Z. Wahrscheinlichkeitstheorie verw. Geb. 5, 317—335 (1966)

Inference in Stochastic Processes. II*

M. M. RAO

Received March 23, 1965, received in revised form May 2, 1966

1. Introduction

This paper is the second part of a study, of inference problems on stochastic processes, started in [*16*] where the m-decision ($m \geqq 2$) problem of certain continuous 'parameter' stochastic processes was treated. In the present paper some discrete parameter (or index) processes will be considered, restricting the attention to certain general estimation problems centering around the asymptotic properties of the estimators. In [*16*] as well as here the processes need not be stationary. As application, certain previously known results on general stochastic difference equations are unified and extended. These relate to the consistency and asymptotic efficiency of the estimators of certain "structural parameters". Some earlier results of this paper were announced in [*17*], and others in [*19*]. Subsequently, the results of the study of other aspects of inference theory of processes, such as the filtering and prediction problems, will be published (cf. [*18*]). The paper can be read independently of [*16*] and the exposition is self-contained.

After preliminaries and notation in the next section, the main problem and the previous work are discussed in Section 3. This reveals the difficulties and limitations of the previously known work, and points out the significance of the problem. In Section 4 a solution to the general problem is provided (Theorems 2 and 3) if there is one unknown parameter in the finite dimensional density functions of the process. When there are several (two or more) parameters in the finite dimensional densities, some (weaker) extensions of the results are obtained (Theorems 4 and 5) for a unified treatment of stable and certain unstable processes. [A process $\{X_n, n > 0\}$ is stable if the stochastic dependence of X_n on X_m decreases to zero as $|m - n|$ increases; otherwise it is unstable. Precise definitions will be given later.] As consequences of these results, estimation problems and their efficiencies are considered in some detail for the "structural parameters" in linear stochastic difference equations because of their practical importance. The first rigorous treatment of such equations was made by MANN and WALD in [*14*]; and some of their results are here extended. The complications appearing in the general study to include the unstable process will become clear in this application. The special treatments will only serve as useful motivations. The main problem itself was partly inspired by WALD's work in [*21*].

It should be noted at this point that the maximum likelihood method is both convenient and natural for estimation problems on processes in the generality in which they are considered. Therefore its study takes the central position in what

---

* Supported in part by the National Science Foundation under grants NSF-G 14832 and GP-1349.

follows. Certain other aspects, than those considered here, were treated by GRENANDER in [6].

2. Notion and Preliminaries

Some notation, definitions and a brief discussion of stochastic convergence are included here for the reader's convenience. They appear several times later.

A random variable (r. v.) is a finite real valued measurable function and a (row or column) vector r. v. is one which has a finite number of r. v.'s as its components. The symbol $P[S]$ is the probability of the event S and $E_\theta(X)$ stands for the expected value (integral) of the r. v. X, when it exists, under the hypothesis that P_θ is the true probability measure. The symbols $X_n \xrightarrow{p} X$, and $X_n \xrightarrow{d} X$ are used respectively to mean that the sequence of r. v.'s $\{X_n\}$ converges in probability to a r. v. X and that the distribution functions (d. f.'s) $F_n(.)$ of X_n converge to $F(\cdot)$, that of X, at all continuity points of the latter. If $\{X_n\}$ and $\{Y_n\}$ are any sequences of r. v.'s, then $X_n \overset{p}{=} Y_n$ means $(X_n - Y_n) \xrightarrow{p} 0$. A sequence of r. v.'s $\{X_n\}$ is said to be *bounded in probability* if, for any $\varepsilon > 0$, there exists an M_ε such that $\overline{\lim\limits_{n\to\infty}}\, P[|X_n| \geqq M_\varepsilon] \leqq \varepsilon$. Clearly such a sequence is always bounded in probability if the means and variances exist and are bounded functions of n.

Let $\{X_n\}$ and $\{Y_n\}$ be two arbitrary sequences of r. v.'s. The following lemmas are known (cf., eg. [15] and the references there).

Lemma 1. *If $X_n \xrightarrow{p} X$ and $Y_n \xrightarrow{p} Y$, then $X_n Y_n \xrightarrow{p} X Y$. Moreover, if*

$$P[Y = 0] = 0\,, \quad \textit{then} \quad (X_n/Y_n) \xrightarrow{p} (X/Y)\,.$$

Lemma 2. *If $X_n \xrightarrow{p} 0$ and $\{Y_n\}$ is bounded in probability, then $X_n Y_n \xrightarrow{p} 0$.*

Lemma 3. *If $X_n \overset{p}{=} Y_n$ and $X_n \xrightarrow{p \text{ or } d} X$ then $Y_n \xrightarrow{d} X$. (As usual, all limits are taken as $n \to \infty$.)*

3. The Problem

Let $\{X_n, n \geqq 1\}$ be a (discrete) stochastic process. All processes and parameters are assumed to be real. If $(X_1, X_2, \ldots, X_n)$ is a set of r. v.'s, then, for each n, let $P_n(x_1, x_2, \ldots, x_n; \theta_1, \ldots, \theta_k)$, or P_n for short, denote their joint d. f. which depends on k parameters. In what follows only those processes whose d. f.'s are absolutely continuous (relative to the Lebesgue measure) with density functions $p_n(x_1, \ldots, x_n; \theta_1, \ldots, \theta_k)$, or p_n for short, will be considered. Also k is assumed to be a (known) fixed positive integer independent of n (i. e., for all p_n)[1]. The problem is to study the asymptotic properties of the maximum likelihood (m. l.) estimators $\hat{\theta}_n$ of θ appearing in p_n.

In the single parameter case (i. e., $k = 1$ above) WALD proved in [21], the following theorem which is the most general result known thus far:

[1] Since the sequence of r. v.'s $\{X_n\}$ is supposed given (i. e. the r. v.'s and the measure space on which they are defined are given) the d. f. 's $\{P_n\}$ always satisfy the compatibility relations $p_n = \int_{-\infty}^{\infty} p_{n+1} dx_{n+1}$. When the sequence of d. f.'s $\{P_n\}$ is supposed given, (which may alternately be assumed) then one has to assume, in addition, these compatibility relations.

Theorem 1. (Wald) *Let the true parameter value, θ, be an interior point of a finite non-degenerate interval A on the θ axis such that the following conditions hold.*

Condition 1. The derivatives $\frac{\partial^i p_n}{\partial \theta^i}$ $(i = 1, 2, 3)$, *exist for all $\theta \in A$, and for all samples $(x_1, \ldots, x_n)$ except perhaps for a set of measure zero. Further,*

$$\int_{\infty}^{\infty} \cdots \int_{\infty}^{\infty} \underset{\theta \in A}{\text{l. u. b.}} \left| \frac{\partial^i p_n}{\partial \theta^i} \right| dx_1, \ldots, dx_n < \infty \,, \; i = 1, 2 \,. \tag{1}$$

Condition 2. For any $\theta \in A$, one has $\lim_{n\to\infty} C_n(\theta) = \infty$, *where* $C_n(\theta) = E_\theta\left[\left(\frac{\partial \log p_n}{\partial \theta}\right)^2\right]$.

Condition 3. For any $\theta \in A$, $\lim_{n\to\infty}\left[\left(\operatorname{Var}_\theta \frac{\partial^2 \log p_n}{\partial \theta^2}\right) / C_n(\theta)\right] = 0$. *('Var' means variance.)*

Condition 4. There exists a positive δ, such that for any $\theta \in A$,

$$\frac{1}{C_n(\theta)} E_\theta\left[\underset{\theta'}{\text{l. u. b.}} \left| \frac{\partial^3 \log p_n(X_1, \ldots, X_n; \theta')}{\partial \theta'^3} \right| \right] \tag{2}$$

is a bounded function of n, where θ' is restricted to the interval $|\theta' - \theta| \leqq \delta$.

Then, the m.l. *equation* $\frac{\partial \log p_n}{\partial \theta} = 0$, *has a root $\hat{\theta}_n$ which is a consistent estimator of θ. Furthermore, any root of the* m. l. *equation which is a consistent estimator of θ is also asymptotically efficient at least in the wide sense. (The concept of wide sense efficiency is given below in Section 5.)*

To see the domain of applicability of this theorem, the case of a first order linear homogeneous stochastic difference equation is considered. This example has important applications. It will be shown that the conditions of Wald's theorem are satisfied if, and only if, the equation is non-explosive (i. e., α in (3) below satisfies $|\alpha| < 1$).

Example. Let the process $\{X_t, t \geqq 1\}$ satisfy the relation (t integer)

$$X_t = \alpha X_{t-1} + u_t \tag{3}$$

where $\alpha \in A$, a bounded non-degenerate interval including the origin, and the sequence of r. v.'s $u_t (t \geqq 1)$ are independent, Gaussian distributed with mean zero and variance unity. Further, let $u_t = 0$ for $t \leqq 0$. To determine the consistency of the m. l. estimator of α it suffices to verify the conditions of Theorem 1.

The density of $X_1, \ldots, X_n$ is given by

$$p_n = p_n(x_1, \ldots, x_n; \alpha) = (2\pi)^{-n/2} \exp\left[- \sum_{t=1}^{n} (x_t - \alpha x_{t-1})^2 / 2\right]. \tag{4}$$

The differentiability conditions clearly are satisfied here. From the fact that A is bounded, follows the existence of the integrals in (1) for the p_n given by (4). Hence Condition 1 of the theorem is satisfied.

