

8. Sull'econometria bayesiana

di Attilio Wedlin

1. Introduzione

Sull'econometria bayesiana molto è già stato scritto per sottolineare la validità della sua impostazione logica e la superiore qualità dei suoi risultati rispetto a quelli ottenibili con procedimenti non bayesiani. Molto lavoro è stato fatto e si continua a fare da parte degli statistici ed econometrici bayesiani per fornire agli utilizzatori modelli operativi e procedure praticabili nella forma di programmi per computer. Tra i molti scritti in proposito mi limito a citare i capitoli 2 e 9 in «Handbook of Econometrics» (vol. I) dovuti ad A. Zellner e rispettivamente a J. H. Drèze e J. F. Richard e gli articoli di A. Zellner [1985a; 1985b].

È mia opinione che su taluni aspetti dell'impostazione bayesiana della statistica e dell'econometria l'economista-utilizzatore non dovrebbe sollevare sostanziali obiezioni; per esempio

1) sul significato di probabilità quale «grado soggettivo di fiducia sulla verità di un evento» piuttosto che quale «frequenza asintotica di successo»;

2) sulla teoria delle probabilità intesa quale «logica dell'incertezza»;

3) sulla natura «aleatoria» di ogni numero (o variabile) non noto per mancanza di informazione;

4) sull'impiego a fini inferenziali e/o decisionali di ogni informazione disponibile, di natura campionaria o extracampionaria (ad esempio quella derivante dalla teoria economica);

5) sull'utilizzazione del teorema di Bayes per l'adeguamento delle valutazioni probabilistiche agli incrementi di informazione che si rendono via via disponibili.

Per contro, su qualche altro aspetto dell'impostazione bayesiana possono sorgere, a mio avviso, legittime perplessità ed è proprio su questi punti che vorrei soffermarmi a riflettere.

Il primo problema riguarda la necessità di esprimere le informazioni extracampionarie disponibili, che solitamente si riferiscono a variabili economiche osservabili, mediante valutazioni probabilistiche sull'insieme dei parametri (non osservabili) del modello formale adottato. Molto spesso manca un legame significativo tra parametri e va-

riabili cosicché la realizzazione di tale obiettivo può incontrare gravi difficoltà. Giova, a questo proposito, riflettere che il passaggio da una distribuzione iniziale sullo spazio dei parametri alla distribuzione finale relativa all'incremento di informazione ottenuto corrisponde al passaggio, per il processo generatore di dati, dalla legge temporale iniziale a quella finale. Come vedremo, quest'ultimo passaggio può effettuarsi, indipendentemente dalle distribuzioni sui parametri, con l'applicazione del teorema delle probabilità composte alle distribuzioni congiunte della legge temporale iniziale [Daboni e Wedlin 1982].

Il secondo problema riguarda la necessità, imposta dall'uso del teorema di Bayes, di lavorare con distribuzioni di probabilità (sull'insieme dei suddetti parametri) in qualunque stato di informazione ci si trovi, e quindi anche quando esso sia molto povero. Si osservi che, con riferimento ad un solo numero aleatorio X , la funzione di ripartizione $F_X(x)$ ne dà la conoscenza completa (almeno in ipotesi di additività numerabile della probabilità); se le informazioni disponibili sono scarse, il soggetto potrebbe essere in grado di esprimere la sua opinione soltanto su alcuni momenti di $F_X(x)$, tipicamente sul valor medio e talvolta sulla varianza, oppure mediante valutazioni del tipo $P(a < X \leq b) < c$ (a, b, c fissati) o altre valutazioni incomplete: nel seguito, per questi casi, parleremo di «valutazioni probabilistiche parziali». È mia opinione che in tali situazioni il soggetto non debba essere forzato ad esprimere comunque la sua personale opinione su un numero aleatorio, o peggio ancora su un insieme di numeri aleatori, mediante una distribuzione di probabilità. Ritengo che una delle caratteristiche salienti dell'impostazione bayesiana consista nella fondamentale libertà dei soggetti di scegliere lo schema formale con il quale realizzare la descrizione dell'incertezza e gli strumenti di espressione della loro personale opinione, avendo come unico vincolo obiettivo il rispetto delle norme di coerenza.

Da questo punto di vista, ho il sospetto che la maggior parte degli schemi operativi elaborati dall'econometria bayesiana possa apparire all'economista-utilizzatore un insieme di «ad hoceries» non molto dissimili da quelli messi a punto dai non-bayesiani.

D'altra parte si obietterà: senza teorema di Bayes non c'è econometria bayesiana e senza distribuzioni di probabilità non si utilizza il teorema di Bayes. Si può rispondere che ciò è solo parzialmente vero: espressa comunque un'opinione, anche mediante una valutazione parziale, su un insieme di numeri aleatori è possibile verificarne la compatibilità intrinseca ed è anche possibile stabilire se ulteriori valutazioni, relative anche a stati di informazione più ampi, riescano coerenti con la prima. Lo strumento matematico che lo consente è un risultato che B. de Finetti ha denominato *teorema fondamentale per le probabilità e le previsioni*.

Il presente saggio si articola come segue: nel paragrafo 2 viene illustrata, su un semplice esempio, l'interpretazione dell'inferenza bayesiana quale trasformazione della legge temporale del processo generatore di dati: trasformazione che denomineremo «procedimento induttivo diretto» in contrapposizione a quello «indiretto» consistente nel passaggio dalla distribuzione parametrica iniziale a quella finale mediante l'applicazione del teorema di Bayes. Nel paragrafo 3, dopo pochi cenni sull'impostazione della teoria delle probabilità secondo B. de Finetti, viene presentato il teorema fondamentale delle probabilità sottolineandone la portata operativa. In particolare, si affronta il problema inferenziale riguardante la sola speranza matematica di un numero aleatorio nell'ipotesi che lo stato di informazione disponibile consenta al soggetto di esprimere la sua opinione soltanto attraverso una valutazione parziale.

2. Procedimento inferenziale bayesiano diretto

Prenderemo le mosse per le nostre considerazioni dall'analisi bayesiana dei sistemi statici ad equazioni simultanee seguendo l'impostazione e la simbologia di [Drèze e Richard 1983], rinviando a tale monografia per tutte le necessarie precisazioni tecniche.

