The Annals of Statistics
1984, Vol. 12, No. 1, 161–171

ASYMPTOTIC CONDITIONAL INFERENCE FOR REGULAR NONERGODIC MODELS WITH AN APPLICATION TO AUTOREGRESSIVE PROCESSES

BY I. V. BASAWA AND P. J. BROCKWELL[1]

LaTrobe University and Colorado State University; and Colorado State University and Kuwait University

A conditional limit theorem is derived for a certain class of stochastic processes whose distributions constitute a nonergodic family. The limit theorem allows us to study the asymptotic behaviour under the conditional model of some standard statistical procedures by making use of results for ergodic families. Explosive Gaussian autoregressive processes are studied in some detail. Here the conditional process is shown to be a nonexplosive Gaussian autoregression bearing a simple relation to the original process. Some optimality results under the conditional model are given for estimators and tests based on the unconditional likelihood.

1. Introduction. Asymptotic inference problems for parameters of nonergodic stochastic processes (see Section 2) have been studied recently by Feigin (1976, 1981), Heyde (1978), Basawa and Scott (1977, 1983), Basawa and Koul (1979, 1983) and others. The main difficulties which arise can be summarized as follows. The usual statistical procedures such as maximum likelihood (ML) estimation and likelihood ratio (LR) tests involve nonstandard limit distributions. The classical efficiency criteria need to be modified to establish the asymptotic optimality of ML estimators (cf. Heyde, 1978) while the LR statistic can be shown to have a suboptimal asymptotic power function (Basawa and Koul, 1983). Also the limit distributions of the ML estimators and the LR statistics for such processes are computationally far from simple.

The regular nonergodic models of Basawa and Koul (1979) can be regarded intuitively as mixtures of ergodic models, so it is natural to attempt to assess the asymptotic performance of estimators and tests relative to the conditional distribution of the process given the mixing random variable. In cases where the mixing random variable is ancillary, this is in accordance with standard statistical practice. (For general discussions of conditional inference see Barndorff-Nielsen (1978), Cox and Hinkley (1974), Lauritzen (1982) and, for applications closely related to ours, Keiding (1974) and Heyde and Feigin (1975).) Under conditions (2.5)–(2.7) following, the conditional model is shown to be locally asymptotically normal, making it a simple matter to check for optimality of estimators and tests in the conditional framework. Since the mixing random variable is usually

Received June 1982; revised September 1983.

[1] Research supported by NSF Grant MCS-8202335.

AMS 1980 *subject classifications.* Primary 62M07, 62M09; secondary 62M10.

Key words and phrases. Nonergodic processes, asymptotic conditionality principle, conditionally locally asymptotically normal families, maximum likelihood estimators, score tests, conditional limit theorem.

unobserved, direct computation of conditional estimates and test functions may be impossible without using a sample approximation. However this problem does not arise when we wish only to establish conditional optimality of unconditional procedures as in Section 5.

In Sections 3, 4 and 5 we study explosive Gaussian autoregressive processes under the conditional model. It is known, Anderson (1959), that the unconditional MLE of the coefficient in the first order case is unconditionally asymptotically normal if and only if it is standardized by the observed Fisher information. Conditionally we show this estimator to be asymptotically normal with the standardization given in Theorem 5.1. Conditional optimality properties follow from standard results for locally asymptotically normal families, LeCam (1960).

2. Conditional inference for regular nonergodic families. The notion of regular nonergodic families of distributions was introduced by Basawa and Koul (1979, 1982). Here we work in a slightly narrower framework but one which is adequate for a variety of applications. Let $\{X_n, n = 1, 2, \cdots\}$ be a stochastic process such that for each n, $X(n) = (X_1, \cdots, X_n)$ has a density $p_n(\cdot; \theta)$ with respect to some product measure μ^n on $(\mathbb{R}^n, \mathscr{B}^n)$. The parameter θ will be assumed to have as its parameter space some open subset Θ of $\mathbb{R}^k$. Let $P_{n,\theta}$ be the probability distribution on $(\mathbb{R}^n, \mathscr{B}^n)$ corresponding to the density $p_n(\cdot; \theta)$ and let P_θ be the probability distribution on $(\mathbb{R}^\infty, \mathscr{B}^\infty)$ of the process $\{X_n\}$. Expectation with respect to P_θ will be denoted by E_θ. Let $\Lambda_n(\psi, \theta)$ denote the loglikelihood ratio,

