

Time series discrimination using the AR metric ¹

La metrica Autoregressiva per la discriminazione di serie storiche

Marcella Corduas

Dipartimento di Scienze Statistiche, Università di Napoli Federico II

e-mail: corduas@unina.it

Summary: In many fields of application one needs to classify a new observed time series as belonging to one of two or more categories. In this paper we propose a discrimination rule based on the AR metric and we investigate its statistical properties.

Keywords: Time Series Discrimination, ARIMA Models, Autoregressive Metric.

1. Introduction

Time series discrimination problems arise in many fields of application such as seismology, medicine, engineering and economics. Two main approaches have been proposed in the literature. The first approach is merely descriptive. It proceeds heuristically by searching for quantities which are good visual discriminators for well-separated populations and have some interpretative meaning within the specific application. The second approach, instead, looks at the time series as a finite realization of a Gaussian stochastic process and reduces it to a multivariate Gaussian vector. Thus, it exploits the known results of discriminant analysis for independent observations to define a discrimination rule. The solutions have been developed both in time and frequency domain in order to alleviate the computational difficulties implied by the former approach. Moreover, the discrimination problem has been investigated in the case of non Gaussian processes and of m -vector Gaussian stationary processes. Shumway (1982), Taniguchi and Kakizawa (2000) provide a wide survey of the methods proposed in the literature. Finally, we recall that Zani (1983) has discussed the problem of taking into account the time shift in time series comparison.

This paper introduces a discriminant rule based on a measure of structural diversity (the Autoregressive metric) between ARIMA processes and examines its statistical properties. This is a univariate approach which is aimed to classify a single time series to one of two categories.

2. The Autoregressive metric

Let Z_t be an invertible $ARIMA(p, d, q)(P, D, Q)_s$ process defined as:

$$\phi(B)\Phi(B^s)\nabla^d\nabla_s^D Z_t = \theta(B)\Theta(B^s)a_t \quad (1)$$

where a_t is a White Noise (WN) process with constant variance $\sigma_a^2 < +\infty$, the polynomials $\varphi(B) = \phi(B)\Phi(B^s)$ and $\vartheta(B) = \theta(B)\Theta(B^s)$ are defined according to the standard

¹This research was supported by the Dipartimento di Scienze Statistiche (University of Naples Federico II) and the Centro per la Formazione in Economia e Politica dello Sviluppo Rurale, Portici.

notation and satisfy the usual uniqueness and admissibility conditions (Box and Jenkins, 1976). The invertibility assumption ensures that Z_t can be represented in terms of its past values according to the $AR(\infty)$ formulation: $\pi(B)Z_t = a_t$ with $\pi(B) = \varphi(B)(1-B)^d(1-B^s)^D\vartheta^{-1}(B) = 1 - \sum_{k=1}^{\infty} \pi_k B^k$ and $\sum_{k=1}^{\infty} |\pi_k| < \infty$. Assuming that a_t is a Gaussian process, given the initial values, the knowledge of the operators: $\varphi(B)$, $\vartheta(B)$ and the WN variance completely characterizes the process Z_t .

The Euclidean distance between the π -weights of the $AR(\infty)$ formulations:

$$d(\pi_x, \pi_y) = \sqrt{\sum_{k=1}^{\infty} (\pi_{x,k} - \pi_{y,k})^2} \quad (2)$$

is a measure of the structural dissimilarity between two $ARIMA$ processes X_t and Y_t . It was introduced by Piccolo (1984, 1990) who discussed its importance and limits for time series comparison. The distance d is well defined because of the absolute convergence of the π -weights sequences and has an interpretation in terms of the forecast functions. Assuming that the orders are known and given the same set of initial values, $d(\pi_x, \pi_y) = 0$ if the models for X_t and Y_t produce the same forecasts. Notice that the metric does not depend on WN variances since they are mere scale factors which do not affect the temporal structure.

In the remaining part of this article we consider d^2 and the estimator $\hat{d}^2$ obtained by replacing the maximum likelihood (ML) estimators for the π -parameters in (2). Corduas (2000) derived the asymptotic distribution of $\hat{d}^2$ under general assumptions. Finally, in real applications, $\hat{d}^2$ is evaluated by truncating the π -sequences at a certain lag m .

3. The discrimination rule

The aim of this paper is to introduce a new discrimination procedure based on the squared AR distance. We investigate the case where a stochastic linear process X_t belongs to one of two categories described by the hypotheses H_1 and H_2 which assume that $X_t \in L$ and is characterized by $\{\pi_j(B), \sigma_j^2\}$, $j = 1, 2$, respectively.