Let $\varphi_n(\alpha) = \frac{\partial \log p_n}{\partial \alpha}$. Then $\varphi_n(\alpha) = \sum_{t=1}^{n} u_t X_{t-1}$ and from (3) $X_t = \sum_{i=1}^{t} \alpha^{i-1} u_{t-i+1}$. Noting that u_t and X_{t-i}, for $i \geqq 1$, are independent, one gets

$$C_n(\alpha) = E_\alpha(\varphi_n^2(\alpha)) = \sum_{t=2}^{n} \sum_{j=1}^{t-1} \alpha^{2(j\quad)} \geqq (n-1)\,. \tag{5}$$

It follows that $C_n(\alpha) \to \infty$, as $n \to \infty$ for all $\alpha \in A$, and Condition 2 holds. Since $\varphi_n''(\alpha') = \frac{\partial^3 \log p_n}{\partial \alpha'^3} = 0$, for all $\alpha' \in A$, Condition 4 is automatically satisfied. Hence Conditions 1, 2 and 4 are satisfied even if $|\alpha| \geqq 1$. Condition 3, on the other hand, will be shown to hold if, and only if, $|\alpha| < 1$.

From (5), it is seen that $\lim_{n\to\infty} \frac{C_n(\alpha)}{n} = (1-\alpha^2)^{-1}$ if $|\alpha| < 1$. It was shown in ([*14*], p. 180) that, for $|\alpha| < 1$, $\mathrm{Var}_\alpha\, \varphi_n'(\alpha) = 0(n)$. These two statements imply the truth of Condition 3 at once. That this condition fails if $|\alpha| \geqq 1$ is seen as follows. Note that $\varphi_n'(\alpha) = -\sum_{t=1}^{n-1} X_t^2$ and $E_\alpha(\varphi_n'(\alpha)) = -C_n(\alpha)$. $\left(\text{Here } \varphi_n'(\alpha) = \frac{\partial \varphi_n(\alpha)}{\partial \alpha}.\right)$ If Condition 3 holds then clearly $\frac{1}{C_n(\alpha)} \sum_{t=1}^{n-1} X_t^2$ must converge in probability to a constant $(= -1)$. But $\lim_{n\to\infty} \frac{C_n(\alpha)}{n^2} = \frac{1}{2}$ if $|\alpha| = 1$, and $X_t = v_1 + \cdots + v_t$ where the $v_t (= \alpha^{i-1} u_{t-i+1})$ are independent Gaussian r. v.'s each with mean zero and variance 1 (since $|\alpha| = 1$). So the above expression converges in probability to a constant if and only if $\frac{1}{n^2} \sum_{t=1}^{n} X_t^2$ converges in the same sense. But this is impossible since by a result of ERDÖS and KAC [*4*], $\frac{1}{n^2} \sum_{t=1}^{n} X_t^2$ has a proper limit distribution. If $|\alpha| > 1$, then $\lim_{n\to\infty} \alpha^{-2n} C_n(\alpha) = (\alpha^2 - 1)^{-1}$, and $\alpha^{-2n} \sum_{t=1}^{n} X_t^2$ converges in probability to a random variable by ([*1*], Theorem 2.1 or [*15*], Lemma 15), which is not a constant. Hence Condition 3 fails if $|\alpha| \geqq 1$. It follows that Theorem 1 can be used to show that the m. l. equation $\varphi_n(\alpha) = 0$ has a root $\hat{\alpha}_n$ which is a consistent estimator of α, of the stochastic equation (3), if $|\alpha| < 1$; it gives no information if $|\alpha| \geqq 1$. The consistency of the m. l. estimator in the latter case (of (3)) was proved by RUBIN [*20*].

The above illustration shows that Condition 3 fails in the unstable case. This condition was used by Wald in his proof to show that $(\varphi_n'(\theta)/C_n(\theta)) \xrightarrow{p} -1$, as $n \to \infty$, where $\varphi_n(\theta) = \left(\frac{\partial \log p_n}{\partial \theta}\right)$ and $C_n(\theta) = E_\theta(\varphi_n^2(\theta))$. This condition may be replaced by the following weaker condition and the conclusions of the theorem remain valid.

Condition 3′. For all θ in A, a finite non-degenerate open interval,

$$\lim_{n\to\infty} E_\theta \left(\frac{Y_n^2}{Y_n^2 + C_n^2(\theta)}\right) = 0; \; Y_n = \varphi_n'(\theta) + C_n(\theta). \tag{6}$$

Note that $E_\theta(Y_n) = 0$ for all n. This condition is the best to show that $[Y_n(\theta)/C_n(\theta)] \xrightarrow{p} 0$, because of

Lemma 4. *A sequence of r.v.'s $\{Z_n\}$, with $E(Z_n) = 0$, converges in probability to zero if, and only if,*

$$\lim_{n\to\infty} E\left(\frac{Z_n^2}{1 + Z_n^2}\right) = 0.$$

(A proof of the lemma may be found in [*5*], Section 20.)

However, for the example given after Theorem 1, in the explosive case, the Condition 3′ also fails, since $[Y_n(\theta)/C_n(\theta)] \xrightarrow{p}$ r.v. $(\neq 0)$. This indicates that a different procedure, not requiring this type of condition, is needed to treat the class of problems which include the unstable cases. A general result containing all these cases is proved in the next section if, in $p_n(x, \theta)$, θ is a scalar parameter.

4. Solution of the Single Parameter Problem

It is convenient to introduce a concept of weak sense efficiency which is meaningful for the unstable cases, but which coincides with the wide sense concept introduced by Wald [*21*] as soon as the process under consideration becomes stable. Since the estimators may not always have limit distributions, in the generality in which they are now considered, the weak sense (or wide sense) concept will be relevant here. (For a discussion on this point, see [*21*].)

Definition 1. A sequence of estimators $\{T_n\}$ of θ is said to be asymptotically efficient in the *weak sense* if there exist two sequences of r.v.'s $\{W_n\}$ and $\{V_n\}$ such that (the W's and V's being defined on the same probability space as T's)

$$\lim_{n\to\infty} E_\theta(W_n) = 0\,, \quad \lim_{n\to\infty} E_\theta(W_n^2) = 1\,,$$

and

$$\lim_{n\to\infty} E_\theta(V_n) = 1\,, \quad \lim_{n\to\infty} P[V_n = 0] = 0\,,$$

implies

$$[C_n(\theta)]^{1/2}(T_n - \theta) \overset{p}{=} \frac{W_n}{V_n}\,, \quad \text{where} \quad C_n(\theta) = E_\theta[\varphi_n^2(\theta)]\,.$$

If the process is stable, this definition coincides with the wide sense concept of Wald, where he needs only one sequence $\{W_n\}$, since then a sequence $\{V_n\}$ always exists. In fact, taking $V_n = 1$, with probability one, this definition becomes identical with the wide sense concept. (See also the next section.) The concept of stability used above is given precisely in the following

Definition 2. Let $p_n(x, \theta)$ be a finite dimensional density of a process

$$\{X_t, t = 0, 1, \ldots\} \quad \text{where} \quad x = (x_1, \ldots, x_n), \theta = (\theta_1, \ldots, \theta_k) \quad \text{and} \quad k(\geqq 1)$$

is fixed. Then the process is said to be *stable* or *unstable* according as $M = 0$ or $0 < M \leqq \infty$, where $M = \lim_{n\to\infty} \max_{i,j} \left(\operatorname{Var}_\theta\left(\frac{\partial^2 \log p_n}{\partial\theta_i\partial\theta_j}\right)/C^2(\theta, n)\right)$, and where $C(\theta, n)$ is the maximum (over all i) of $C_{ii}(\theta, n) = E_\theta[\partial \log p_n/\partial\theta_i]^2$, $i, j = 1, \ldots, k$.

For example, if $\{X_n\}$ is an "m-dependent" sequence of r.v.'s [7], with two moments then it turns out that $M = 0$, so the sequence is stable by this definition. [Here and elsewhere M is a generic constant.]

Remark. In case $M = 0$, it follows, by the Tshebyshev inequality, that $\frac{\partial^2 \log p_n}{\partial\theta_i\partial\theta_j} \overset{p}{=} E_\theta\left(\frac{\partial^2 \log p_n}{\partial\theta_i\partial\theta_j}\right)$. Such a sequence of r.v.'s $\left\{\left(\frac{\partial^2 \log p_n}{\partial\theta_i\partial\theta_j}\right)\right\}$ is included in the concept of the *stable* sequence introduced by Gnedenko and Kolmogorov ([*15*]; Section 22). Thus the above definition may be considered as a specialized version of the classical concept.

Theorem 2. *Suppose the finite dimensional density $p_n(x_1, \ldots, x_n; \theta)$, or p_n for short, of the process $\{X_t, t > 0\}$ depending on a parameter θ satisfies the following conditions.*

Condition (a). $\frac{\partial p_n}{\partial \theta}, \frac{\partial^2 p_n}{\partial \theta^2}$ *exist for all θ in A and for almost all x, where A is a finite non-degenerate open interval. Further the absolute values of these functions are dominated by $G_n(\cdot)$ and $H_n(\cdot)$ where G_n and H_n are integrable on the cartesian n-space.*

Condition (b). $C_n(\theta) = E_\theta[(\varphi_n(\theta))^2]$ *exists and* $\lim_{n\to\infty} C_n(\theta) = \infty$, *for all $\theta \in A$, where* $\varphi_n(\theta) = \frac{\partial \log p_n}{\partial \theta}$.