$$\Lambda_n(\psi, \theta) = \ln\{p_n(X(n); \psi)/p_n(X(n); \theta)\}, \tag{2.1}$$

and assume the existence for each $\theta \in \Theta$ of

$$I_{nj}(\theta) = -E_\theta\left[\frac{\partial^2 \ln p_n(X(n); \theta)}{\partial\theta_j^2}\right], \quad j = 1, \cdots, k, \tag{2.2}$$

$$\Delta_{nj}(\theta) = I_{nj}^{-1/2}(\theta)\,\frac{\partial \ln p_n(X(n); \theta)}{\partial\theta_j}, \quad j = 1, \cdots, k, \tag{2.3}$$

and

$$G_{nij}(\theta) = -I_{ni}^{-1/2}(\theta)\,\frac{\partial^2 \ln p_n(X(n); \theta)}{\partial\theta_i \partial\theta_j}\, I_{nj}^{-1/2}(\theta), \quad i, j = 1, \cdots, k, \tag{2.4}$$

where $0 < I_{nj}(\theta)$ and $I_{nj}(\theta) \uparrow \infty$ as $n \to \infty$ for each j and θ. Assume also that for each $h \in \mathbb{R}^k$ and each $\theta \in \Theta$ we have under P_θ as $n \to \infty$,

$$\Lambda_n(\theta + I_n^{-1/2}(\theta)h, \theta) = h^\top \Delta_n(\theta) - \tfrac{1}{2} h^\top G_n(\theta) h + o_p(1), \tag{2.5}$$

$$G_n(\theta) \to_p W(\theta), \tag{2.6}$$

and

$$(\Delta_n(\theta), G_n(\theta)) \to_d (W^{1/2}(\theta)Z, W(\theta)) \tag{2.7}$$

where Z is a $k \times 1$ vector with independent standard normal components while $\Delta_n(\theta)$ is the $k \times 1$ vector $[\Delta_{nj}(\theta)]_{j=1,\cdots,k}$, $I_n(\theta)$ is the diagonal matrix $[I_{nj}(\theta)\delta_{ij}]_{i,j=1,\cdots,k}$, $G_n(\theta) = [G_{nij}(\theta)]_{i,j=1,\cdots,k}$ and $W(\theta)$ is a positive semidefinite random matrix independent of Z.

If the distribution of $W(\theta)$ is degenerate then the family $\{p_n(\cdot;\theta)\}$ is locally asymptotically normal (L.A.N) at θ in the sense of Le Cam (1960). When $W(\theta)$ has a nondegenerate distribution the family is said to be nonergodic.

Supercritical branching processes provide examples of nonergodic families. Consider for example the Bienaymé-Galton-Watson process $\{X_n, n = 1, 2, \cdots\}$ with $X_0 = 1$, $E\, X_1 = \theta > 1$ and $E\, z^{X_1} = z/[\theta - (\theta - 1)z]$. It is well-known that $X_n/\theta^n \rightarrow_{\text{a.s.}} W$ where W has the standard exponential distribution and that, conditional on W, $X_j - X_{j-1}$, $j = 1, 2, \cdots$, are independent Poisson random variables with means $W(\theta^j - \theta^{j-1})$. Some straightforward calculations for this process (see Basawa, 1981a) show that the assumptions (2.5) – (2.7) are all satisfied with $k = 1$,

$$I_n(\theta) = \frac{\theta^n - 1}{\theta(\theta - 1)^2}, \quad \Delta_n(\theta) = \frac{\theta - 1}{\theta^n - 1}[\theta(X_n - 1) - (\theta - 1)\textstyle\sum_1^n X_j],$$

$$G_n(\theta) = \frac{1}{\theta(\theta^n - 1)}[\theta^2(X_n - 1) - (\theta - 1)^2 \textstyle\sum_1^n X_j],$$

and

$$W(\theta) = W.$$

In this case it is natural to assess inference procedures conditionally on W since (a) W is ancillary for θ, (b) $\{X_n\}$ has a particularly simple probabilistic structure for given W, and (c) under the conditional probability measure given W the unconditionally nonergodic family becomes locally asymptotically normal and standard asymptotic theory applies.