Let $\mathbf{X}_{(n)} = \{X_1, \dots, X_n\}'$ be a finite stretch of the series X_t , the problem is to classify $\mathbf{X}_{(n)}$ into one of two known categories. Since d^2 represents a dissimilarity measure, then $d^2(\hat{\pi}_x, \pi_2) \geq d^2(\hat{\pi}_x, \pi_1)$ implies that the sequence $\hat{\pi}_x = \{\hat{\pi}_{x,k}, k = 1, \dots, m\}$ is closer to $\pi_1 = \{\pi_{1,k}, k = 1, \dots, m\}$ than $\pi_2 = \{\pi_{2,k}, k = 1, \dots, m\}$.

We propose the following rule: to assign $\mathbf{X}_{(n)}$ to H_1 (or H_2) according as $D > 0$ (or $D \leq 0$), where $D = d^2(\hat{\pi}_x, \pi_2) - d^2(\hat{\pi}_x, \pi_1)$.

Given the properties of the AR metric, the discriminant function D is aimed to match the observed time series to the category whose temporal structure is closer. In order to assess the performance of the D -rule we need to derive the asymptotic distribution and misclassification probabilities. First, we recall that, under H_j , $j = 1, 2$, asymptotically $\sqrt{n}(\hat{\pi}_x - \pi_j) \sim N(0, \Sigma_j)$ where $\Sigma_j = \sigma_j^2 \Gamma_{m,j}^{-1}$ being $\Gamma_{m,j}$ the Toeplitz matrix of order m of the process specified in H_j .

It is easy to see that: $d^2(\hat{\pi}_x, \pi_2) = d^2(\hat{\pi}_x, \pi_1) + d^2(\pi_1, \pi_2) + 2(\pi_1 - \pi_2)'(\hat{\pi}_x - \pi_1)$ and, similarly, $d^2(\hat{\pi}_x, \pi_1) = d^2(\hat{\pi}_x, \pi_2) + d^2(\pi_2, \pi_1) + 2(\pi_2 - \pi_1)'(\hat{\pi}_x - \pi_2)$. Of course, $d^2(\pi_1, \pi_2) = d^2(\pi_2, \pi_1)$.

Then, it can be shown that asymptotically:

- i) under H_1 , $\frac{\sqrt{n}}{2}\{D - d^2(\boldsymbol{\pi}_1, \boldsymbol{\pi}_2)\} \sim N(0, v_1^2)$ where $v_1^2 = (\boldsymbol{\pi}_1 - \boldsymbol{\pi}_2)' \boldsymbol{\Sigma}_1 (\boldsymbol{\pi}_1 - \boldsymbol{\pi}_2)$;
 ii) under H_2 , $\frac{\sqrt{n}}{2}\{D + d^2(\boldsymbol{\pi}_2, \boldsymbol{\pi}_1)\} \sim N(0, v_2^2)$ where $v_2^2 = (\boldsymbol{\pi}_2 - \boldsymbol{\pi}_1)' \boldsymbol{\Sigma}_2 (\boldsymbol{\pi}_2 - \boldsymbol{\pi}_1)$.

We can, now, evaluate the misclassification probabilities assuming $d^2(\boldsymbol{\pi}_2, \boldsymbol{\pi}_1) > 0$:

$$P(2|1) = P(D \leq 0 \mid H_1) = \Phi\left(-\frac{\sqrt{n}}{2} \frac{d^2(\boldsymbol{\pi}_1, \boldsymbol{\pi}_2)}{v_1}\right); \quad (3)$$

$$P(1|2) = P(D > 0 \mid H_2) = 1 - \Phi\left(\frac{\sqrt{n}}{2} \frac{d^2(\boldsymbol{\pi}_2, \boldsymbol{\pi}_1)}{v_2}\right). \quad (4)$$

Notice that $\lim_{n \rightarrow \infty} P(2|1) = \lim_{n \rightarrow \infty} P(1|2) = 0$, that is the discriminant statistic is consistent in the sense that the misclassification probabilities tend to zero as $n \rightarrow \infty$. Moreover, if $d^2(\boldsymbol{\pi}_1, \boldsymbol{\pi}_2) = 0$, as one would expect, $P(2|1) = P(1|2) = 0.5$.

In real applications, the π -weight sequences $\{\pi_{1,k}\}$ and $\{\pi_{2,k}\}$ associated to H_1 and H_2 are not known. Then, they have to be estimated using learning samples. Suppose that for $j = 1, 2$ we have s_j observed time series $x_{i,t}^{(j)}, t = 1, \dots, n, i = 1, \dots, s_j$, correctly classified with respect to the two categories defined by H_j . These represent the training samples. We denote with $M_i^{(j)}$ the ARIMA model fitted to the i -th time series belonging to the j -th category. Then, we suggest to use the estimated π -weight sequence of the *centroid* model of each group of observed time series in place of the unknown $\{\pi_{1,k}\}$ and $\{\pi_{2,k}\}$. The *centroid* model is the one which minimizes the sum of the AR distances with respect to all the other models included in the group.

