Ann. Inst. Statist. Math.
Vol. 40, No. 3, 507-520 (1988)

# ON STOCHASTIC ESTIMATION

JENS-PETER KREISS

*Institut für Mathematische Stochastik, Universität Hamburg, Bundesstrasse 55, 2000 Hamburg 13, West Germany*

(Received September 29, 1986; revised July 20, 1987)

**Abstract.** We consider a local random searching method to approximate a root of a specified equation. If such roots, which can be regarded as estimators for the Euclidean parameter of a statistical experiment, have some asymptotic optimality properties, the local random searching method leads to asymptotically optimal estimators in such cases. Application to simple first order autoregressive processes and some simulation results for such models are also included.

*Key words and phrases*: Autoregressive process, local asymptotic normality, Monte Carlo, parameter estimation, stochastic search.

## 1. Introduction

In this paper we deal with a stochastic estimation procedure for a Euclidean parameter in statistical experiments. The main example for which these methods are developed is a stochastic process of autoregressive type. The method of stochastic estimation was introduced by Beran and Millar (1987). The main contribution of the present paper is the idea that a root of a specified equation can be achieved by a local random searching method. Since such roots, which may be regarded as estimators for the parameter of interest, have some optimality properties, the local random searching method leads to asymptotically optimal estimators in such situations. The methods used in the proofs of this paper follow those of Beran and Millar (1987), but the situation to which their ideas are applied is different.

This paper is organized as follows: Section 2 introduces our main example of autoregressive (AR) processes and contains some useful properties for easy reference. In Section 3 we give the main results for the described local random searching method for arbitrary statistical experiments, while Section 4 contains applications to the autoregressive case and some simulation results. Finally we prove a version of a convergence lemma needed in Section 3.

## 2. Autoregressive processes and some characteristics

Consider real valued and stationary random variables $(X_t; t \in \boldsymbol{Z} = \{0, \pm 1, \pm 2, \dots\})$ which are solutions of the following stochastic difference equation

$$X_t = \sum_{i=1}^{p} a_i X_{t-i} + e_t, \qquad t \in \boldsymbol{Z}\,. \tag{2.1}$$

The errors $(e_t; t \in \boldsymbol{Z})$ consist of independent and identically distributed (i.i.d.) random variables with zero mean and positive variance $\sigma^2$. Further we assume that this model has the LAN-property (local asymptotic normality). Assumptions which guarantee this property are contained in Kreiss (1987). Especially we need the existence of an absolute continuous density $f$ of the distribution of $e_1$, which has to have finite Fisher-Information $I(f) = \int (f'/f)^2 f d\lambda$.

In the sequel we use the following notations:

$$\theta = (a_1, \dots, a_p) \in \Theta = \left\{ (a_1, \dots, a_p) \in \boldsymbol{R}^p \,|\, 1 - \sum_{i=1}^{p} a_i z^i \text{ has no zeros with magnitude less or equal to one} \right\}.$$

This definition of the parameter space ensures that stationary solutions $(X_t; t \in \boldsymbol{Z})$ of (2.1) exist (cf. Fuller (1976), Theorem 2.6.1).

$$P_{n,\theta} = \mathscr{L}(X_{1-p}, \dots, X_n | \theta), \qquad \dot{\varphi} = -f'/f, \qquad X(j-1) = (X_{j-1}, \dots, X_{j-p})^T,$$

$\Gamma(\theta) = (E_\theta X_s X_t)_{s,t=1,\dots,p}$ denotes the $p \times p$-covariance-matrix of the process if $\theta$ is the underlying parameter.

The LAN-property is equivalent to: For all sequences $\{\theta_n\} \subset \Theta$ for which $\sqrt{n}(\theta_n - \theta)$ stays bounded, we have (2.2) and (2.3).

$$\log \frac{dP_{n,\theta_n}}{dP_{n,\theta}} - \sqrt{n}(\theta_n - \theta)\, \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \dot{\varphi}(e_j) X(j-1) + \frac{1}{2} \sqrt{n}(\theta_n - \theta) \Gamma(\theta) I(f) \sqrt{n}(\theta_n - \theta)^T \to 0\,, \tag{2.2}$$

in $P_{n,\theta}$-probability, where $e_j = e_j(\theta) = X_j - \theta X(j-1)$.

$$\mathscr{L}\left( \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \dot{\varphi}(e_j) X(j-1) - I(f) \Gamma(\theta) \sqrt{n}(\theta_n - \theta)^T \,|\, P_{n,\theta_n} \right) \tag{2.3}$$

$$\Rightarrow \mathcal{N}(0, \Gamma(\theta) I(f)),$$

where "$\Rightarrow$" denotes weak convergence.

From Kreiss (1987) we have that, under some additional regularity conditions on the score-function $\dot{\varphi}$, asymptotically optimal estimators (in the local asymptotic minimax-sense) $\{\hat{\theta}_n^{opt}\}$ for $\theta$ may be characterized through

$$\Delta_n(\hat{\theta}_n^{opt}) \to 0 \quad \text{in } P_{n,\theta}\text{-probability},$$

where

$$\Delta_n(\theta) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \dot{\varphi}(X_j - \theta X(j-1)) X(j-1). \tag{2.4}$$

Also $M$-estimators $\{\hat{\theta}_n^M\}$ for $\theta$ in this autoregressive setup may be defined through

$$\Psi_n(\hat{\theta}_n^M) \to 0 \quad \text{in } P_{n,\theta}\text{-probability},$$

where

$$\Psi_n(\theta) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \psi(X_j - \theta X(j-1)) X(j-1), \tag{2.5}$$

and $\psi: \boldsymbol{R} \to \boldsymbol{R}$ denotes a suitable score-function (cf. Kreiss (1985)).

