1}^k \mathcal{D}(\alpha_i(\cdot|u)) dH(u)$$

Anche per questo processo valgono interessanti proprietà:

$$a) E(P_i(A)) = \int_U \frac{\alpha_i(A|u)}{\alpha_i(\mathcal{R}|u)} dH(u)$$

$$b) \alpha_i(A|u) = 0 \Rightarrow P_i(A) = 0 \quad \text{q.c.}$$

Inoltre vale la solita proprietà di chiusura.

7.2. Applicazione del modello lineare

Consideriamo il solito modello

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

e supponiamo che \underline{y} ed X siano così partizionati.

$$\underline{Y} = \begin{bmatrix} \underline{Y}_1 \\ \vdots \\ \underline{Y}_k \end{bmatrix} \quad \text{con } \underline{Y}_i = \begin{bmatrix} Y_{i1} \\ \vdots \\ Y_{in_i} \end{bmatrix} \quad i = 1, 2, \dots, k$$

dove $\sum_{j=1}^k n_j = n$ ed n_j eventualmente nullo.

$$X = \begin{bmatrix} X_1 \\ \dots \\ X_k \end{bmatrix} \quad \text{con } X_i = \frac{1}{n_i} \underline{x}_i' = \begin{bmatrix} \underline{x}_i' \\ \vdots \\ \underline{x}_i' \end{bmatrix}$$

con $\underline{x}_i' = (x_{i1}, \dots, x_{ip})$

$$\underline{\varepsilon} = \begin{bmatrix} \underline{\varepsilon}_1 \\ \vdots \\ \underline{\varepsilon}_k \end{bmatrix} \quad \text{con } \underline{\varepsilon}_i' = (\varepsilon_{i1}, \dots, \varepsilon_{in_i})$$

In tal modo vengono individuate k classi di elementi scambiabili, in corrispondenza di k diverse conformazioni dei regressori. Per quanto detto nel paragrafo 2 si ha:

$$P\{\underline{Y}_1 \leq \underline{\xi}_1, \dots, \underline{Y}_k \leq \underline{\xi}_k \mid \underline{X}, \underline{\beta}, F_1, \dots, F_k\} = \\ = \prod_{j=1}^{n_1} F_1(\xi_{1j}) \prod_{j=1}^{n_2} F_2(\xi_{2j}) \dots \prod_{j=1}^{n_k} F_k(\xi_{kj}).$$

Supponiamo di non conoscere la forma funzionale di F_1, \dots, F_k ed assumiamo che queste siano selezionate da PMPPD di parametro $\underline{\alpha}$. (E' chiaro che la scelta di tale parametro deve essere confacente alle opinioni iniziali del ricercatore).

Allo scopo di poter effettuare alcuni confronti specificheremo $\underline{\alpha}$ come segue:

$$\alpha_i(\xi \mid \underline{\beta}, \underline{x}_i', \sigma_i) = \alpha(\mathcal{R}) \Phi\left(\frac{\xi - \underline{x}_i' \underline{\beta}}{\sigma_i}\right)$$

dove $\alpha_i(t \mid \underline{\beta}) \stackrel{\text{def}}{=} \alpha_i((-\infty, t] \mid \underline{\beta})$ e Φ è la f.r. della normale standardizzata.

E' necessario ora assegnare una distribuzione ai parametri $\underline{\beta}$ e σ_i . In analogia a quanto fatto precedentemente supporremo σ_i noto e $\underline{\beta}$ distribuito normalmente.

Per determinare la distribuzione finale di $\underline{\beta}$ utilizzeremo un risultato di Antoniak (1974) e Cifarelli - Regazzini (1978) che presentiamo in una forma semplificata corrispondente alle nostre ipotesi.

Sia $\underline{Y}_1, \dots, \underline{Y}_k$ un campione estratto da un PMPPD con parametro

$\alpha(\cdot|u)$ e distribuzione dei pesi $H(u)$, con α misura assolutamente continua rispetto alla misura di Lebesgue. Organizziamo, ora, le osservazioni di ciascuna classe in forma di variabile statistica.

$$\begin{array}{l} y_{i1}, \dots, y_{i r_i} \\ n_{i1}, \dots, n_{i r_i} \end{array} \quad \sum_{j=1}^{r_i} n_{ij} = n_i$$

Si ha allora

$$dH(\underline{u} | \underline{y}_1, \dots, \underline{y}_k) \propto \prod_{i=1}^k \frac{1}{\alpha_i(\mathbb{R} | \underline{u})} \prod_{j=1}^{r_i} \frac{d\alpha_i}{d\mu}(y_{ij} | \underline{u}) dH(\underline{u})$$

dove r_i indica il numero di osservazioni distinte per ogni classe.

E' immediato verificare che per il modello lineare si ottiene:

$$H^{(p)}(\underline{\beta} | \underline{y}_1, \dots, \underline{y}_k, \underline{x}'_1, \dots, \underline{x}'_k, \sigma_\beta^2) \propto$$

$$\propto \prod_{i=1}^k \prod_{j=1}^{r_i} \phi\left(\frac{y_{ij} - \underline{x}'_i \underline{\beta}}{\sigma_i}\right) H^{(p)}(\underline{\beta}) \propto$$

$$\propto \exp\left\{-\frac{1}{2} [(\tilde{\underline{y}} - \tilde{\underline{X}}\underline{\beta})' \tilde{C}_1^{-1} (\tilde{\underline{y}} - \tilde{\underline{X}}\underline{\beta}) + (\underline{\beta} - \underline{1} \mu_\beta) \sigma_\beta^{-2}]\right\}$$

Tale espressione coincide con quella data per il modello parametrico tranne che per il fatto che vengono considerate solo le osservazioni distinte di ogni classe: $\tilde{\underline{y}}$ (e le corrispondenti matrici $\tilde{\underline{X}}$ e \tilde{C}_1).