4. Simulation results and concluding remarks

To study the performance of the proposed rule, we consider the $AR(1)$ Gaussian process: $X_t = \phi X_{t-1} + a_t$ with $a_t \sim WN(0, 1)$ and specify the following hypotheses:

$$H_1 : X_t = \phi_1 X_{t-1} + a_t; \quad H_2 : X_t = \phi_2 X_{t-1} + a_t \text{ with } \phi_2 = \phi_1 + h/\sqrt{n}.$$

For each value of $\phi_1 = 0.2, 0.4, 0.6, 0.8$ several H_2 are examined by allowing h to vary in $[-3, 3]$. This choice represents the less favourable situation in which a discriminant rule has to work, since the hypotheses H_1 and H_2 are contiguous. Also, only the values $\phi_1 > 0$ are considered since in this case for $\phi_1 < 0$ the results would be specular.

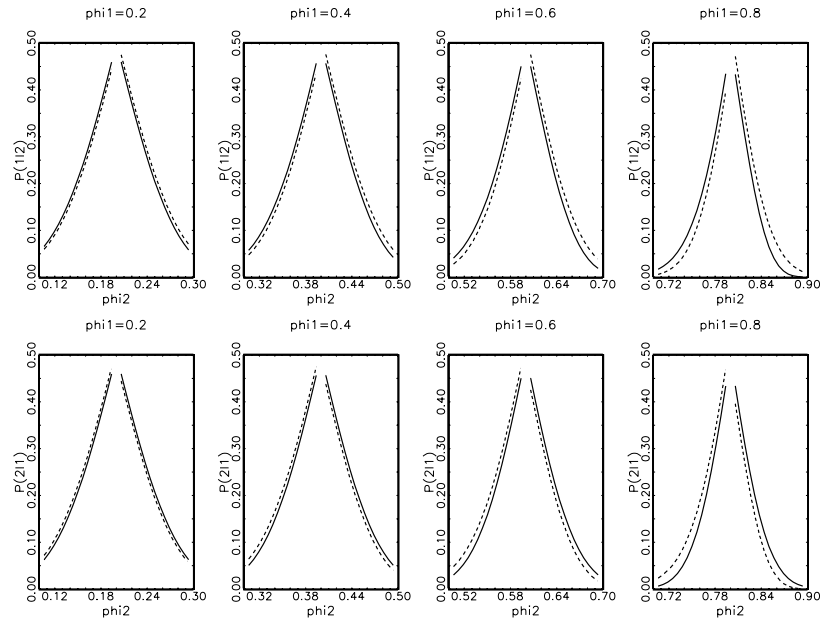
For given H_1 and H_2 , we have generated 1000 independent time series of $n = 1024$ observations. Then, we have applied the proposed rule for discrimination and evaluated the misclassification probabilities $P(2|1)$ and $P(1|2)$. The simulated values are very close to the asymptotic values, as it is reasonable to expect given the formulation of D . Moreover, we have generated 200 training samples of 50 time series for each hypotheses and we have evaluated the empirical misclassification probabilities by using the *centroid* estimated model. For $|h| > 1.5$, the results are encouraging being the largest discrepancy between empirical and theoretical probabilities less than 0.04.

Finally, we compare the asymptotic misclassification probabilities of D-rule with those of the approximated quadratic discriminant rule derived in the frequency domain by Shumway (1982, pp. 17-18). As we see from Figure 1, the limiting error probabilities

are very close. Some differences are observed for larger ϕ_1 .

Further studies are needed in order to generalize these empirical findings. In conclusion, we underly the fact that the proposed rule is reliable, easy to implement and it provides a general framework to extend time series discrimination to other form of models such as *ARFIMA* models or fields of application as *data mining*.

Figure 1. $P(1|2)$ (1st row) and $P(2|1)$ (2nd rows); Quadratic discriminant function (dashed line); D-rule (solid line)



References

- Box G.E.P., Jenkins G.M. (1976) *Time series analysis: forecasting and control*, rev. edition, Holden-Day, San Francisco.
- Corduas M. (2000) La metrica Autoregressiva tra modelli ARIMA: una procedura in linguaggio GAUSS, *Quaderni di Statistica*, 2, 1-37.
- Piccolo D. (1984) Una topologia per la classe dei processi ARIMA, *STATISTICA*, XLIV, 47-59.
- Piccolo D. (1990) A distance measure for classifying ARIMA models, *Journal of Time Series Analysis*, 11, 153-164.
- Shumway R.H. (1982) Discriminant analysis for time series, in *Handbook of Statistics*, eds. P.R. Krishnaiah and L.N. Kanal, North Holland, New York, vol.2, 1-46.
- Taniguchi M., Kakizawa Y. (2000) *Asymptotic theory of statistical inference for time series*, Springer-Verlag, New York.
- Zani S. (1983) Osservazioni sulle serie storiche multiple e l'analisi dei gruppi, in *Analisi moderna delle serie storiche*, ed. D. Piccolo, F. Angeli, Milano, 263-274.