Asymptotic inference for discrete vector AR processes

By MÁTYÁS ARATÓ (Debrecen)

Dedicated to my friend Zoltán Daróczy on his 50th birthday

Abstract: We consider a first order autoregressive (AR) vector process which has stationary behavior and fulfils equation (1). The least squares estimate of the matrix parameter Q is investigated when the observation interval, n, tends to infinity. It is proved that the approximate distribution of the estimate depends on Q and the functional $\hat{Q}$ in (4) can be approximated by the help of continuous time Gauss—Markov process (6′).
Keywords: autoregression, least squares estimate, stochastic difference equation, martingale difference, Gauss—Markov process.

1. Results and comparisons

Let us consider an autoregressive vector model

(1)
$$\underline{\xi}(t) = Q\underline{\xi}(t-1)+\underline{\varepsilon}(t), \quad t = 1, 2, \dots, n,$$

with

$$E\underline{\xi}(t) = E\underline{\varepsilon}(t) = 0, \quad \underline{\xi}(0) = 0.$$

Here, $\underline{\xi}(t)$ is the observation at time t, $\underline{\varepsilon}(t)$ is the random disturbance and the matrix Q is unknown. We shall assume that $\underline{\varepsilon}(t)$ is a martingale difference sequence with respect to the σ-fields F_t, $F_t \subseteqq F_{t+1}$, such that, $\forall \alpha > 0$, in probability

(2)
$$\frac{1}{n}\sum_{t=1}^{n} E\big(\underline{\varepsilon}(t)^* \underline{\varepsilon}(t)\, I_{(\|\varepsilon(t)\| > n^{1/2}\alpha} | F_{t-1}\big) \to 0, \quad as \quad n \to \infty,$$

(3)
$$\frac{1}{n}\sum_{t=1}^{n} E\big(\underline{\varepsilon}(t)^* \underline{\varepsilon}(t) | F_{t-1}\big) \to I, \quad as \quad n \to \infty.$$

The unknown parameter Q is customarily estimated by its least squares estimate

(4)
$$\hat{Q}_n = \sum_{t=1}^{n} \underline{\xi}(t)\,\underline{\xi}^*(t-1)\Big(\sum_{t=1}^{n-1} \underline{\xi}(t)\,\underline{\xi}^*(t)\Big)^{-1}.$$

If the $\underline{\varepsilon}(t)$'s are normally distributed, $\hat{Q}_n$ is the maximum likelihood estimator and $\underline{\xi}(t)$ is an elementary Gaussian process (see [2]). The stochastic differential equation related to (1) is

(5)
$$d\underline{\xi}(t) = A\underline{\xi}(t)\,dt + d\underline{\omega}(t), \quad 0 \leqq t \leqq T, \quad \underline{\xi}(0) = 0,$$

where $Q=e^{A\cdot\Delta}$, and the least squares estimate of A is given by

$$\hat{A}=\Big[\int_0^T d\underline{\xi}(t)\underline{\xi}^*(t)\Big]\Big[\int_0^T \underline{\xi}(t)\,\underline{\xi}^*(t)\Big]^{-1}. \tag{6}$$

Note that every process $\xi(t)$, $0\leq t\leq T$, satisfying the equation

$$d\underline{\xi}(t)=A\underline{\xi}(t)\,dt+d\underline{\omega}(t),\quad \underline{\omega}(t) \text{ is a Wiener process},\quad \underline{\xi}(0)=0,$$

can be transformed to the form

$$d\underline{\eta}(s)=A_0\underline{\eta}(s)\,ds+d\underline{\overline{\omega}}(s),\quad 0\leq s\leq 1,\quad \underline{\xi}(0)=0, \tag{6'}$$

by

$$\underline{\eta}(s)=\underline{\xi}(s)/\sqrt{T},\quad A_0=A\cdot T,\quad s=\frac{t}{T},\quad \underline{\overline{\omega}}(s)=\frac{\underline{\omega}(s)}{\sqrt{T}}. \tag{7}$$

So in the following we may assume that $T=1$ and the observation interval is $0\leq t\leq 1$.