Il discorso vale in generale sia per i modelli con matrice di varianze e covarianze note sia per quelli in cui tale matrice non è nota.

I risultati ottenuti permettono di trattare in maniera sufficientemente agevole anche un modello in cui le ipotesi di partenza siano molto più deboli di quelle tradizionalmente adottate (la co-

noscenza della forma analitica della distribuzione delle osservazioni). Nel caso particolare da noi analizzato basterà escludere le osservazioni coincidenti e quindi operare in maniera analoga al modello parametrico.

Sottolineiamo ancora che la scelta di α è assolutamente soggettiva e che comunque α non rappresenta altro che una opinione iniziale, opinione che le osservazioni possono mutare anche in maniera radicale.

La (7.1) mostra che la scelta del modello non parametrico (ai fini della determinazione della distribuzione di β) non comporta nessun aggravio dal punto di vista dei calcoli rispetto al corrispondente modello parametrico.

In conclusione risulterà conveniente al ricercatore non impegnarsi in maniera dogmatica circa la distribuzione delle osservazioni a meno che non abbia sufficienti informazioni.

Riferimenti bibliografici

Il processo di Dirichlet: Ferguson (1973), (1974), Antoniak (1974), Cifarelli-Regazzini (1978).

Il modello lineare nell'impostazione non parametrica: Cifarelli-Muliere-Scarsini (1980).

8. PROBLEMI DI PREVISIONE NEL MODELLO LINEARE

Nell'analisi finora condotta abbiamo considerato gli aspetti ipotetici del problema statistico. Nella sostanza abbiamo rivolto la nostra attenzione alla verifica della attendibilità di leggi che governano un dato fenomeno.

Esamineremo ora il problema della previsione di nuove osservazioni sulla base di altre passate, problema che, a nostro avviso, è meglio individuabile nei suoi connotati probabilistici, potendo essere ridotto alla determinazione della funzione "predittiva":

$$\Pr\{Y_{n+1}' \leq Y_{n+1}, \dots, Y_{n+r}' \leq Y_{n+r} \mid Y_1 = y_1, \dots, Y_n = y_n\}. \quad (8.1)$$

Per poter valutare la probabilità di eventi futuri sulla base di eventi passati (noti) occorrerà stabilire un legame logico di dipendenza tra gli eventi tutti (passati e futuri).

Vedremo ora come sia possibile ricavare la (8.1) allorchè le osservazioni siano estratte da una successione $\{Y_n\}$ scambiabile. Dal teorema de Finetti discende:

$$\begin{aligned} & \Pr\{Y_{n+1}' \leq Y_{n+1}, \dots, Y_{n+r}' \leq Y_{n+r} \mid Y_1 = y_1, \dots, Y_n = y_n\} = \\ & = \int_F \prod_{j=1}^r F(y_{n+j}) \, d\mu(F \mid y_1, \dots, y_n) \end{aligned} \quad (8.2)$$

dove $\mu(F \mid y_1, \dots, y_n)$ può essere calcolata mediante il teorema di Bayes (2.3).

Chiaramente la (8.2) nel caso di problemi parametrici diventa

$$\begin{aligned} & \Pr\{Y_{n+1}' \leq Y_{n+1}, \dots, Y_{n+r}' \leq Y_{n+r} \mid Y_1 = y_1, \dots, Y_n = y_n\} = \\ & = \int_{\mathcal{R}^n} \prod_{j=1}^r F(y_{n+j} \mid \theta) \, dH(\theta \mid y_1, \dots, y_n). \end{aligned}$$

Una posizione estrema, all'interno dell'ipotesi di scambiabilità, è quella in cui la successione $\{Y_n\}$ è composta da n.a. stocasticamente indipendenti.

Tale assunzione porta a:

$$\begin{aligned} & \Pr\{Y_{n+1} \leq Y_{n+1}', \dots, Y_{n+r} \leq Y_{n+r}' \mid Y_1 = y_1, \dots, Y_n = y_n\} = \\ & = P\{Y_{n+1} \leq Y_{n+1}', \dots, Y_{n+r} \leq Y_{n+r}'\}. \end{aligned} \quad (8.3)$$

Vale dire: gli eventi passati non forniscono alcuna informazione circa quelli futuri.

(La (8.3) implica che la distribuzione F di ciascun n.a. della successione $\{Y_n\}$ (nel modello parametrico il parametro) sia completamente nota, il che significa che non vi è necessità di fare inferenza.

Nel caso di parziale scambiabilità avremo:

$$\begin{aligned} & \Pr\{Y_{1,n_1+1} \leq Y_{1,n_1+1}', \dots, Y_{1,n_1+r_1} \leq Y_{1,n_1+r_1}', \dots \\ & \dots Y_{k,n_k+1} \leq Y_{k,n_k+1}', \dots, Y_{k,n_k+r_k} \leq Y_{k,n_k+r_k}' \mid \\ & \mid Y_{1,1} = y_{1,1}, \dots, Y_{k,n_k} = y_{k,n_k}\} = \\ & = \int_{F^{(k)}} \prod_{i=1}^k \prod_{j_i=1}^{n_i} F(y_{i,j_i}) \, d\mu(F \mid \underline{y}_1, \dots, \underline{y}_k) \end{aligned} \quad (8.4)$$

La (8.4) sarà utilizzata per la determinazione della predittiva nel modello lineare.

La funzione predittiva svolge un ruolo assai importante nella soluzione di molti problemi di analisi multivariata.

In particolare il fine ultimo di molti modelli econometrici è la determinazione (in senso probabilistico) del valore che assume-

ranno alcune variabili endogene in corrispondenza di valori prefissati di altre variabili esogene, questo anche al fine di risolvere problemi di controllo.

Mostreremo che la trattazione di un problema previsivo nel modello lineare seppur unitaria nell'impostazione conduce a risultati sensibilmente differenti a seconda che il modello sia parametrico o non parametrico.

