Change Recognition in a Model Structure of Noisy Autoregressive Process

Konstantin N. Nechval, Nicholas A. Nechval & Edgars K. Vasermanis

Department of Mathematical Statistics, University of Latvia
Raina Blvd 19, LV-1050 Riga, Latvia
Fax: +371-7034702, E-mail: nechval@junik.lv

Abstract

In this paper, we model the noise as an autoregressive (AR) process with unknown parameters. A special case of a general class of problems concerned with recognition of changes of model structure of stochastic processes is considered. It is assumed that a stochastic process can be written in the form of an autoregressive model, which has numerous applications, ranging from industrial quality control to edge recognition in images and the diagnosis of faults in computer communication networks. The approach, which is taken here, is to apply the theory of generalized likelihood ratio testing for composite hypothesis testing. A procedure is used for calculating the exact likelihood function for the autoregressive (AR) process. The decision rule is based on the generalized likelihood ratio statistics.

Keywords: noise, autoregressive model, change recognition, test

1. Introduction

The autoregressive (AR) model has proved to be quite useful in communications and signal processing. For example, in spectral analysis, it provides spectral estimates with high resolution even for rather short record length. In many procedures of adaptive segmentation of nonstationary signal into "homogeneous" parts (where the times of change in the signal spectrum might indicate significant events which are to be monitored (as e.g. with seismic signals)), the signal is modeled by a Gaussian distributed AR process. In several picture processing, an elastic registration of distorted pictures is based on the AR model.

In this paper, the problem of recognizing possible changes in structure of a given autoregressive model is considered. This problem is a special case of a general class of problems concerned with recognition of changes of model structure of stochastic processes.

The approach, which is taken here, is to apply the theory of generalized likelihood ratio testing for composite hypothesis testing. A procedure is used for calculating the exact likelihood function for the AR process. The decision rule is based on the generalized likelihood ratio statistic. The test obtained is invariant to intensity changes in the noise background and achieves a fixed probability of a false alarm. Thus, operating in accordance to the local noise situation, the test is adaptive. In addition, it can be shown that the test is uniformly most powerful invariant.

2. Problem Statement

In this section, the problem of change recognition in an AR model is formulated as a test of the following hypotheses:

$$H_0: x_n = \sum_{j=1}^{p} a_j x_{n-j} + w_n \qquad \text{(unchanged model)},$$

$$H_1: x_n = \sum_{j=1}^{p} a_j x_{n-j} + \sum_{j=p+1}^{p+q} s_j x_{n-j} + w_n \qquad \text{(changed model)}, \tag{1}$$

where it is assumed that, under H_0, $\{x_n\}$, n = ... , -1, 0, 1, ... , is an AR zero-mean Gaussian process of known order p; $\mathbf{a}=(a_1, a_2, \ldots, a_p)'$ is a $p \times 1$ column vector of the AR process parameters, the

w_n's are Gaussian independent identically distributed random variables, with mean zero and variance σ^2, $\mathbf{s}=(s_{p+1}, s_{p+2}, \ldots, s_{p+q})'$ is a column vector of order q. Let us call $r_i=E\{x_i x_1\}$ for $i\geq 1$, the covariance of the process, and

$$\boldsymbol{\Gamma} = \begin{pmatrix} r_1 & r_2 & \cdots & r_p \\ r_2 & r_3 & \cdots & r_{p-1} \\ \cdots & \cdots & \cdots & \cdots \\ r_p & r_{p-1} & \cdots & r_1 \end{pmatrix} \quad (2)$$

the p × p positive definite Toeplitz covariance matrix of the process, and $\mathbf{x}_n=(x_{n-1}, \ldots, x_{n-p})'$ the p × 1 column vector of the previous p data points at instant n. Then the joint density of $\mathbf{x}_{p+1}=(x_p, \ldots, x_1)'$ is given by

$$f_p(\mathbf{x}_{p+1};\boldsymbol{\theta}) = \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Gamma}|^{1/2}} \times \exp(-\mathbf{x}'_{p+1}\boldsymbol{\Gamma}^{-1}\mathbf{x}_{p+1}/2), \quad (3)$$

where $\boldsymbol{\theta}=(\mathbf{a}',\sigma^2)'$. The matrix $\boldsymbol{\Gamma}$ and the vector $\boldsymbol{\theta}$ are related through the Yule-Walker equations [Box and Jenkins, 1970]. The calculation of the joint distribution of the set of measurements $(x_1, \ldots, x_N)$ given $\boldsymbol{\theta}$, a vector which contains the unknown parameters $a_1, \ldots, a_p$, σ^2, gives (for N>p)