The distributions of the statistics $\int_0^1 d\underline{\xi}(t)\underline{\xi}^*(t)$, $\int_0^1 \underline{\xi}(t)\underline{\xi}(t)\,dt$ are given (see [7]) and they depend on A. If $A=0$ (i.e. $Q=I$) we get (compare with [3])

$$\hat{A}=\Big(\int_0^1 d\underline{\omega}(t)\underline{\omega}^*(t)\Big)\Big(\int_0^1 \underline{\omega}(t)\underline{\omega}^*(t)\,dt\Big)^{-1}. \tag{8}$$

The main purpose of this paper is to prove that the approximate distribution of $\hat{Q}_n$, when $n\to\infty$, depends on Q (or on A), i.e., on a family of distributions. Such a phenomenon can be seen in the binomial case when $(x-np)(np(1-p))^{-1/2}$ has a Poisson approximation with parameter λ if $np\sim\lambda$, and a normal approximation if p is fixed and $n\to\infty$, respectively.
When Q is fixed and $|Q-\lambda I|=0$ has solutions λ_i with $|\lambda_i|<1$ then it is well known that (see [1], [2], [6], [8]).

$$\sqrt{n}(\hat{Q}_n-Q)\to N(0,\overline{B}^{-1}(0)),\quad \overline{B}^{-1}(0)=\begin{pmatrix} B^{-1}(0) & 0 & \dots & 0\\ 0 & B^{-1}(0) & \dots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \dots & B^{-1}(0)\end{pmatrix},$$

or

$$(\hat{Q}_n-Q)\Big(\sum_{t=1}^n \underline{\xi}_{t-1}\underline{\xi}^*_{t-1}\Big)^{1/2}\to N(O,I), \tag{10}$$

in distribution, where the steady state covariance of $\underline{\xi}(t)B(0)$ is the solution of equation

$$B(0)=QB(0)Q^*+I, \tag{11}$$

(or $AB(0)+B(0)A^*=-I$). This was first proved by Mann and Wald [8]. We get the Yule—Walker equations ([1], [6]). But this convergence is not uniform in $|\lambda_i|<1$ (it depends on $B(0)$, i.e. on Q). In recent years, there has been considerable interest in the asymptotic properties of $\hat{Q}_n$ when the λ_i are close or equal to one (see [3], [10], [12]). The simplest case is $Q=I$ (or $A=0$). In the one-dimensional case White proved (see [3], [13]) that if $Q=1$ then

$$\Big(\sum_{t=1}^{n} \xi^2(t-1)\Big)^{1/2}(\hat{\varrho}_n-1) \to \frac{1}{2}\big(\omega^2(1)-1\big)\Big/\int_0^1 \omega^2(t)\,dt,$$

where

$$\hat{\varrho}_n = \sum_{t=1}^{n} \xi_t\,\xi_{t-1}\Big/\sum_{t=1}^{n} \xi_{t-1}^2.$$

This is a special case of (6). Our principal result is the following.

Theorem 1. *Let Q be a matrix with characteristic roots $|\lambda_i|\leqq 1$. For $t=1,2,\dots,n$ suppose $\underline{\xi}(t)$ satisfies* (1), *with $\underline{\xi}(0)=0$; and $\underline{\varepsilon}(t)$ satisfies* (2), (3). *Then, as $n\to\infty$,*

$$(12)\quad (\hat{Q}_n-Q)\Big(\sum_{t=1}^{n} \underline{\xi}(t-1)\underline{\xi}^*(t-1)\Big)^{1/2} \to \int_0^1 \big(d\underline{\omega}(t)\underline{\xi}^*(t)\big)\Big(\int_0^1 \underline{\xi}(t)\underline{\xi}^*(t)\,dt\Big)^{-1/2} =$$

$$= \Big[\int_0^1 d\underline{\xi}(t)\underline{\xi}^*(t)-A_0\int_0^1 \underline{\xi}(t)\underline{\xi}^*(t)\,dt\Big]\Big(\int_0^1 \xi(t)\xi^*(t)\,dt\Big)^{-1/2},$$

in distribution, where $\underline{\xi}(t)$ fulfils (6′) *and $\underline{\omega}(t)$ is a standard Brownian motion.*