8.1. Modello parametrico: varianza nota

Dato il solito modello a tre stadi:

$$\underline{y} | \underline{\beta}_1 \sim N(\underline{x}_1 \underline{\beta}_1, \sigma^2 I)$$

$$\underline{\beta}_1 | \underline{\beta}_2 \sim N(\underline{x}_2 \underline{\beta}_2, C_2)$$

$$\underline{\beta}_2 | \underline{\beta}_3 \sim N(\underline{x}_3 \underline{\beta}_3, C_3)$$

si ha:

$$\underline{\beta}_1 | \underline{y} \sim N(\underline{\beta}^*, V) \tag{8.5}$$

dove $\underline{\beta}^*$ si determinano sulla base delle espressioni (6.2), (6.3), (6.4).

Indicato con $\dot{\underline{y}}$ un vettore ($n \times 1$) di osservazioni future effettuate in corrispondenza della matrice $\dot{\underline{x}}_1$

$$\dot{\underline{y}} | \underline{\beta}_1 \sim N(\dot{\underline{x}}_1 \underline{\beta}_1, \sigma^2 I) \tag{8.6}$$

si ha:

$$\dot{\underline{y}} | \underline{y} \sim N(\dot{\underline{x}}_1 \underline{\beta}^*, \sigma^2 I + \dot{\underline{x}}_1 V \dot{\underline{x}}_1') \tag{8.7}$$

La (8.7) si ottiene moltiplicando la (8.5) per la (8.6) ed integrando rispetto a $\underline{\beta}_1$. La (8.7) si ottiene altresì svolgendo

considerazioni analoghe a quelle presentate nell'osservazione del punto 6.2. Cioè:

$$\underline{\dot{Y}} = \underline{\dot{X}} \underline{\beta} + \underline{\dot{\epsilon}} \quad , \quad \underline{\dot{\epsilon}} | \underline{Y} \sim N(\underline{0} , \sigma^2 \underline{I})$$

$$\underline{\beta} = \underline{\beta}^* + \underline{\eta} \quad , \quad \underline{\eta} | \underline{Y} \sim N(\underline{0} , \underline{V})$$

quindi

$$\underline{\dot{Y}} = \underline{\dot{X}} \underline{\beta}^* + \underline{\dot{X}} \underline{\eta} + \underline{\dot{\epsilon}}$$

Poichè $\underline{\dot{X}} \underline{\eta} + \underline{\dot{\epsilon}}$ è una funzione lineare di vettori aleatori normali ed indipendenti dato \underline{Y} , si ha

$$\underline{\dot{X}} \underline{\eta} + \underline{\dot{\epsilon}} | \underline{Y} \sim N(\underline{0} , \underline{\dot{X}} \underline{V} \underline{\dot{X}}' + \sigma^2 \underline{I})$$

da cui la (8.7).

Come si può vedere la distribuzione predittiva è normale ed ha media uguale al prodotto della matrice dei regressori per il vettore media della distribuzione finale dei parametri.

La matrice di varianze e covarianze è uguale alla somma di due componenti: la prima corrisponde alla matrice di varianze e covarianze delle osservazioni condizionatamente ai parametri, la seconda alla matrice di varianze e covarianze indotta dai parametri.

E' immediato notare che la previsione mantiene la struttura lineare inizialmente ipotizzata, infatti la media della predittiva è linearmente legata ai regressori, essendo i coefficienti della combinazione lineare dati dalle medie di $\underline{\beta} | \underline{Y}$.

Le osservazioni effettuate nelle diverse classi hanno tutte il medesimo peso nella determinazione della predittiva (anche se riferite a classi diverse). Queste due ultime considerazioni sono diretta conseguenza della struttura parametrica del modello. Infatti nella formulazione iniziale abbiamo ipotizzato la normalità

delle osservazioni. Tale ipotesi non viene modificata dal verificarsi di qualunque risultato sperimentale.

L'ipotesi forte di normalità delle osservazioni si traduce quindi in una ipotesi forte di linearità del modello vale a dire in un legame lineare tra le diverse classi di scambiabilità.

8.2. Modello parametrico: varianza non nota.

Dato il modello:

$$\begin{aligned} \underline{Y} | \underline{\beta}_1, \sigma^2 &\sim N(\underline{X}_1 \underline{\beta}_1, \sigma^2 \underline{I}) \\ \underline{\beta}_1 | \sigma^2 &\sim N(\underline{\beta}_0, \sigma^2 \underline{C}_\beta) \\ \sigma^2 &\sim \text{I Ga} (m/2, \gamma/2), \end{aligned}$$

si avrà, dalla (8.7)

$$\underline{\dot{Y}} | \underline{Y}, \sigma^2 \sim N(\underline{\dot{X}}_1 \underline{\beta}^*, \sigma^2 (\underline{I} + \underline{\dot{X}}_1 \underline{V} \underline{\dot{X}}_1')) \quad (8.8)$$

dove

$$\begin{aligned} \underline{\beta}^* &= (\underline{X}_1' \underline{X}_1 + \underline{C}_\beta^{-1})^{-1} (\underline{X}_1' \underline{Y} + \underline{C}_\beta^{-1} \underline{\beta}_0) \\ \underline{V} &= (\underline{X}_1' \underline{X}_1 + \underline{C}_\beta^{-1})^{-1} \end{aligned}$$

Moltiplicando la (8.8) per $p(\sigma^2 | \underline{Y})$ ed integrando rispetto a σ^2 , si ottiene che $\underline{\dot{Y}} | \underline{Y}$ è distribuito secondo una t multivariata con $(m+n)$ gradi di libertà con valore atteso $\underline{\dot{X}}_1 \underline{\beta}^*$ e matrice di varianze e covarianze

$$\frac{Z}{m+n-2} (\underline{\dot{X}}_1 \underline{V} \underline{\dot{X}}_1' + \underline{I});$$

l'espressione di Z è data dalla (5.11).

