Mathematica Balkanica

Mathematical Society of South-Eastern Europe
A quarterly published by
the Bulgarian Academy of Sciences – National Committee for Mathematics

Mathematica
Balkanica

New Series Vol. 6, 1992, Fasc. 2

Innovations, Sufficient Statistics and Maximum Likelihood in ARMA Models

Georgi N. Boshnakov

Presented by P. Kenderov

Applying the innovations algorithm in a new way we derive a new expression for the likelihood function of the ARMA processes and propose formulas for the computation of the sufficient statistics for ARMA models with fixed moving average parameters.

1. Introduction

Let $\{X_t, t\in Z\}$ be a zero mean ARMA (p, q) process, $\Phi(B)X_t=\theta(B)\varepsilon_t$. Here $\{\varepsilon_t, t\in Z\}$ is an i. i. d. sequence of Gaussian $N(0, \sigma^2)$ random variables, the polynomials $\Phi(z)=1-\Phi_1 z-\ldots-\Phi_p z^p$, and $\theta(z)=1+\theta_1 z+\ldots+\theta_q z^q$, have no common factors and their zeros are in the regions $|z|>1$ and $|z|\geq 1$, respectively. Let n be an integer, $n>\max(p, q)$, and $X=(X_1,\ldots, X_n)'$.

It is known that when $q\neq 0$ the distribution of X does not admit any sufficient statistic whose dimension is independent of n (see M. Arato (1961)). B. W. Dickinson (1982) showed that if the moving average parameters are fixed then there exists a sufficient statistic whose dimension does not depend on n. The proof of Dickinson is in fact an existence one, though in pinciple it shows how to compute that statistic.

In this paper we give a new expression for the likelihood function of an ARMA process. The main term in this expression is quadratic in the autoregressive parameters. Its coefficients are functions of the moving average parameters only. Explicit formulas for the sufficient statistic (in the case of fixed moving average parameters) are obtained from the likelihood function.

The derivation is based on the innovations algorithm (P. S. Brockwell, R. A. Davis (1987)). We apply it to the ARMA case in a new way. The idea is to transform the sample with the autoregressive operator and begin the innovations algorithm from the side of the new autocovariance matrix which does not depend on Φ. We apply the autoregressive operator in the forward direction. Only the lower right corner (with size $\max(p, q)\times\max(p, q)$) of the covariance matrix of the transformed sample depends on the autoregressive parameters. Thus, roughly speaking, only the last $\max(p, q)$ steps in the innovations algorithm depend on them.

2. Notations and definitions

Let us define

$$W_t=\begin{cases}\Phi(F)X_t, & 1\leqq t\leqq n-p,\\ X_t, & n-p+1\leqq t\leqq n,\end{cases}\qquad \hat{W}_t=\begin{cases}0, & t=1,\\ E(W_t\,|\,W_s,\ 1\leqq s\leqq t-1), & 2\leqq t\leqq n.\end{cases}$$

Here F is the forward shift operator ($FX_t=X_{t+1}$). We will use the following notations, along with those in the introduction: $W=(W_1,\dots,W_n)'$, $\hat{W}=(\hat{W}_1,\dots,\hat{W}_n)'$, $\Phi=(\Phi_1,\dots,\Phi_p)'$, $\theta=(\theta_1,\dots,\theta_q)'$; $v_t=\operatorname{var}(W_{t+1}-\hat{W}_{t+1})$, $r_t=v_t/\sigma^2$ (r_t does not depend on σ^2), $t=0,\dots,n-1$; Γ – the covariance matrix of W with entries $\Gamma(t,s)=E(W_tW_s)$;

$$Y_k=\begin{cases}(X_k,\dots,X_{k+p-1})', & 1\leqq k\leqq n-p,\\ (0,\dots,0)', & n-p+1\leqq k\leqq n,\end{cases}$$

$$s(t)=\begin{cases}\min(t,q), & 1\leqq t\leqq\min(n-1,\ n-p+q),\\ t-n+p, & n-p+q+1\leqq t\leqq n-1.\end{cases}$$

It is easy to see that

$$\Gamma(t,s)=\begin{cases}R_q(t-s), & s,t=1,\dots,n-p,\\ R_x(t-s), & n-p+1\leqq s,\ t\leqq n,\\ R_c(t-s), & 1\leqq t\leqq n-p,\ n-p+1\leqq s\leqq n,\\ R_c(s-t), & 1\leqq s\leqq n-p,\ n-p+1\leqq t\leqq n,\end{cases}$$

where $R_x(t-s)=E(X_tX_s)$; $R_c(t-s)=E(W_tX_s)$, $1\leqq t\leqq n-p$, $n-p+1\leqq s\leqq n$; and $R_q(.)$ is the autocovariance function of a $MA(q)$ process with the same $\theta(z)$ and σ^2.