Condition (c). If $\varphi_n'(\theta) = \frac{\partial \varphi_n(\theta)}{\partial \theta}$ *(which exists), then φ_n' satisfies a Lipschitz condition of order α. More explicitly, for a given $\beta > 0$, there exists $0 < \alpha \leqq 1$ such that* $|\varphi_n'(\theta) - \varphi_n'(\theta')| \leqq |\theta - \theta'|^\alpha M_n(\theta, \theta')$, *for almost all x, where* $\lim_{n\to\infty} E_\theta\left(\text{l.u.}\,b_{\theta'} \frac{M_n(\theta,\theta')}{C_n(\theta)}\right) < \infty$, *for all $\theta, \theta' \in A$ satisfying $|\theta - \theta'| < \beta$.*

Condition (d). Given $0 < \delta < 1$, there is an $\varepsilon_\delta > 0$, such that for all $\theta \in A$ one has $\lim_{n\to\infty} P\left[\left|\frac{\varphi_n'(\theta)}{C_n(\theta)}\right| \geqq \varepsilon_\delta\right] > 1 - \delta$.

Condition (e). If P_θ is the measure generated by $p_n(x, \theta)$ on the sample space, then $P_{\theta_1} = P_{\theta_2}$ implies $\theta_1 = \theta_2$ and that P_θ is a continuous function in θ, (i. e., the variation of $(P_{\theta_1} - P_{\theta_2})$ tends to zero as $(\theta_1 - \theta_2) \to 0$).

Then the m.l. equation $\varphi_n(\theta) = 0$ has a root $\hat{\theta}_n$ which satisfies the condition $(\hat{\theta}_n - \theta) \xrightarrow{p} 0$ *as $n \to \infty$,* (i. e., *$\hat{\theta}_n$ is a consistent estimator of θ).*

Remark. In the work of WALD [*21*] and CRAMÉR [*3*], as well as in most earlier studies the Condition (e) was implicitly assumed. This may be seen in their proofs. The absence of an explicit statement, when taken literally as being unnecessary, leads to difficulties, as was pointed out by KRAFT and LECAM in [*10*].

Proof. Using part (a) of the hypothesis, $\varphi_n(\theta)$ may be expanded by the Taylor's formula as

$$\varphi_n(\theta) = \varphi_n(\theta^0) + (\theta - \theta^0)\varphi_n'(\theta^0) + (\theta - \theta^0)U_n(\theta) \tag{7}$$

where θ^0 is the true parameter value in A, $\varphi_n(\theta^0) = \varphi_n(\theta)|_{\theta=\theta^0}$; and $U_n(\theta) = \varphi_n'(\theta^0 + \delta_n(\theta - \theta^0)) - \varphi_n'(\theta^0)$ with $0 < \delta_n < 1$. From the same hypothesis it follows that, for all θ in A,

$$E_\theta(\varphi_n(\theta)) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \frac{\partial \log p_n}{\partial \theta} p_n dx_1 \ldots dx_n = \frac{\partial}{\partial \theta}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} p_n dx_1 \ldots dx_n = 0\,.$$

Similarly $E(\varphi_n^2(\theta)) = C_n(\theta) = -E_\theta(\varphi_n'(\theta))$. Consider (7) in the following form (which can clearly be written),

$$\begin{aligned}\frac{\varphi_n(\theta)}{\varphi_n'(\theta^0)} &= \left[\frac{C_n(\theta^0)}{\varphi_n'(\theta^0)}\right]\frac{\varphi_n(\theta^0)}{C_n(\theta^0)} + (\theta - \theta^0) + \left[\frac{C_n(\theta^0)}{\varphi_n'(\theta^0)}\right]\frac{U_n(\theta)}{C_n(\theta^0)}(\theta - \theta^0)\,. \\ &= B_n + (\theta - \theta^0) + \bar{B}_n(\theta - \theta^0)\,,\end{aligned} \tag{8}$$

where the B_n's are defined by the corresponding terms on the right side. Now the B_n's are simplified as follows.

First consider B_n. It is clear that $E_{\theta^0}\left(\frac{\varphi_n(\theta^0)}{C_n(\theta^0)}\right) = 0$, and $\mathrm{Var}_\theta\left(\frac{\varphi_n(\theta^0)}{C_n(\theta^0)}\right) = [C_n(\theta^0)]^{-1}$, the latter tends to zero as $n \to \infty$ by Condition (b). It therefore follows that $[\varphi_n(\theta^0)/C_n(\theta^0)] \xrightarrow{p} 0$. From Condition (d), since $[\varphi_n'(\theta^0)/C_n(\theta^0)]$ is bounded away from zero in probability, one obtains the following. For any $\varepsilon > 0$, there exists a positive number L_ε such that

(9) $$\overline{\lim_{n\to\infty}} P\left[\left|\frac{C_n(\theta^0)}{\varphi_n'(\theta^0)}\right| \geqq L_\varepsilon\right] \leqq \varepsilon .$$

Hence from the preceding and Lemma 2 it follows that $B_n \xrightarrow{p} 0$.

Next consider $\bar{B}_n$. Using Condition (c) the second term of $\bar{B}_n$ can be written for almost all x, as follows: $(0 < \alpha \leqq 1)$

(10) $$\left|\frac{U_n(\theta)}{C_n(\theta^0)}\right| \leqq |\theta - \theta^0|^\alpha \frac{M_n(\theta, \theta^0)}{C_n(\theta^0)} ,$$

where for $|\theta - \theta^0| < \beta$, $[M_n(\theta, \theta^0)/C_n(\theta^0)]$ is bounded in probability. Consequently (8) can be rewritten, using (10), as

(11) $$\frac{\varphi_n(\theta)}{\varphi_n'(\theta^0)} = B_n + (\theta - \theta^0) + (\theta - \theta^0)^{1+\alpha} \tilde{B}_n$$

where $\tilde{B}_n = [C_n(\theta^0)/\varphi_n'(\theta^0)][M_n(\theta, \theta^0)/C_n(\theta^0)]\tilde{\delta}_n$ in which $|\tilde{\delta}_n| < 1$. It follows, with (9), that $\tilde{B}_n$ is bounded in probability for all $|\theta - \theta^0| < \beta$.

Let $\varepsilon_1, \varepsilon_2$ be any two small positive numbers less than β. If $f_1 = P[|B_n| \geqq \varepsilon_1^{1+\alpha}]$, then the result on the asymptotic behavior on B_n established above shows that there exists an $n(\varepsilon_1, \varepsilon_2)$ such that $n \geqq n(\varepsilon_1, \varepsilon_2)$ implies $f_1 \leqq \varepsilon_2/2$. Also from the result on $\tilde{B}_n$ of the preceding paragraph it follows that there is an L_{ε_2} and an $n_1(\varepsilon_2)$ such that, if $f_2 = P[|\tilde{B}_n| \geqq L_{\varepsilon_2}]$ and $n \geqq n_1(\varepsilon_2)$ then $f_2 \leqq \varepsilon_2/2$. Let $n_0(\varepsilon_1, \varepsilon_2) = \max(n(\varepsilon_1, \varepsilon_2), n_1(\varepsilon_2))$, and let

$$S = \{X = (X_1, \ldots, X_n) : |B_n| < \varepsilon_1^{1+\alpha}, |\tilde{B}_n| < L_{\varepsilon_2}\} .$$

If S' is the complement of S, then for $n \geqq n_0(\varepsilon_1, \varepsilon_2)$,

(12) $$P[S'] = P[(|B_n| \geqq \varepsilon^{1+\alpha}) \cup (|\tilde{B}_n| \geqq L_{\varepsilon_2})] \leqq f_1 + f_2 \leqq \varepsilon_2 .$$

Hence $P[S] > 1 - \varepsilon_2$. Consequently for $\theta = \theta^0 \pm \varepsilon_1$, the first and last terms on the right side of (11) are less than $(1 + L_{\varepsilon_2})\varepsilon_1{}^{1+\alpha}$ with probability greater than $1 - \varepsilon_2$ for $n \geqq n_0(\varepsilon_1, \varepsilon_2)$. If now ε_1 is chosen such that $(1 + L_{\varepsilon_2})\varepsilon_1{}^\alpha < 1$, then the whole expression on the right side of (11) is determined, in sign, for $\theta = \theta^0 \pm \varepsilon_1$ by the second term, i. e.,

(13) $$\begin{aligned} \frac{\varphi_n(\theta)}{\varphi_n'(\theta^0)} &> 0 \quad \text{if} \quad \theta = \theta^0 + \varepsilon_1 , \\ &< 0 \quad \text{if} \quad \theta = \theta^0 - \varepsilon_1 . \end{aligned}$$

Since by Condition (a) $\varphi_n(\theta)$ is differentiable in θ for almost all x, it is continuous in θ, and hence $[\varphi_n(\theta)/\varphi_n'(\theta^0)] = 0$ has a root $\hat{\theta}_n$ in the interval $(\theta^0 - \varepsilon_1, \theta^0 + \varepsilon_1)$ if $n > n_0(\varepsilon_1, \varepsilon_2)$ with probability $> 1 - \varepsilon_2$, in view of the continuity of the probability measures P_θ by Condition (e) of the hypothesis. Since $\varepsilon_1, \varepsilon_2$ are arbitrary the theorem is proved.

