Annales Academiæ Scientiarum Fennicæ
Series A. I. Mathematica
Volumen 17, 1992, 111–116

ON REPARAMETRIZATION AND ASYMPTOTICALLY OPTIMAL MINIMAX ESTIMATION IN A GENERALIZED AUTOREGRESSIVE MODEL

S.M. Pergamentshikov and A.N. Shiryaev

Steklov Mathematical Institute, Vavilov Street 42, 117966 Moscow GSP-1, Russia

1. Formulation of the problem

Denote by

$$\mathbf{E}^n = \big(\mathbf{R}^n, B(\mathbf{R}^n); \mathbf{P}^n_\theta, |\theta| \le 1\big)$$

the statistical experiment corresponding to a sequence of observations $(x_1, x_2, \ldots, x_n) \in \mathbf{R}^n$, which satisfy the following (first order) autoregressive model

$$x_k = \theta x_{k-1} + \epsilon_k, \tag{1}$$

where $x_0 = 0$ and (ϵ_k) is a sequence of independent normal random variables, $\epsilon_k \sim N(0,1)$, or which satisfy the generalized model

$$x_k = \theta_k x_{k-1} + \epsilon_k, \tag{2}$$

where $x_0 = 0$ and (θ_k) is a sequence of independent identically distributed random variables with $\mathbf{E}\theta_k = \theta$ and $\mathbf{D}^2\theta_k = \sigma^2$. We assume that the sequences $(\epsilon_k), (\theta_k)$ are independent and that σ^2 is known. The parameter θ is unknown, but we have $\theta^2 + \sigma^2 \le 1$. It is clear that the model (2) includes the model (1) when $\sigma^2 = 0$.

When studying asymptotical ($n \to \infty$) properties of estimators $\hat{\theta}_n$ of the unknown parameter θ one important method is based on the transition from the original experiment $\mathbf{E}^n$ to a specially reparametrized experiment $\mathscr{E}^n$, which converges (in a certain way) to a simple "limiting" experiment $\mathscr{E}$. This allows us to apply, say, asymptotical minimax theorems, from which we obtain results of the form $\liminf R(\mathscr{E}^n) \ge R(\mathscr{E})$, where R is the minimax risk, the key concept in obtaining properties of asymptotic optimality of the corresponding estimators.

The purpose of this paper is to present basic steps to these ideas in connection to the models (1) and (2). Below we show, in particular, that (in a certain class to be specified below) sequential maximum likelihood estimators are asymptotically minimax. We have considered the model (1) (i.e. the case $\sigma^2 = 0$) in [1].

doi:10.5186/aasfm.1992.1714

2. Localization and reparametrization

Consider statistical experiments

$$\mathbf{F}^n = (\Omega^n, F^n; \mathbf{P}^n_\theta, \theta \in \Theta^n),$$

$n \geq 1$, where Θ^n are open subsets in $\mathbf{R}^d$, $\Theta^n \subset \Theta^{n+1}$. Instead of using the experiment $\mathbf{E}^n$ to estimate the true value of the parameter θ ("the true theory") in many cases we can assume that the observer has some idea where the true value θ "is" or "might be". For example, such an idea might be an assumption that "the true value θ belongs to a neighbourhood of the point θ_0", and this will be the null approximation. (In many cases the value θ_0 is based on preliminary "crude" estimators. It may be given by a random variable, e.g. by an estimator based on preliminary observations.)

Assuming that the true value θ belongs to a neighbourhood of a fixed point θ_0 we shall write θ in the form

$$\theta = \theta_0 + \alpha\phi(n, \theta_0), \qquad \alpha \in \mathbf{R}^d, \tag{3}$$

where $\phi = \phi(n, \theta_0)$ must be of the form $\phi(n, \theta_0) \to 0$, $n \to \infty$ ("the larger the number of the observations the closer the approximation to θ_0 under the null hypothesis"). We can view the parameter α as a new parameter to be estimated.

The corresponding reparametrization of the experiment $\mathbf{E}^n$ (transition with the help of localization from the "θ-model" to the "α-model") gives a new experiment $\mathscr{E}^n = \mathscr{E}^n(\theta_0)$ with $\mathscr{E}^n = (\Omega^n, F^n; \mathbf{P}^n_\alpha, \alpha \in \mathscr{A}^n)$, where

$$\mathscr{A}^n = \{\alpha \in \mathbf{R}^d : \theta_0 + \alpha\phi(n, \theta_0) \in \Theta^n\}$$

and $\mathbf{P}^n_\alpha = \mathbf{P}^n_{\theta_0 + \alpha\phi(n,\theta_0)}$.