Remark. $E(W_tX_s)=0$ when $s-t>q$, $1\leqq t\leqq n-p$, $n-p+1\leqq s\leqq n$. $R_q(k)=0$ for $k>q$. Therefore, when $q\geqq p$, Γ is a band matrix with band $2q+1$. If $q<p$ then Γ is a band matrix having a $p\times p$ block in the lower right corner. Moreover the entries $\Gamma(t,s)$ of Γ do not depend on Φ when $1\leqq s$, $t\leqq n-p$. Thus, the number of the elements of Γ which depend on Φ, does not depend on n.

3. Results

Lemma 1 is an innovations algoritm application which takes into account the special pattern of Γ. Note that the sums in Lemma 1 involve at most $\max(q,\ p-1)$ terms.

Lemma 1. (*Innovations algorithm for W*). *The following relations hold*:

$$v_0=\Gamma(1,1),\quad \hat{W}_1=0,$$

$$\hat{W}_{t+1}=\sum_{i=1}^{s(t)}\theta_{ti}(W_{t+1-i}-\hat{W}_{t+1-i}),\qquad 1\leqq t\leqq n-1,$$

$$v_t=\Gamma(t+1,\ t+1)-\sum_{i=t-s(t)}^{t-1}\theta^2_{tt-i}v_i,\qquad 1\leqq t\leqq n-1,$$

where $\theta_{tt-k}=0,\ k=0,\ 1,\dots,\ t-s(t)-1$ *and*

$$\theta_{tt-k}=\frac{1}{v_k}(\Gamma(t+1,k+1)-\sum_{i=t-s(t)}^{k-1}\theta_{kk-i}\theta_{tt-i}v_i),\qquad k=t-s(t),\dots,\ t-1.$$

Theorem 1. *$W_t-\hat{W}_t=\alpha_t+\beta_t'\Phi$, where β_t is a p-dimensional vector, α_t is a scalar, $\alpha_1=X_1$, $\beta_1=-Y_2$,*

$$\alpha_t=X_t-\sum_{i=1}^{s(t-1)}\theta_{t-1i}\alpha_{t-i}\quad \beta_t=-(Y_{t+1}+\sum_{i=1}^{s(t-1)}\theta_{t-1i}\ \beta_{t-i}),\qquad 2\leqq t\leqq n.$$

Corollary 1. *α_i and β_i, $i=1,\dots,\ n-p$, do not depend on Φ and σ^2.*

Theorem 2. a) *the log-likelihood function of X is given by*

$$l(X)=-\frac{n}{2}\log(2\pi)-\frac{n}{2}\log(\sigma^2)-l_1(X)-l_2(X),$$

where

$$l_1(X)=\frac{1}{2}\sum_{i=0}^{n-p-1}\log(r_i)+\frac{1}{2\sigma^2}\sum_{i=1}^{n-p}(W_i-\hat{W}_i)^2/r_i,$$

$$l_2(X)=\frac{1}{2}\sum_{i=n-p}^{n-1}\log(r_i)+\frac{1}{2\sigma^2}\sum_{i=n-p+1}^{n}(W_i-\hat{W}_i)^2/r_i,$$

b) $l_1(X)=\frac{1}{2}\sum_{i=0}^{n-p-1}\log(r_i)+S_1(X)/(2\sigma^2)$, *where $S_1(X)=A+B'\Phi+\Phi'C\Phi$, A is a scalar, B is $p\times 1$, C is a $p\times p$ matrix. A, B, C depend only on θ and X by the relations:*

$$A=\sum_{i=1}^{n-p}\alpha_i^2/r_i,\quad B'=2\sum_{i=1}^{n-p}\alpha_i\beta_i'/r_i,\quad C=\sum_{i=1}^{n-p}\beta_i\beta_i'/r_i.$$

Theorem 3. *The quantities A, B, C, α_i, β_i, $i=n-p,\dots,n-p-q+1$, $X_{n-p+1},\dots,\ X_n$, form a sufficient statistic for Φ, σ^2 when θ is fixed.*
The dimension of this statistic is at most $p^2/2+(5/2+q)p+q+1$.

4. Proofs

Lemma 1. follows from Proposition 5.2.2. of P. S. Brockwell and R. A. Davis (1987) and the following properties of the coefficients in the innovations algorithm. For clarity they are formulated as lemmas.

Let R, R_1, and R_2 be nonsingular covariance matrices. Let v_0, $v_0^{(1)}$, $v_0^{(2)}$, $\theta_{k,j}$, $\theta_{k,j}^{(1)}$, $\theta_{k,j}^{(2)}$, v_k, $v_k^{(1)}$, $v_k^{(2)}$, $k=1,\dots,\ t$, $j=0,\dots,\ k-1$, be the coefficients in the innovations algorithm for R, R_1 and R_2, respectively (see P. J. Brockwell and R. A. Davis (1987)).