An important property of the consistent m.l. estimator is given in

Theorem 3. *Every consistent m.l. estimator $\hat{\theta}_n$ of θ, established in Theorem 2, is asymptotically efficient in the weak sense.*

Proof. Let $\hat{\theta}_n$ be a root of $\varphi_n(\theta) = 0$ which is a consistent estimator of θ as given by Theorem 2. Setting $\varphi_n(\theta) = 0$ in (8), and rearranging, the following equation obtains.

$$W_n = V_n Z_n + \varrho_n Z_n \,, \tag{14}$$

where

$$V_n = \varphi_n'(\theta^0)/C_n(\theta^0)\,, Z_n = [C_n(\theta^0)]^{1/2}(\hat{\theta}_n - \theta^0)\,, W_n = -\varphi_n(\theta^0)/[C_n(\theta^0)]^{1/2}\,,$$

and $\varrho_n = U_n(\hat{\theta}_n)/C_n(\theta^0)$. By hypothesis $(0 < \alpha \leqq 1)$

$$|\varrho_n| \leqq |\hat{\theta}_n - \theta^0|^{\alpha} \frac{M_n(\hat{\theta}_n, \theta^0)}{C_n(\theta^0)} \tag{15}$$

and the right side terms are such that the last factor is bounded in probability and $\hat{\theta}_n \overset{p}{\to} \theta^0$. So $\varrho_n \overset{p}{\to} 0$. This means $V_n \overset{p}{=} (V_n + \varrho_n)$. But $E_{\theta^0}(W_n) = 0$ and $E_{\theta^0}(W_n^2) = 1$, for all n, which implies, by a remark in the penultimate paragraph of Section 2, that $\{W_n\}$ is bounded in probability. Thus from (14) it follows that the right side is bounded in probability. Since $E_{\theta^0}(V_n) = 1$ and by Condition (d) of Theorem 2, $\{V_n\}$ is bounded away from zero in probability, it must be true [in (14)] that $\{Z_n\}$ is bounded in probability as well. [It may be of interest to note that the preceding statement implies that $(C_n(\theta^0))^{1/2}$ is the correct normalizing factor for $(\hat{\theta}_n - \theta^0)$.] Hence

$$(W_n - V_n Z_n) \overset{p}{\to} 0\,. \tag{16}$$

The properties established for $\{V_n\}$ and $\{W_n\}$ above satisfy the conditions of Definition 1, so that

$$[C_n(\theta^0)]^{1/2}(\hat{\theta}_n - \theta^0) \overset{p}{=} [W_n/V_n]. \tag{17}$$

Thus the consistent estimator $\hat{\theta}_n$ is asymptotically efficient in the weak sense, completing the proof.

Remark. If only (c) is assumed, as here, Condition (d) of Theorem 2 cannot be dropped, as easy counter-examples show. It may also be noted that the conditions given in Theorem 2 are different from, and much weaker than, the classical ones (see the application below). Another set of conditions (variant of the above) was given in [*19*]. Several different sets (and even weaker) conditions can be produced. *The point of the result here is that it seems to be the first of its kind which deals with the stable and unstable cases together.*

Application. Consider the first order stochastic equation $X_t = \alpha X_{t-1} + u_t$, given by (3) of Section 3 with the same assumptions as there. From (4), the likelihood equation is given by [expanding $\varphi_n(\cdot)$ around α_0, in A, the true value]

$$\varphi_n(\alpha) = \sum_{t=1}^{n} u_t X_{t-1}\big|_{\alpha=\alpha_0} - (\alpha - \alpha_0)\sum_{t=1}^{n} X_{t-1}^2\,. \tag{18}$$

For this the Conditions (a) and (b) of Theorem 2 have been checked in the preceding section. Conditions (c) and (e) are trivial in the present case.

The Condition (d) is that $\sum_{t=1}^{n} X_{t-1}^2/C_n^2(\alpha)$ is bounded away from zero (in probability) if $|\alpha| \geqq 1$ or $|\alpha| < 1$. If $|\alpha| = 1$, this result is a consequence of the results of ERDÖS and KAC [4], and if $|\alpha| > 1$, it follows from the fact that it converges stochastically to a r.v. ($\neq 0$) as shown in Lemma 15 of [15]. If $|\alpha| < 1$, $[\sum_{t=1}^{n} X_{t-1}^2 / C_n^2(\alpha)] \overset{p}{=} 1$ (by Tshebyshev's theorem, [3]). Consequently Condition (d) is also satisfied for all α in A, a bounded open non-degenerate interval containing, say, $(-2, 2)$. Hence by Theorem 2, it follows that the m. l. estimator $\hat{\alpha}_n$ of α, a root of $\varphi_n(\alpha) = 0$, is consistent and, by Theorem 3, it is asymptotically efficient in the weak sense. Since there is only one root here, $\hat{\alpha}_n$ is also unique.

The part on consistency of $\hat{\alpha}_n$ for this particular process was proved by RUBIN [20] using a special method.

Remarks. It is also of interest to consider a constant term, β_0 in the above example. However, this case cannot be subsumed under Theorem 1, because there will be two parameters. Several difficulties arise in the consideration of this multiparameter case. Certain matrices appearing in the proofs corresponding to reciprocals used in Theorems 2 and 3, for instance, become singular for large n, if the process is unstable (cf. Section 6). Hence the multiparameter extensions of these results in this generality seem rather difficult with the present methods. In the following section, some extensions will be given, for a certain class of unstable and all stable processes, which may be used for some cases of the above problems. Using special methods, certain more general, unstable (or explosive) processes satisfying kth order ($k \geqq 2$) linear stochastic difference equations have been studied in [15], but unfortunately no general methods are available at present.

5. Some Extensions to the k-parameter Problem

At the outset it is convenient to state the concept of efficiency in the wide sense, of the estimators, in the vector case. (See the discussion preceding Definition 1 above.)

Definition 3. A sequence of (row) vector estimators $\{\hat{\theta}_n\}$ of $\theta = (\theta_1, \ldots, \theta_k)$, of $p_n(x, \theta)$, is said to be asymptotically efficient in the wide sense, if there exists another sequence of vector r. v.'s $\{W_n\}$ such that $\lim_{n\to\infty} E_\theta(W_n) = 0, \lim_{n\to\infty} E_\theta(W_n' W_n) = I_k$, the identity matrix of order k (prime denotes transposition), and that $(\hat{\theta}_n - \theta) B_n(\theta) \overset{p}{=} W_n$ where $B_n^2(\theta) = \Gamma_n(\theta) = (C_{ij}(\theta, n))$ with $C_{ij}(\theta, n) = E[\varphi_i(\theta, n)\varphi_j(\theta, n)]$, $i, j = 1, \ldots, k$. Here $\varphi_i(\theta, n) = \frac{\partial\varphi(\theta, n)}{\partial\theta_i}$ and $\varphi(\theta, n) = \log p_n(x, \theta)$ (and similarly $\varphi_{ij}(\theta, n)$ is the second order mixed partial derivative) which are assumed to exist.

If the $\{W_n\}$ have a limt normal (multidimensional) distribution, then the efficiency defined above coincides with the (classical) *strict sense* concept, [3]. If $\Gamma({}_n\theta)$ is singular, such a $\{W_n\}$ may not exist, and thus this definition will be of interest whenever $\Gamma_n(\theta)$ is non-singular, in which case $\Gamma_n(\theta)$ is positive definite. In any case it is positive semidefinite (and symmetric) so that the positive (semi-) definite $B_n(\theta)$ is uniquely defined.

With the notations used above, the following result, which applies to the stable processes, can be given.

Theorem 4. *Suppose that $\theta \in A$, where A is a finite non-degenerate open interval in the k-dimensional Euclidean space. Let p_n satisfy the following conditions:*

Condition I. $\frac{\partial p_n}{\partial \theta_i}, \frac{\partial^2 p_n}{\partial \theta_i \partial \theta_j}$, $i, j = 1, 2, \ldots, k$, *exist for almost all x and for all $\theta \in A$, and that the absolute values of these functions are dominated by $G_n(\cdot)$ and $H_n(\cdot)$ for all i, j, where $G_n(\cdot)$ and $H_n(\cdot)$ are integrable on the cartesian n-space.*

Condition II. $C(\theta, n) = \max_i C_{ii}(\theta, n)$ *exists and $C(\theta, n) \to \infty$ as $n \to \infty$, all θ in A.*

Condition III. For all $\theta \in A$, $\lim_{n\to\infty} [C(\theta, n)]^{-1} \Gamma_n(\theta)$ exists as a nonsingular matrix.

Condition IV. For any $\theta \in A$, and all i, j, $\lim_{n\to\infty} \operatorname{Var}_\theta \left[\frac{\varphi_{ij}(\theta, n)}{C(\theta, n)}\right] = 0.$

Condition V. For any θ and θ' in A, and a given $\beta > 0$ and almost all x, $\varphi_{ij}(\theta, n)$ satisfies: $|\varphi_{ij}(\theta, n) - \varphi_{ij}(\theta', n)| \leqq [\sum_{i=1}^{k} (\theta_i' - \theta_i)^2]^{\alpha/2} M_{ij}(\theta, \theta', n)$, *for* $0 < \alpha \leqq 1$, *where* $\lim_{n\to\infty} E_\theta \left[\text{l. u. b.} \frac{M_{ij}(\theta, \theta', n)}{C(\theta, n)}\right] < \infty$ *whenever* $\sum_{i=1}^{k} (\theta - \theta_i')^2 < \beta$.