$$\begin{aligned} &f(x_1,\ldots,x_N;\boldsymbol{\theta}) \\ &= f(x_N;\boldsymbol{\theta}, x_1,\ldots,x_{N-1})f(x_1,\ldots,x_{N-1};\boldsymbol{\theta}) \\ &= f(x_N;\boldsymbol{\theta}, x_{N-p},\ldots,x_{N-1})f(x_1,\ldots,x_{N-1};\boldsymbol{\theta}) \\ &= f(x_N;\boldsymbol{\theta}, x_{N-p},\ldots,x_{N-1})f(x_{N-1};\boldsymbol{\theta}, x_{N-p-1},\ldots,x_{N-2}) \\ &\quad \ldots f(x_{p+1};\boldsymbol{\theta}, x_1,\ldots,x_p)f_p(x_1,\ldots,x_p;\boldsymbol{\theta}) \\ &= f^c(x_{p+1},\ldots,x_N;\boldsymbol{\theta}, x_1,\ldots,x_p)f_p(x_1,\ldots,x_p;\boldsymbol{\theta}), \end{aligned} \quad (4)$$

where

$$f^c(x_{p+1},\ldots,x_N;\boldsymbol{\theta}, x_1,\ldots,x_p) = \frac{1}{\left(2\pi\sigma^2\right)^{(N-p)/2}} \times \exp\left(-\sum_{n=p+1}^{N}[x_n - \mathbf{a}'\mathbf{x}_n]^2/(2\sigma^2)\right). \quad (5)$$

In the above, $f_p(\cdot)$ is the marginal density of the first p observations, while $f^c(\cdot)$ is the conditional density of the remaining observations given the first p observations.

The above problem can be recast as

$$H_0 : \boldsymbol{\theta}' = \left(\mathbf{0}', \boldsymbol{\theta}'_{p+1}\right)$$

$$H_1 : \boldsymbol{\theta}' = \left(\boldsymbol{\theta}'_q, \boldsymbol{\theta}'_{p+1}\right), \quad \boldsymbol{\theta}_q \neq \mathbf{0}, \quad (6)$$

where $\boldsymbol{\theta}'_{p+1}=(\mathbf{a}', \sigma^2)$, $\boldsymbol{\theta}_q= \mathbf{s}$. In either case the dimensions of $\boldsymbol{\theta}_{p+1}$ and $\boldsymbol{\theta}_q$ are p+1 and q, respectively. It is well known that there is no uniformly most powerful (UMP) test for (6). Yet the generalized maximum likelihood ratio (GMLR) test is widely preferred because of its nice asymptotic (large sample size) properties such as consistency, unbiasedness, and constant false alarm rate (CFAR). It is also called the uniformly most powerful invariant (UMPI) test since it exhibits the UMP property among the class of tests which are invariant to a natural set of transformations [Lehmann, 1959].

3. GMLR Test

The GMLR test for testing (6) is to decide H_1 if

$$l_G^{\bullet} = \frac{L(\hat{\boldsymbol{\theta}}_q, \hat{\boldsymbol{\theta}}_{p+1})}{L(\mathbf{0}, \hat{\boldsymbol{\theta}}_{p+1})} > h^{\bullet} \tag{7}$$

for some threshold $h^{\bullet}$, where L is the likelihood function,

$$L(\boldsymbol{\theta}_q, \boldsymbol{\theta}_{p+1}) = f(x_1, \ldots, x_N; \boldsymbol{\theta}_q, \boldsymbol{\theta}_{p+1}), \tag{8}$$

$\hat{\boldsymbol{\theta}}_{p+1}$ is the MLE of $\boldsymbol{\theta}_{p+1}$ assuming H_0 is true while $\hat{\boldsymbol{\theta}}_q$ and $\hat{\boldsymbol{\theta}}_{p+1}$ are joint MLE′s of $\boldsymbol{\theta}_q$ and $\boldsymbol{\theta}_{p+1}$ assuming H_1 is true. $\hat{\boldsymbol{\theta}}_{p+1}$ is found by maximizing $L(\mathbf{0},\boldsymbol{\theta}_{p+1})$ over $\boldsymbol{\theta}_{p+1}$. Similarly, $\hat{\boldsymbol{\theta}}_q, \hat{\boldsymbol{\theta}}_{p+1}$ are obtained by maximizing $L(\boldsymbol{\theta}_q, \boldsymbol{\theta}_{p+1})$ over $\boldsymbol{\theta}_q$ and $\boldsymbol{\theta}_{p+1}$.