This example suggests the possible desirability of conditioning on the mixing random variable $W(\theta)$ in the general framework of the assumptions (2.5) – (2.7). Such a suggestion is made in the monograph of Basawa and Scott (1983). We show below (Theorem 2.1) that under the assumptions (2.5) – (2.7), conditioning on $W(\theta)$ does indeed have the expected effect of reducing the nonergodic family to a locally asymptotically normal family, thereby making it a relatively simple matter to study the asymptotic performance of estimators and tests under the conditional model. A detailed study is made of such conditioning applied to an explosive Gaussian autoregressive process which, like the supercritical branching process, is nonergodic. The conditioned process has a particularly nice structure, namely that of a nonexplosive autoregression related in a simple way to the original one. In Section 5 some standard statistical procedures are shown to be optimal when assessed in terms of the conditional model.

THEOREM 2.1 *Let $\{X_n\}$ be a process satisfying (2.5) – (2.7) and let $P_{\theta|w}$ be a conditional distribution under P_θ for $\{X_n\}$ given $W(\theta) = w$. Then under $P_{\theta|w}$ as $n \rightarrow \infty$,*

$$\text{(2.8)} \qquad \Lambda_n(\theta + I_n^{-1/2}(\theta)h, \theta) = h^{\top}\Delta_n(\theta) - \tfrac{1}{2}h^{\top}wh + o_p(1), \quad \forall\, h \in \mathbb{R}^k,$$

$$G_n(\theta) \to_p w \tag{2.9}$$

and

$$(\Delta_n(\theta), G_n(\theta)) \to_d (w^{1/2}Z, w), \tag{2.10}$$

(*except possibly for* $w \in N(\theta)$ *where* $P_\theta(W(\theta) \in N(\theta)) = 0$).

PROOF. To establish (2.10) it clearly suffices to show that

$$E_\theta[\exp(is^\top \Delta_n(\theta)) \mid W(\theta)] \to_{\text{a.s.}} \exp(-\tfrac{1}{2}s^\top W(\theta)s) \tag{2.11}$$

for all real $(k \times 1)$ vectors s. By Skorokhod's theorem (Billingsley, 1971, page 7) condition (2.7) implies the existence, for each θ, of a probability space on which is defined a sequence of random vectors $\{(\Delta_n^*(\theta), W^*(\theta))\}$ and a random variable $\Delta^*(\theta)$ such that $(\Delta_n^*(\theta), W^*(\theta)) =_d (\Delta_n(\theta), W(\theta))$ and $(\Delta_n^*(\theta), W^*(\theta)) \to_{\text{a.s}} (\Delta^*(\theta), W^*(\theta))$.

The dominated convergence theorem for conditional expectations and condition (2.7), which specifies the distribution of $(\Delta^*(\theta), W^*(\theta))$, then give

$$E[\exp(is^\top \Delta_n^*(\theta)) \mid W^*(\theta)] \to E(\exp(is^\top \Delta^*(\theta)) \mid W^*(\theta)] = \exp(-\tfrac{1}{2}s^\top W(\theta)s),$$

thus establishing (2.11).

To prove (2.8), observe that conditions (2.5) and (2.6) imply that the indicator function of the set $S(\theta) = \{ \mid \Lambda_n - h^\top \Delta_n + \tfrac{1}{2}h^\top Wh \mid > \varepsilon\}$ converges in probability to zero under P_θ as $n \to \infty$ for each $\varepsilon > 0$. Applying the dominated convergence theorem for conditional expectations again gives $E_\theta[I_{S(\theta)} \mid W] \to 0$ as $n \to \infty$, which is equivalent to (2.8).

REMARK. The conditions (2.5) – (2.7) may be satisfied when $I_n(\theta)$ is not defined precisely as in (2.2). The conclusion of Theorem 2.1 however remains valid since it relies only on (2.5) – (2.7). We make use of this fact in the following sections to simplify the statement of the results for autoregressive processes.