Condition VI. If P_θ is the measure generated by the $p_n(x, \theta)$'s on the sample space, then $P_\theta = P_{\theta^*}$ only if $\theta = \theta^*$ and that the total variation of $(P_\theta - P_{\theta^*}) \to 0$ as $\sum_{i=1}^{k} (\theta_i - \theta_i^*)^2 \to 0$.

Then, the m.l. equation $0 = \varphi_n(\theta) = (\varphi_i(\theta, n), i = 1, \ldots, k)$ has a (vector) root $\hat{\theta}_n$ which is a consistent estimator of θ, and which is asymptotically efficient in the wide sense.

Corollary 4.1. *If the matrix $\Psi_n(\theta) = [\varphi_{ij}(\theta, n), i, j = 1, 2, \ldots, k]$ has the property that $[C(\theta^0, n)]^{-1} \Psi_n(\theta)$ is negative definite for all $\theta \in A$ a non-degenerate open convex interval and for all large n with probability one, then the consistent m.l. estimator $\hat{\theta}_n$ of θ is also unique.*

The proofs of the theorem and its corollary run on classical lines. As the algebra is complicated and the hypothesis is somewhat different, the essential steps in the proof will be briefly sketched below.

Remark. If the r.v.'s $\{X_n\}$ are independent and identically distributed, and $C_{ii}(\theta, 1)$ is positive for all i, then Conditions II and IV are always satisfied. Note that for Condition III to hold it is necessary that $C_{ii}(\theta, n)/C_{jj}(\theta, n) = O(1)$, so that $C(\theta, n)$ of the theorem may be replaced by $C_{ii}(\theta, n)$.

Sketch of Proof. Let $\tilde{\varphi}_n(\theta^0)$ and $\Psi_n(\theta^0)$ be the (random) quantities given in the theorem evaluated at $\theta = \theta^0$, the true parameter value. By Condition I and Taylor's expansion,

$$\tilde{\varphi}_n(\theta) = \tilde{\varphi}_n(\theta^0) + (\theta - \theta^0)\Psi_n(\theta^0) + (\theta - \theta^0) U_n(\theta) \tag{19}$$

where $U_n(\theta) = \Psi_n(\theta^0 + \delta_n(\theta - \theta^0)) - \Psi_n(\theta^0)$, and $0 < \delta_n < 1$. From the same condition it follows that

$$E_{\theta^0}(\tilde{\varphi}_n(\theta^0)) = 0, \quad \text{and} \quad E_{\theta^0}[\Psi_n(\theta^0)] = -\Gamma_n(\theta^0),$$

where $\Gamma_n(\theta)$ is given in Definition 3. Let n be large. Now post-multiply (19) by $\Gamma_n(\theta^0)^{-1}$ (cf. Condition III), so that

$$(20) \quad \tilde{\varphi}_n(\theta)\Gamma_n^{-1}(\theta^0) = \tilde{\varphi}_n(\theta^0)\Gamma_n^{-1}(\theta^0) + (\theta - \theta^0)\Psi_n(\theta^0)\Gamma_n^{-1}(\theta^0) + \\ + (\theta - \theta^0)U_n(\theta)\Gamma_n^{-1}(\theta).$$

The stochastic limits of various terms on the right of (20) must be considered. (All limits are taken as $n \to \infty$.)

First, consider $\tilde{\varphi}_n(\theta^0)\Gamma_n^{-1}(\theta^0)$. It was noted that $E_{\theta^0}(\tilde{\varphi}_n(\theta^0))\Gamma_n^{-1}(\theta^0) = 0$. Let $\varepsilon_i > 0$ be given, $i = 1, \ldots, k$, and $\bar{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_k)$. By Conditions III and IV, (absolute value of a matrix or vector means absolute value for each element)

$$P\left[\frac{1}{C(\theta^0, n)} |\tilde{\varphi}_n(\theta^0)| \geqq \varepsilon\right] = P[|\varphi_i(\theta^0, n)| \geqq \varepsilon_i C(\theta^0, n),\ i = 1, \ldots, k] \\ \leqq C^{-2}(\theta^0, n) \sum_{i=1}^{k} \frac{C_{ii}(\theta^0, n)}{\varepsilon_i^2} \to 0,$$

which implies

$$(21) \qquad \tilde{\varphi}_n(\theta^0)\Gamma_n^{-1}(\theta^0) \xrightarrow{p} 0.$$

Next consider $\Psi_n(\theta^0)\Gamma_n^{-1}(\theta^0)$. Since $E_{\theta^0}(\Psi_n(\theta^0)\Gamma_n^{-1}(\theta^0)) = -I_k$, to show that $\Psi_n(\theta^0)\Gamma_n^{-1}(\theta^0) \xrightarrow{p} -I_k$, it suffices to show (by Condition III) that $\Psi_n(\theta^0)$ $C^{-1}(\theta^0, n) \overset{p}{=} -\Gamma_n(\theta^0)C^{-1}(\theta^0, n)$ (element-wise). But this follows from Condition IV, so that

$$(22) \qquad \Psi_n(\theta^0)\Gamma_n^{-1}(\theta^0) \xrightarrow{p} -I_k.$$

Finally, consider $U_n(\theta)\Gamma_n^{-1}(\theta^0)$. Since $[C(\theta^0, n)]^{-1}\Gamma_n(\theta^0)$ is, by Condition III, non-singular for large n, it suffices to consider $[C(\theta^0, n)]^{-1}U_n(\theta)$.

If $u_{ij}(\theta, n)$ is the (i, j)th term of $U_n(\theta)$, then by Condition V, $|u_{ij}(\theta, n)| \leqq \|\theta - \theta^0\|^\alpha M_{ij}(\theta, \theta', n)$, for almost all x, where $\|\theta - \theta^0\|$ is the Euclidean norm $(= [\sum_{i=1}^{k}(\theta_i - \theta_i)^2]^{1/2})$. From this it follows easily that

$$(23) \qquad \|(\theta - \theta^0)U_n(\theta)/C(\theta^0, n)\| \leqq \|\theta - \theta^0\|^{1+\alpha} \left\| \left(\frac{M_{ij}(\theta, \theta^0, n)}{C(\theta^0, n)}\right) \right\|,$$

for almost all x. The last norm symbol on the right side of (23) is the matrix norm $(= [\text{trace}(AA')]^{1/2}$ for any matrix A). The hypothesis on $[M_{ij}(\theta, n)/C(\theta, n)]$ of Condition V implies that it is bounded in probability and hence the first and last terms on the right side of (20), in norm, are together bounded by $(1 + \lambda_\varepsilon)\|\theta - \theta^0\|^{1+\alpha}$, for some $\lambda_\varepsilon > 0$, with large probability $(> 1 - \varepsilon)$ where $\varepsilon > 0$ is arbitrary. So the final argument of Theorem 1 is applicable.

Thus from (21)—(23) it follows that $\tilde{\varphi}_n(\theta) = 0$ has a (vector) root in the interval $(\theta^0 - \eta, \theta^0 + \eta)$ where $\eta = (\eta_1, \ldots, \eta_k)$, $\eta_i > 0$ being arbitrary, with probability tending to one. Hence there exists a root $\hat{\theta}_n$ of $\tilde{\varphi}_n(\theta) = 0$, which is a consistent estimator of θ.

The proof of the last part of the theorem proceeds as that of Theorem 3. Let $\hat{\theta}_n$ be a consistent estimator of θ and a root of $\tilde{\varphi}_n(\theta) = 0$. Then (20) can be written as,

$$(24)\qquad 0 = \tilde{\varphi}_n(\theta^0) B_n^{-1}(\theta^0) + (\hat{\theta}_n - \theta^0) B_n(\theta^0) B_n^{-1}(\theta^0) \Psi_n \theta^0) B_n^{-1}(\theta^0) + (\hat{\theta}_n - \theta^0) B_n\ (\theta^0) B_n^{-1}(\theta^0) U_n(\hat{\theta}_n) B_n^{-1}(\theta^0),$$

where $B_n(\theta^0)$ is the unique square root of $\Gamma_n(\theta^0)$. Set $Y_n = \tilde{\varphi}_n(\theta^0) B_n^{-1}(\theta^0)$, $Z_n = (\hat{\theta}_n - \theta^0) B_n(\theta^0)$, and $V_n = B_n^{-1}(\theta^0) U_n(\hat{\theta}_n) B_n^{-1}(\theta^0)$.

From (22), it follows that

$$B_n^{-1}(\theta^0)\ \Psi_n(\theta^0) B_n^{-1}(\theta^0) \overset{p}{=} B_n^{-1}(\theta^0) I_k B_n(\theta^0) = -I_k .$$

From (23) and the fact that $\hat{\theta}_n \xrightarrow{p} \theta$, it is seen that $V_n \xrightarrow{p} 0$. Now (24) can be rewritten as

$$(25)\qquad Y_n = Z_n(I_k + V_n) .$$

But $E_{\theta^0}(Y_n) = 0$, and $E_{\theta^0}(Y_n' Y_n) = I_k$ which implies that $\{Y_n\}$ is a (vector) r. v. bounded in probability and from (25) the same must hold for $\{Z_n\}$. So $Y_n \overset{p}{=} Z_n$. Thus from Definition 3, it follows that $\hat{\theta}_n$ is asymptotically efficient in the wide sense, since the sequence $\{W_n\}$ of that definition may be identified with the $\{Y_n\}$ sequence here. This completes the sketch of proof.