La forma strutturale per un sistema del tipo suddetto può essere espressa secondo le

$$[1] \quad \begin{matrix} (T, m) \\ Y \end{matrix} \cdot \begin{matrix} (m, m) \\ B \end{matrix} + \begin{matrix} (T, n) \\ Z \end{matrix} \cdot \begin{matrix} (n, m) \\ \Gamma \end{matrix} = \begin{matrix} (T, m) \\ U \end{matrix}$$

$$[2] \quad p(U) \sim f_{MN}(0; \Sigma \otimes I_T)$$

ove nella [1] si sono indicate entro parentesi le dimensioni delle matrici Y e Z , relative alle variabili endogene e rispettivamente esogene, e delle matrici parametriche non note B e Γ , oltre che della matrice U dei disturbi non osservabili. Chiaramente, le righe t -esime ($t = 1, \dots, T$) di Y e Z contengono le osservazioni statistiche sulle variabili dei due tipi nel periodo t -esimo.

La [2] assegna alla matrice dei disturbi U una distribuzione di probabilità congiunta con densità normale matriciale di parametri 0 e $\Sigma \otimes I_T$, essendo generalmente non nota la matrice di dispersione Σ .

Supponendo invertibile la matrice B è possibile considerare la forma ridotta del modello costituito dalle [1] e [2] espressa dalle

$$[3] \quad Y = Z \cdot (-\Gamma B^{-1}) + U \cdot B^{-1} = Z \cdot \Pi + V$$

$$[4] \quad p(V) \sim f_{MN}(0; \Omega \otimes I_T)$$

ove la matrice Ω è definita dall'uguaglianza $\Omega = (B')^{-1} \cdot \Sigma \cdot B^{-1}$.

La [3] costituisce un modello di regressione multipla multivariata in quanto è $Z \cdot \Pi = E(Y/Z, \Pi, \Omega)$; inoltre la [4] implica che Y abbia la seguente distribuzione congiunta, subordinata alla generica ipotesi sulle determinazioni di Z, Π e Ω :

$$[5] \quad p(Y/Z, \Pi, \Omega) \sim f_{MN}(Z \cdot \Pi; \Omega \otimes I_T).$$

Supponendo che la caratteristica della matrice Z sia uguale a n e che sia $n+m < T$ si ha che la verosimiglianza di Π ed Ω , relativa alle osservazioni disponibili Y e Z , implicata dalla [5] è espressa dalla

$$[6] \quad l(\Pi, \Omega/Y, Z) \propto (\det \Omega)^{-T/2} \cdot \exp \left\{ -\frac{tr \Omega^{-1}}{2} \cdot [(\Pi - \hat{\Pi})' \cdot Z'Z \cdot (\Pi - \hat{\Pi}) + W] \right\},$$

avendo posto

$$\hat{\Pi} = (Z'Z)^{-1} \cdot Z' \cdot Y,$$

$$W = Y' \cdot [I - Z(Z'Z)^{-1}Z'] \cdot Y.$$

Il procedimento induttivo bayesiano sulle matrici parametriche incognite Π e Ω consiste, com'è noto, nell'assegnare a tali matrici una distribuzione congiunta iniziale di densità $p_o(\Pi, \Omega)$ e nel determinare la corrispondente densità finale mediante il teorema di Bayes:

$$[7] \quad p(\Pi, \Omega/Y, Z) \propto p_o(\Pi, \Omega) \cdot l(\Pi, \Omega/Y, Z).$$

È anche noto che se $p_o(\Pi, \Omega)$ viene scelta entro la famiglia Normale-Wishart inversa, accade che la densità finale appartiene ancora a questa famiglia di distribuzioni che risulta quindi «chiusa» rispetto alla trasformazione [7] o, in altri termini, autoriproduttore. È bene sottolineare che la nozione di famiglie di distribuzioni autoriproduttrici, sebbene rivesta un'importanza notevole nell'applicazione degli schemi bayesiani ai problemi concreti, non è altrettanto importante dal punto di vista concettuale: il punto essenziale è la adeguatezza o meno della distribuzione iniziale a rappresentare l'effettiva opinione dell'operatore. L'enfasi che molti utilizzatori pongono nell'uso delle distribuzioni autoriproduttrici e, in situazioni di informazione molto scarsa, delle distribuzioni «non informative» ci sembra eccessiva e potenzialmente pericolosa in quanto può indurre a sopravvalutare la convenienza matematica a scapito di aspetti più sostanziali, come avviene nel cosiddetto approccio «bayesiano empirico».

Con riferimento alla [7], e più precisamente alla densità $p_o(\Pi, \Omega)$ che vi compare, intendiamo soffermarci su due considerazioni critiche: la prima, che sarà sviluppata in questo paragrafo, concerne l'apparente inevitabilità di compendiare le informazioni disponibili anteriormente all'osservazione statistica (Y, Z) in una distribuzione di probabilità riguardante i parametri incogniti non osservabili Π e Ω piuttosto che gli elementi osservabili del fenomeno studiato sui quali l'operatore ha un'esperienza diretta. La seconda considerazione, che tratteremo nel prossimo paragrafo, concerne la presunta necessità di esprimere le informazioni iniziali, qualunque sia la loro consistenza, mediante un complesso di valutazioni organico (cioè consistente e completo) costituente una distribuzione di probabilità anche quando l'operatore fosse disponibile ad esprimere soltanto delle valutazioni parziali, riguardanti per esempio solo alcuni momenti di quella distribuzione.

Ritornando alla prima questione, e cioè alla scelta di una densità iniziale per i parametri incogniti del modello adottato, un possibile approccio decisionale potrebbe essere il seguente: poiché su quei parametri non osservabili è difficile acquisire informazioni dirette atte a guidare la scelta di $p_o(\Pi, \Omega)$ ci si potrebbe chiedere qual è o quali sono le densità iniziali esprimenti uno stato di conoscenza corrispondente a quello che possediamo sulle variabili osservabili del modello. Si osservi che ad una fissata densità iniziale $p_o(\Pi, \Omega)$ per i parametri Π e Ω corrisponde, per coerenza, la densità

$$[8] \quad p_o(Y/Z) = \int p(Y/Z, \Pi, \Omega) \cdot p_o(\Pi, \Omega) d\Pi d\Omega$$

per le variabili osservabili; denomineremo «legge temporale iniziale» il complesso di distribuzioni di probabilità per le variabili osservabili legate alla $p_o(\Pi, \Omega)$ dalla relazione [8]. Allora la validità della scelta di $p_o(\Pi, \Omega)$ può essere giudicata dalla misura in cui la corrispondente densità $p_o(Y/Z)$ traduce le nostre informazioni iniziali sulle variabili osservabili.