3. The explosive Gaussian autoregressive process. We restrict attention for the moment to the explosive Gaussian autoregressive process

$$X_n - \theta X_{n-1} = Z_n, \quad n = 1, 2, \cdots, \quad |\theta| > 1, \tag{3.1}$$

where $X_0 = 0$ and $\{Z_n\}$ is a sequence of independent standard normal random variables.

In the notation of Section 2 we easily find that

$$p_n(X(n); \theta) = (2\pi)^{-n/2}\exp\{-\tfrac{1}{2} \textstyle\sum_{j=1}^n (X_j - \theta X_{j-1})^2\}. \tag{3.2}$$

Instead of defining $I_n(\theta)$ by equation (2.2), which gives the cumbersome expression $(\theta^2 - 1)^{-2}[\theta^{2n} - \theta^2 - (n-1)(\theta^2 - 1)]$, we shall use the asymptotically equivalent definition,

$$I_n(\theta) = \theta^{2n}/(\theta^2 - 1)^2, \tag{3.3}$$

which gives, with the definitions (2.3) and (2.4),

$$\Delta_n(\theta) = (\theta^2 - 1)\theta^{-n} \textstyle\sum_1^n X_{j-1}(X_j - \theta X_{j-1}) \tag{3.4}$$

and

$$G_n(\theta) = (\theta^2 - 1)^2\theta^{-2n} \textstyle\sum_1^n X_{j-1}^2. \tag{3.5}$$

With this choice of $I_n(\theta)$, it is known that the basic conditions (2.5)–(2.7) are satisfied. To see this note first that there is a random variable $Y(\theta)$ such that

$$\sqrt{\theta^2 - 1}\, \theta^{-n}X_n \to_{\text{a.s.}} Y(\theta) \quad \text{as} \quad n \to \infty. \tag{3.6}$$

This result follows at once from the fact that

$$E_\theta[\theta^{-n-1}X_{n+1} - \theta^{-n}X_n]^2 = \theta^{-2(n+1)}$$

(see e.g. Cramer and Leadbetter (1967), condition (3.5.6)). Moreover it is clear that $Y(\theta)$ is normally distributed with mean 0 and variance 1. A strengthened version of (2.6) can now be established as follows:

$$\begin{aligned} G_n(\theta) &= (\theta^2 - 1)^2\theta^{-2n} \textstyle\sum_{j=1}^n X_{j-1}^2 \\ &= (\theta^2 - 1)^2 \textstyle\sum_{j=1}^\infty (\theta^{-n+j}X_{n-j})^2 I_{[1,n]}(j)\theta^{-2j} \\ &\to_{\text{a.s.}} (\theta^2 - 1)Y(\theta)^2 \textstyle\sum_{j=1}^\infty \theta^{-2j} \quad \text{as} \quad n \to \infty, \end{aligned}$$

the last step being a consequence of the dominated convergence theorem since for almost every realization $\{\theta^{-k}X_k(\omega)\}$ there is a bound $K(\omega) < \infty$ such that $|\theta^{-k}X_k(\omega)| \le K(\omega)$ for all k. Hence

$$G_n(\theta) \to_{\text{a.s.}} Y(\theta)^2 \quad \text{as} \quad n \to \infty. \tag{3.7}$$

A completely analogous argument shows that

$$\Delta_n(\theta) = Y(\theta)(1 - \theta^{-2})^{1/2} \textstyle\sum_{j=0}^{n-1} \theta^{-j}Z_{n-j} \to_{\text{a.s.}} 0. \tag{3.8}$$

The random variables $Y(\theta)$ and $\sum_{j=0}^{n-1} \theta^{-j}Z_{n-j}$ are joint normal with zero means and covariance

$$\begin{aligned} E_\theta(Y(\theta) \textstyle\sum_{j=0}^{n-1} \theta^{-j}Z_{n-j}) &= \lim_{m\to\infty}E_\theta[(\theta^{-m} \textstyle\sum_{i=0}^{m-1} \theta^{i}Z_{m-i})(\sum_{j=0}^{n-1} \theta^{-j}Z_{n-j})]\sqrt{\theta^2 - 1} \\ &= n\theta^{-n}\sqrt{\theta^2 - 1} \to 0 \quad \text{as} \quad n \to \infty. \end{aligned}$$