Remark. The above argument shows that the matrix $B_n(\theta^0)$ is the correct normalizing factor for $(\hat{\theta}_n - \theta^0)$ in case Z_n has a limit distribution.

Proof of Corollary 4.1. If $\hat{\theta}_n$ were not unique, let $\bar{\theta}_n$ be another consistent m. l. estimator of θ; i.e., $\tilde{\varphi}_n(\bar{\theta}_n) = 0$ and $\tilde{\varphi}_n(\hat{\theta}_n) = 0$, or equivalently

$$C^{-1}(\theta^0, n)\, \varphi_n(\hat{\theta}_n) = 0 \quad \text{and} \quad C^{-1}(\theta^0, n)\, \tilde{\varphi}_n(\bar{\theta}_n) = 0 .$$

But, for n sufficiently large, $\hat{\theta}_n$ and $\bar{\theta}_n$ both lie in the interior of A, for otherwise the consistency hypothesis will be violated. Since A is convex, it contains also the line segment joining $\hat{\theta}_n$ and $\bar{\theta}_n$. Consequently, by the k-dimensional Rolle's theorem, there exists a value of θ, say θ^*, in A such that $C^{-1}(\theta^0, n) \times \det[\Psi_n(\theta^*)] = 0$. ('det' stands for determinant.) But by hypothesis, for all θ in A and all large n, $C^{-1}(\theta^0, n) \Psi_n(\theta)$ is negative definite with probability one. Therefore, $C^{-1}(\theta^0, n) \det[\Psi_n(\theta^*)] = 0$ can occur only with probability zero. This implies that $\hat{\theta}_n$ and $\bar{\theta}_n$, for large n, become equal with probability tending to unity.

Alternate proof (due to the referee). If $\hat{\theta}_n$ and $\bar{\theta}_n$ are two distinct consistent estimators, let $g(\lambda) = C^{-1}(\theta^0) \varphi(\hat{\theta}_n + \lambda(\bar{\theta}_n - \hat{\theta}_n))$, where $0 \leq \lambda \leq 1$ and $\varphi(\cdot)$ is the log-likelihood function. Then $g(\cdot)$ takes its minimum at an interior point λ_0 of the interval. This is because $g(\lambda) < g(0)$ for small enough $\lambda > 0$ and $g(\lambda) < g(1)$ for λ close to 1, $(\lambda < 1)$, which is a consequence of the hypothesis that $\Psi_n(\theta) = (\varphi_{ij}(\theta))$ is negative definite, a. e. Clearly $g''(\lambda_0) \geqq 0$. But a simple computation yields

$$g''(\lambda_0) = C^{-1}(\theta^0) \sum_{i,j=1}^{k} \varphi_{ij}(\hat{\theta}_n + \lambda_0(\bar{\theta}_n - \hat{\theta}_n))\, (\bar{\theta}_i - \hat{\theta}_i)\, (\bar{\theta}_j - \hat{\theta}_j) < 0 ,$$

a.e., since $\hat{\theta}_n \neq \bar{\theta}_n$. [Hence $\hat{\theta}_i\,(\bar{\theta}_i)$ is the ith component of $\theta_n\,(\theta_n)$].

The contradiction contained in the precading two sentences proves the result.

Remark. If the X_n are identically distributed independent r.v.'s, the result of this corollary reduces to that of [*2*]. The above theorem in that case can be sharpened along the lines of [*11*] and [*12*]. Again various other sets of conditions may be given, and moreover the corresponding results for the independent r. v.'s of [*13*] may be generalized.

The following result is a partial extension of Theorem 2, to k-parameters, which includes some unstable processes. The notations of the above theorem will be used without further explanation.

Theorem 5. *Let $p_n(x_1, \ldots, x_n; \theta_1, \ldots, \theta_k)$, or p_n for short, be the finite dimensional densities, of the process, depending on a parameter $\theta = (\theta_1, \ldots, \theta_k)$ which belongs to a finite non-degenerate open cell A in the Euclidean k-space. Suppose p_n satisfies the following conditions:*

Condition (i). *Same as Condition* I *of Theorem* 4.

Condition (ii). *$C_{ii}(\theta, n)$ exists and $C_{ii}(\theta, n) \to \infty$ as $n \to \infty$, $i = 1, \ldots, k$, all $\theta \in A$.*

Condition (iii). *Same as Condition* V *of Theorem* 4.

Condition (iv). *Given $0 < \delta < 1$, there exists an $\varepsilon_\delta > 0$ such that for all $\theta \in A$, one has $\overline{\lim_{n\to\infty}} P[|\det(\Psi_n(\theta)\Gamma_n^{-1}(\theta))| \geqq \varepsilon_\delta] > 1 - \delta$, where* 'det' *stands for the determinant of the matrix, and Ψ_n and Γ_n were defined before.*

Condition (v). *Same as Condition* VI *of Theorem* 4.

Then the m.l. equation $\tilde{\varphi}_n(\theta) = 0$ has a (vector) root $\hat{\theta}_n$ which satisfies $(\hat{\theta}_n - \theta^0) \xrightarrow{p} 0$; i.e., $\hat{\theta}_n$ is a consistent estimator of θ. Moreover, such a $\hat{\theta}_n$ is also efficient in the weak sense (where the latter concept is an obvious vector analogue of Definition 1).

The proof of this result is obtained with a judicious mixture of the arguments (and methods) of Theorems 2, 3 and 4 and need not be repeated here. It may be noted that this result includes the consistency part of Theorem 4. This and the fact that Theorem 5 is not a full extension of Theorem 2 will become clear in the following section.

6. Applications to Linear Stochastic Equations

Consider a stochastic process $\{X_t, t \geqq 1\}$ which satisfies, for each integer t, the following conditions.

Condition 1. $X_t = \alpha_1 X_{t-1} + \cdots + \alpha_k X_{t-k} + u_t$, $(-\infty < \alpha_i < \infty)$, where $(\alpha_1, \ldots, \alpha_k) \varepsilon A$, a bounded non-degenerate cell in the Euclidean k-space, and the u_t are independent Gaussian r. v.'s each having mean zero and unit variance.

Condition 2. The k roots $m_1, \ldots, m_k$ of the characteristic equation

$$m^k - \alpha_1 m^{k-1} - \cdots - \alpha_k = 0\,, \tag{26}$$

are simple (i. e., $m_i \neq m_j$ if $i \neq j$).

Condition 3. For non-positive t, $u_t = 0$.

It is required to examine the consistency of the m.l. estimators $\hat{\alpha}_n$ of α. Since the u_t are independent Gaussian r. v.'s, the density of X_t is given by $p_n = p_n(x_1, \ldots, x_n; \alpha_1, \ldots, \alpha_k) = (2\pi)^{-n/2} \exp[-\sum_{t=1}^{n}(x_t - \cdots - \alpha_k x_{t-k})^2/2]$. To find the consistency of $\hat{\alpha}_n$, one has only to verify the conditions of Theorems 4 and 5. It should be pointed out that this problem is of considerable importance in many applications and numerous special studies have been made (cf., [*14*], [*8*], [*20*], [*1*], [*15*] among others) with special methods of attack. In what follows, a unified account of the problem emerges.

Since the p_n is a Gaussian density and A is bounded, Conditions I and VI are satisfied here. The other conditions of Theorem 4 will be verified first.

Now Conditions II and III can be checked together. With $n > k$, define

$$\tilde{\varphi}_n(\alpha) = (\partial \log p_n / \partial \alpha_i, i = 1, \ldots, k) = (\varphi_i(\alpha, n), i = 1, \ldots, k)$$
$$\Psi_n(\alpha) = (\partial^2 \log p_n / \partial \alpha_i \partial \alpha_j, i, j = 1, \ldots, k) = (\varphi_{ij}(\alpha, n), i, j = 1, \ldots, k).$$

Substituting for p_n, these expressions become

$$\varphi_i(\alpha, n) = \sum_{t=i+1}^{n} u_t X_{t-i}, \tag{27}$$

$$\varphi_{ij}(\alpha, n) = -\sum_{t=[i,j]+1}^{n} X_{t-i} X_{t-j}, [i, j] = \max(i, j). \tag{28}$$

Here it is required to compute (E stands for E_α below)

$$C_{ij}(\alpha, n) = E[\varphi_i(\alpha, n) \varphi_j(\alpha, n)] = -E(\varphi_{ij}(\alpha, n)).$$

To simplify the right side of $C_{ij}(\alpha, n)$, X_t has to be expressed in terms of u_t, and for this some properties of difference equations are needed.

Since the roots of (26) are simple, from the theory of difference equations (cf., [*9*] p. 564, or [*14*][2], p. 178), it follows that

$$X_t = a_1(t) u_1 + a_2(t) u_2 + \cdots + a_t(t) u_t,$$

where

$$a_r(t) = a(t - r) = \sum_{q=1}^{k} \lambda_q m_q^{t-r}, r = 1, \ldots, t, \quad \text{with} \quad \sum_{q=1}^{k} \lambda_q = 1.$$

Here the constants λ_q are the solutions of the equations (because $u_t = 0$, for t non-positive implies that $X_t = 0$, t non-positive),

$$\delta_{1t} = \sum_{q=1}^{k} \lambda_q m_q^{t-1}, t = 1, 0, -1, \ldots, -(k-2),$$

where, if for any q an $m_q = 0$, the corresponding λ_q is taken as zero, and $\delta_{1t} = 1$ or 0 according as $t = 1$ or $t \neq 1$. Hence, one obtains

$$X_t = \sum_{r=1}^{t} \sum_{q=1}^{k} \lambda_q m_q^{t-r} u_r. \tag{29}$$

If the roots m_i of (26) are allowed to be multiple also, then the λ_q of (29) will be functions of (polynomials in) t, and the computations become more complicated.