Anche per il modello con varianza incognita quindi, a patto di

scegliere una distribuzione iniziale coniugata per i parametri valgono proprietà analoghe a quelle ricordate per il modello con varianza nota.

Se nel modello con matrice di varianze e covarianze non nota si fosse assunta una distribuzione del tipo normal-gamma non coniugata che conduce ad una distribuzione marginale finale di $\underline{\beta}$ del tipo double-t allora la determinazione della distribuzione predittiva sarebbe stata molto difficoltosa. A tutt'oggi non ci consta che ne sia stata fornita espressione in forma chiusa.

8.3. Modello non parametrico.

Nei numeri precedenti abbiamo mostrato gli effetti della ipotesi di normalità del modello statistico sulla funzione predittiva. In questo paragrafo esamineremo le conseguenze della rimozione di una tale ipotesi. Assumeremo in particolare che le osservazioni siano estratte da un processo mistura di prodotti di processi di Dirichlet.

Ricordiamo alcuni risultati:

Se $P \in \mathcal{D}(\alpha)$ ed F è la f.r. corrispondente a P allora

$$\Pr\{Y \leq y\} = \frac{\alpha(-\infty, y]}{\alpha(\mathcal{R})} = F_0(y)$$

La funzione predittiva quando si sono effettuate n osservazioni è:

$$\begin{aligned} \Pr\{Y_{n+1} \leq y \mid Y_1 = y_1, \dots, Y_n = y_n\} = \\ = \frac{\alpha(\mathcal{R})}{\alpha(\mathcal{R}) + n} F_0(y) + \frac{n}{\alpha(\mathcal{R}) + n} \hat{F}_n(y) \end{aligned}$$

essendo $\hat{F}_n(y)$ la f.r. empirica

$$\text{Se } (P_1, P_2, \dots, P_k) = \underline{p} \in \int_U \prod_{i=1}^k \mathcal{D}(\alpha_i(\cdot|u)) dH(u)$$

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

$$\begin{aligned} \text{Pr}\{Y_{1, n_1+1} \leq Y_1, \dots, Y_{k, n_k+1} \leq Y_k \mid \underline{y}_1, \underline{y}_2, \dots, \underline{y}_k\} = \\ = \int_U \prod_{i=1}^k \left\{ \frac{\alpha_i(\mathcal{R}|u)}{\alpha_i(\mathcal{R}|u) + n_i} \frac{\alpha_i(y_i|u)}{\alpha_i(\mathcal{R}|u) + n_i} + \frac{n_i}{\alpha_i(\mathcal{R}|u) + n_i} \hat{F}_{i, n_i}(y_i) \right\} \\ dH(u | \underline{y}_1, \dots, \underline{y}_k) \end{aligned}$$

in cui \hat{F}_{i, n_i} è la f.r. empirica determinata con le osservazioni della i .esima classe, y_{i1}, \dots, y_{in_i} ed $H(u | \underline{y}_1, \dots, \underline{y}_k)$ è la f.r. finale di $\underline{\beta}$ di cui abbiamo detto nel paragrafo 8.2.

Per semplicità di riferimento supponiamo che la distribuzione finale per $\underline{\beta} | \underline{y}_1, \dots, \underline{y}_k$ sia normale con media:

$$\underline{\beta}_0^* = (\tilde{X} \tilde{\Sigma}^{-1} \tilde{X} + \sigma_\beta^{-2} I)^{-1} (\tilde{X} \tilde{\Sigma}^{-1} \underline{y} + \sigma_\beta^2 \underline{1} \beta_0)$$

e matrice di varianze e covarianze

$$V_0^* = (\tilde{X} \tilde{\Sigma}^{-1} \tilde{X} + \sigma_\beta^{-2} I)^{-1}$$

dove

$$\tilde{\Sigma} = \begin{bmatrix} \sigma_1^2 I_{r1} & & & \\ & \sigma_2^2 I_{r2} & & \\ & & \dots & \\ & & & \sigma_k^2 I_{rk} \end{bmatrix}$$

La f.r. predittiva della generica marginale Y_i è:

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

ed integrando si ha:

$$\psi_i(Y|\underline{Y}) = \frac{\alpha(\mathcal{R})}{\alpha(\mathcal{R}) + n_i} \phi\left(\frac{y - \underline{x}'_i \underline{\beta}_0^*}{v_{ii}}\right) + \frac{n_i}{\alpha(\mathcal{R}) + n_i} \hat{F}_{i, n_i}(y) \quad (8.9)$$

in cui v_{ii} è l'elemento i.esimo sulla diagonale della matrice V_0^* , cioè

$$v_{ii} = \sigma_i^2 + \underline{x}'_i V_0^* \underline{x}_i$$

La (8.9) è una mistura di due f.r.: quella normale (analoga a quella del modello parametrico) e quella empirica delle osservazioni della classe i.esima. Le osservazioni di tutte le classi intervengono nella determinazione della f.r. predittiva attraverso la prima componente. Tutte le osservazioni della i.esima classe intervengono anche nella seconda componente, per cui la $\psi_i(Y|\underline{Y})$ concentra masse di probabilità in corrispondenza di tutte le osservazioni della classe a cui si riferisce.