The likelihood ratio for problem (1) has the form

$$l_G^{\bullet} = \frac{f(x_1, \ldots, x_N; \hat{\boldsymbol{\theta}}_q, \hat{\boldsymbol{\theta}}_{p+1})}{f(x_1, \ldots, x_N; \mathbf{0}, \hat{\boldsymbol{\theta}}_{p+1})}$$

$$= \frac{f^c(x_{p+1}, \ldots, x_N; \hat{\boldsymbol{\theta}}_q, \hat{\boldsymbol{\theta}}_{p+1}, x_{-q+1}, \ldots, x_1, \ldots, x_p)}{f^c(x_{p+1}, \ldots, x_N; \mathbf{0}, \hat{\boldsymbol{\theta}}_{p+1}, x_1, \ldots, x_p)}$$

$$\times \frac{f_p(x_{-q+1}, \ldots, x_1, \ldots, x_p; \hat{\boldsymbol{\theta}}_q, \hat{\boldsymbol{\theta}}_{p+1})}{f_p(x_1, \ldots, x_p; \mathbf{0}, \hat{\boldsymbol{\theta}}_{p+1})}. \tag{9}$$

The second factor is dropped for ease of computation. A heuristic justification for ignoring the second term is that its contribution to $l_G^{\bullet}$ will be negligible when N is large and the two hypotheses are close to each other. With this simplification, the test is equivalent to deciding H_1 if

$$l_G^{\circ} = \frac{f^c(x_{p+1}, \ldots, x_N; \hat{\boldsymbol{\theta}}_q, \hat{\boldsymbol{\theta}}_{p+1}, x_{-q+1}, \ldots, x_1, \ldots, x_p)}{f^c(x_{p+1}, \ldots, x_N; \mathbf{0}, \hat{\boldsymbol{\theta}}_{p+1}, x_1, \ldots, x_p)} > h^{\circ}, \tag{10}$$

where

$$f^c(x_{p+1}, \ldots, x_N; \hat{\boldsymbol{\theta}}_q, \hat{\boldsymbol{\theta}}_{p+1}, x_{-q+1}, \ldots, x_1, \ldots, x_p)$$

$$\max_{\mathbf{s}, \mathbf{a}, \sigma^2} \left(\frac{1}{(2\pi\sigma^2)^{(N-p)/2}} \times \exp\left(-\sum_{n=p+1}^{N} [x_n - \mathbf{a}'\mathbf{x}_n - \mathbf{s}'\mathbf{x}_{n-p}]^2 / (2\sigma^2) \right) \right), \tag{11}$$

$$f^c(x_{p+1}, \ldots, x_N; \mathbf{0}, \hat{\boldsymbol{\theta}}_{p+1}, x_1, \ldots, x_p)$$

$$\max_{\mathbf{a}, \sigma^2} \left(\frac{1}{(2\pi\sigma^2)^{(N-p)/2}} \times \exp\left(-\sum_{n=p+1}^{N} [x_n - \mathbf{a}'\mathbf{x}_n]^2 / (2\sigma^2) \right) \right), \tag{12}$$

$\mathbf{x}_{n-p} = (x_{n-p-1}, \ldots, x_{n-p-q})$. It can be shown that l_G° is equivalent finally to the statistic

$$v = \frac{N-2p-q}{V_{H_1}} \frac{V_{H_0} - V_{H_1}}{q},$$

(13)

where

$$V_{H_0} = \min_{\mathbf{a}} \sum_{n=p+1}^{N} \left(x_n - \mathbf{a}'\mathbf{x}_n \right)^2,$$

(14)

$$V_{H_1} = \min_{\mathbf{a},\mathbf{s}} \sum_{n=p+1}^{N} \left(x_n - \mathbf{a}'\mathbf{x}_n - \mathbf{s}'\mathbf{x}_{n-p} \right)^2,$$

(15)

Here the following theorem clearly holds.

Theorem 1. Under H_1 , the statistic v is subject to a noncentral F-distribution with k_1=q and k_2=N-2p-q degrees of freedom, the probability density function of which is

$$f_{H_1}(v;k_1,k_2,\lambda) = \sum_{j=0}^{\infty} e^{-\lambda/2} \frac{(\lambda/2)^j}{j!}$$

$$\times \frac{[k_1+2j]^{(k_1+2j)/2} k_2^{k_2/2}}{B\left(\frac{k_1+2j}{2}, \frac{k_2}{2}\right)}$$

$$\times \frac{v^{(k_1+2j-2)/2}}{[k_2+(k_1+2j)v]^{(k_1+2j+k_2)/2}},$$

$$0 < v < \infty,$$

(16)

with a noncentrality parameter $\lambda \geq 0$ given by

$$\lambda = \mathbf{s}'\left(\mathbf{X}_1'\mathbf{X}_1 - \mathbf{X}_1'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}_1\right)\mathbf{s}/\sigma^2,$$

