

Un programma per la rappresentazione dei modelli ARIMA

Programmo-
teca

Marcella Corduas, Centro di Specializzazione e Ricerche Economico Agrarie per il Mezzogiorno, Portici, Università di Napoli

In questo articolo viene presentato un programma di elaborazione predisposto per il calcolo di una matrice di distanza, definita su un insieme di modelli ARIMA, e per la soluzione di un problema di « scaling » multidimensionale classico per la loro rappresentazione.

1. Introduzione

La metrica definita sulla classe dei modelli ARIMA trova un primo campo d'applicazione nel confronto e nella classificazione di serie storiche.

In quest'articolo viene presentato un programma FORTRAN che rende operativo tale approccio. Esso costituisce il primo elemento di una procedura che, nell'ambito di un problema di classificazione, produca informazioni sufficienti all'individuazione di gruppi di modelli simili, modelli anomali, rappresentativi, etc... La « formulazione » che viene proposta, sebbene completa ed operante, è diretta espressione delle esigenze di verifica che l'attuale stadio della ricerca ha generato; di conseguenza essa si offre come oggetto immediato di ulteriori ed interessanti sviluppi, soprattutto in connessione con le numerose tecniche statistiche applicabili nella definizione dei clusters.

Il programma è rivolto al calcolo della matrice di distanze su un insieme di modelli $\{M_1, M_2, \dots, M_m\}$, appartenenti alla classe dei modelli ARIMA invertibili, secondo una metrica prescelta, ed alla soluzione di un problema di scaling classico per la loro rappresentazione. Nel prosieguo, quindi, supporremo nota la metodologia sviluppata in Box e Jenkins (1970), e la problematica connessa all'istituzione di una metrica sullo spazio dei processi ARIMA, come definita in Piccolo (1983), riportando, in sintesi, le caratteristiche del programma.

2. Descrizione del programma

Il programma è stato implementato su un elaboratore UNIVAC 1100/80 utilizzando il linguaggio Fortran. Esso si articola in alcune unità di cui riportiamo le funzioni nel seguente schema:

Sottoprogramma	Scopo
FOLD	dati i vettori dei coefficienti di due polinomi, di ordine r ed s , fornisce il vettore dei coefficienti del polinomio prodotto
POLYDV	effettua la divisione tra due polinomi, dati i vettori dei coefficienti ed il grado del polinomio quoziente
INDEX	calcola alcuni indici caratteristici su una successione data di pesi $\{a_h\}$ $h = 1, 2, \dots, n$
DIST	date le successioni $\{a_h\}_r$ $r = 1, 2, \dots, m$ calcola la matrice di distanze $\Delta_k = \{d_{ij}\}$ secondo la metrica $d_{ij} = (\sum_{h=1}^n a_{ih} - a_{jh} ^k)^{1/k}$, k intero positivo
MULT	calcola (*) autovalori ed autovettori di un'opportuna trasformazione di Δ_k , ed alcuni indici di « fit ».

(*) Utilizza a tal fine un'unità disponibile nella biblioteca di programmi NAG (Numerical Algorithms Group)

Come noto, la distanza tra modelli ARIMA proposta in Piccolo (1983) è calcolata come la distanza, secondo la metrica di Minkowski, tra le successioni di pesi $\{\pi_j\}$ che caratterizzano la formulazione AR(∞) di ciascun modello. Di conseguenza, il programma, dati gli operatori dell' i -mo modello definito come:

$$\phi(B) \Phi(B^s) Z_t = \theta(B) \Theta(B^s) a_t$$

procede al calcolo della successione $\{\pi_j\}_i$ $j = 1, 2, \dots, n$, mediante la nota relazione:

$$\pi(B) = \phi(B) \Phi(B^s) [\theta(B) \Theta(B^s)]^{-1};$$

a tale successione associa, inoltre, alcuni indicatori che possono orientare in una prima classificazione dei dati secondo quanto esplicitato in Corduas (1983), Viene così fornita una misura di trend, $\Sigma \pi_j$, un indice del contenuto informativo del processo, $\Sigma |\pi_j|$, una misura della sua non invertibilità, $\Sigma \pi_j^2$, con il corrispettivo indice normalizzato M^2 .