Si noti che:

$$\psi_i(Y|\underline{Y}) \rightarrow \Phi\left(\frac{y - \underline{x}'_i \underline{\beta}_0^*}{v_{ii}}\right) \quad \text{se } \alpha(\mathcal{R}) \rightarrow +\infty.$$

e

$$\psi_i(Y|\underline{Y}) \rightarrow \hat{F}_{i, n_i}(y) \quad \text{se } \alpha(\mathcal{R}) \rightarrow 0.$$

L'espressione generale della predittiva per due o più osservazioni

(di classi diverse) si complica notevolmente essendo data dalla mistura di diverse componenti (assolutamente continue, discrete e miste). E' possibile tuttavia dare agevolmente l'espressione della media di queste distribuzioni.

$$E(Y_{1,n_1+1}, \dots, Y_{k,n_1+1} | \underline{y}_1, \dots, \underline{y}_k) = \\ = \Pi \hat{\beta}_0^* + (I - \Pi) \underline{\bar{y}}$$

essendo

$$\Pi_{ii} = \frac{\alpha(\mathcal{R})}{\alpha(\mathcal{R}) + n_i} \quad e \quad \Pi_{ij} = 0 \quad \text{per } i \neq j$$

e

$$\bar{y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}$$

E' immediato notare che in questa espressione non viene preservata la struttura lineare del modello nel senso che la media della distribuzione predittiva non è linearmente legata ai regressori. Al crescere del numero delle osservazioni, inoltre, acquista sempre più peso la f.r. empirica che è svincolata da qualsiasi struttura lineare tanto che le k funzioni \hat{F}_i sono tra loro indipendenti.

Riferimenti bibliografici

Problemi di previsione statistica: Aitchison-Dunsmore (1975) de Finetti (1971), Geisser (1971).

Per il modello lineare (non parametrico) Cifarelli-Muliere-Scarsini (1981).

9. ANALISI DELLA VARIANZA

Una significativa particolarizzazione del modello lineare generale è data dai modelli di analisi della varianza, modelli in cui si considerano osservazioni provenienti da diverse classi, con lo scopo di mettere a confronto le medie di queste classi.

Il modello si formula nel modo seguente:

$$\begin{bmatrix} y_{11} \\ \cdot \\ \cdot \\ \cdot \\ y_{1n_1} \\ \dots \\ y_{21} \\ \cdot \\ \cdot \\ \cdot \\ y_{2n_2} \\ \dots \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ y_{k1} \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ y_{kn_k} \end{bmatrix} = \begin{bmatrix} 1 \\ \cdot \\ \cdot \\ \cdot \\ 1 \\ \dots \\ 1 \\ \cdot \\ \cdot \\ \cdot \\ 1 \\ \dots \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ 1 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \cdot \\ \theta_k \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \cdot \\ \cdot \\ \cdot \\ \epsilon_{1n_1} \\ \dots \\ 2 \ 1 \\ \cdot \\ \cdot \\ \cdot \\ \epsilon_{2n_2} \\ \dots \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \epsilon_{k1} \\ \cdot \\ \cdot \\ \cdot \\ \epsilon_{kn_k} \end{bmatrix}$$

$$\underline{Y} = A \underline{\theta} + \underline{\epsilon}$$

Consideriamo ora il consueto modello gerarchico ed assumiamo che i parametri siano tra loro scambiabili.

$$\underline{y} | \underline{\theta}_1 \sim N (A_1 \underline{\theta}_1, C_1)$$

$$\underline{\theta}_1 | \theta_2 \sim N (\underline{1}_k \theta_2, C_2)$$

$$\theta_2 \sim \text{diffusa}$$

con

$$C_1 = \begin{bmatrix} \sigma_1^2 I & & & & \\ & \sigma_2^2 I & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & \sigma_k^2 I \end{bmatrix}$$

e

$$C_2 = \sigma_\beta^2 I$$

Per le cose dette nel paragrafo 6.2 si ha:

$$\underline{\theta}_1 | \underline{y} \sim N (\hat{V}_1 \hat{\underline{b}}_1, \hat{V}_1)$$

con

$$\hat{V}_1 = \{A_1' C_1^{-1} A_1 + C_2^{-1} - C_2^{-1} \underline{1} (\underline{1}' C_2^{-1} \underline{1})^{-1} \underline{1} C_2^{-1}\}^{-1}$$

$$\hat{\underline{b}}_1 = A_1' C_1^{-1} \underline{y}$$

Indichiamo la media della distribuzione finale con $\theta^* = V_1^{-1} \underline{b}_1$.

Semplici calcoli portano a verificare che \hat{V}_1 è una matrice di elementi:

$$v_{ii} = \frac{m_i}{\sigma_I^2} + \frac{1}{\sigma_\beta^2} - \frac{1}{k \sigma_\beta^2}$$

$$v_{ij} = -\frac{1}{k \sigma_\beta^2} \quad i \neq j$$

Quindi \hat{V}_1^{-1} avrà elementi:

$$v^{ii} = \sigma_\beta^2 (1 - \omega_i) (1 + \sum_{i \neq j} \omega_j) (\sum \omega_i)^{-1}$$

$$v^{ij} = \sigma_\beta^2 (1 - \omega_i) (1 - \omega_j) (\sum \omega_i)^{-1} \quad i \neq j$$

$$\text{con } \omega_i = n_i \sigma_\beta^2 (n_i \sigma_\beta^2 + \sigma_i^2)^{-1} \quad i = 1, 2, \dots, n.$$

Moltiplicando la matrice \hat{V}_1^{-1} per $A_1' C_1^{-1} \underline{y}$ otteniamo:

$$\theta_i = \frac{n_i \sigma_\beta^2 \bar{y}_i + \sigma_i^2 \bar{\bar{y}}}{n_i \sigma_\beta^2 + \sigma_i^2} \quad i = 1, 2, \dots, n$$

$$\text{dove } \bar{\bar{y}} = (\sum \omega_i)^{-1} \sum \omega_i \bar{y}_i$$

I risultati ora esposti sono dovuti a Lindley (1971) che si limitò al caso $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2 = \sigma_A^2$.

Lo stesso modello è stato studiato in ambito non parametrico da Cifarelli (1979).

Lo studio dei modelli dell'analisi della varianza a due criteri è stato condotto da Smith (1973a) e da Lindley (1974). Nel seguito illustreremo un modello di analisi della varianza a due criteri di cui i modelli di Lindley e Smith possono essere considerati casi particolari.