Alcuni autori sottolineano il fatto che in presenza di una consistente informazione campionaria può risultare sufficiente anche una scelta approssimativa della $p_o(\Pi, \Omega)$ in quanto la densità finale $p(\Pi, \Omega/Y, Z)$ riesce, in tali condizioni, sostanzialmente determinata dalla funzione di verosimiglianza: pur concordando su questa considerazione, non riteniamo che essa sia decisiva per la risoluzione del problema sollevato.

Il procedimento inferenziale diretto, sul quale vogliamo attirare l'attenzione del lettore, si basa sulla semplice osservazione che la trasformazione $p_o(\Pi, \Omega) \rightarrow p(\Pi, \Omega/Y, Z)$ ottenuta mediante l'applicazione del teorema di Bayes corrisponde alla trasformazione della legge temporale iniziale per il processo generatore di dati (Y_t/Z_t) , $t \geq 1$,

nella legge temporale finale per il processo di osservazione residuo $(Y_t/Z_t)_t$, con $t > T$. Quest'ultima trasformazione si realizza analiticamente con l'applicazione del teorema delle probabilità composte alle distribuzioni della legge temporale iniziale, come si vedrà tra breve.

In altri termini, il procedimento inferenziale bayesiano può applicarsi direttamente alle distribuzioni di probabilità delle variabili osservabili prescindendo dalle distribuzioni dei parametri non osservabili, almeno nell'ipotesi che non siano proprio queste ultime l'oggetto principale dell'inferenza. Riteniamo di qualche utilità illustrare questo approccio su un esempio particolarmente semplice e cioè sul primo dei modelli aggregati analizzati da T. Haavelmo nell'ormai storico articolo *Methods of Measuring the Marginal Propensity to Consume*. L'analisi bayesiana di tale modello è stata effettuata da T. J. Rothenberg [1963] e V. K. Chetty [1968]; la trattazione che segue si scosta da quelle degli autori citati per la diversa scelta dei parametri subordinanti e per l'uso di una distribuzione iniziale informativa.

Il modello consiste delle due equazioni ben note

$$\begin{cases} C_t = \alpha Y_t + \beta + U_t \\ Y_t = C_t + Z_t \end{cases} \quad \begin{array}{l} C_t = \text{consumo aggregato} \\ Y_t = \text{reddito disponibile} \\ Z_t = \text{spesa non per consumi} \end{array}$$

e dall'assunzione che, subordinatamente ad ogni realizzazione del processo $(Z_t)_t$, il processo stocastico delle perturbazioni aleatorie $(U_t)_t$ sia normale con funzione valor medio identicamente nulla e funzione di covarianza $\text{Cov}(U_{t+h}, U_t) = v \cdot \vartheta_{o,h}$, essendo $\vartheta_{o,h}$ il simbolo di Kroenecker.

Indichiamo le equazioni della forma ridotta con le

$$[9] \quad Y_t = a \cdot Z_t + b + E_t,$$

$$C_t = a^* \cdot Z_t + b + E_t,$$

ove $a = (1 - \alpha)^{-1}$ è il moltiplicatore del reddito disponibile,

$$b = a \cdot \beta, \quad a^* = a \cdot \alpha \quad \text{ed} \quad E_t = a \cdot U_t.$$

Il processo stocastico $(Y_t/Z_t)_t$, come anche quello $(C_t/Z_t)_t$, è costituito da numeri aleatori che, subordinatamente ad ogni possibile determinazione di a , b ed $s = \text{Var}(E_t/Z_t) = a^2 \cdot v$, risultano indipendenti e dotati della stessa distribuzione con densità $f_{Y/Z}(y_t/z_t, a, b, s) \sim N(az_t + b; s)$.

Ne scende che la funzione di verosimiglianza dei parametri aleatori a , b ed s relativa all'osservazione statistica (y, z) è espressa dalla

$$[10] \quad l(a, b, s/y, z) = \prod_{i=1}^T f_{Y/Z}(y_i/z_i, a, b, s) = \\ = (2\pi s)^{-T/2} \cdot \exp\{-[\nu k + (\varphi - \hat{\varphi})' Z' Z (\varphi - \hat{\varphi})]/2s\}$$

ove si è posto $\varphi = (a, b)'$, $Z = (1, z)$ essendo $I = (1, 1, \dots, 1)'$, ed inoltre

$$\hat{\varphi} = (Z'Z)^{-1} \cdot Z' \cdot y,$$

$$k = \nu^{-1} \cdot (y - Z \cdot \hat{\varphi})' \cdot (y - Z \cdot \hat{\varphi}),$$

$$\nu = T - 2.$$

L'espressione analitica della [10] rivela che la famiglia coniugata naturale di distribuzioni per la terna di parametri aleatori (a, b, s) è costituita dalle distribuzioni del tipo «normale bivariata - gamma inversa» di densità

$$[11] \quad g(a, b, s) = g(\varphi, s) \propto s^{-(\nu+2)/2-1} \cdot \\ \cdot \exp\{-[\gamma\vartheta + (\varphi - \mu)' \cdot H^{-1} \cdot (\varphi - \mu)]/2s\}$$

Indicando con $g_o(a, b, s)$ la densità della distribuzione iniziale e con $\gamma_o, \vartheta_o, \mu_o$ ed H_o i suoi parametri, quella della distribuzione finale, data dalla

$$[12] \quad g_T(a, b, s/y, z) \propto g_o(a, b, s) \cdot l(a, b, s/y, z),$$

è ancora dello stesso tipo funzionale ed i suoi parametri sono [si veda Daboni, Wedlin 1982, 144 e Zellner 1971, 71]:

$$H_T = (H_o^{-1} + Z'Z)^{-1}$$

$$\mu_T = H_T \cdot (H_o^{-1} \cdot \mu_o + Z' \cdot y)$$

$$\gamma_T = \gamma_o + T$$

$$\vartheta_T = (\gamma_o \vartheta_o + \mu_o' H_o^{-1} \mu_o + y'y - \mu_T' H_T^{-1} \mu_T) / \gamma_T.$$