Lemma 2. *If R_1 is proportional to $R_2 (R_1 = c R_2,\ c>0)$ then $\theta^{(1)}_{k,j} = \theta^{(2)}_{k,j}$, and $v^{(1)}_k = c v^{(2)}_k$, for all k, j.*

Lemma 3. *Let t and $m (t>m)$ be such that $R(t+1, k)=0$ for $k=1,\dots, t-m$. Then*

$$\theta_{tt-k}=0 \qquad k=0, 1,\dots, t-m-1,$$

$$\theta_{tt-(t-m)}=\theta_{tm}=R(t+1,\ t-m+1)/v_k, \qquad k=t-m,$$

$$\theta_{tt-k}=\Big(R(t+1,\ k+1)-\sum_{i=t-m}^{k-1}\theta_{kk-i}\theta_{tt-i}v_i\Big)/v_k, \quad k=t-m+1,\dots, t-1,$$

$$v_t=R(t+1,\ t+1)-\sum_{i=t-m}^{t-1}\theta^2_{tt-i}v_i.$$

In the last two formulas the number of the summands is not greater than m. We omit the proofs of Lemma 2 and Lemma 3 since they are straightforward.

To prove Theorem 1 we note that

$$W_1-\hat W_1=W_1=X_1-\sum_{i=1}^{p}\Phi_i X_{1+i}=X_1-Y'_2\Phi=\alpha_1+\beta'_1\Phi,$$

where $\alpha_1=X_1$, $\beta_1=-Y_2$. Let $t\in[2, n]$. Suppose that $W_i-\hat W_i=\alpha_i+\beta'_i\Phi$, $i=1,\dots, t-1$. Then by Lemma 1

$$W_t-\hat W_t=W_t-\sum_{i=1}^{s(t-1)}\theta_{t-1i}(W_{t-i}-\hat W_{t-i})$$

$$=X_t-Y'_{t+1}\Phi-\sum_{i=1}^{s(t-1)}\theta_{t-1i}(\alpha_{t-i}+\beta'_{t-i}\Phi)$$

$$=\Big(X_t-\sum_{i=1}^{s(t-1)}\theta_{t-1i}\alpha_{t-i}\Big)-\Big(Y_{t+1}+\sum_{i=1}^{s(t-1)}\theta_{t-1i}\beta_{t-i}\Big)'\Phi.$$

By induction we obtain Theorem 1.

Corollary 1 is obtained now from Theorem 1, Lemma 2, and the properties of Γ. Part a) of Theorem 2 holds because the Jacobian of the transformation from X to W is unity, and the two vectors are Gaussian.

We get part b) of Theorem 2. from the equations

$$\sum_{i=1}^{n-p}(W_t-\hat W_t)^2/r_t=\sum_{i=1}^{n-p}(\alpha_t+\beta'_t\Phi)^2/r_t=\sum_{i=1}^{n-p}(\alpha_t^2+2\alpha_t\beta'_t\Phi+\Phi'\beta_t\beta'_t\Phi)/r_t.$$

The assertion of Theorem 3 follows from Theorem 1, Corollary 1, and Theorem 2.

5. Applications

Some problems lead to the estimation of ARMA models with fixed moving average parameters. For example, let $\theta(z)$ be a polynomial of degree q with all roots on the unit circle, $Y_t=\theta(B)X_t$,

$$\begin{cases} H_0=(Y_t \text{ is } \mathrm{AR}(p)), \\ H_1=(Y_t \text{ is } \mathrm{ARMA}(p, q), \text{ with moving average operator } \theta(\ )). \end{cases}$$

By testing the hypothesis H_0 against H_1 one can decide whether to include the nonstationary operator $\theta()$ in the autoregressive model.

The application to the case of unknown moving average parameters is also possible, at least for large n. The sufficient statistic can be used to speed up the maximization of the likelihood function along the autoregressive coordinates.

References

1. M. Arato.On the Sufficienct Statistics for Stationary Gaussian Random Processes. *Theory Prob. Appl.*, **6**, 1961, 199-201.
2. P. J. Brockwell, R. A. Davis. Time Series: Theory and Methods. Springer Series in Statistics, 1987.
3. B. W. Dickinson. Sufficient Statistics for Stationary Discrete-time Gaussian Random Processes. *J. Time Series Anal.*, **3**, **3**, 1982, 165-168.

Institute of Mathematics
Computer Stochastics Laboratory
Bulgarian Academy of Sciences
P. O. B. 373
1090 Sofia
BULGARIA

Received 11.05.1991