Bollettino Unione Matematica Italiana

Sezione A – La Matematica nella Società e nella Cultura

Letizia Magnoni

Determinazione dell'indice di produzione industriale mediante l'uso di metodi Bayesiani

Bollettino dell'Unione Matematica Italiana, Serie 8, Vol. **2-A**—*La Matematica nella Società e nella Cultura* (1999), n.1S (Supplemento Tesi di Dottorato), p. 185–188.

Unione Matematica Italiana

<http://www.bdim.eu/item?id=BUMI_1999_8_2A_1S_185_0>

Bollettino dell'Unione Matematica Italiana, Unione Matematica Italiana, 1999.

La matematica nella Società e nella Cultura
Bollettino U. M. I.
(8) 2-A Suppl. (1999), pag. 185-188

Determinazione dell'indice di produzione industriale mediante l'uso di metodi Bayesiani.

LETIZIA MAGNONI

Lo scopo del lavoro della tesi è stato quello di applicare i metodi della statistica Bayesiana in un campo, l'economia, nel quale questo approccio è scarsamente praticato. In particolare, è stato esaminato il problema della determinazione dell'indice di produzione industriale usando le osservazioni sul consumo di energia elettrica. Tale relazione tra i due indici è frutto di uno studio della Banca d'Italia che si propone di determinare il valore dell'indice di produzione industriale in anticipo rispetto a quello fornito dall'ISTAT [2]. In questo lavoro è stata ripercorsa questa strada, usando però un approccio diverso al problema statistico, quello della statistica Bayesiana; in questo senso presentiamo un adattamento del noto filtro di Kalman-Bucy [3] la cui implementazione ha prodotto risultati più che soddisfacenti. Si è inoltre dimostrato che l'algoritmo proposto risulta analiticamente equivalente a quello proposto in [4], quando ci si limiti alla determinazione del valore atteso e della varianza dell'indice in considerazione.

La relazione che lega l'indice di produzione industriale ed il consumo di energia elettrica determina essenzialmente un problema di filtraggio stocastico, che da un punto di vista Bayesiano risulta essere di tipo non-lineare, nelle cui equazioni sono presenti due parametri incogniti. Lavorando in questa nuova ottica, non possiamo stimare il valore «vero» di tali parametri in quanto, potendo essi assumere un'infinità continua di valori, il punto di vista Bayesano considera questa stima come una scommessa avente probabilità di successo nulla. Per questa ragione, abbiamo considerato i parametri come delle variabili aleatorie dotate di densità a priori nota e si è poi provveduto, mediante l'utilizzo delle osservazioni di energia, ad aggiornare tale funzione di densità (densità a posteriori), utilizzando il teorema di Bayes [1].

I risultati raggiunti sono pienamente soddisfacenti anche alla luce di un loro confronto con quelli ottenuti utilizzando metodi di statistica frequentista.

1. – Presentazione dell'algoritmo.

In questo primo approccio abbiamo considerato una esemplificazione del modello utilizzato dalla Banca d'Italia; in particolare:

$$x_t = \alpha \cdot x_{t-1} + w_t$$

$$y_t = \beta \cdot x_t + v_t$$

dove x_t è la variabile di stato, y_t la variabile osservabile, $\{v_t\}$ e $\{w_t\}$ sono processi di rumore bianco gaussiano indipendenti tra loro e da x_0 che consideriamo gaussiana con media e varianza note; inoltre α e β sono i parametri incogniti e si suppone abbiano densità a priori nota $f_0(\alpha, \beta)$. Questo modello non può, da un punto di vista Bayesiano, essere considerato lineare in quanto i parametri, essendo incogniti, devono anch'essi essere considerati facenti parte del vettore di stato. Va da sè che il modello è condizionatamente lineare, cioè, se conoscessimo i valori assunti dalle variabili aleatorie α e β, il modello sarebbe di tipo lineare. Proprio per recuperare la condizionata linearità delle equazioni che descrivono il modello, proponiamo questo adattamento del noto filtro di Kalman-Bucy; indichiamo con $E[x_t/y^t]$ il valore atteso del processo x al tepmo t condizionatamente alle osservazioni effettuate sulla variabile osservabile fino al tempo t incluso, otteniamo:

$$E[x_t/y^t] = \int_\alpha \int_\beta m_t(\alpha, \beta) \cdot f(\alpha, \beta/y^t)\, d\alpha d\beta ,$$

dove la funzione $m_t(\alpha, \beta)$ è quella ottenuta dall'applicazione dell'algoritmo ricorsivo del filtro di Kalman-Bucy, relativamente alla determinazione del valore atteso della variabile di stato, condizionatamente alle osservazioni $y^t = (y_1, y_2, \ldots, y_t)$ ed ai valori dei parametri α e β; la funzione $f(\alpha, \beta/y^t)$ è la densità a posteriori congiunta dei parametri, che resta determinata dalle ipotesi di indipendenza fatte sul modello. In modo analogo possiamo ottenere il momento secondo della variabile di stato, date le osservazioni fino al tempo t:

$$E[x_t^2/y^t] = \int_\alpha \int_\beta [P_t(\alpha, \beta) + m_t^2(\alpha, \beta)] \cdot f(\alpha, \beta/y^t) d\alpha d\beta,$$

dove $P_t(\alpha, \beta)$ è la funzione, che rappresenta la varianza, fornita dall'applicazione del filtro di Kalman-Bucy.