Assumiamo

$$Y_{ijr} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijr}$$

con $i = 1, \dots, a$; $j = 1, \dots, b$; $r = 1, \dots, n$; e

$$\begin{aligned} \text{a) } \sum_i \alpha_i &= \sum_j \beta_j = 0 \\ \text{b) } \sum_i \gamma_{ij} &= 0 \quad \forall i \\ \sum_j \gamma_{ij} &= 0 \quad \forall j \end{aligned} \tag{9.1}$$

Nel seguito supporremo che vi siano k osservazioni per ogni cella ij , si avrà pertanto $r = 1, 2, \dots, k.a.b.$

Sia, ora

$$\underline{y} | \underline{\theta}_1 \sim N(A_1 \underline{\theta}_1, C_1)$$

dove

$$\begin{aligned} \underline{\theta}'_1 &= [\mu, \alpha_1, \dots, \alpha_a, \beta_1, \dots, \beta_b, \gamma_{11}, \dots, \gamma_{ab}] , \\ C_1 &= \sigma^2 I. \end{aligned}$$

e

A_1 è una matrice di elementi 0, 1, -1 tale che siano soddisfatte le assunzioni a) e b).

Assegniamo ora a $\underline{\theta}_1$ una distribuzione iniziale che tenga conto delle condizioni a) e b) e che esprima una ipotesi di scambiabilità degli α_i e dei β_i . L'assunzione di scambiabilità implica che vi siano inizialmente medie e varianze uguali per gli α_i e β_i e uguali covarianze tra ogni coppia (α_i, α_j) e (β_i, β_j) .

La condizione b) impedisce invece di assumere i γ_{ij} scambiabili.

Ipotizzeremo quindi:

$$\text{Cov} (\gamma_{ij}, \gamma_{qr}) = z \quad \text{se } i \neq q, \quad j \neq r$$

$$\text{Cov} (\gamma_{ij}, \gamma_{qr}) = v \quad \text{se } i \neq q, \quad j = r$$

$$\text{Cov} (\gamma_{ij}, \gamma_{qr}) = u \quad \text{se } i = q, \quad j \neq r$$

$$\text{Cov} (\gamma_{ij}, \gamma_{qr}) = s \quad \text{se } i = q, \quad j = r$$

Pertanto

$$\underline{\theta}_1 | \underline{\theta}_2 \sim N (A_2 \underline{\theta}_2, C_2)$$

con

$$A_2 \underline{\theta}_2 = \begin{bmatrix} \omega \\ \underline{0} \end{bmatrix}$$

e

$$C_2 = \begin{bmatrix} \sigma_\mu^2 & \underline{0}' & \underline{0}' & \underline{0}' \\ \underline{0} & \sigma_\alpha^2 (I_a - \frac{1}{a} J_a) & 0 & 0 \\ \underline{0} & 0 & \sigma_\beta^2 (I_b - \frac{1}{b} J_b) & 0 \\ \underline{0} & 0 & 0 & \sigma_\gamma^2 (I_{ab} - \frac{1}{b} (I_a \otimes J_b) - \frac{1}{a} (J_a \otimes I_b) + \frac{1}{ab} J_{ab}) \end{bmatrix}$$

C_2 , come si vede è una matrice diagonale a blocchi. Il primo blocco è uno scalare. Gli altri tre blocchi sono singolari. Nell'ultimo blocco sono singolari tutti i sottoblocchi quadrati ottenuti considerando le a righe e le corrispondenti colonne relative a $\gamma_{i.}$, o quelli ottenuti considerando le b righe e le corrispondenti colonne relative a $\gamma_{.j}$.

Evidentemente $\mu, \alpha, \beta, \gamma$ sono mutuamente indipendenti.

Sottolineiamo che

$\text{rango} (A_1) = a \cdot b < (a+1) (b+1)$; l'identificazione di $\underline{\theta}_1$ è ottenu

Indicando con $\underline{\theta}^* = D_1 \underline{b}_1$ e sostituendo nella (9.3) e (9.4) i valori precedentemente indicati di $A_1, C_1, C_2, \underline{\theta}_2$ e ponendo $\sigma_\mu^{-2} \rightarrow 0$ si ottiene

$$\underline{\theta}^* = \begin{bmatrix} \underline{\mu}^* \\ \underline{\alpha}^* \\ \underline{\beta}^* \\ \underline{\gamma}^* \end{bmatrix} \quad D_1 = \begin{bmatrix} \Sigma_u^* & & & \\ & \Sigma_\alpha^* & & \\ & & \Sigma_\beta^* & \\ & & & \Sigma_\gamma^* \end{bmatrix}$$

con

$$u^* = \bar{y}_{..}$$

$$\alpha_i^* = k b \sigma_\alpha^2 (k b \sigma_\alpha^2 + \sigma^2)^{-1} (\bar{y}_{i.} - \bar{y}_{..})$$

$$\beta_j^* = k a \sigma_\beta^2 (k a \sigma_\beta^2 + \sigma^2)^{-1} (\bar{y}_{.j} - \bar{y}_{..})$$

$$\gamma_{ij}^* = k \sigma_\gamma^2 (k \sigma_\gamma^2 + \sigma^2)^{-1} (\bar{y}_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})$$

e

$$\Sigma_u^* = \frac{\sigma^2}{k a b}$$

$$\Sigma_\alpha^* = \sigma^2 \sigma_\alpha^2 (k b \sigma_\alpha^2 + \sigma^2)^{-1} (I_a - \frac{1}{a} J_a)$$

$$\Sigma_\beta^* = \sigma^2 \sigma_\beta^2 (k a \sigma_\beta^2 + \sigma^2)^{-1} (I_b - \frac{1}{b} J_b)$$

$$\Sigma_\gamma^* = \sigma^2 \sigma_\gamma^2 (k \sigma_\gamma^2 + \sigma^2)^{-1} (I_{ab} - \frac{1}{b} (I_a \otimes J_b) - \frac{1}{a} (J_a \otimes I_b) + \frac{1}{ab} J_{ab}).$$