---

[2] On page 178 of MANN and WALD [*14*], equation (11) is $\varphi_r(t) = \sum_{i=1}^{k} \lambda_i p_i^{t-r+1}$. The right side should be $\sum_{i=1}^{k} \lambda_i p_i^{t-r}$, since $\varphi_t(t) = 1$ should hold when $k = 1$. This correction does not affect the main results of [*14*].

Thus, the variable λ_q case suggests the study of processes satisfying Condition 1 but with α_i as functions of t. This general problem is quite realistic in non-stationary processes, and some instances of it have been studied in the past [8]. However, the computations needed are lengthy in this general case. So, for the present, Condition 2 of this section will be assumed.

From (29), on noting that $E(X_t) = 0$ and the u_t are independent, the simplifications proceed as follows:

$$E(X_{t-i}X_{t-j}) = \sum_{r=1}^{t-[i,j]} \Big(\sum_{q=1}^{k} \lambda_q m_q^{t-i-r}\Big)\Big(\sum_{q=1}^{k} \lambda_q m_q^{t-j-r}\Big), \tag{29'}$$

and hence, after reduction, one obtains

$$C_{ij}(\alpha, n) = \sum_{q=1}^{k} \sum_{q'=1}^{k} \lambda_q \lambda_{q'} \frac{m_q^{|i-j|}}{(1 - m_q m_{q'})} \left(n - [i, j] \frac{m_q m_{q'} (m_q m_{q'})^{n-[i,j]+1}}{(1 - m_q m_{q'})}\right), \tag{30}$$

provided $m_q m_{q'} \neq 1$. It will be seen below that Theorem 4 is a multi-dimensional extension of Wald's theorem but not that of Theorem 2.

Case 1. Let $|m_q| < 1, q = 1, \ldots, k$. Then from (30), it follows that

$$\lim_{n\to\infty} \frac{C_{ij}(\alpha, n)}{n} = \sum_{q=1}^{k} \sum_{q'=1}^{k} \lambda_q \lambda_{q'} \frac{m_q^{|i-j|}}{(1 - m_q m_{q'})} = D_{ij}(\text{say}).$$

Since $\sum_{q=1}^{k} \lambda_q = 1$, and m_q are distinct, it is seen that $\lim_{n\to\infty} C_{ii}(\alpha, n) = \infty, i = 1, \ldots, k$. If $\max_i C_{ii}(\alpha, n) = C(\alpha, n)$, then $C(\alpha, n) = O(n)$ and it is found that $\lim_{n\to\infty} \frac{1}{C(\alpha, n)} \Gamma_n(\alpha) = (D_{ij})$ is a non-singular matrix. (For example, taking $k = 2$, one readily verifies that $D_{11}D_{12} - D_{12}^2 \neq 0$, because $|m_q| < 1$.)

Thus in this case Conditions II and III are satisfied.

Case 2. Let $|m_q| = 1$ for at least one $q (q = 1, \ldots, k)$. Suppose that $m_1 = 1$ and $|m_j| < 1, j = 2, \ldots, k$. Then (30) becomes, on simplification,

$$C_{ij}(\alpha, n) = \frac{\lambda_1^2}{2} (n - [i, j]) (n - [i, j] + 1) + O(n).$$

It follows that

$$\lim_{n\to\infty} \frac{C_{ij}(\alpha, n)}{n^2} = \frac{\lambda_1^2}{2}, \; i, j = 1, 2, \ldots, k.$$

If $m_1 = 1$, $m_2 = -1$ and $|m_{i'}| < 1, i' = 3, \ldots, k$, then from a similar analysis it is seen that

$$\lim_{n\to\infty} \frac{C_{ij}(\alpha, n)}{n^2} = \frac{\lambda_1^2 + \lambda_2^2}{2}, \; i, j = 1, \ldots, k.$$

Finally, if all combinations occur, say $m_1 = 1$, $m_2 = -1$, $m_3 = e^{i\theta_1}$ so that for some q_1, $m_{q1} = e^{-i\theta_1}$, etc., the above discussion reveals that, for all i, j, (since $m_3 m_{q_1} = 1$, one starts from (29') instead of (30))

$$\lim_{n\to\infty} \frac{C_{ij}(\alpha, n)}{n^2} = \frac{\lambda_1^2 + \lambda_2^2 + \lambda_3 \lambda_{q1}}{2}.$$

Thus summarizing the above for the case $|m_i| = 1$, it is found that $C^{-1}(\alpha, n)\Gamma_n(\alpha)$ tends to a singular matrix of rank one. Consequently Condition III fails.

Case 3. Suppose $\varrho = m_1$, $|\varrho| > 1$ and $|\varrho| >$ maximum $|m_j|$, for $j \geqq 2$. From (30), one can easily identify the matrix of Condition III to be the matrix of the expected values of the r. v.'s considered in Lemma 15 of [*15*]. Consequently, in this case also, the matrix in Condition III tends to a singular matrix of rank one, and Condition III fails. (Of course, this statement can be directly verified, without the above reference, by an independent computation.)

Out of the above cases, one has to determine those that satisfy Condition IV. Since Condition III has failed in Cases 2 and 3 above, the theorem is not applicable to those cases, and hence only the Case 1, $|m_q| < 1$ for all q, need be considered in the following.

In the case $|m_q| < 1$, however, $\mathrm{Var}_\alpha\, \varphi_{ij}(\alpha, n)$ has been computed in [14; p. 180], and was found to be

$$\mathrm{Var}_\alpha\, \varphi_{ij}(\alpha, n) = O(n - [i, j]) = O(n)\,.$$

Since $\mathrm{Var}_\alpha\, (\varphi_{ij}(\alpha, n))/C^2(\alpha, n) = O(1/n) \to 0$, as $n \to \infty$, Condition IV holds. Condition V is trivial in the present case.

Thus in the case $|m_q| < 1$, $q = 1, \ldots, k$, the Conditions I—VI of Theorem 4 are satisfied. If $|m_q| \geqq 1$, Conditions III and IV are not satisfied and hence this theorem is not applicable in the unstable case. However, a certain subclass of unstable processes are covered by Theorem 5, as seen below.

Case 4. Let $|m_q| > 1$ for $q = 1, \ldots, k$. Then from the fact that $\sum_{q=1}^{k} \lambda_q = 1$, and (30), it follows that $C_{ii}(\alpha, n) \geqq (n - i)$ so that Condition (iii) of Theorem 5 is satisfied. Conditions (i), (iii) and (v) hold in the present case also. This follows easily in all cases whatever the magnitudes of the roots are. The remaining Condition (iv) also holds in the present case since the roots are distinct and lie outside the unit circle. This follows from the results of ([*1*], Theorem 3.1 and Corollary 3.1), since $\Psi_n(\alpha)$ here is just B_n of ([*1*], p. 682) and the α there is the $k \times k$ matrix given by

$$\alpha = \begin{bmatrix} \alpha_1, \ldots, \alpha_k \\ 1, 0, \ldots, 0 \\ 0, 1, \ldots, 0 \\ 0, \ldots, 1, 0\,. \end{bmatrix}.$$

Thus in case $|m_q| > 1$, $q = 1, \ldots, k$, the Conditions (i)—(v) of Theorem 5 are satisfied. However, if some roots of the characteristic equation (26) lie outside and some inside or on the unit circle then (by ([*15*], Remark 2 on p. 216) the Condition (iv) of Theorem 5 fails and in this case the theorem is not applicable. In summary, the results of [*14*] and certain others on consistency may be obtained from Theorems 4 and 5. More precisely the following result holds.

Theorem 6. *Suppose the Conditions 1 to 3 (given at the beginning of this section) on the process* $\{X_n, n \geqq 1\}$ *are satisfied. If either* $|m_q| < 1$, *or* $|m_q| > 1$, *where* m_q, $q = 1, \ldots, k$, *are the roots of* (26), *then the m.l. estimator* $\hat{\alpha}_n$ *of* $\alpha = (\alpha_1, \ldots, \alpha_k)$ *is consistent and asymptotically efficient (in the wide sense in the first case and in the weak sense in the second case). Moreover, the consistent estimator* $\hat{\alpha}_n$ *is also unique.*

For the last part of the theorem one notes that the likelihood equation has only one root in the present case.

Note. The α_n are known to have a limit normal distribution when $|m_q| < 1$, $i = 1, \ldots, k$, (for the stable processes), [14], so that the efficiency in that case becomes a strict sense one. It may be noted that the result on the efficiency of $\hat{\alpha}_n$, was possible because of the foregoing results, and such an analysis could not be given in ([14], [1], or [15]).

7. Final Remarks

If Condition II of Theorem 4 is strengthened, the conclusion there on the efficiency of the consistent estimator may also be strengthened. For this, consider

Condition II'. If $C_{ii}(\theta, n) = E_\theta(\partial \log p_n/\partial\theta_i)^2$, then $C_{ii}(\theta, n) \to \infty$ with n, and for every integer r (independent of n) the correlation between

$$\frac{\partial \log p_n}{\partial\theta_i} \quad \text{and} \quad \frac{\partial \log p_{n+r}}{\partial\theta_i}$$

tends to one as $n \to \infty$, uniformly in r.