Se, per brevità, denotiamo con $g_i(a, b, s)$, $i=0, T$, la densità relativa allo stato di informazione iniziale ($i=0$) oppure a quello finale ($i=T$) allora le corrispondenti distribuzioni congiunte appartenenti alla legge temporale del processo (Y_i/Z_i) , si determinano mediante le (cfr. la formula [8]):

$$[13] \quad f_i(y_{h+1}, \dots, y_{h+n}/z_{h+1}, \dots, z_{h+n}) = \int f(y/z, a, b, s) \cdot g_i(a, b, s) da db ds = \int \prod_{j=h+1}^{h+n} f(y_j/z_j, a, b, s) \cdot g_i(a, b, s) da db ds$$

essendo h un intero non negativo arbitrario se $i=0$ e invece $h \geq T$ se $i=T$. Queste distribuzioni risultano essere di tipo Student t multivariate con densità congiunta data, per $i=0, T$, da [si veda Daboni e Wedlin 1982, 151 e Zellner 1971]:

$$[14] \quad f_i(y/z) \propto [\gamma_i + (y - Z\mu_i)'K_i(y - Z\mu_i)]^{-(\gamma_i+n)/2},$$

ove

$$K_i = \vartheta_i^{-1} \cdot [I - Z \cdot (H_i^{-1} + Z'Z) \cdot Z'].$$

Enunciamo ora la seguente

PROPOSIZIONE 1: indicata con L_0 la legge temporale iniziale costituita dalla famiglia compatibile di densità congiunte finite-dimensionali $f_0(y/z)$, quella L_T corrispondente all'osservazione statistica $(y^*, z^*) = ((y_1^*, \dots, y_T^*)', (z_1^*, \dots, z_T^*)')$ è costituita dalle densità congiunte $f_T(y_{T+1}, \dots, y_{T+n}/z_{T+1}, \dots, z_{T+n})$ espresse dai rapporti

$$\frac{f_0(y_1^*, \dots, y_T^*, y_{T+1}, \dots, y_{T+n}/z_1^*, \dots, z_T^*, z_{T+1}, \dots, z_{T+n})}{f_0(y_1^*, \dots, y_T^*/z_1^*, \dots, z_T^*)}$$

di densità congiunte appartenenti alla legge temporale L_0 .

La dimostrazione si consegue facilmente sostituendo nella espressione della densità

$$f_T(y/z) = \int [\prod_{j=T+1}^{T+n} f(y_j/z_j, a, b, s)] \cdot g_T(a, b, s) da db ds$$

il fattore $g_T(a, b, s)$ con la sua espressione desunta dall'applicazione del teorema di Bayes.

Pertanto, se il modello stocastico costituito dalle equazioni di regressione lineare [9], dall'ipotesi sulla normalità della densità subordinata $f_{Y/Z}(y_i/z_i, a, b, s)$ e dall'assunzione [11] per la densità iniziale $g_0(a, b, s)$ venisse riformulato equivalentemente in termini delle densità [14] di L_0 , la inferenza statistica relativa all'incremento di informazione (y^*, z^*) potrebbe realizzarsi mediante il procedimento bayesiano indicato nella Proposizione 1 che fa intervenire la norma di coerenza nella forma del

teorema delle probabilità composte anziché in quella del teorema di Bayes. Tale procedimento, che può qualificarsi «diretto» in contrapposizione a quello «indiretto» costituito dalla trasformazione $g_0(a, b, s) \rightarrow g_T(a, b, s/y^*, z^*)$ realizzata dalla [12], è sempre proponibile da un punto di vista teorico, qualunque sia la struttura probabilistica del processo generatore di dati. Quando tale struttura è scambiabile esso riesce particolarmente agevole; sussiste infatti la seguente:

PROPOSIZIONE 2: se il processo generatore di dati è scambiabile e se nell'approccio indiretto esiste una famiglia coniugata della verosimiglianza, allora L_0 è costituita da distribuzioni congiunte la cui forma funzionale è invariante nel passaggio alle corrispondenti distribuzioni subordinate.

Per la dimostrazione basta osservare che se la distribuzione sullo spazio dei parametri subordinanti è autoriproduttrice le distribuzioni della legge temporale corrispondente che si ottengono, per l'ipotesi di scambiabilità, attraverso la mistura [13] devono necessariamente avere la stessa forma funzionale per ogni i .

3. Valutazioni probabilistiche parziali

La seconda considerazione critica che vogliamo fare a partire dalla [7] concerne la presunta necessità di esprimere le informazioni extracampionarie, qualunque sia la loro consistenza, mediante un complesso di valutazioni probabilistiche costituente una distribuzione di probabilità, appunto la $p_0(\Pi, \Omega)$.

La nostra opinione è che l'operatore possa condurre a buon fine, e in modo convinto, tale operazione soltanto quando l'ammontare di quelle informazioni sia rilevante. Si pensi, per esempio, che l'assegnazione ad un numero aleatorio X di una distribuzione di probabilità, di cui indichiamo con $F(x)$ la funzione di ripartizione, corrisponde all'assegnazione degli infiniti momenti $E(X^n)$, $n \geq 1$. Ora è invece verosimile che quando l'informazione disponibile su X è scarsa si possa avere una qualche idea solo sul valore di alcuni suoi momenti, per esempio su quelli del valor medio $E(X)$ e del momento secondo $E(X^2)$; oppure potremmo, per esempio, essere convinti soltanto del sussistere delle disuguaglianze $0 \leq E(X) \leq 5$ e $E(X^2) < 12$.

In queste o in altre simili circostanze non ci sembra di nessuna utilità forzare l'interessato a «completare» comunque la sua valutazione allo scopo di specificare la richiesta distribuzione $F(x)$. E se invece di un solo numero o parametro aleatorio occorre formulare una valutazione per più numeri o parametri aleatori simultaneamente, come nel caso della $p_0(\Pi, \Omega)$, il suddetto problema si complica tremendamente.