Our aim is to construct an estimator which is easy to compute and which is asymptotically equivalent up to order $1/\sqrt{n}$ to an estimator which fulfils (2.4) or (2.5). This aim is achieved by a local random searching method as defined in Sections 3 and 4.

The results are also of practical interest, since it is often difficult to compute the exact solution of $\Delta_n(\theta) \equiv 0$ and to prove $\sqrt{n}$-consistency for such a solution, for example.

## 3. Stochastic estimation for statistical experiments

In this section we do not refer to the special model considered in Section 2, but we consider the following sequence $E_n = (\mathcal{X}_n, \mathcal{B}_n, (P_{n,\theta}; \theta \in \Theta \subset \boldsymbol{R}^k))$ of experiments. Here $\mathcal{X}_n$ denotes an arbitrary metric space and $\mathcal{B}_n$ the corresponding Borel $\sigma$-algebra. $\Theta$ is an open subset of $\boldsymbol{R}^k$.

$$T_n: \mathcal{X}_n \times \Theta \to \boldsymbol{R}^m, \tag{3.1}$$

denotes a statistic, i.e., $T_n$ is measurable with respect to $\mathscr{B}_n \otimes (\Theta \cap \boldsymbol{B}_k)$, which is continuous in $\theta$ for each fixed $x \in \mathscr{X}_n$. Here $\boldsymbol{B}_k$ denotes the $k$-dimensional Borel $\sigma$-algebra on $\boldsymbol{R}^k$.

Assume that we are interested in estimators $\{\hat{\theta}_n^{opt}\}$, which have the following property

$$T_n(\cdot, \hat{\theta}_n^{opt}(\cdot)) \to 0 \quad \text{in } P_{n,\theta}\text{-probability}. \tag{3.2}$$

In the independent situation examples are well-known, while more complicated examples for dependent observations are considered in Sections 2 and 4 of the present paper.

Our aim is to improve a sequence of estimators $\{\bar{\theta}_n\}$ which is $\sqrt{n}$-consistent, via simulation, in such a way that the improved version has property (3.2), and is asymptotically equivalent to $\{\hat{\theta}_n^{opt}\}$ up to order $1/\sqrt{n}$. To do so, we use a local random searching procedure defined as follows:

Step 1: Simulate random variables $t_1, \ldots, t_{j_n}$ according to $\mu_n(x_n, \cdot)$, which denotes a Markov kernel on $\mathscr{X}_n \times \boldsymbol{B}_k$, where $x_n$ represents the observation.

Step 2: As an improved estimator use ($\|\cdot\|$ denotes the Euclidean norm on $\boldsymbol{R}^k$)

$$\hat{\theta}_n = \bar{\theta}_n + n^{-1/2} t_{k_n},$$

if and only if

$$\| T_n(x_n, \bar{\theta}_n + n^{-1/2} t_{k_n}) \| = \min_i \{ \| T_n(x_n, \bar{\theta}_n + n^{-1/2} t_i) \| : 1 \le i \le j_n \}.$$

Beran and Millar (1987) introduce a similar local random searching method to (among other things) compute the MLE in complicated parametric situations. The main contribution of this paper is the idea that the solution of an approximative equation $T_n(\theta) \equiv 0$ can be achieved by a local random searching procedure as defined above. This result is the content of the following theorem. The method of proof follows that of Beran and Millar (1987).

THEOREM 3.1. *Assume*:

(A1) *There exists a sequence $\{\bar{\theta}_n\}$ of estimators, such that $\sqrt{n}(\bar{\theta}_n - \theta)$ stays bounded in $P_{n,\theta}$-probability.*

(A2) $\sup_{\theta_1} \{ \| T_n(\theta_1) - T_n(\theta) + \Sigma\sqrt{n}(\theta_1 - \theta)^T \| : \|\theta_1 - \theta\| \le C/n^{1/2-\alpha} \} \to 0,$

*in $P_{n,\theta}$-probability as $n \to \infty$, for a suitable $m \times k$-matrix $\Sigma$ with rank $k$ and a suitable constant $\alpha > 0$ (we say that $T_n$ admits an asymptotic expansion).*

(A3) *$\mu_n(x_n, \cdot)$ converges in $P_{n,\theta}$-probability to $\mu_o$, a probability measure on $\boldsymbol{R}^k$, which is continuous and gives positive mass to each open* *set. Further* $\mu_n\{x|\,\|x\| > c_0 \cdot n^{\alpha}\} = 0$, *for all large n.*