Il programma procede, poi, alla costruzione della matrice di distanza sulle successioni di pesi, secondo la metrica già definita, e calcola un indice di variabilità dell'insieme dei modelli analizzati rispetto a ciascun elemento M_i , dato da:

$$V_i = \frac{\sum_{h=1}^m d(M_i, M_h)}{(m - 1)}$$

che, nel lavoro citato, si dimostra poter essere utilizzato quale prima informazione al fine di individuare elementi anomali ed elementi rappresentativi di un gruppo. Inoltre, l'utente può specificare la richiesta di operare con una tecnica di « scaling » sulla matrice di distanze, ottenendo così le coordinate di ciascun modello in uno spazio R^m e i due più comuni indicatori di « fit » della soluzione di « scaling » a tale matrice (Mardia, 1979; Piccolo, 1984).

Il programma prevede, infine, la possibilità di analizzare l'insieme iniziale di modelli, nel caso in cui esso appartenga alla classe ARMA, secondo un approccio alternativo. Ciò si traduce, dal punto di vista operativo, nel calcolo delle successioni $\{\psi_j\}$ della formulazione MA(∞) di ciascun modello derivabile dall'identità: $\psi(B)\pi(B) = 1$ e nel calcolo, quindi, di una distanza δ tra tali successioni. L'algoritmo risulta, perciò, inalterato anche se ovviamente il significato degli indicatori prodotti sarà ridefinito in funzione della metrica diversa.

Sui dati d'ingresso viene effettuato un controllo iniziale al fine di fornire all'utente delle diagnostiche sul tipo di errore in cui eventualmente sia incorso. Lo schema di input e delle informazioni prodotte può essere così descritto:

INPUT	OUTPUT
— numero dei modelli	
— numero dei pesi	— Successioni $\{\pi_j\}$ e/o $\{\psi_j\}$
— metrica prescelta: intervallo $[a, b]$ in cui è definito k altrimenti in default $k = 2$)	— Indicatori sulle successioni
— analisi prescelta ARMA e/o ARI- MA	— Matrice di distanze
— attivazione dello « scaling »	
— controllo sul tipo di stampa (completa o di sintesi) (per ciascuno modello ripetere la sequenza:)	— Indici di variabilità
— titolo	— Soluzione classica dell'MDS
— ordine degli operatori	
— parametri corrispondenti « nome » « Lag » « valore »	

3. Ulteriori sviluppi

Il programma, come già anticipato, è suscettibile di numerosi interventi in funzione di una migliore caratterizzazione di ciascun modello e di una più ampia disponibilità di metodi di classificazione. In tal senso la nostra attenzione è rivolta allo studio ed all'applicazione di diversi algoritmi di clustering per produrne una scelta che sia sufficiente a rispondere alle diverse esigenze dell'utente. Altri interventi si presentano, invece, di più immediata realizzazione: ci riferiamo all'inserimento di un supporto di grafica che renda più leggibile i risultati prodotti ed alla realizzazione di una subroutine per il calcolo dello spettro parametrico di un modello ARMA che consentirebbe di verificare alcune relazioni duali su spazi metrici alternativi.

Il presente lavoro è stato svolto nell'ambito del « Progetto DESEC », promosso dal Centro di Specializzazione e Ricerche Economico agrarie per il Mezzogiorno, Portici, Università di Napoli, con il contributo finanziario del C.N.R.

Riferimenti bibliografici

- Box G.E.P. e Jenkins G.M. (1970), *Time Series Analysis: Forecasting and Control*, (rev. ed. 1976), Holden-Day, S. Francisco.
- Corduas M. (1983), « Distanza tra modelli: problemi metodologici ed indici statistici », RS 9/83, Centro di Specializzazione e Ricerche, Portici, Università di Napoli, in corso di pubblicazione su « Statistica », 3, 1984.
- Mardia K.V. - Kent J.T. - Bibby J.M. (1979), *Multivariate Analysis*, Academic Press, New York.
- Piccolo D. (1983), « Una topologia per la classe dei processi ARIMA »; RS 8/83, Centro di Specializzazione e Ricerche, Portici, Università di Napoli; in corso di pubblicazione su « Statistica », 1, 1984.
- Piccolo D. (1984), « Una rappresentazione multidimensionale per modelli statistici dinamici », *Atti XXXII Riunione Scientifica S.I.S.*, Sorrento, vol. II, 149-160.

Summary

Starting from a set of ARIMA models, fitted to m time series, the computer program estimates the π - weights characterizing each model, calculates the distance matrix, and by the classic approach to multidimensional scaling provides the co-ordinates of points which represent the set of models in a convenient space.