Anche in questo caso le medie della distribuzione finale sono medie ponderate tra lo stimatore dei minimi quadrati e l'opinione iniziale. Dato che la distribuzione iniziale per μ è diffusa, μ^* coincide con lo stimatore dei minimi quadrati; per $\alpha_i, \beta_i, \gamma_{ij}$ l'opinione iniziale è zero ed il peso dell'esperienza empirica è in ogni caso proporzionale a $\frac{k}{\sigma^2}$, il peso dell'opinione ini-

ziale è:

$$\begin{array}{ll} \text{per } \alpha & \propto \frac{1}{b\sigma_{\alpha}^2} \\ \text{per } \beta & \propto \frac{1}{a\sigma_{\beta}^2} \\ \text{per } \gamma & \propto \frac{1}{\sigma_{\gamma}^2} \end{array}$$

IL citato modello di Smith (1973) è un caso particolare di questo esposto, ottenibile eliminando il blocco delle interazioni. Si può notare, tuttavia, che poichè i blocchi sono stati assunti indipendenti i risultati relativi alla parte comune coincidono.

Lindley (1974) invece nell'analisi di un modello a due criteri considera l'effetto congiunto riga colonna, strutturando la matrice di varianze e covarianze in modo analogo a quanto da noi fatto per il blocco dei γ_{ij} .

Poichè nel modello di Lindley i coefficienti non sono sottoposti a vincoli, la matrice delle loro varianze e covarianze è non singolare.

Riferimenti bibliografici

Impostazione bayesiana di analisi della varianza: ad un criterio: Box e Tiao (1968), Lindley (1971), Cifarelli (1979);
a due criteri: Lindley (1974), Smith (1973).

B I B L I O G R A F I A

- [1] J. Aitchison - J.R. Dunsmore (1975): *Statistical Prediction Analysis*, Cambridge University Press, Cambridge.
- [2] C. Antoniak (1974): Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *Ann. Statist.*, 2, 1152-1174.
- [3] J. Bibby (1974): Minimum mean square error estimation, ridge regression, and some unanswered questions. *Proc. 9th. European Meeting statisticians Budapest* (1972) reprinted in J. Gani, K. Sarkadi, I. Vincze *Progress in Statistics*, 107-121.
- [4] G.E. Box e G.C. Tiao (1968): Bayesian estimation of means for the random effect model, *J. Am. Statist. Ass.*, 63, 174-181.
- [5] G.E.P. Box - G.C. Tiao (1973): *Bayesian Inference in Statistical Analysis*, Addison-Wesley, Reading Massachusetts
- [6] L.D. Brown (1966): On the admissibility of invariant estimators of one or more location parameters. *Ann. Math. Statist.*, 37, 1087-1136.
- [7] D. M. Cifarelli (1979): Impostazione bayesiana di un problema dell'analisi della varianza con approccio non parametrico.
Ist. Mat. Fin. Un. Torino,

- [8] D.M. Cifarelli - P. Muliere - M. Scarsini (1981): Il modello lineare nell'approccio bayesiano non parametrico, Quaderni dell'Istituto G. Castelnuovo, Roma.
- [9] D.M. Cifarelli - E. Regazzini (1978): Problemi statistici non parametrici in condizioni di scambiabilità parziale; impiego di medie associative. Ist. Mat. Fin. Un. Torino.
- [10] A. Cohen (1966): All admissible linear estimates of the mean vector. *Ann. Math. Statist.*, 37, 458-463.
- [11] B. de Finetti (1931): Funzione caratteristica di un fenomeno aleatorio. *Mem. Acc. Lincei*, 4, 86-133.
- [12] B. de Finetti (1937): La prévision: ses lois logiques, ses sources subjective, *Ann. Inst. H. Poincaré*, VII, 1-68.
- [13] B. de Finetti (1938): Sur la condition d'équivalence partielle, VI Colloque Genève, *Act. Sc. Ind.*, n. 739, Paris, 5-18.
- [14] B. de Finetti (1959): La probabilità e la statistica nei rapporti con l'induzione, secondo i diversi punti di vista. In corso C.I.M.E. su Induzione e Statistica, *Varenna, Cremonese, Roma*.
- [15] B. de Finetti (1970): *Teoria delle probabilità (sintesi introduttiva con appendice critica)*, Giulio Einaudi editore, Torino.

- [16] B. de Finetti (1971): Probabilità di una teoria e probabilità dei fatti, In *Studi di Probabilità e Ricerca operativa in onore di Giuseppe Pompilj*, 86-101, Tipografia Oderisi.
- [17] B. de Finetti - L.J. Savage (1962): Sul modo di scegliere le probabilità iniziali, *Biblioteca del Metron* 1, 81-154.
- [18] M.H. De Groot (1970): *Optimal Statistical Decisions*, Mc Graw-Hill, New York.
- [19] J.H. Dreze (1976): Bayesian Regression analysis using poly-t densities, In A. Aykac e C. Brumat (eds), *New Developments in the Applications of Bayesian Methods*, 153-184, North-Holland.
- [20] W. Edwards - H. Lindman - L.J. Savage (1963): Bayesian statistical inference for psychological research. *Psych Review*, 70, 193-242.
- [21] T.S. Ferguson (1973): A Bayesian analysis of some non parametric problems. *Ann. Statist*, 1, 209-230.
- [22] T.S. Ferguson (1974): Prior distributions on spaces of probability measures. *Ann. Statist*, 2, 615-619.
- [23] S. Geisser (1971): The inferential use of predictive distributions, In B.P. Godambe and D.A. Sprott (eds), *Foundations of Statistical Inference*, Holt, Rinehardt e Winston, Toronto, 456-469.
- [24] P.K. Goel - M.H. De Groot (1981): Information about hyperparameters in hierachcal models, *J. Am. Statist. Assoc.*, 76, 140-147.