*Assume that a* $\sqrt{n}$*-consistent sequence* $\{\hat{\theta}_n^{opt}\}$ *of estimators exists, which satisfies* (3.2). *Suppose* $j_n \to \infty$ *as* $n \to \infty$. *Then for all* $\varepsilon > 0$ *and all* $n \in \boldsymbol{N} = \{1, 2, \dots\}$, $\Omega_{n\varepsilon} \in \mathscr{B}_n \otimes \boldsymbol{B}_{kj_n}$, $P_{n,\theta} \otimes \mu_n^{j_n}(\Omega_{n\varepsilon}) \geq 1 - \varepsilon$, *exists such that for all sequences* $\{(x^n, t_1^n, \dots, t_{j_n}^n)\}_n$, $(x^n, t_1^n, \dots, t_{j_n}^n) \in \Omega_{n\varepsilon}$: (3.3)

$$\sqrt{n}(\hat{\theta}_n(x^n, t_1^n, \dots, t_{j_n}^n) - \hat{\theta}_n^{opt}) \to 0 \quad as \quad n \to \infty . \tag{3.3}$$

*Remark.* According to the paper of Lohse (1987) the assertion of the above theorem is equivalent to

$$\sqrt{n}(\hat{\theta}_n - \hat{\theta}_n^{opt}) \to 0, \quad \text{as} \quad n \to \infty, \quad \text{in } P_{n,\theta} \otimes \mu_n^{j_n}\text{-probability} ,$$

if $\{\hat{\theta}_n\}$ is measurable.

The proof of the theorem follows from the following three lemmas.

LEMMA 3.1. *For* $W_n(x_n; u) := T_n(x_n, \hat{\theta}_n^{opt}(x_n) + n^{-1/2}u)$, *which is an element of* $C = C(\boldsymbol{R}^k, \boldsymbol{R}^m)$ *for* $x_n \in \mathscr{X}_n$ *fixed, we have*

$$\sup_{\|u\| \leq n^{\alpha} \cdot c} \| W_n(x_n; u) + \Sigma u \| \xrightarrow{n \to \infty} 0 \quad \text{in } P_{n,\theta}\text{-probability} . \tag{3.4}$$

*This implies, of course,* $W_n \xrightarrow{C} -\Sigma u$ *in* $P_{n,\theta}$*-probability.*

PROOF. First, because of continuity

$$\begin{aligned}
&\sup\{\| T_n(x_n, \hat{\theta}_n^{opt} + n^{-1/2}u) - T_n(x_n, \hat{\theta}_n^{opt}) + \Sigma u \| : \|u\| \leq n^{\alpha} \cdot c\} \\
&\quad = \| T_n(x_n, \hat{\theta}_n^{opt} + n^{-1/2}\hat{u}_n(x_n)) - T_n(x_n, \hat{\theta}_n^{opt}) + \Sigma \hat{u}_n(x_n) \|, \quad \text{say} \\
&\quad \leq \| T_n(x_n, \hat{\theta}_n^{opt} + n^{-1/2}\hat{u}_n) - T_n(x_n, \theta) + \Sigma \sqrt{n}(\hat{\theta}_n^{opt} + n^{-1/2}\hat{u}_n - \theta) \| \\
&\qquad + \| T_n(x_n, \hat{\theta}_n^{opt}) - T_n(x_n, \theta) + \Sigma \sqrt{n}(\hat{\theta}_n^{opt} - \theta) \| \\
&\quad \xrightarrow{n \to \infty} 0 \quad \text{in } P_{n,\theta}\text{-probability} ,
\end{aligned}$$

because of (A2). Together with (3.2) this implies assertion (3.4). To see the convergence in $C$, consider Fahrmeir (1973), Satz 2.1b. □

For each $A \in \boldsymbol{B}_k$ define the following two empirical measures

$$\hat{\mu}_n(t_1, \dots, t_{j_n}; A) = \frac{1}{j_n} \sum_{\nu=1}^{j_n} 1_A(t_\nu) , \tag{3.5}$$

and

$$\hat{\mu}_{nc}(A) = \hat{\mu}_n(\{a - \sqrt{n}(\bar{\theta}_n - \hat{\theta}_n^{opt}) | a \in A\}) . \tag{3.6}$$

$\hat{\mu}_{nc}$ is a randomly centered version of $\hat{\mu}_n$ with support $\{t_\nu + \sqrt{n}(\bar{\theta}_n - \hat{\theta}_n^{opt}) | \nu \in \{1,\dots,j_n\}\}$. For these discrete measures the following results hold:

LEMMA 3.2.
(i) $\hat{\mu}_n \Rightarrow \mu_o$ *in* $P_{n,\theta} \otimes \mu_n^{j_n}$-*probability.*
(ii) *For all* $\varepsilon > 0$ *and all* $n \in \boldsymbol{N}$, $\Omega_{n\varepsilon} \in \mathscr{B}_n \otimes \boldsymbol{B}_{kj_n}$ *exists with* $P_{n,\theta} \otimes \mu_n^{j_n}(\Omega_{n\varepsilon}) \geq 1 - \varepsilon$ *such that for all sequences* $\{(x^n, t_1^n, \dots, t_{j_n}^n)\}_n$, $(x^n, t_1^n, \dots, t_{j_n}^n) \in \Omega_{n\varepsilon}$:

$$\{\hat{\mu}_{nc}(\cdot)\}_{n\in N} \quad \textit{is tight} . \tag{3.7}$$

PROOF. (i) is shown in Beran *et al.* (1987). From (i) we have, according to the paper of Lohse (1987), the existence of a sequence $\Omega_{n\varepsilon}^* \in \mathscr{B}_n \otimes \boldsymbol{B}_{kj_n}$, $P_{n,\theta} \otimes \mu_n^{j_n}(\Omega_{n\varepsilon}^*) \geq 1 - \varepsilon$ such that for all sequences $\{(x^n, t_1^n, \dots, t_{j_n}^n)\}_{n\in N}$, $(x^n, t_1^n, \dots, t_{j_n}^n) \in \Omega_{n\varepsilon}^*$:

$$\hat{\mu}_n(t_1^n, \dots, t_{j_n}^n; \cdot) \Rightarrow \mu_o .$$

Now define ($|\cdot|$ denotes the sup-norm on $\boldsymbol{R}^k$)

$$\Omega_{n\varepsilon} = (\{x^n | \sqrt{n}|\bar{\theta}_n(x^n) - \theta| \leq M_\varepsilon, \sqrt{n}|\hat{\theta}_n^{opt}(x^n) - \theta| \leq M_\varepsilon\} \times \boldsymbol{R}^{kj_n}) \cap \Omega_{n\varepsilon}^* .$$

Because of $\sqrt{n}$-consistency of $\{\bar{\theta}_n\}$ and $\{\hat{\theta}_n^{opt}\}$ we get $P_{n,\theta} \otimes \mu_n^{j_n}(\Omega_{n\varepsilon}) \geq 1 - 2\varepsilon$. Now we have: For each sequence $\{(x^n, t_1^n, \dots, t_{j_n}^n)\}_n$, $(x^n, t_1^n, \dots, t_{j_n}^n) \in \Omega_{n\varepsilon}$ and for all $\delta > 0$ a compact set $K \in \boldsymbol{B}_k$ exists, such that $\hat{\mu}_n(t_1^n, \dots, t_{j_n}^n; K) \geq 1 - \delta$ for all $n$ (without loss of generality assume $K = [-\kappa, \kappa]^k$). Let $K_\varepsilon^* = [-\kappa - 2M_\varepsilon, \kappa + 2M_\varepsilon]^k$ and observe

$$K \subset \{k - \sqrt{n}(\bar{\theta}_n(x^n) - \hat{\theta}_n^{opt}(x^n)) | k \in K_\varepsilon^*\} .$$

From this follows

$$\hat{\mu}_n(\{k - \sqrt{n}(\bar{\theta}_n(x^n) - \hat{\theta}_n^{opt}(x^n)) | k \in K_\varepsilon^*\}) \geq \hat{\mu}_n(K) \geq 1 - \delta ,$$

for all $n \in N$, so that (3.7) holds. □

For the discrete probability measure $\hat{\mu}_{nc}$ on $\boldsymbol{R}^k$ and a continuous function $g$: $\boldsymbol{R}^k \to \boldsymbol{R}^m$ define

$$\Lambda(\hat{\mu}_{nc}, g) = t_{k_n} + \sqrt{n}(\bar{\theta}_n - \hat{\theta}_n^{opt}) , \tag{3.8}$$

if $\|g(t_{k_n} + \sqrt{n}(\bar{\theta}_n - \hat{\theta}_n^{opt}))\| = \min_\nu \{\|g(t_\nu + \sqrt{n}(\bar{\theta}_n - \hat{\theta}_n^{opt}))\| : 1 \leq \nu \leq j_n\}$. Then we finally have:

LEMMA 3.3. *For each* $\varepsilon > 0$ *and* $n \in N$ *a sequence* $\Omega_{n\varepsilon} \in \mathscr{B}_n \otimes \boldsymbol{B}_{k \cdot j_n}$ *exists with* $P_{n,\varepsilon} \otimes \mu_n^{j_n}(\Omega_{n\varepsilon}) \geq 1 - \varepsilon$, *such that for all sequences* $\{(x^n, t_1^n, \ldots, t_{j_n}^n)\}_n$, $(x^n, t_1^n, \ldots, t_{j_n}^n) \in \Omega_{n\varepsilon}$:

$$\Lambda(\hat{\mu}_{nc}, W_n) \to 0 \quad \text{as} \quad n \to \infty .$$

PROOF. For $\varepsilon > 0$ and $n \in N$ choose $\Omega_{n\varepsilon}$ such way that for all sequences $\{(x^n, t_1^n, \ldots, t_{j_n}^n)\}_n$, $(x^n, t_1^n, \ldots, t_{j_n}^n) \in \Omega_{n\varepsilon}$ the assertions of Lemmas 3.1 and 3.2 hold and that $\sqrt{n}(\bar{\theta}_n(x^n) - \theta_n^{opt}(x^n))$ stays bounded (cf. Lohse (1987)).

Under these assumptions we can conclude that each subsequence $\{n_1\} \subset \{n\}$ contains a further subsequence $\{m\} \subset \{n_1\}$ such that for a suitable probability measure $\mu$ on $\boldsymbol{R}^k$

$$\hat{\mu}_m(t_1^m, \ldots, t_{j_m}^m, \{a - \sqrt{m}(\bar{\theta}_m(x^m) - \theta_m^{opt}(x^m)) | a \in \cdot\}) \Rightarrow \mu , \tag{3.9}$$

and

$$\kappa_m := \sqrt{m}(\bar{\theta}_m(x^m) - \theta_m^{opt}(x^m)) \to \kappa_o \quad \text{as} \quad n \to \infty . \tag{3.10}$$