The second term in (3.8) therefore converges in distribution to $Y(\theta)Z$ where Z and $Y(\theta)$ are independent standard normal random variables. Combining (3.7) and (3.8), we obtain

$$(\Delta_n(\theta), G_n(\theta)) \to_d (Y(\theta)Z, Y^2(\theta)). \tag{3.9}$$

Conditions (2.6) and (2.7) are therefore satisfied with $W(\theta) = Y(\theta)^2$. The validity of (2.5) is a consequence of the fact that the densities (3.2) constitute a globally curved exponential family.

Instead of conditioning on $W(\theta) = Y(\theta)^2$, it is more convenient for the autoregressive process to condition on $Y(\theta)$. The proof of Theorem 2.1 carries over essentially without change to establish that if $P_{\theta|y}$ is a conditional distribution for $\{X_n\}$ given $Y(\theta) = y$, then under $P_{\theta|y}$ as $n \to \infty$, the results (2.8) – (2.10) hold with $w^{1/2}$ replaced by y. In Section 4 we shall derive the conditional distribution of $\{X_n\}$ given $Y(\theta)$ by showing that the process $\{X_n - E(X_n | Y(\theta))\}$ conditional on $Y(\theta)$ is a nonexplosive Gaussian autoregression. Although this result is not essential for our study, we give the result since it is of interest in itself.

A similar analysis can be carried out for the kth order autoregressive process, AR(k), defined by

$$(3.10)\qquad X_n - \beta_1 X_{n-1} - \beta_2 X_{n-2} - \cdots - \beta_k X_{n-k} = Z_n, \quad n = 1, 2, \cdots,$$

where $Z_1, Z_2, \cdots$ are i.i.d. $N(0, 1)$ variates and $X_n = 0$, $1 - k \le n \le 0$. An alternative representation of (3.10) is

$$(3.11)\qquad (1 - \theta B) \textstyle\prod_{j=1}^{k-1} (1 - \theta_j B) X_n = Z_n$$

where B is the backward shift operator with $B^k X_n = X_{n-k}$, and $(\theta_1, \theta_2, \cdots, \theta_{k-1}, \theta)$ are certain functions of $\beta = (\beta_1, \beta_2, \cdots, \beta_k)$ determined by equating the left-hand sides of (3.10) and (3.11). Assume that

$$|\theta| > \max\{1, |\theta_1|, |\theta_2|, \cdots, |\theta_{k-1}|\}.$$

It can be shown as for the AR(1) process that $\sqrt{\theta^2 - 1}\,\theta^{-n} X_n$ converges almost surely to a random variable $Y(\theta)$ where $Y(\theta)$ again has the distribution $N(0, 1)$.

Using the notation of Section 2 we find that

$$p_n(X(n); \beta) = (2\pi)^{-n/2} \exp\{-\tfrac{1}{2} \textstyle\sum_{j=1}^n (X_j - \beta_1 X_{j-1} - \cdots - \beta_k X_{j-k})^2\}.$$

Defining $I_n(\theta)$ as in (3.3) and $I_{nj}(\theta) = I_n(\theta)$, $j = 1, \cdots, k$, we obtain

$$(3.12)\qquad \Delta_n(\beta) = I_n^{-1/2}(\theta)(S_{n1}(\beta), S_{n2}(\beta), \cdots, S_{nk}(\beta))^{\top}$$

with

$$S_{ni}(\beta) = \textstyle\sum_{r=1}^n X_{r-i}(X_r - \beta_1 X_{r-1} - \cdots - \beta_k X_{r-k}), \qquad i = 1, \cdots, k.$$

The $(k \times k)$ matrix $G_n(\beta)$ has components

$$(3.13)\qquad G_{nij}(\beta) = I_n^{-1}(\theta) \textstyle\sum_{r=1}^n X_{r-i} X_{r-j}.$$