- [25] A.S. Goldberger (1964): *Econometric Theory*, J. Wiley, New York.
- [26] M. Goldstein - A.F.M. Smith (1974): Ridge-type estimators for regression analysis, *J. Roy. Statist. Soc. Ser. B.* 36, 284-291.
- [27] E. Hewitt - L.J. Savage (1955): Symmetric measures on Cartesian products, *Trans. Amer. Math. Soc.*, 80, 470-501.
- [28] A.E. Hoerl - R.W. Kennard (1970a): Ridge regression: biased estimation for nonorthogonal problems. *Technometrics*, 12, 55-67.
- [29] A.E. Hoerl - R.W. Kennard (1970b): Ridge regression: application to nonorthogonal problems. *Technometrics*, 12, 69-82.
- [30] A.E. Hoerl - R.W. Kennard (1975): A note on a power generalization of ridge regression, *Technometrics*, 17, 269.
- [31] W. James - C. Stein (1961): Estimation with quadratic loss. In *Proc. 4th Berkeley Symp. Math. Statist. Prob.* 1, 361-379 Univ. of California Press, Berkeley.
- [32] H. Jeffreys (1961): *Theory of Probability* (3rd ed), Clarendon Press, Oxford.
- [33] J. Johnston (1972): *Econometric Methods*, 2nd ed. Mc Graw-Hill, New York.

- [34] J.F.C. Kingman (1978): Uses of exchangeability, *Ann. Probability*, 6, 183-197.
- [35] T. Leonard (1975): A Bayesian approach to the linear model with unequal variances, *Technometrics*, 17, 95-102.
- [36] D.V. Lindley (1962): Discussion on Professor Stein's paper. *J. Roy. Statist. Soc. Ser. B*, 24, 285-287.
- [37] D.V. Lindley (1965): *Introduction to Probability and Statistics from a Bayesian Viewpoint*, Vol. I, II. Cambridge University Press.
- [38] D.V. Lindley (1971): *Bayesian Statistics; A Review*, Philadelphia, SIAM.
- [39] D.V. Lindley (1971): *Making Decision*, J. Wiley, New York.
- [40] D.V. Lindley (1971): The estimation of many parameters. In V.P. Godambe - D.A. Sprott (eds) *Foundations of Statistical Inference*, 435-455, Holt, Rinehart & Winston, Toronto.
- [41] D.V. Lindley (1974): A Bayesian solution for two-way analysis of variance. *Proc. 9th. European Meeting of Statistician Budapest* (1972) reprinted in J. Gani, K. Sarkadi, I; Vincze (eds). *Progress in Statistics*, 475-497.
- [42] D.V. Lindley - M.R. Novick (1981): The role of exchangeability in inference, *Ann. Statist.*, 9, 45-58.
- [43] D.V. Lindley - A.F.M. Smith (1972): Bayes estimates for the linear model (with discussion). *J. Roy. Statist. Soc. Ser. B*, 34, 1-41.

- [44] G.S. Maddala (1977): *Econometrics*, Mc Graw-Hill, New York.
- [45] L.S. Mayer - T.A. Willke (1973): On biased estimation in linear models. *Technometrics*, 15, 497-508.
- [46] H. Raiffa - R. Schlaifer (1961): *Applied Statistical Decision Theory*, Graduate School of Business Administration, Harvard University.
- [47] C.R. Rao (1973): *Linear Statistical Inference and Its Applications*, 2nd ed. J. Wiley, New York.
- [48] C.R. Rao (1975): Simultaneous estimation of parameters in different linear models and applications to biometric problems. *Biometrics*, 31, 545-553.
- [49] C.R. Rao (1976): Estimation of parameters in the linear model, *Ann. Statist*, 4, 1023-1037.
- [50] C.R. Rao - S.K. Mitra (1971): *Generalized Inverse of Matrices and Its Applications*, J. Wiley, New York.
- [51] J.F. Richard - H. Tompa (1980): On the evaluation of Poly-t density functions, *Journal of Econometrics*, 12, 335-351.
- [52] L.J. Savage (1959): La probabilità soggettiva nei problemi pratici della statistica. In corso C.I.M.E. su Induzione e Statistica, Varenna, Cremonese, Roma.
- [53] L.J. Savage (1962): *Bayesian Statistics. In Decision and Information Process*. Mac Millan, New York,

- [54] A.F.M. Smith (1973a): Bayes estimates in one-way and two-way models.
Biometrika, 60, 319-329.
- [55] A.F.M. Smith (1973b): A general Bayesian linear model, *J. Roy. Statist. Soc. Ser. B*, 35, 67-75.
- [56] C. Stein (1956): Inadmissibility of the usual estimator for the mean of a multivariate normal distribution.
Proc. 3rd. Berkeley Symp. on Math. Statist and Prob. I, 197-206, University of California Press, Berkeley.
- [57] C. Stein (1962): Confidence sets for the mean of a multivariate normal distribution. *J. Roy Statist. Soc. Ser. B*, 24, 265-296.
- [58] G.C. Tiao - A. Zellner (1964): Bayes's theorem and the use of prior knowledge in regression analysis, *Biometrika*, 51, 219-230.
- [59] S. Wilks (1962): *Mathematical Statistics*, J. Wiley, New York.
- [60] A. Zellner (1971): *An Introduction to Bayesian Inference in Econometrics*, J. Wiley, New York.
- [61] A. Zellner - W. Vandaele (1974): Bayes-Stein estimators for k-means regression and simultaneous equation models.
In S.E. Fienberg - A. Zellner (eds), *Studies in Bayesian Econometrics and Statistics*, 627-651, North-Holland.