If $\hat{\alpha}^n$ is an estimator (in the "α-model") and $\hat{\theta}^n = \theta_0 + \hat{\alpha}^n\phi(n, \theta_0)$ is the corresponding estimator in the "θ-model", then

$$\theta - \hat{\theta}^n = \phi(n, \theta_0)(\alpha - \hat{\alpha}^n). \tag{4}$$

Hence, if the loss function $W = W_\theta(a)$ (here θ is the true value of the parameter and a is an "estimate" or a "solution") has the form $W = W(\theta - a)$ then

$$\mathbf{E}^n_\theta W\big(\phi^{-1}(n, \theta_0)(\theta - \hat{\theta}^n)\big) = \mathbf{E}^n_\alpha W(\alpha - \hat{\alpha}^n). \tag{5}$$

It follows that for any $b > 0$

$$\sup_{\{\theta : |\theta - \theta_0| \leq b\phi(n,\theta_0)\}} \mathbf{E}^n_\theta W\big(\phi^{-1}(n, \theta_0)(\theta - \hat{\theta}^n)\big) = \sup_{\{\alpha : |\alpha| \leq b\}} \mathbf{E}^n_\alpha W(\alpha - \hat{\alpha}^n). \tag{6}$$

The above method of transition to the "α-model" depends, of course, on the choice of the sequence $\phi = \phi(n, \theta_0)$. A meaningful choice of this sequence depends on the conditions which guarantee that the sequence of experiments $\mathscr{E}^n$ will converge (in some sense) to a "limiting" experiment $\mathscr{E}$. If the limiting experiment has a simple structure and if the problem of estimating the parameter α has "good" solutions in it, then we can hope that the corresponding result is also true for the experiments $\mathscr{E}^n$, at least for large n.

The requirement "convergence of $\mathscr{E}^n$ to some non-trivial experiment $\mathscr{E}$" automatically leads to such a choice of the norming sequence $\phi = \phi(n, \theta_0) \to 0$, $n \to \infty$, that the family of measures $(\mathbf{P}^n_{\theta_0+\alpha\phi(n,\theta_0)})_{n\geq 1}$ must get closer to the family of measures $(\mathbf{P}^n_{\theta_0})_{n\geq 1}$ for each possible value of the parameter α. Formally this is expressed by saying that the family $(\mathbf{P}^n_{\theta_0+\alpha\phi(n,\theta_0)})_{n\geq 1}$ is contiguous to the family $(\mathbf{P}^n_{\theta_0})_{n\geq 1}$.

3. Reparametrization and minimax estimation in the model (2)

Put $\eta_k = (\theta_k - \theta)/\sigma$. Then the model (2) can be written as

$$x_k = \theta x_{k-1} + (\epsilon_k + \sigma\eta_k)x_{k-1}. \tag{7}$$

Furthermore, for $n \geq 1$ denote $\alpha = (1-\theta)n$, $\Sigma = \sigma\sqrt{n}$, and for $k/n \leq t < (k+1)/n$ define

$$X^{(n)}(t) = \frac{x_k}{\sqrt{n}}, \qquad W^{(n)}(t) = \frac{1}{\sqrt{n}}\sum_{i\leq k}\epsilon_i \qquad \text{and} \qquad A^{(n)}(t) = \frac{k}{n}.$$

It then follows that

$$dX^{(n)}(t) = -\alpha X^{(n)}(t)\,dA^{(n)}(t) + \sqrt{1+\Sigma^2\big(X^{(n)}(t)\big)^2}\,dW^{(n)}(t), \tag{8}$$

where $X^{(n)}(0) = W^{(n)}(0) = 0$ and $(\alpha, \Sigma) \in \mathscr{A}^{(n)}$ with

$$\mathscr{A}^{(n)} = \big\{(\alpha,\Sigma) : \alpha = (1-\theta)n, \Sigma = \sigma\sqrt{n}; \theta^2 + \sigma^2 \leq 1\big\}.$$

Denote by $P^{(n)}_{(\alpha,\Sigma)}$ the probability distribution of the process $X^{(n)}$ considered as a random element in the Skorohod space $(\mathbf{D}, D, (D_t)_{t\geq 0})$ consisting of right continuous functions $x = \big(x(t)\big)_{t\geq 0}$ having left limits equipped with σ-algebras

$$D_t = \bigcap_{s>t}\sigma\{x : x(u), u \leq s\}, \qquad D = \bigvee_t D_t.$$

In this framework define the (filtered) experiment

$$\mathscr{E}^{(n)} = \big\{(\mathbf{D}, D, (D_t)_{t\geq 0}); P^{(n)}_{(\alpha,\Sigma)}, (\alpha,\Sigma) \in \mathscr{A}^{(n)}\big\}.$$