Remark. If $E\underline{\varepsilon}(t)\underline{\varepsilon}^*(t)=B_\varepsilon$ the least squares estimate of Q is given by

$$(13)\qquad \hat{Q}_n = \Big(\sum_{t=1}^{n} \underline{\xi}(t)B_\varepsilon^{-1}\underline{\xi}^*(t-1)\Big)\Big(\sum_{t=0}^{n-1} \underline{\xi}(t)B_\varepsilon^{-1}\underline{\xi}^*(t)\Big)^{-1}$$

and

(14)

$$(\hat{Q}_n-Q)\Big(\sum_{t=1}^{n} \underline{\xi}(t-1)B_\varepsilon^{-1}\underline{\xi}^*(t-1)\Big)^{1/2} \to \int_0^1 \big(d\underline{\omega}(t)B_\varepsilon^{-1}\underline{\xi}^*(t)\big)\Big(\int_0^1 \underline{\xi}(t)B_\varepsilon^{-1}\underline{\xi}^*(t)\,dt\Big)^{-1/2}.$$

PROOF. We follow the method of [1] which was used for Gaussian processes. Minimizing in Q the following expression

$$\sum_{t=1}^{n} \big(\underline{\xi}(t)-Q\underline{\xi}(t-1)\big)\big(\underline{\xi}(t)-Q\underline{\xi}(t-1)\big)^* = \sum_{t=1}^{n} \underline{\varepsilon}(t)\underline{\varepsilon}^*(t)$$

we get

$$(15)\qquad \sum_{t=1}^{n} \big(\underline{\xi}(t)-\hat{Q}\underline{\xi}(t-1)\big)\underline{\xi}^*(t-1) = 0.$$

Further, from (1) (multiplying by $\underline{\xi}^*(t-1)$ and summing up)

$$(16)\qquad \sum_{t=1}^{n} \big(\underline{\xi}(t)-Q\underline{\xi}(t-1)\big)\underline{\xi}^*(t-1) = \sum_{t=1}^{n} \underline{\varepsilon}(t)\underline{\xi}^*(t-1).$$

The difference of (15) and (16) and gives

$$(\hat{Q}_n - Q)\Big[\sum_{t=1}^{n} \underline{\xi}(t-1)\underline{\xi}^*(t-1)\Big] = \sum_{t=1}^{n} \underline{\varepsilon}(t)\underline{\xi}^*(t-1).$$

Using transformation (7) for $\underline{\xi}(t)$ and $\underline{\varepsilon}(t)$ one gets a discrete process in $0 \leqq s \leqq 1$. In view of the central limit theorem for martingales ([4], [5], [12]) under conditions (2), (3)

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{k} \underline{\varepsilon}(t) \to \underline{\omega}(s), \quad 0 \leqq s \leqq 1, \quad \frac{k}{n} \to s,$$

and, ([11]),

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \underline{\varepsilon}(t)\underline{\xi}^*(t-1) \to \int_0^1 d\underline{\omega}(t)\underline{\xi}^*(s).$$

On the other hand if $n \to \infty$ equation (1) tends to (6′) and $\underline{\xi}(t)$ tends to the solution of (6′) with a standard Wiener process.