We have for each $(s, t] \subset \boldsymbol{R}^k$: (Assume that $\|\kappa_o - \kappa_m\| \leq \delta$ holds for $m \geq m_o(\delta)$)

$$\begin{aligned}
&|\hat{\mu}_m\{a - \kappa_m | s < a \leq t\} - \hat{\mu}_m\{a - \kappa_o | s < a \leq t\}| \\
&= |\hat{\mu}_m(s - \kappa_m, t - \kappa_m] - \hat{\mu}_m(s - \kappa_o, t - \kappa_o]| \\
&\leq \hat{\mu}_m(s - \kappa_o - \delta, t - \kappa_o + \delta] - \hat{\mu}_m(s - \kappa_o, t - \kappa_o] \qquad \text{for } m \geq m_o \\
&= \hat{\mu}_m(s - \kappa_o - \delta, s - \kappa_o] + \hat{\mu}_m(t - \kappa_o, t - \kappa_o + \delta] \\
&\leq \delta \ \text{ for } m \text{ large enough, because of Lemma 3.2 (i) and the continuity of } \mu_o .
\end{aligned}$$

From this and Lemma 3.2 (i) we conclude

$$\hat{\mu}_m\{a - \kappa_m | a \in \cdot\} \Rightarrow \mu_o\{a - \kappa_o | a \in \cdot\} . \tag{3.11}$$

Additionally we have (cf. (A3)):

$$\begin{aligned}
\text{supp}\,(\hat{\mu}_{mc}) := \{t_\nu + \sqrt{m}(\bar{\theta}_m(x^m) - \theta_m^{opt}(x^m)) &: \nu = 1, \ldots, j_m\} \\
&\subset \{x | \ \|x\| \leq c_o \cdot m^a + 2\kappa_o\}.
\end{aligned}$$

Moreover, Lemma 3.1 yields

(3.12)
$$\sup_u \{\| W_n(x^n; u) + \Sigma u \|, u \in K_n\} \xrightarrow{n\to\infty} 0 \qquad \text{for} \quad K_n = \{u | \ \|u\| \le Cn^{\alpha}\},$$

and $\Sigma u = 0 \Leftrightarrow u = 0$, because of (A2).

The convergence lemma (cf. Appendix) implies

(3.13)
$$\Lambda(\hat{\mu}_{mc}, W_m) \to 0, \quad \text{as} \quad n \to \infty .$$

Since for each subsequence $\{n_1\} \subset \{n\}$ there exists a further subsequence $\{m\} \subset \{n_1\}$ such that (3.13) holds, the whole sequence $\Lambda(\hat{\mu}_{nc}, W_n)$ converges to zero. □

Now we are ready to give

PROOF OF THEOREM 3.1. $\hat{\mu}_{nc}$ has the following support $\{t_\nu + \sqrt{n} \cdot(\bar{\theta}_n - \hat{\theta}_n^{opt}): \nu = 1,\dots,j_n\}$. From the definition of $W_n$ (cf. Lemma 3.1), we see

(3.14)
$$\Lambda(\hat{\mu}_{nc}, W_n) = t_{k_n} + \sqrt{n}(\bar{\theta}_n(x_n) - \hat{\theta}_n^{opt}(x_n)) ,$$

where $t_{k_n}$ is defined in Step 2 at the beginning of this section. Recall the definition of $\{\hat{\theta}_n\}$ to observe

$$\Lambda(\hat{\mu}_{nc}, W_n) = \sqrt{n}(\hat{\theta}_n - \hat{\theta}_n^{opt}) .$$

Now Lemma 3.3 gives the assertion. □

*Remark.* Under suitable regularity conditions it is often possible to prove the following central limit theorem for $\{\hat{\theta}_n^{opt}\}$:

(3.15)
$$\mathscr{L}(\sqrt{n}(\hat{\theta}_n^{opt} - \theta) | P_{n,\theta}) \Rightarrow \mathscr{N}(0, \Gamma) ,$$

for a positive definite $k \times k$-matrix $\Gamma$. Theorem 3.1 says that for the stochastic estimator $\{\hat{\theta}_n\}$, which is defined on $\mathscr{X}_n \times \boldsymbol{R}^{k \cdot j_n}$, the same asymptotic distribution appears, more exactly

(3.16)
$$\mathscr{L}(\sqrt{n}(\hat{\theta}_n - \theta) | P_{n,\theta} \otimes \mu_{n}^{j_n}) \Rightarrow \mathscr{N}(0, \Gamma) ,$$

if $\{\hat{\theta}_n\}$ is measurable.

The following section is devoted to an application of the above result to a simple autoregressive model of first order.

## 4. Application to first order autoregression

In this section we consider real valued random variables $\{X_t; t \in \boldsymbol{Z}\}$ which are stationary solutions of

$$X_t = \theta X_{t-1} + e_t, \quad t \in \boldsymbol{Z}, \quad |\theta| < 1\,, \tag{4.1}$$

where the i.i.d. random variables $\{e_t;\, t \in \boldsymbol{Z}\}$ are as in Section 2. $\Theta = (-1, 1)$. We are interested in estimators $\{\hat{\theta}_n^{opt}\}$ which fulfil

$$\Delta_n(\hat{\theta}_n^{opt}) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \dot{\varphi}(X_j - \hat{\theta}_n^{opt} X_{j-1}) X_{j-1} \to 0 \quad \text{in } P_{n,\theta}\text{-probability}\,, \tag{4.2}$$

or

$$\Psi_n(\hat{\theta}_n^{opt}) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \psi(X_j - \hat{\theta}_n^{opt} X_{j-1}) X_{j-1} \to 0 \quad \text{in } P_{n,\theta}\text{-probability}\,, \tag{4.3}$$

where $\psi: \boldsymbol{R} \to \boldsymbol{R}$ denotes a suitable score-function, e.g., $\psi \equiv \arctan$.