Corresponding to (3.7) we have the result

$$(3.14)\qquad G_n(\beta) \to_{\text{a.s.}} Y(\theta)^2 \Sigma,$$

where Σ is the $(k \times k)$ covariance matrix with rank 1 and components

$$\Sigma_{ij} = \theta^{2-i-j}, \quad i, j = 1, \cdots, k,$$

and $Y(\theta)$ has the distribution $N(0, 1)$. The analogue of (3.9) is

$$(3.15)\qquad (\Delta_n(\beta), G_n(\beta)) \to_d (Y(\theta)\Sigma^{1/2} Z, Y(\theta)^2 \Sigma)$$

where Z is a $(k \times 1)$ vector of independent standard normal components, independent of Y. Proofs of (3.14) and (3.15) are given by Basawa and Koul (1979).

It can easily be verified that the conditions (2.5) – (2.7) are satisfied. For an arbitrary real $(k \times 1)$ vector h, with $I_n(\theta)$ as in (3.3) and with $\beta = (\beta_1, \beta_2, \cdots,$

$\beta_k)^\top$, we have

$$\Lambda_n(\beta + I_n^{-1/2}(\theta)h, \beta) = h^\top \Delta_n(\beta) - ½h^\top G_n(\beta)h, \tag{3.16}$$

where $\Delta_n(\beta)$ and $G_n(\beta)$ are defined in (3.12) and (3.13) respectively. The AR(k) model under consideration thus satisfies the requirements of a regular nonergodic family. Notice that the limits in (3.14) and (3.15) depend on β only through θ. As in the case of the AR(1) process, we shall condition on $Y(\theta) = \sqrt{1 - \theta^2}\,\lim_{n\to\infty}\theta^{-n}X_n$ rather than on $W(\theta) = Y(\theta)^2\Sigma$. In the next section we derive the conditional distribution of the process given $Y(\theta)$.

4. The conditioned autoregressive process. Consider the explosive Gaussian autoregressive process defined in (3.11). We shall show that, conditionally on $Y(\theta) = \lim_{n\to\infty}\theta^{-n}\sqrt{\theta^2 - 1}\,X_n$, the sequence

$$W_n = X_n - E(X_n \mid Y), \quad n = 1, 2, \cdots, \tag{4.1}$$

is distributed as the autoregressive process,

$$\begin{cases} (\theta - B)\prod_{j=1}^{k-1}(1 - \theta_j B)W_n = Z_n, & n = 1, 2, \cdots, \\ W_n = 0, & -k + 1 \le n \le 0. \end{cases} \tag{4.2}$$

We establish the result first in the case $k = 1$.

THEOREM 4.1. *If $\{X_n, n = 1, 2, \cdots\}$ is the Gaussian* AR(1) *process,*

$$\begin{cases} (1 - \theta B)X_n = Z_n, \\ X_0 = 0 \end{cases} \tag{4.3}$$

with $\{Z_n\}$ an i.i.d. *standard normal sequence and $|\theta| > 1$, then conditional on $Y(\theta) = \lim_{n\to\infty}\theta^{-n}\sqrt{\theta^2 - 1}\,X_n$, the sequence*

$$W_n = X_n - E(X_n \mid Y) = X_n - (\theta^n - \theta^{-n})Y/\sqrt{\theta^2 - 1}, \tag{4.4}$$

is distributed as the autoregressive process,

$$\begin{cases} (\theta - B)W_n = Z_n \\ W_0 = 0. \end{cases} \tag{4.5}$$

PROOF. First note from (4.3) that

$$\theta^n Y = \sqrt{\theta^2 - 1}\,X_n + Z_{n+1}^* \tag{4.6}$$

where $Z_{n+1}^* = \sqrt{\sigma^2 - 1}\,\sum_1^\infty \theta^{-j}Z_{n+j}$ is distributed as $N(0, 1)$ independently of $Z_1, \cdots, Z_n$. Writing equations (4.3) and (4.6) in matrix form, we see that

$$A\begin{bmatrix} \mathbf{X}_n \\ Y \end{bmatrix} = \begin{bmatrix} \mathbf{Z}_n \\ Z_{n+1}^* \end{bmatrix} \tag{4.7}$$

where $\mathbf{X}_n^{\mathrm{T}} = [X_1, \cdots, X_n]$, $\mathbf{Z}_n^{\mathrm{T}} = [Z_1, \cdots, Z_n]$ and

$$A = \begin{bmatrix} 1 & & & & \\ -\theta & 1 & & & \\ & \ddots & \ddots & & \\ & & -\theta & 1 & \\ & & & -(\theta^2-1)^{1/2} & \theta^n \end{bmatrix}. \tag{4.8}$$