In questo paragrafo intendiamo soffermarci a riflettere sulla possibilità di utilizzare correttamente, a scopo inferenziale, valutazioni probabilistiche «parziali», cioè valutazioni incomplete rispetto a quella esaustiva costituita da una distribuzione di probabilità. Come è stato accennato già nel paragrafo 1, lo strumento teorico necessario è il «teorema fondamentale delle probabilità» di B. de Finetti la cui potenzialità operativa è stata illustrata in Bruno e Giglio [1980] e soprattutto in Lad, Dickey, Rahman [1987]. La nostra esposizione, basata essenzialmente su quest'ultima monografia, richiede qualche cenno sulla logica dell'incertezza secondo B. de Finetti.

L'approccio di B. de Finetti alla probabilità e all'induzione statistica risente fortemente della sua sensibilità e del suo interesse verso le applicazioni della matematica a problemi concreti, particolarmente in campo economico e, più specificamente, in campo economico-assicurativo. Realisticamente, l'impostazione della sua analisi dell'incertezza è di tipo dinamico in quanto

a) considera che lo stato di informazione del soggetto si modifica nel tempo rendendo necessaria una revisione delle valutazioni;

b) considera che gli interessi del soggetto relativi al problema in oggetto possono variare nel tempo rendendo eventualmente necessario un approfondimento dell'analisi.

Quest'ultimo motivo di dinamismo è ben evidente nel seguente passaggio che riportiamo da de Finetti [1937-1938]: «Come il problema generale della logica del certo consiste nel riconoscere per quali eventi e in che modo si può concludere se sono veri o falsi quando si conosca l'esito di n eventi dati $E_1, \dots, E_n$, così per la logica del probabile il problema generale consiste nel riconoscere per le quali eventi e in che modo il valore della probabilità risulta determinato quando tale valore sia assegnato per gli eventi $E_1, \dots, E_n$ ».

Si fa evidentemente riferimento all'eventualità che dopo una prima fase dell'analisi in cui il soggetto si è limitato a considerare i soli eventi $E_1, \dots, E_n$ riguardanti il problema in esame e a formulare le valutazioni di probabilità $P(E_1), \dots, P(E_n)$, sorgano in lui ulteriori curiosità inducendolo ad estendere la sua analisi anche ad altri eventi. Per il caso particolare in cui ai primi n eventi si aggiunga un solo altro evento E il suddetto problema trova la sua soluzione nel seguente risultato [si veda de Finetti 1970, 133]:

Teorema fondamentale delle probabilità. Date le probabilità $P(E_i)$, $i=1, \dots, n$, di un numero finito di eventi, la probabilità di un altro evento E , $P(E)$:

1) risulta determinata se E è linearmente dipendente dagli E_i ;

2) altrimenti può esserle attribuito, coerentemente, un qualunque valore di un intervallo chiuso $[p', p'']$.

Questo teorema costituisce in effetti un caso particolare del seguente risultato noto come

Teorema del prolungamento delle probabilità [si veda de Finetti 1949]. Siano ε ed ε' insiemi di eventi di cardinalità qualunque con $\varepsilon \subset \varepsilon'$ e $P(\cdot)$ una probabilità coerente su ε ; allora esiste sempre una probabilità coerente su ε' la cui restrizione su ε coincide con $P(\cdot)$.

Il problema dell'individuazione concreta dell'intervallo $[p', p'']$ per $P(E)$ nel Teorema fondamentale può essere affrontato operativamente in termini di programmazione lineare [si veda Bruno, Giglio 1980 e Lad, Dickey, Rahman 1987] nel modo seguente. Indicati con $Q_1, \dots, Q_s$ i costituenti possibili della partizione generata dagli eventi $E_1, \dots, E_n, E$ (e si tratterà dei prodotti logici $E'_1 \cdot E'_2 \cdot \dots \cdot E'_n \cdot E'$ non impossibili, ove gli apici indicano che il valore logico di ogni evento è fissato) risulta

$$E_{n+1} = A \cdot Q,$$

o più in dettaglio

$$E_{n+1} = \begin{bmatrix} E_1 \\ \vdots \\ E_n \\ \hline E \end{bmatrix} = \begin{bmatrix} a_{11} & \dots & a_{1s} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{ns} \\ \hline a_{n+1,1} & \dots & a_{n+1,s} \end{bmatrix} \cdot \begin{bmatrix} Q_1 \\ \vdots \\ \vdots \\ Q_s \end{bmatrix},$$

ove gli elementi a_{ij} sono uguali a 0 oppure 1 a seconda che Q_j contenga l'evento E_i affermato oppure negato e ove $s \leq 2^n$.

Passando alle probabilità corrispondenti si ha, per la linearità di $P(\cdot)$:

$$\begin{bmatrix} P(E_1) \\ \vdots \\ P(E_n) \\ P(E) \end{bmatrix} = A \cdot \begin{bmatrix} P(Q_1) \\ \vdots \\ \vdots \\ P(Q_s) \end{bmatrix}$$

cosicché, assegnate che siano le $P(E_i)$, $i=1, \dots, n$, le prime n uguaglianze, assieme alla condizione $P(Q_1) + \dots + P(Q_s) = 1$ e alle $P(Q_j) \geq 0$, $j = 1, \dots, s$, costituiscono l'insieme dei vincoli del problema di programmazione lineare. L'obiettivo consiste invece nella minimizzazione o nella massimizzazione (per l'individuazione di p' e, rispettivamente, p'') della $P(E) = \sum_{j=1}^s a_{n+1,j} \cdot P(Q_j)$.

Valutazioni $P(E)$ che soddisfano la $p' \leq P(E) \leq p''$ sono tutte ammissibili per l'evento E in relazione alle valutazioni $P(E_1), \dots, P(E_n)$; come caso particolare può essere $p'=p''$ (se E dipende linear-

mente degli E_i) mentre invece se il problema di programmazione lineare non ha soluzione per $P(E)$ ciò significa che già la valutazione delle probabilità sugli E_i non è coerente e perciò va riformulata più accuratamente.