It is known (see [2]) that the only solution of (6′) is a Gauss—Markov type process and

$$\frac{1}{n} \sum_{t=1}^{n} \underline{\xi}(t-1)\underline{\xi}^*(t-1) \to \int_0^1 \underline{\xi}(s)\underline{\xi}^*(s)\, ds,$$

with a Gauss—Markov process $\underline{\xi}(t)$. It is known (see [7], [9], [2]), that

$$\psi_T(A_0, C) = E_{A_0} \exp\Big\{\int_0^1 \underline{\xi}^*(t) C \underline{\xi}(t)\, dt\Big\}$$

has the form

$$\psi_T(A_0, C) = e^{-\mathrm{Trace}\, D} \cdot |I - 2\tilde{D}\tilde{\Gamma}(1)|^{-1/2},$$

here the symmetric matrix D and $\tilde{a}$ satisfy equations

$$DA_0 + A_0^* - 2DD = C,$$

$$2D = A_0 - \tilde{a},$$

and $\Gamma(t)$ is defined by

$$\Gamma(t) = e^{\tilde{a}t}\Gamma(0)e^{\tilde{a}^* t} + \int_0^t e^{\tilde{a}s} \cdot e^{\tilde{a}^* s}\, ds, \quad \Gamma(0) = E(\underline{\xi}(0)\underline{\xi}^*(0)).$$

$\tilde{D}$ and $\tilde{\Gamma}(t)$ are hypermatrices

$$\tilde{D} = \begin{pmatrix} -D & 0 \\ 0 & D \end{pmatrix}, \quad \tilde{\Gamma}(t) = \begin{pmatrix} \Gamma(0) & \Gamma(0)e^{\tilde{a}^* t} \\ e^{\tilde{a}t}\Gamma(0) & \Gamma(t) \end{pmatrix}.$$

The distribution function of $\int_0^1 \underline{\xi}^*(t) C \underline{\xi}(t)\, dt$ is calculated and tabulated only in the one dimensional case (see [2]).

References

[1] M. Arató, On estimators of parameters of processes fulfilling linear stochastic differential equations. *Studia Sci. Math. Hungar.* **5** (1970), 11—15 (in Russian).
[2] M. Arató, Linear stochastic systems with constant coefficients. *Lecture Notes in Control and Information Sci., Springer* (1982), No. **45**, pp. 310.
[3] N. H. Chan and C. Z. Wei, Asymptotic inference for nearly nonstationary AR (1) process, *Annals of Stat.* **15** (1987), No. 3, 1050—1063.
[4] P. Hall and C. Heyde, Martingale limit theory and its applications, *Academic Press, New-York* (1980).
[5] I. S. Helland, Central limit theorems for martingales with discrete or continuous time, *Scand. J. Statist.* **9** (1982), 79—94.
[6] M. G. Kendall and A. Stuart, The advanced theory of statistics, Vol. III., *Ch. Griffin and Hafner* (1966), New-York, London.
[7] K. Koncz, On the parameter estimation of a diffusional type with constant coefficients, *Analysis Math.* **13** (1987), 75—91.
[8] H. Mann and A. Wald, On the statistical treatment of linear stochastic difference equations, *Econometrica* **11** (1943), 173—220.
[9] A. A. Novikov, Estimation of parameters of diffusion processes, *Studia Sci. Math. Hungar.* **7** (1972), 201—209 (in Russian).
[10] M. M. Rao Probability theory and its applications, *Academic Press, New-York* (1975).
[11] H. Rootzén, Limit distributions for the error in approximations of stochastic integrals. *Annals of Probab.* **8** (1980), 244—251.
[12] S. Veres, Asymptotic distributions of likelihood ratios for overparametrized ARMA process, *Journal Time Series Analysis* **8** (1987), No. 3, 345—357.
[13] J. S. White, The limiting distribution of the serial correlation coefficient in the explosive case, *Annals. Math. Statist.* **29** (1958), 1188—1197.

MÁTYÁS ARATÓ
KOSSUTH UNIVERSITY
H—4010, MATH. INSTITUTE
DEBRECEN, HUNGARY

(Received June 22, 1988)