As an initial estimator, which will be improved by the stochastic procedure defined below, we use the least squares estimator:

$$\bar{\theta}_n(\tilde{x}_n) = \frac{\sum_{j=1}^{n} x_j x_{j-1}}{\sum_{j=1}^{n} x_{j-1}^2}\,, \tag{4.4}$$

if observations $\tilde{x}_n = (x_0, \ldots, x_n)$ are available.

It is well-known that this estimator fulfils (4.2), if and only if the errors $e_t$ are normally distributed. To improve $\{\bar{\theta}_n\}$ in non-normal situations we proceed as follows $(0 < \alpha < 1/4)$.

Step 1: Simulate $t_1, \ldots, t_{j_n}$ according to $\mu_n$, where $\mu_n$ denotes the distribution of

$$U_n = \begin{cases} \tau \cdot X, & |X| \le n^{\alpha} \\ 0, & \text{otherwise} \end{cases}, \quad X \sim \mathcal{N}(0, 1), \quad \tau > 0\,.$$

Step 2: As an improved estimator use $\hat{\theta}_n = \bar{\theta}_n + n^{-1/2} t_{k_n}$, if $|\Delta_n(\hat{\theta}_n)| = \min_i \{|\Delta_n(\bar{\theta}_n + n^{-1/2} t_i)| : 1 \le i \le j_n\}$. Note that the distribution according to which we draw the random variables $t_i$ does not depend on the observations $\tilde{x}_n$.

Of course we use $\Psi_n(\theta)$ instead of $\Delta_n(\theta)$, if we are searching for $M$-estimators, i.e., estimators which fulfil (4.3).

The remaining task of this section will be to establish assumptions (A1)–(A3) (cf. Theorem 3.1) in order to ensure that our improved estimator $\{\hat{\theta}_n\}$ behaves asymptotically as well as $\{\hat{\theta}_n^{opt}\}$.

Since (A1) holds and $\mu_n \Rightarrow \mathcal{N}(0, \tau^2)$ we have to prove

$$\sup_{\theta_1}\left\{|\Delta_n(\theta_1) - \Delta_n(\theta) + \Sigma\sqrt{n}(\theta_1 - \theta)| : |\theta_1 - \theta| \le \frac{C}{n^{1/2-\alpha}}\right\} \to 0 \tag{4.5}$$

in $P_{n,\theta}$-probability .

To see this assume that $\dot{\varphi} = -f'/f$ is absolute continuous, that $\dot{\varphi}'$ obeys a global Lipschitz-condition, and $E_\theta X_0^4 < \infty, \int \dot{\varphi}'(x)^2 f(x)dx < \infty$. We have

$$\begin{aligned}
&|\Delta_n(\theta_1) - \Delta_n(\theta) + \Sigma\sqrt{n}(\theta_1 - \theta)| \\
&= \left|(\theta - \theta_1)\frac{1}{\sqrt{n}}\sum_{j=1}^n \int_0^1 \dot{\varphi}'(x_j - \theta x_{j-1} - \lambda(\theta_1 - \theta)x_{j-1})d\lambda\, x_{j-1}^2 \right.\\
&\qquad\qquad\qquad\qquad \left. - \Sigma\sqrt{n}(\theta - \theta_1)\right| \\
&\le |n^{1/2-\alpha}(\theta - \theta_1)|\left[\left|\frac{1}{n^{1-\alpha}}\sum_{j=1}^n\{\dot{\varphi}'(x_j - \theta x_{j-1}) \cdot x_{j-1}^2 - \Sigma\}\right|\right.\\
&\qquad\qquad \left. + \left|\frac{O(1)}{n^{1-\alpha}}\sum_{j=1}^n |x_{j-1}|^3 (\theta_1 - \theta)\right|\right].
\end{aligned}$$

Since

$$\begin{aligned}
&E_\theta\left(\frac{1}{n^{1-\alpha}}\sum_{j=1}^n\{\dot{\varphi}'(e_j(\theta))x_{j-1}^2 - \int \dot{\varphi}'(x)f(x)dx\, x_{j-1}^2\}\right)^2 \\
&= \frac{1}{n^{2(1-\alpha)}}\sum_{j=1}^n E_\theta\left\{\dot{\varphi}'(e_j(\theta)) - \int\dot{\varphi}'(x)f(x)dx\right\}^2 E_\theta X_0^4 \\
&= o(1)\,, \\
&E_\theta\frac{1}{n^{3/2-2\alpha}}\sum_{j=1}^n |x_{j-1}|^3 = O(n^{2\alpha-1/2}) = o(1)\,,
\end{aligned}$$

because $\alpha < 1/4$, and

$$n^\alpha \cdot \left(\frac{1}{n}\sum_{j=1}^n x_{j-1}^2 - E_\theta X_0^2\right) \to 0 \quad \text{in } P_{n,\theta}\text{-probability}\,,$$

we obtain (4.5) with $\Sigma = \int \dot{\varphi}'(x)f(x)dx E_\theta X_0^2$. Usually $\int \dot{\varphi}'(x)f(x)dx = I(f)$.