From (4.7) it follows at once that the inverse of the covariance matrix of $[\mathbf{X}_n, Y]^{\mathrm{T}}$ is

$$A^{\mathrm{T}}A = \begin{bmatrix} 1+\theta^2 & -\theta & & & & \\ -\theta & 1+\theta^2 & -\theta & & & \\ & \ddots & \ddots & \ddots & & \\ & & -\theta & 1+\theta^2 & -\theta & \\ & & & -\theta & \theta^2 & -\theta^n(\theta^2-1)^{1/2} \\ & & & & -\theta^n(\theta^2-1)^{1/2} & \theta^{2n} \end{bmatrix}. \tag{4.9}$$

Deleting the last row and column we obtain the inverse of the conditional covariance matrix of $\mathbf{X}_n$ given Y which we immediately recognize as the inverse of the covariance matrix of the AR(1) defined by (4.5).

The conditional expectation of X_n given Y is found to be

$$\begin{aligned} E(X_n \mid Y) &= Y \operatorname{Cov}(X_n, Y)/\operatorname{Var} Y \\ &= Y \lim_{m\to\infty}[\operatorname{Cov}(\theta^{-m}X_m, X_n)/\operatorname{Var}(\theta^{-m}X_m)]/\sqrt{\theta^2-1} \\ &= (\theta^n - \theta^{-n})Y/\sqrt{\theta^2-1}. \end{aligned} \tag{4.10}$$

Since for every n the conditional distribution given Y of the vector $\{X_m - E(X_m \mid Y)\}_{m=1,\cdots,n}$ is Gaussian with mean zero and covariance matrix corresponding to the autoregressive process defined by (4.5), the proof is complete.

COROLLARY 4.1. *If $\{X_n\}$ is the* AR(k) *process defined by*

$$\begin{cases} (1-\theta B)\prod_{j=1}^{k-1}(1-\theta_j B)X_n = Z_n, & n = 1, 2, \cdots, \\ X_n = 0, & -k+1 \le n \le 0, \end{cases} \tag{4.11}$$

with $|\theta| > \max\{1, |\theta_1|, \cdots, |\theta_{k-1}|\}$ and $\{Z_n\}$ an i.i.d. *standard normal sequence, then conditional on $Y = \lim_{n\to\infty}\theta^{-n}\sqrt{\theta^2-1}\,X_n$, the process $W_n = X_n - E(X_n \mid Y)$ is distributed as the autoregression defined by* (4.2).

PROOF. The process $\{\prod_{j=1}^{k-1}(1-\theta_j B)X_n\} = \{U_n\}$ satisfies the conditions of Theorem 4.1, so conditional on $\lim_{n\to\infty}\theta^{-n}\sqrt{\theta^2-1}\,U_n = \prod_{j=1}^{k-1}(1-\theta^{-1}\theta_j)Y$, the process $\{V_n\} = \{U_n - E(U_n \mid Y)\}$ is distributed as the AR(1), $(\theta - B)V_n = Z_n$, $n = 1, 2, \cdots$, with $V_0 = 0$. This proves the corollary.

5. Conditional inference for the autoregressive process. We now apply the foregoing results to the autoregressive process defined by (3.10) and (3.11). (The notation of Section 3 will be used throughout.) We have seen already that the process satisfies the conditions of Theorem 2.1. Notice also that the unconditional maximum likelihood estimator $\hat{\beta}_n$ of β satisfies

(5.1) $$G_n(\beta)I_n^{1/2}(\theta)(\hat{\beta}_n - \beta) - \Delta_n(\beta) = 0.$$

Application of Theorem 2.1 together with standard results for L.A.N. models now yields the following theorem.