By (8) it is natural to expect that the experiment $\mathscr{E}^{(n)}$ converges (in a certain sense) to the experiment

$$\mathscr{E} = \{(\mathbf{D}, D, (D_t)_{t\geq 0}); P_{(\alpha,\Sigma)}, (\alpha,\Sigma) \in \mathscr{A}\},$$

where $\mathscr{A} = \{(\alpha,\Sigma) : \alpha \geq 0, \Sigma \geq 0\}$ and $P_{(\alpha,\Sigma)}$ is the probability distribution of a diffusion process $X = \big(X(t)\big)_{t\geq 0}$ satisfying

$$dX(t) = -\alpha X(t)\,dt + \sqrt{1+\Sigma^2 X^2(t)}\,dW(t),$$

where $W = \big(W(t)\big)_{t\geq 0}$ is the standard Wiener process.

For $(\alpha,\Sigma) \in \mathscr{A}^{(n)}$ by

$$Z^{(n)}_{(\alpha,\Sigma)}(t) = \frac{dP^{(n)}_{(\alpha,\Sigma)}}{dP^{(n)}_{(0,\Sigma)}}(t)$$

we denote the density of the restriction of the measure $P^{(n)}_{(\alpha,\Sigma)}$ to D_t with respect to the restriction of the measure $P^{(n)}_{(0,\Sigma)}$ to D_t. The density $Z_{(\alpha,\Sigma)}(t)$, $(\alpha,\Sigma) \in \mathscr{A}$, is defined analogously for the experiment $\mathscr{E}$.

In [2] it is proved that for fixed $\Sigma \geq 0$, $p \geq 1$ and $q \geq 1$ we have

$$\begin{aligned} &\mathscr{L}\Big\{Z^{(n)}_{(\alpha_j,\Sigma)}(t_i); i = 1,\dots,p, j = 1,\dots q|P^{(n)}_{(0,\Sigma)}\Big\} \\ &\quad \xrightarrow{w} \mathscr{L}\Big\{Z_{(\alpha_j,\Sigma)}(t_i); i = 1,\dots,p, j = 1,\dots,q|P_{(0,\Sigma)}\Big\} \end{aligned} \tag{10}$$

i.e., the experiments $\mathscr{E}^{(n)}$ converge weakly to the experiment $\mathscr{E}$.

Define stopping times

$$T^{(n)}(\Sigma) = \inf\Big\{\frac{k}{n} : \sum_{\frac{i}{n}\leq\frac{k}{n}} \frac{\big[X^{(n)}((i-1)/n)\big]^2\Delta}{1+\Sigma^2\big[X^{(n)}((i-1)/n)\big]^2} \geq 1\Big\} \qquad (\text{here } \Delta = \frac{1}{n})$$

and

$$T(\Sigma) = \inf\Big\{t : \int_0^t \frac{X^2(s)}{1+\Sigma^2 X^2(s)}\,ds \geq 1\Big\}.$$

Corresponding to (10) it is possible to prove that

$$\begin{aligned} &\mathscr{L}\Big\{Z^{(n)}_{(\alpha_j,\Sigma)}\big(T^{(n)}(\Sigma)\big); j = 1,\dots,q|P^{(n)}_{(0,\Sigma)}\Big\} \\ &\quad \xrightarrow{w} \mathscr{L}\Big\{Z_{(\alpha_j,\Sigma)}\big(T(\Sigma)\big); j = 1,\dots,q|P_{(0,\Sigma)}\Big\}. \end{aligned} \tag{11}$$

But we have

$$\int_0^{T(\Sigma)} \frac{X^2(s)}{1+\Sigma^2 X^2(s)}\,ds = 1 \qquad (P_{(0,\Sigma)}\text{-a.s.})$$

yielding

$$\int_0^{T(\Sigma)} \frac{X(s)dW(s)}{\sqrt{1+\Sigma^2 X^2(s)}} \stackrel{d}{=} \xi,$$

where $\xi \sim N(0,1)$. Hence

(12)

$$Z_{(\alpha,\Sigma)}(T(\Sigma)) = \exp\Big\{ -\alpha \int_0^{T(\Sigma)} \frac{X(s)dW(s)}{\sqrt{1+\Sigma^2 X^2(s)}} + \frac{\alpha^2}{2}\int_0^{T(\Sigma)} \frac{X^2(s)ds}{1+\Sigma^2 X^2(s)}\Big\}$$
$$\stackrel{d}{=} e^{-\alpha\xi+\frac{\alpha^2}{2}},$$