The same results, under similar regularity conditions, hold true, if $\Delta_n$ is replaced by $\Psi_n$.

To see that the proposed procedure works quite well we add some simulation results. We simulate AR(1)-models for the following four densities of the distribution of the errors $e_t$:

$$f_1(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2},$$

$$f_2(x) = \frac{0.1}{\sqrt{20\pi}} e^{-x^2/20} + \frac{0.9}{\sqrt{2\pi}} e^{-x^2/2},$$

$$f_3(x) = \frac{0.5}{\sqrt{2\pi}} (e^{-(x-3)^2/2} + e^{-(x+3)^2/2}),$$

$$f_4(x) = 0.5\, e^{-|x|}, \qquad x \in \boldsymbol{R}.$$

We compare the behaviour of the proposed stochastic estimator for several values of $j_n$ and $\tau^2$ (cf. construction of $\{\hat{\theta}_n\}$) with the usual least squares (LS)-estimator $\{\bar{\theta}_n\}$, by tabulating empirical 95% confidence-intervals for $\sqrt{n}(\hat{\theta}_n - \theta)$, $\sqrt{n}(\bar{\theta}_n - \theta)$, respectively (cf. Step 2 and (4.4)). To judge the efficiency of the local random searching procedure, we compare the obtained results with the behaviour of a one-step Newton approximation with initial value $\bar{\theta}_n$ to find roots of $\Delta_n$.

As values for $\theta$ we use 0.5 and 0.8, while the length of the simulated time series is $n = 50$ or $n = 100$. For all simulations the Monte Carlo repetition number is 3000.

As can be seen the values of $\tau^2$ do not have great influence, while increasing $j_n$ improves the results a bit. Comparison with one-step Newton approximation, starting with the LS-estimator, shows that the stochastic procedure is as good as a one-step Newton iteration and that the choices $j_n = 20$, $\tau^2 = 2$ are suitable for the situation considered in Table 1.

From Table 2 we again see that the proposed stochastic procedure can

Table 1. $n=50$, $\theta=0.5$, $\Delta_n(\theta)\equiv 0$.

| | $f_1$ | $f_2$ | $f_3$ | $f_4$ |
|---|---|---|---|---|
| LS-estimator | (−1.96, 1.39) | (−1.99, 1.34) | (−2.00, 1.37) | (−1.99, 1.35) |
| stochastic estimator | | | | |
| $j_n=10$, $\tau^2=1$ | (−1.96, 1.39) | (−1.78, 1.24) | (−0.87, 0.67) | (−1.79, 1.36) |
| $j_n=10$, $\tau^2=2$ | (−1.96, 1.39) | (−1.79, 1.25) | (−0.83, 0.70) | (−1.75, 1.35) |
| $j_n=10$, $\tau^2=3$ | (−1.96, 1.39) | (−1.81, 1.26) | (−0.82, 0.74) | (−1.70, 1.36) |
| $j_n=15$, $\tau^2=1$ | (−1.96, 1.39) | (−1.78, 1.22) | (−0.78, 0.63) | (−1.77, 1.27) |
| $j_n=15$, $\tau^2=2$ | (−1.96, 1.39) | (−1.79, 1.23) | (−0.76, 0.65) | (−1.77, 1.30) |
| $j_n=15$, $\tau^2=3$ | (−1.96, 1.39) | (−1.78, 1.24) | (−0.77, 0.66) | (−1.79, 1.28) |
| $j_n=20$, $\tau^2=2$ | (−1.96, 1.39) | (−1.69, 1.22) | (−0.69, 0.60) | (−1.67, 1.29) |
| one-step Newton | (−1.96, 1.39) | (−1.74, 1.21) | (−0.68, 0.57) | (−1.69, 1.36) |

Table 2. $n=100$, $\theta=0.8$, $\Delta_n(\theta)\equiv 0$.

| | $f_1$ | $f_2$ | $f_3$ | $f_4$ |
|---|---|---|---|---|
| LS-estimator | $(-1.54,\ 0.85)$ | $(-1.52,\ 0.83)$ | $(-1.56,\ 0.87)$ | $(-1.59,\ 0.84)$ |
| stochastic estimator | | | | |
| $j_n=10$, $\tau^2=1$ | $(-1.54,\ 0.85)$ | $(-1.31,\ 0.78)$ | $(-0.62,\ 0.48)$ | $(-1.37,\ 0.78)$ |
| $j_n=15$, $\tau^2=2$ | $(-1.54,\ 0.85)$ | $(-1.28,\ 0.79)$ | $(-0.54,\ 0.47)$ | $(-1.29,\ 0.77)$ |
| one-step Newton | $(-1.54,\ 0.85)$ | $(-1.31,\ 0.75)$ | $(-0.49,\ 0.38)$ | $(-1.36,\ 0.75)$ |

compete with one-step Newton approximation, but is not really better. Finally let us mention that the local random searching method is in no case worse than the usual least squares procedure, but sometimes considerably better.

## 5. Comments

Of great theoretical and practical interest are methods to construct estimators with property (4.2) and which make no use of the shape parameter $f$ or $\phi$ (so-called ***adaptive procedures***). Thus it is an interesting result that all results of this paper hold true even if we replace in Step 2 $\Delta_n$ by $\hat{\Delta}_n$, where $\phi$ is replaced by a consistent estimator. We will deal with this problem in a subsequent paper. Likewise it is possible to apply the proposed stochastic procedure to the more complicated ARMA situation.