Passando da eventi a numeri aleatori, la nozione di probabilità viene sostituita, nell'impostazione di B. de Finetti, da quella di *previsione*; diciamo subito che per numeri aleatori limitati X la previsione $P(X)$ coincide con la speranza matematica o valor medio $E(X)$, mentre invece se X ha rango illimitato il valor medio è soltanto una delle determinazioni possibili per la previsione $P(X)$ [si veda per esempio de Finetti 1970]. Naturalmente nelle applicazioni l'ipotesi che i numeri aleatori siano limitati non impone alcuna restrizione. I problemi di prolungamento delineati per gli eventi si ripropongono anche per i numeri aleatori allo stesso modo. Enunciamo quindi, rinviando a Holzer [1985] e a de Finetti [1970] per ulteriori approfondimenti, il

Teorema fondamentale della previsione. Date le previsioni $E(X_i)$, $i=1, \dots, n$, di un numero finito di numeri aleatori limitati, la previsione di un altro numero aleatorio X , $E(X)$:

- 1) risulta determinata se X è linearmente dipendente dagli X_i ;
- 2) altrimenti può esserle attribuito, coerentemente, un qualunque valore di un intervallo chiuso $[p', p'']$.

Ci limitiamo ad affermare che, come per le probabilità, esiste un teorema di prolungamento per le previsioni del tutto analogo, rispetto al quale il precedente risultato è un caso particolare; esiste infine un teorema di prolungamento per le previsioni subordinate [per il quale rinviamo a Holzer 1985] che richiameremo nel seguito.

Anche in questo caso, per la linearità della previsione $E(\cdot)$, gli estremi p' e p'' dell'intervallo di ammissibilità per $E(X)$ si possono operativamente determinare risolvendo un problema di programmazione lineare analogo a quello già descritto *a condizione però che tutti i numeri aleatori considerati abbiano un numero finito di determinazioni possibili*.

Accettando tale ipotesi, che nelle applicazioni è del tutto naturale, si può scrivere:

$$X_{n+1} = \begin{bmatrix} X_n \\ X \end{bmatrix} = \begin{bmatrix} C_{n,s} \\ c_{n+1} \end{bmatrix} \cdot Q_s$$

e passando alle previsioni (o valori medi)

$$\begin{bmatrix} E(X_n) \\ E(X) \end{bmatrix} = \begin{bmatrix} C_{n,s} \\ c_{n+1} \end{bmatrix} \cdot P(Q_s)$$

ove ora Q_s contiene i costituenti possibili della partizione prodotto generata dai numeri aleatori $X_1, \dots, X_n, X$ (e cioè tutti gli s prodotti

logici non impossibili riguardanti i valori degli $n+1$ numeri aleatori considerati) mentre la matrice $C_{n,s}$ ed il vettore riga c_{n+1} hanno come elementi i valori possibili dei numeri aleatori corrispondenti.

L'insieme dei vincoli del problema di programmazione lineare è costituito dalle equazioni $E(X_n) = C_{n,s} \cdot P(Q_s)$ oltre che dalle $P(Q_1) + \dots + P(Q_s) = 1$ e $P(Q_j) \geq 0, j=1, \dots, s$; l'obiettivo consiste invece nella minimizzazione o nella massimizzazione della funzione lineare delle $P(Q_j)$ espressa dalla $E(X) = c_{n+1} \cdot P(Q_s)$.

Valutazioni $E(X)$ che soddisfano $p' \leq E(X) \leq p''$ sono tutte ammissibili per il numero aleatorio X in relazione alle valutazioni $E(X_1), \dots, E(X_n)$.

Fin qui i dati del problema erano costituiti da valutazioni numeriche precise $E(X_i), i = 1, \dots, n$, ma, come si è già accennato, può accadere che il soggetto non sia in grado di esprimere la sua opinione sugli X_i in un modo così puntuale. In base all'informazione in suo possesso egli potrebbe trovarsi nella condizione di affermare soltanto che è $E(X_i) \leq E(X_j)$ per qualche coppia di valori dell'indice i, j , oppure che è $a_i \leq E(X_i) \leq b_i$, con a_i e b_i fissati, per alcuni valori dell'indice i , o ancora che è $a_1 E(X_1) + a_2 E(X_2) = b \dots$ sostituendo cioè alle valutazioni puntuali $E(X_i), i = 2, \dots, n$, valutazioni meno precise, o sfumate, o parziali, del tipo indicato o di altro ancora.

Si pone naturalmente il problema se sia possibile e come utilizzare razionalmente tali opinioni sfumate sugli X_i allo scopo di determinare per $E(X)$ l'intervallo ammissibile $[p', p'']$ corrispondente. La possibilità teorica di realizzare tale obiettivo è garantita dal teorema di prolungamento dianzi citato, qualunque sia il modo in cui vengono espresse le valutazioni. Limitando invece opportunamente queste modalità riesce possibile determinare concretamente l'intervallo $[p', p'']$. Si osservi a tal fine che la disuguaglianza $E(X_i) \leq E(X_j)$ equivale al vincolo lineare $E(X_i) - E(X_j) \leq 0$; che la doppia limitazione $a_i \leq E(X_i) \leq b_i$ corrisponde ovviamente alla coppia di condizioni $-E(X_i) \leq -a_i$ e $E(X_i) \leq b_i$; che l'uguaglianza $a_1 E(X_1) + a_2 E(X_2) = b$ equivale alle due disuguaglianze $a_1 E(X_1) + a_2 E(X_2) \leq b$ e $-a_1 E(X_1) - a_2 E(X_2) \leq -b$.

Si può allora dimostrare che [si veda Lad, Dickey, Rahman 1987] assumendo che le valutazioni parziali del soggetto sugli X_i possano esprimersi mediante disuguaglianze lineari con coefficienti noti del tipo

$$\sum_{k=1}^K d_{ki} \cdot E(X_i) \leq d_k, \quad k=1, \dots, K$$

l'individuazione dell'intervallo corrispondente $[p', p'']$ per $E(X)$ può essere ottenuta risolvendo il problema di programmazione lineare costituito dal sistema di vincoli

$$[15] \quad D \cdot E(X_n) = D \cdot C_{n,s} \cdot P(Q_s) \leq d,$$

che compendia le [14], oltre alle condizioni $\sum_{j=1}^s P(Q_j) = 1$ e $P(Q_j) \geq 0$, ($j=1, \dots, s$) e dalla funzione obiettivo $c_{n+1} \cdot P(Q_s)$.

Tutte le considerazioni svolte finora si riferivano a valutazioni coerenti fatte dal soggetto in un medesimo stato di conoscenza. Se quest'ultimo subisce una variazione perché si rendono disponibili nuove informazioni, costituite per esempio da un evento H che si sa, o si suppone, essere vero, si devono affrontare problemi di coerenza di tipo diverso da quelli già esaminati: problemi di coerenza induttiva che, in presenza di valutazioni espresse da una distribuzione di probabilità, vengono affrontati mediante la norma di coerenza espressa dai teoremi di Bayes o delle probabilità composte.