where "d" means equivalence of distributions (with respect to the measure $P_{(0,\Sigma)}$). Formula (12) shows that the "stopped" (at the time $T(\Sigma)$) experiment $\mathscr{E}$ is a Gaussian shift experiment. According to (11) the "stopped" (at the time $T^{(n)}(\Sigma)$) experiments $\mathscr{E}^{(n)}$ converge weakly to $\mathscr{E}$. From this it follows that it is plausible to apply the general asymptotical minimax theorem of Hajek and LeCam (see [4] or [5]), which states that for any continuous symmetrical function W

$$\liminf_n \inf_{\tilde\alpha\in A^n} \sup_{(\alpha,\Sigma)\in\mathscr{A}^n} \mathbf{E}^{(n)}_{(\alpha,\Sigma)} W(\tilde\alpha-\alpha) \geq \mathbf{E}W(\xi), \tag{13}$$

where A^n is a class of estimates of the parameter α, which are based on the observations $X^{(n)}(t)$ for $t \leq T^{(n)}(\Sigma)$. With the notation of model (2) define

$$\tau(h,\sigma) = \inf\Big\{n\geq 1 : \sum_{k=1}^{n} \phi_k^2 \geq h)\Big\},$$

where $\phi_k^2 = (x_k^2)/(1+\sigma^2 x_k^2)$. Then, with $\Sigma = \sigma\sqrt{n}$, we have $T^{(n)}(\Sigma) = (\tau(n^2,\sigma))/n$.

From this, the relation $\alpha = (1-\theta)n$ and (13) it follows that

$$\liminf_{h\to\infty} \inf_{\tilde\theta\in\tilde\Theta^h} \sup_{\theta^2+\sigma^2\leq 1} \mathbf{E}^{(n)}_{(\theta,\sigma)} W(\sqrt{n}(\tilde\theta-\theta)) \geq \mathbf{E}W(\xi), \tag{14}$$

where $\tilde\Theta^h$ is a class of estimates of the parameter θ, based on the observations $x_1, x_2, \dots, x_{\tau(h,\sigma)}$.

We consider then sequential maximum likelihood estimators

$$\hat\theta_{\tau(h,\sigma)} = \frac{\sum_k^{\tau(h,\sigma)}(x_{k-1}x_k)/(1+\sigma^2 x_{k-1}^2)}{\sum_k^{\tau(h,\sigma)} \phi_{k-1}^2}.$$

From [1], [2] and [3] it follows that

$$\lim_{n\to\infty} \sup_{\theta^2+\sigma^2\le 1} \left| P_{(\theta,\sigma)} \left(\sqrt{n}(\hat\theta_{\tau(h,\sigma)} - \theta) \le x \right) - \Phi(x) \right| \to 0, \tag{15}$$

where

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2} dy$$

and $P_{(\theta,\sigma)}$ is the distribution of the sequence $(x_k)_{k\ge 1}$, given (θ,σ). Combining (14) and (15) we get the following result.

Theorem. *In the model* (2) *the sequential estimators* $\hat\theta_{\tau(h,\sigma)}$, $h \to \infty$, *are asymptotically minimax for any continuous bounded symmetric loss function* $W = W(a)$, $W(0) = 0$, *that is*

$$\lim_{h\to\infty} \inf_{\tilde\theta\in\tilde\Theta^h} \sup_{\theta^2+\sigma^2\le 1} \mathbf{E}_{(\theta,\sigma)} W\big(\sqrt{h}(\tilde\theta - \theta)\big) = \lim_{h\to\infty} \sup_{\theta^2+\sigma^2\le 1} \mathbf{E}_{(\theta,\sigma)} W\big(\sqrt{h}(\hat\theta_{\tau(h,\sigma)} - \theta)\big).$$

References

[1] GREENWOOD, P.E., and A.N. SHIRYAEV: Sequential estimation for first order autoregressive models. - Stochastics (to appear).

[2] PERGAMENTSHIKOV, S.M., and A.N. SHIRYAEV: On the sequential estimation of a parameter of stochastic difference equations with random coefficients. - Theory Probab. Appl. (to appear).

[3] LAI, T.L., and D. SIEQMUND: Fixed accuracy estimation of an autoregressive parameter. - Ann. Statist. 11, 1983, 478–485.

[4] LECAM, L.: Asymptotic methods in statistical decision theory. - Springer-Verlag, Berlin, 1986.

[5] STRASSER, H.: Mathematical theory of statistics. - De Gruyter, Berlin, 1985.