## Acknowledgements

The author is grateful to Professor R. Beran for a preprint of the paper Beran and Millar (1987) and to a referee, whose suggestions led to considerable improvement of the presentation of the original version of the paper.

## Appendix

Here we prove a convergence lemma needed in Section 3.

LEMMA A.1. *Let $\{m_n\}$ be a sequence of discrete probability measures on $\boldsymbol{R}^k$, with support* supp $(m_n)=\{p_1^n,\dots,p_{j_n}^n\}\subset\boldsymbol{R}^k$, *respectively. Further let $K_1\subset K_2\subset\cdots$ be a sequence of compact sets with* supp $(m_n)\subset K_n$ *for all large $n$ and $g_n$: $\boldsymbol{R}^k\to\boldsymbol{R}^m$ a sequence of continuous functions. Define*

$$\Lambda(m_n,g_n)=p_{\nu_n}^n,\quad \textit{if}\quad \|g_n(p_{\nu_n}^n)\|=\min_{\nu}\{\|g_n(p_\nu^n)\|:1\le\nu\le j_n\}\,.$$

*If*

(i) *$m_n\Rightarrow m_o$, and the probability measure $m_o$ gives positive mass to* *every open set*

(ii) $\sup_t \{\|g_n(t) - g_o(t)\|, t \in K_n\} \to 0$, *as* $n \to \infty$, *for* $g_o \in C(\boldsymbol{R}^k, \boldsymbol{R}^m)$,

*with* $g_o(x) = 0 \Leftrightarrow x = 0$ *and* $\|g_o(x)\| \geq \delta > 0$ *for all large* $x$, *then*

(a) $\Lambda(m_n, g_n) \to 0$,

(b) $g_n(\Lambda(m_n, g_n)) \to 0$, *as* $n \to \infty$.

PROOF. Let

$$[f]_n = \min_i \{\|f(p_i^n)\| : i = 1,\dots,j_n\} = \min_i \{\|f(p_i^n)\| : i = 1,\dots,j_n; p_i^n \in K_n\},$$

for all large $n$, and consider

$$\begin{aligned} |[g_n]_n - [g_o]_n| &= |\,\|g_n(p_{\nu_n}^n)\| - \|g_o(p_{\mu_n}^n)\|\,|, \quad \text{say},\\ &\leq \max_i \{|\,\|g_n(p_i^n)\| - \|g_o(p_i^n)\|\,| : i = 1,\dots,j_n; p_i^n \in K_n\}\\ &\leq \sup_t \{\|g_n(t) - g_o(t)\|; t \in K_n\}\\ &\to 0 \quad \text{as} \quad n \to \infty, \end{aligned}$$

because of (ii). Consider the following set $A_\varepsilon = \{x | \, \|g_o(x)\| < \varepsilon\}$. Since $g_o$ is continuous, $A_\varepsilon$ is an open set, and $0 \in A_\varepsilon$. From (i) we have $\liminf_n m_n(A_\varepsilon) \geq m_o(A_\varepsilon) > 0$, so that for large $n$ $[g_o]_n < \varepsilon$. From this and the above we obtain $[g_n]_n \to 0$, as $n \to \infty$. This is (b).

To see (a) consider

$$B_\varepsilon^n = \{x | \, \|g_n(x)\| < [g_n]_n + \varepsilon, x \in K_n\}.$$

For large $n$ and small $\varepsilon > 0$, we have because of (b) and (ii):

$$\begin{aligned} \emptyset \neq B_\varepsilon^n &\subset \{x | \, \|g_n(x)\| < 2\varepsilon, x \in K_n\}\\ &\subset \{x | \, \|g_o(x)\| < 3\varepsilon\} \overset{\varepsilon \downarrow 0}{\to} \{0\}, \qquad \text{cf. (ii)}. \end{aligned}$$

Since $[g_n]_n = \|g_n(p_{\nu_n}^n)\|$, say, we have $p_{\nu_n}^n \in B_\varepsilon^n$, for $n$ large enough and all $\varepsilon > 0$. From this we conclude $p_{\nu_n}^n = \Lambda(m_n, g_n) \to 0$, as $n \to \infty$, which completes the proof. □

## REFERENCES

Beran, R. and Millar, P. W. (1987). Stochastic estimation and testing, *Ann. Statist.*, **15**, 1131–1154.

Beran, R., LeCam, L. and Millar, P. W. (1987). Convergence of stochastic empirical measures, *J. Multivariate Anal.*, **23**, 159–168.

Fahrmeir, L. (1973). Konvergenz stetiger stochastischer Prozesse mit polnischen Zustands- und Parameterraum, *Z. Wahrsch. Verw. Gebiete*, **28**, 37–42.

Fuller, W. A. (1976). *Introduction to Statistical Time Series*, Wiley, New York.

Kreiss, J.-P. (1985). A note on *M*-estimation in stationary ARMA processes, *Statist. Decisions*, **3**, 317–336.

Kreiss, J.-P. (1987). On adaptive estimation in stationary ARMA processes, *Ann. Statist.*, **15**, 112–133.

Lohse, K. (1987). Consistency of the Bootstrap, *Statist. Decisions*, **5**, 353–366.