La questione che ora intendiamo affrontare concerne invece l'uso di valutazioni probabilistiche parziali in procedimenti inferenziali; ci limiteremo ancora a considerare numeri aleatori con un insieme finito di determinazioni possibili. Presentiamo innanzitutto la trattazione di Lad, Dickey e Rahman [1987] del problema della determinazione dei valori ammissibili per il valor medio subordinato $E(X_{n+1}/H)$ relativamente ad una valutazione probabilistica parziale concernente i numeri aleatori $X_1, \dots, X_n$. Commenteremo nel seguito alcuni aspetti di questo problema inferenziale.

Si supponga che il soggetto abbia espresso una sua valutazione probabilistica iniziale sui numeri aleatori del vettore X_n mediante il sistema di disuguaglianze [15] e che successivamente sia stato informato della verità dell'evento H connesso in qualche modo al problema studiato, e conseguentemente ai numeri aleatori $X_1, \dots, X_n, X_{n+1}$. Se indichiamo con Q_r^* il vettore formato dai costituenti possibili della partizione prodotto generata dagli elementi di X_n e dai tre ulteriori numeri aleatori X_{n+1} , $|H|$ e $X_{n+1} \cdot |H|$, ove si è denotato con $|H|$ l'indicatore dell'evento H che costituisce l'incremento di informazione, poniamo che sia

$$\begin{bmatrix} X_{n+1} \\ |H| \\ X_{n+1} \cdot |H| \end{bmatrix} = \begin{bmatrix} c_{n+1} \\ c_{n+2} \\ c_{n+3} \end{bmatrix} \cdot Q_r^*$$

per cui, passando ai valori medi, si ottiene

$$\begin{bmatrix} E(X_{n+1}) \\ E(|H|) \\ E(X_{n+1} \cdot |H|) \end{bmatrix} = \begin{bmatrix} c_{n+1} \\ c_{n+2} \\ c_{n+3} \end{bmatrix} \cdot P(Q_r^*).$$

Ovviamente gli elementi del vettore Q_r della [15] sono unioni (somme logiche) opportune degli elementi del più ampio vettore Q_r^* di modo che la [15] può essere riformulata in termini del nuovo vettore di costituenti possibili Q_r^* secondo la

$$[15'] \quad D \cdot E(X_n) = D \cdot C_{n,r}^* \cdot P(Q_r^*) \leq d^*.$$

Infine, si ha:

$$[16] \quad E(X_{n+1}/H) = E(X_{n+1} \cdot |H|) / E(|H|) = c_{n+3} P(Q_r^*) / c_{n+2} P(Q_r^*).$$

Con riferimento allo schema ora illustrato siamo interessati ad individuare, al variare di $P(Q_r^*)$ entro l'insieme convesso determinato dai vincoli [15'], oltre che da quelli soliti sulle probabilità $P(Q_r^*)$, i valori minimo e massimo per il rapporto [16]. L'esistenza di un intervallo non vuoto di valori possibili per $E(X_{n+1}/H)$ è stabilita da un teorema di prolungamento per previsioni subordinate a cui si è già accennato e del quale ci limitiamo a riportare [Holzer 1985] il solo enunciato: «Se D e D_1 indicano insiemi di numeri aleatori subordinati X/H , essendo X un numero aleatorio limitato ed H un evento possibile, con $D \subset D_1$, e se $E(\cdot)$ indica una previsione subordinata coerente su D allora esiste su D_1 una previsione subordinata coerente, estensione di $E(\cdot)$ ».

Nel caso che sia $P(H) > 0$ la determinazione concreta dell'intervallo può effettuarsi risolvendo il problema di programmazione *non* lineare costituito dai vincoli lineari espressi dalla [15'], dalle solite condizioni sulle $P(Q_r^*)$ e da una funzione obiettivo non lineare data dall'ultimo membro della [16]. Ci limitiamo ad affermare che questo problema non lineare può essere riconvertito in un equivalente problema di programmazione lineare in cui la funzione obiettivo assume la forma $c_{n+3} \cdot P(Q_r^*) - g[c_{n+2} \cdot P(Q_r^*)]$, ove il parametro g è tale da rendere nulla la differenza nel punto di ottimo.

Si osservi che nel caso ora considerato si individua l'intervallo dei valori possibili per $E(X_{n+1}/H)$, che denotiamo con $[e'_{\min}, e'_{\max}]$, corrispondente alla valutazione sui numeri aleatori $X_1, \dots, X_n$ espressa dalla [15'] in assenza di qualsiasi valutazione diretta concernente X_{n+1} . Indirettamente, però, le [15'] determinano un insieme di valori coerenti per $E(X_{n+1})$ costituito dall'intervallo $[e_{\min}, e_{\max}]$ con estremi dati dal minimo e massimo valore per il prodotto $c_{n+1} \cdot P(Q_r^*)$ cosicché, per quanto concerne la speranza matematica o previsione di X_{n+1} , l'effetto inferenziale dell'incremento di informazione H può essere rappresentato con la trasformazione dell'intervallo $[e_{\min}, e_{\max}]$ in quello $[e'_{\min}, e'_{\max}]$.

È necessario sottolineare che quest'ultimo intervallo per $E(X_{n+1}/H)$ non è invariante rispetto ad una particolare scelta di

$E(X_{n+1})$ entro il primo intervallo. Se il soggetto è in grado di scegliere un valore puntuale entro $[e_{\min}, e_{\max}]$, e sia e^* , questa scelta costituisce un vincolo che va ad aggiungersi alle [15'] condizionando i valori ammissibili di $E(X_{n+1}/H)$ entro un intervallo diverso da quello $[e'_{\min}, e'_{\max}]$. In altre parole, se nello stato di informazione iniziale il soggetto valuta $E(X_{n+1}) = e^*$ e se questa scelta è compatibile con le valutazioni [15'], cioè risulta $e_{\min} \leq e^* \leq e_{\max}$, l'incremento di informazione costituito dalla conoscenza dell'evento H costringe, per coerenza, la speranza matematica $E(X_{n+1}/H)$ entro l'intervallo $[e'_{\min}(e^*), e'_{\max}(e^*)]$ i cui estremi dipendono dal valore e^* oltre che dagli altri vincoli già considerati.