Now the following result holds.

Theorem 7. *The consistent m.l. estimator given by Theorem 4 (with Condition* II *replaced by* II' *above) has a joint limit distribution with mean vector θ and covariance matrix* $\lim_{n\to\infty} (C_n(\theta)\Gamma_n^{-1}(\theta))$; *(i.e., $(\hat{\theta}_n - \theta)B_n(\theta)$ has a limit d.f.).*

Remark. If the limit distribution of $\hat{\theta}_n$ (the m.l. estimator) is also Gaussian then they are efficient in the *strict sense*. However, further conditions are needed for the latter conclusion.

Proof. In Theorem 4 it is shown that (cf. (25) and the following)

$$(Y_n - Z_n) \xrightarrow{p} 0, \quad \text{as} \quad n \to \infty,$$

where $Y_n = \tilde{\varphi}_n(\theta^0)B_n^{-1}(\theta^0)$, $B_n^2(\theta) = \Gamma_n(\theta)$, and $Z_n = (\hat{\theta}_n - \theta^0)B_n(\theta^0)$. If $Y_n \xrightarrow{d} Y$, then by Lemma 3, $Z_n \xrightarrow{d} Y$. Consequently to prove this theorem, it suffices to show that $Y_n \xrightarrow{d} Y$.

First note that $E_{\theta^0}(Y_n) = 0$ and $E_{\theta^0}(Y_n' Y_n) = I_k$, for all n. Next consider

$$\begin{aligned} E_{\theta^0}(Y_{n+r}' - Y_n')(Y_{n+r} - Y_n) &= E_{\theta^0}(Y_{n+r}' Y_{n+r}) + E_{\theta^0}(Y_n' Y_n) - 2E_{\theta^0}(Y_{n+r}' Y_n) \\ &= 2I_k - 2E_{\theta^0}(Y_{n+r}' Y_n). \end{aligned} \tag{31}$$

By Condition III of the theorem $[C(\theta^0, n)]^{-1}\Gamma_n(\theta^0) \to$ a non-singular matrix $= D(\theta^0)$, say. Furthermore, $[C(\theta^0, n)]^{-1}E_{\theta^0}(\tilde{\varphi}_{n+r}'(\theta^0)\tilde{\varphi}_n(\theta^0)) =$

$$\left([C(\theta^0, n)]^{-1}E_{\theta^0}\left(\frac{\partial \log p_{n+r}}{\partial\theta_i}\frac{\partial \log p_n}{\partial\theta_j}\right)\right)$$

and, by Condition II', it follows that

$$[C(\theta^0, n)]^{-1}E_{\theta^0}\left(\frac{\partial \log p_{n+r}}{\partial\theta_i} - \frac{\partial \log p_n}{\partial\theta_i}\right)^2 \to 0$$

uniformly in r, for each $i = 1, \ldots, k$. Consequently, for n large and any fixed positive r,

$$[C(\theta^0, n)]^{-1}E_{\theta^0}[\tilde{\varphi}_{n+r}'(\theta^0)\tilde{\varphi}_n(\theta^0)] \sim [C(\theta^0, n)]^{-1}\Gamma_n(\theta^0) \to D(\theta^0),$$

where "$\sim$" stands for asymptotic equality. Hence Conditions II′ and III together imply that

$$E_{\theta^0}(Y'_{n+r}Y_n) = B^{-1}_{n+r}(\theta^0)E_{\theta^0}(\tilde{\varphi}'_{n+r}(\theta^0)\tilde{\varphi}_n(\theta^0))\,B_n^{-1}(\theta^0) \to I_k\,,$$

uniformly in r, as $n \to \infty$. Therefore the right side of (31) tends to the zero matrix, uniformly in r, as $n \to \infty$.

The above two facts imply that Y_n converges in mean square to a r.v. Y, so that $Y_n \xrightarrow{p} Y$, with mean vector zero, and covariance matrix I_k. Consequently

$$Z_n = (\hat{\theta}_n - \theta^0)\,[C(\theta^0, n)]^{1/2}B_n(\theta^0)\,[C(\theta^0, n)]^{-1/2} \xrightarrow{d} Y\,.$$

However, this is equivalent to the statement that $\hat{\theta}_n$ has a limit distribution with mean vector θ^0 and covariance matrix $D(\theta^0)$ (when normalized by $[C(\theta^0, n)]^{1/2}$). This completes the proof.

Now to show that this limit distribution of Y_n is Gaussian, further regularity conditions, on p_n, should be imposed, so that one may work, for instance, with the right side of

$$p_n(x_1, \ldots, x_n; \theta) = p_1(x_1; \theta)p_2(x_2 \mid x_1; \theta) \ldots p_n(x_n \mid x_1, \ldots, x_{n-1}; \theta)\,, \tag{32}$$

and consider the type of reasoning used in the central limit theorems for m-dependent r.v.'s (cf., e.g., [*7*]). However, the precise conditions are not yet available. Also the study of [*13*] for stable processes, with the above results, would be useful.

A more general form of Theorem 5 can also be obtained. For instance the following statement holds. (The notation of Theorem 4 will be used.)

Proposition. *Suppose Conditions* (i), (ii), (iii) *and* (iv) *of Theorem* 5 *hold and let Condition* (iv) be replaced by

Condition (iv)′. $\tilde{\varphi}_n(\theta)\Psi_n^{-1}(\theta) \xrightarrow{p} 0$ *as* $n \to \infty$ *for all* $\theta \in A$. *Then the m.l. equation* $\tilde{\varphi}_n(\theta) = 0$ *has a root* $\hat{\theta}_n$ *which is a consistent estimator of* θ.

The proof of this result follows as stated in the remarks after Theorem 5. This result applies to and includes more cases of the unstable processes, (for instance the result of [*15*], Part II is implied). However, this type of result is unsatisfactory since the essential point of the study on the asymptotic properties of m.l. estimators is to obtain certain verifiable conditions in applications. Thus the verification of Condition (iv)′ above for more general stochastic difference equations (e. g. those of [*15*]) often involves independent and lengthy computations. A better extension of Theorem 2 will therefore be of interest.

References

[*1*] Anderson, T. W.: On asymptotic distributions of estimates of parameters of stochastic difference equations. Ann. math. Statistics **30**, 676—687 (1959).

[*2*] Chanda, K. C.: A note on the consistency and maxima of the roots of likelihood equations. Biometrika **41**, 56—61 (1954).

[*3*] Cramér, H.: Mathematical methods of statistics. Princeton: Princeton University Press 1946.

[*4*] Erdös, P., and M. Kac: On certain limit theorems in the theory of probability. Bull. Amer. math. Soc. **52**, 292—302 (1946).

[*5*] Gnedenko, B. V., and A. N. Kolmogorov: Limit distributions for sums of independent random variables (Translation). Cambridge, Mass.: Addison-Wesley 1954.

[6] GRENANDER, U.: Stochastic processes and statistical inference. Ark. Mat. **1**, 195—277 (1950).
[7] HOEFFDING, W., and H. ROBBINS: The central limit theorem for dependent random variables. Duke math. J., **15**, 773—780 (1948).
[8] HURWICZ, L.: Variable parameters in stochastic processes. Cowles Commission Monograph, 10, New York: Wiley 1950.
[9] JORDAN, C.: Calculus of finite differences. Budapest: Röttig and Romwalter 1939.
[10] KRAFT, C., and L. LECAM: A remark on the roots of the maximum likelihood equation. Ann. math. Statistics **27**, 1174—1177 (1956).
[11] KULLDORFF, G.: On the conditions for consistency and asymptotic efficiency of maximum likelihood estimates. Skand. Aktuarietidskr. **40**, 129—144 (1958).
[12] LECAM, L.: On some asymptotic properties of maximum likelihood estimates and related Bayes' estimates. Univ. California Publ. Statistics **1**, 277—330 (1953).
[13] — On the asymptotic theory of estimation and testing hypotheses. Proc. Third Berkeley Sympos. math. Statist. Probab. **1**, 129—156 (1956).
[14] MANN, H. B., and A. WALD: On the statistical treatment of linear stochastic difference equations. Econometrica **11**, 173—220 (1943).
[15] RAO, M. M.: Consistency and limit distributions of estimators of parameters in explosive stochastic difference equations. Ann. math. Statistics **32**, 195—218 (1961).
[16] — Inference in stochastic processes — I. Teor. Verojatn. Primen. 8, 282—298 (1963).
[17] — Inference in stochastic processes — II: Asymptotic properties of maximum likelihood estimators. Notices Amer. math. Soc. **6**, 849 (1959).
[18] — Inference in stochastic processes — III: Nonlinear prediction, filtering and sampling theorems. (To appear.)
[19] — Some infernece theorems in stochastic processes. Bull. Amer. math. Soc. **69**, 72—77 (1963).
[20] RUBIN, H.: Consistency of maximum likelihood estimates in the explosive case. Cowles Commission Monograph, **10**, 356—364. New York: Wiley 1950.
[21] WALD, A.: Asymptotic properties of the maximum likelihood estimate of an unknown parameter of a discrete stochastic process. Ann. math. Statistics **19**, 40—46 (1948).

Carnegie Institute of Technology
Pittsburgh, Pennsylvania