(17)

where $\mathbf{X}_1$ is an (N-p) × q matrix and $\mathbf{X}$ is an (N-p) × p matrix, respectively,

$$\mathbf{X}_1 = \begin{pmatrix} x_0 & x_{-1} & \cdots & x_{-q+1} \\ x_1 & x_0 & \cdots & x_{-q+2} \\ \cdots & \cdots & \cdots & \cdots \\ x_{N-p-1} & x_{N-p-2} & \cdots & x_{N-p-q} \end{pmatrix},$$

(18)

$$\mathbf{X} = \begin{pmatrix} x_p & x_{p-1} & \cdots & x_1 \\ x_{p+1} & x_p & \cdots & x_2 \\ \cdots & \cdots & \cdots & \cdots \\ x_{N-1} & x_{N-2} & \cdots & x_{N-p} \end{pmatrix}.$$

(19)

Under H_0, when λ=0, (16) reduces to a standard F-distribution with k_1 and k_2 degrees of freedom,

$$f_{H_0}(v;k_1,k_2) = \frac{k_1^{k_1/2} k_2^{k_2/2}}{B\left(\frac{1}{2}k_1, \frac{1}{2}k_2\right)}$$

$$\times \frac{v^{(k_1-2)/2}}{[k_2+k_1 v]^{(k_1+k_2)/2}},$$

$$0 < v < \infty .$$

(20)

Proof. The proof follows by applying the known results of linear regression analysis [Nechval, 1982, 1984] and being straightforward it is omitted here.

The GMLR test of H_0 versus H_1 is given by

$$v\begin{cases} > h, & \text{then} \quad H_1 \\ \le h, & \text{then} \quad H_0 \end{cases},$$

(21)

where the threshold h is equal to $F^{\alpha}_{k_1,k_2}$, the upper $100\alpha\%$ point of the central F_{k_1,k_2} distribution with k_1 and k_2 degrees of freedom.

The probability of deciding H_1 when H_0 is true (also called the probability of false alarm) is given by

$$P_{FA}(h) = \Pr\{v > h; H_0\} = \int_h^{\infty} f_{H_0}(v; k_1, k_2)dv$$

$$= \frac{k_1^{k_1/2} k_2^{k_2/2}}{B\left(\frac{1}{2}k_1, \frac{1}{2}k_2\right)} \int_h^{\infty} \frac{v^{(k_1-2)/2}}{[k_2 + k_1 v]^{(k_1+k_2)/2}} dv.$$

(22)

Note that $P_{FA}(h)$ is a function only of the integers N, p, and q and threshold h. Hence for these parameters held fixed the recognition test in (18) has a constant false alarm rate.

The probability of correctly deciding H_1 (called the power of the test) is

$$P_D(h) = \Pr\{v > h; H_1\} = \int_h^{\infty} f_{H_1}(v; k_1, k_2, \lambda)dv$$

$$= \sum_{j=0}^{\infty} e^{-\lambda/2} \frac{(\lambda/2)^j}{j!}$$

$$\times \frac{[k_1 + 2j]^{(k_1+2j)/2} k_2^{k_2/2}}{B\left(\frac{k_1+2j}{2}, \frac{k_2}{2}\right)}$$

$$\times \int_h^{\infty} \frac{v^{(k_1+2j-2)/2}}{[k_2 + (k_1+2j)v]^{(k_1+2j+k_2)/2}} dv.$$

(23)

4. Conclusion

The authors hope that this work will stimulate further investigation using the approach on specific applications to see whether obtained results with it are feasible for realistic applications and may be extended to provide existence results for similar problems.

Acknowledgments

This research was supported in part by Grant No. 01.0031 and Grant No. 02.0918 from the Latvian Council of Sciences and the National Institute of Mathematics and Informatics of Latvia.

References

Box, G. and Jenkins, G. (1970). *Time Series Analysis - Forecasting and Control*. Holden-Day, San Francisko, CA..

Lehmann, E.L. (1959). *Testing Statistical Hypotheses*. Wiley, New York.

Nechval, N.A. (1982). *Modern Statistical Methods of Operations Research*. RCAEI, Riga.

Nechval, N.A. (1984). *Theory and Methods of Adaptive Control of Stochastic Processes*. RCAEI, Riga.