In quest'ultimo caso l'inferenza, o l'apprendimento dall'esperienza, sulla speranza matematica di X_{n+1} è rappresentata dalla trasformazione del valore numerico e^* nell'intervallo $[e'_{\min}(e^*), e'_{\max}(e^*)]$.

A prima vista può sembrare alquanto paradossale che un incremento di informazione porti dall'unica valutazione iniziale e^* ad un intervallo di infiniti valori possibili per $E(X_{n+1}/H)$: ciò perché, di primo acchito, si è portati ad identificare l'aumento di informazione su un numero aleatorio con la diminuzione della dispersione della sua distribuzione di probabilità e perché la suddetta situazione può ingannevolmente essere interpretata come passaggio da una distribuzione concentrata su e^* ad una distribuzione uniforme sull'intervallo $[e'_{\min}(e^*), e'_{\max}(e^*)]$. Evitando accuratamente tali suggestioni, questo intervallo va interpretato soltanto come totalità dei valori ammissibili per il valor medio subordinato $E(X_{n+1}/H)$ che riescono consistenti con la valutazione espressa dalla [15'] e dalla $E(X_{n+1}) = e^*$.

Riteniamo di dover concludere qui queste nostre considerazioni critiche sull'approccio bayesiano della Statistica e dell'Econometria. Chiaramente, il nostro obiettivo non è quello di argomentare contro tale approccio, del quale siamo assertori convinti, ma piuttosto quello di far intravedere la possibilità di evitare l'applicazione meccanica dello schema

$$\begin{array}{ccccc} \text{Valutazione} & & \text{Valutazione} & & \text{Funzione} \\ \text{probabilistica} & \propto & \text{probabilistica} & \times & \text{di} \\ \text{finale} & & \text{iniziale} & & \text{verosimiglianza} \end{array}$$

nei casi in cui esso sia giudicato di difficile applicazione, garantendo peraltro la consistenza logica dei procedimenti alternativi adottati.

Ci siamo soffermati su due soli di questi casi: il primo concerne la difficoltà di effettuare valutazioni probabilistiche su parametri incogniti non direttamente osservabili. È stata proposta una reinterpretazione del teorema di Bayes visto come relazione tra due leggi temporali, quella iniziale e quella finale, del processo generatore di dati; secondo noi, oltre che dal punto di vista concettuale, essa può risultare utile anche dal punto di vista operativo.

Il secondo caso esaminato riguarda l'uso di valutazioni probabilistiche parziali in procedimenti inferenziali coerenti; si tratta di un problema noto in letteratura come «robustness of bayesian analysis». In proposito ci sembra di particolare interesse l'approccio operativo che prende lo spunto da Lad, Dickey, Rahman [1987] e che si basa sul teorema fondamentale delle probabilità di B. de Finetti. Ci siamo limitati a considerare l'adeguamento coerente della funzione valor medio del processo generatore di dati, ma l'analisi potrebbe essere estesa anche a momenti di ordine superiore. Il tema abbisogna evidentemente di uno svolgimento più ampio ed approfondito che cercheremo di realizzare in altra occasione.

L'autore desidera ringraziare i proff. L. Crisma e S. Holzer per i loro preziosi suggerimenti, ma ovviamente si assume la completa responsabilità del presente lavoro.

Riferimenti bibliografici

- Bruno, G. e Giglio, A. (1980), *Applicazione del metodo del semplice al teorema fondamentale per le probabilità nella concezione soggettivistica*, in «Statistica», 40, 3.
- Crisma, L. (1987), *Dalla certezza all'incertezza: aspetti dinamici in una impostazione soggettiva*, in *Atti del Convegno su Incertezza ed Economia*, Trieste, Edizioni LINT Trieste.
- (in preparazione) *Elementi di probabilità soggettiva*, Trieste, Edizioni LINT.
- Chetty, V. K. (1968), *Bayesian Analysis of Haavelmo's Models*, in «Econometrica», vol. 36, n. 3-4.
- Daboni, L. e Wedlin, A. (1982), *Statistica. Un'introduzione all'impostazione neo-bayesiana*, Torino, UTET.
- De Finetti B., (a.a. 1937-1938), *Calcolo delle Probabilità*, Università di Padova.
- (1949), *Sull'impostazione assiomatica del calcolo delle probabilità*, Annali dell'Università di Trieste.
- (1970), *Teoria delle Probabilità*, Torino, Einaudi.
- Drèze, J. H. e Richard, J. F. (1983), *Bayesian Analysis of Simultaneous Equation Systems*, in *Handbook of Econometrics*, vol. I, a cura di Griliches Z. e Intriligator M. D., Amsterdam, North-Holland P.C.
- Holzer, S. (1985), *On Coherence and Conditional Prevision*, Bollettino U.M.I., Serie VI, Vol. IV-C, n. 1.
- Lad, F. R., Dickey, J. M. e Rahman, M. A. (1987), *The Fundamental Theorem of Prevision*, Technical Rep. n. 506 School of Statistics, University of Minnesota.
- Rahman, M. A. (1987), *Computable Reformulations and Extensions of de Finetti's Fundamental Theorem of Probability*, Ph. D. Thesis in Mathematics and Statistics, State University of New York.

- Rothenberg, T. J. (1963), *A Bayesian Analysis of Simultaneous Equation Systems*, Econometric Institute Rep. 6315, Erasmus University of Rotterdam.
- Zellner, A. (1971), *An Introduction to Bayesian Inference in Econometrics*, New York, J. Wiley.
- (1983), *Statistical Theory and Econometrics*, in *Handbook of Econometrics*, vol. I, a cura di Griliches Z. e Intriligator M. D., Amsterdam, North-Holland P.C.
- (1984), *Basic Issues in Econometrics*, The University of Chicago Press.
- (1985a), *Bayesian Econometrics*, in «Econometrica», vol. 53, n. 2.
- (1985b), *Bayesian Statistics in Econometrics*, in *Bayesian Statistics 2*, a cura di Bernardo J. M., DeGroot M. H., Lindley D.V., Smith A.F.M., Amsterdam, North-Holland.