Per il modello in esame le due funzioni m_t e P_t sono:

$$m_t(\alpha, \beta) = E[x_t/\alpha, \beta, y^t] = \frac{\alpha \cdot m_{t-1}(\alpha, \beta) + \alpha^2 \cdot \beta \cdot y_t \cdot P_{t-1}(\alpha, \beta) + \beta \cdot y_t}{\alpha^2 \cdot \beta^2 \cdot P_{t-1}(\alpha, \beta) + \beta^2 + 1} ,$$

e per la varianza

$$P_t(\alpha, \beta) = Var[x_t/\alpha, \beta, y^t] = \frac{\alpha^2 \cdot P_{t-1}(\alpha, \beta) + 1}{\alpha^2 \cdot \beta^2 \cdot P_{t-1}(\alpha, \beta) + \beta^2 + 1} ,$$

con le condizioni iniziali $m_0(\alpha, \beta) = 0$ e $P_0(\alpha, \beta) = 1$ (con y^0 e y_0 si intende nessuna osservazione).

Vediamo infine come possiamo determinare la densità a posteriori dei parametri.

Applicando più volte il teorema di Bayes e considerando le ipotesi di indipendenza fatte, otteniamo:

$$f(\alpha, \beta/y^{t}) = f(\alpha, \beta/y_t, y^{t-1}) \propto f(y_t/\alpha, \beta, y^{t-1}) \cdot f(\alpha, \beta/y^{t-1}) \propto$$
$$\propto \prod_{i=1}^{t} f(y_i/\alpha, \beta, y^{i-1}) \cdot f_0(\alpha, \beta),$$

dove $f_0(\alpha, \beta)$ è la densità a priori dei parametri e le funzioni di densità condizionata $f(y_i/\alpha, \beta, y^{i-1})$ sono, per ogni $i \in \{1, 2, \ldots, t\}$, gaussiane aventi media $\alpha \cdot \beta \cdot m_{i-1}(\alpha, \beta)$ e varianza $\alpha^2 \cdot \beta^2 \cdot P_{i-1}(\alpha, \beta) + \beta^2 + 1$.

Si è dunque fornito un algoritmo ricorsivo per la determinazione della media e della varianza condizionate della variabile di interesse.

Rimane ora il problema della effettiva valutazione degli integrali ottenuti, dal momento che, per la complessità delle funzioni coinvolte, essi non possono essere risolti per via analitica. Per questo scopo è stato utilizzato un metodo probabilistico per la risoluzione numerica [4] che si basa sulla discretizzazione della funzione di densità a priori dei parametri; tale tipo di risoluzione è stata preferita in quanto per essa è dimostrata la proprietà di robusteza rispetto a piccole perturbazioni del modello.

2. – Applicazione dei risultati teorici.

Il problema della determinazione dell'indice di produzione industriale mediante l'osservazione dei dati relativi al consumo di energia elettrica si può considerare di filtraggio stocastico, dal momento che l'ISTAT rende noto tale valore con circa due mesi di ritardo, mentre i dati relativi al consumo di energia elettrica sono disponibili tempestivamente.

Nell'applicazione si è considerata come unità temporale un mese, la variabile di stato x_t come il logaritmo naturale del quoziente dell'indice di produzione industriale al tempo t e quello al tempo precedente, cioè:

$$x_t = ln \frac{Prod_t}{Prod_{t-1}},$$

ed analogamente la variabile osservata y_t come il logaritmo naturale del quoziente del consumo di energia elettrica al tempo t e quello al tempo $t-1$, cioè:

$$y_t = ln \frac{Elett_t}{Elett_{t-1}}.$$

I rumori sono stati considerati bianchi gaussiani con varianza 0.01.

L'analisi è stata fatta usando i dati di consumo di energia elettrica (fonte ENEL) dal Gennaio 1985 al Gennaio 1991; i primi 36 dati sono stati usati per la prima previsione e successivamente, aggiungendo una osservazione alla volta, sono state compiute le successive previsioni.

Nel confronto con un metodo di previsione frequentista (che usa essenzialmente la regressione lineare) abbiamo notato che le nostre previsioni sono fortemente penalizzate dai dati relativi al mese di Agosto, nel quale si ha un brusco crollo della produzione industriale non dovuto a problemi di ciclo economico. Se infatti includiamo anche le previsioni relative a tale mese nel calcolo dell'errore medio quadratico compiuto, si ottiene un valore pari a 0. 071, valore molto alto rispetto al corrispondente frequentista che risulta essere 0.033. Se invece escludiamo da tale calcolo le previsioni ottenute in corrispondenza del mese di Agosto, otteniamo che l'errore medio quadratico si riduce a 0. 0059 mentre il suo corrispondente frequentista rimane dello stesso ordine di grandezza, 0.031.

BIBLIOGRAFIA

[1] BILLINGSLEY P., *Probability and Measure*, John Wiley and Sons, (1995).
[2] BODO G., SIGNORINI L.F., *Short-Term Forecasting of the Industrial Production Index*, International Journal of Forecasting, **3** (1987).
[3] HARVEY A.C., *Forecasting, structural time series models and the Kalman filter*, Cambridge University Press (1985).
[4] RUNGGALDIER W.J., VISENTIN C., *Combined Filter and Parameter Estimation: Approximations and Robustness*, Automatica, **26** (1990), 401-404.

Dipartimento di Matematica «R. Magari», Università di Siena; e-mail: magnoni@unisi.it
Dottorato in Logica Matematica e Informatica Teorica (sede amministrativa: Siena) - Ciclo VIII
Direttore della ricerca: Prof. M. Mirolli, Università di Siena