THEOREM 5.1. *For the* AR(k) *process defined by* (3.10) *and* (3.11), *we have, under the conditional measure* $P_\theta(\cdot \mid Y(\theta) = y)$,

(i) $$\Delta_n(\beta) \to_d N_k(0, y^2\Sigma) \quad as \quad n \to \infty,$$
$$G_n(\beta) \to_p y^2\Sigma \quad as \quad n \to \infty,$$
$$\hat{\beta}_n \to_p \beta \quad as \quad n \to \infty,$$
$$and \quad \theta^n(\hat{\beta}_n - \beta) \to_d Z\,\frac{\sigma^2(1-\theta^{-2})^2}{y(1-\theta^{-2k})}\,(1, \theta^{-1}, \dots, \theta^{-k+1})^{\mathsf{T}},$$
where Z is a standard normal random variable.

(ii) $\hat{\beta}_n$ *is asymptotically optimal in the following sense. Let T_n be any estimator of β for which there exists a random vector $T(\beta)$ such that for each $h \in \mathbb{R}^k$ the distribution under $P_{\beta+I_n^{-1/2}(\theta)h}$ of $(G_n(\beta), I_n^{1/2}(\theta)(T_n - \beta - I_n^{-1/2}(\theta)h))$ converges weakly to that of $(Y(\theta)^2\Sigma, T(\beta))$ as $n \to \infty$. Then, denoting by Φ the standard normal* c.d.f., *we have for each $\delta \geq 0$*
$$\lim_{n\to\infty} P_\theta\left(\frac{\theta \mid y \mid \sqrt{1-\theta^{-2k}}}{(\theta^2-1)^{3/2}}\,\theta^n \mid T_n - \beta \mid < \delta \mid Y = y\right) \leq 2\Phi(\delta) - 1,$$
with equality holding when $T_n = \hat{\beta}_n$. (Here $\mid T_n - \beta \mid$ denotes the Euclidean norm of $T_n - \beta$.)

(iii) *If $k = 1$ and if ϕ_n is any test function for testing $\theta = \theta_0$ against $\theta > \theta_0$ such that for each $y \in \mathbb{R}$ and $\alpha \in (0, 1)$,*
$$\lim_{n\to\infty} E_{\theta_0}(\phi_n(X(n)) \mid Y(\theta) = y) = \alpha,$$
then for the sequence of alternatives $\theta_n(h) = \theta_0 + I_n^{-1/2}(\theta_0)h$ with $h > 0$ we have
$$\limsup_{n\to\infty} E_{\theta_n(h)}(\phi_n(X(n)) \mid Y(\theta) = y) \leq 1 - \Phi(z_{1-\alpha} - h \mid y \mid),$$
where $z_{1-\alpha}$ is the $(1-\alpha)$-quantile of Φ. Equality is achieved by the test
$$\phi_n^*(X(n)) = \begin{cases} 1 & \text{if } \left(\sum_1^n X_{j-1}^2\right)^{1/2}(\hat{\theta}_n - \theta_0) > z_{1-\alpha}, \\ 0 & \text{otherwise}, \end{cases}$$
where $\hat{\theta}_n$ is the unconditional maximum likelihood estimator of θ.

The results of Theorem 5.1 can be derived as indicated by using Theorem 2.1

and standard results for L.A.N. families. A more direct approach however can be made via the conditional process derived in Section 4. We illustrate by establishing the very first result in the case $k = 1$, namely $\Delta_n(\theta) \to_d N(0, y^2)$ under the conditional measure. We have, under $P_\theta(\cdot \mid Y = y)$,

$$
\begin{aligned}
I_n^{-1/2}(\theta) & \frac{\partial}{\partial\theta} \log p_n(X(n); \theta) \\
&= I_n^{-1/2}(\theta) \textstyle\sum_1^n (X_j - \theta X_{j-1})X_{j-1} \\
&= I_n^{-1/2}(\theta) \textstyle\sum_1^n (W_j - \theta W_{j-1})(W_{j-1} + \theta^{j-1}(\theta^2 - 1)^{-1/2} y) + o(1) \\
&= I_n^{-1/2}(\theta) \textstyle\sum_1^n W_j W_{j-1} - \theta I_n^{-1/2}(\theta) \sum_1^n W_{j-1}^2 \\
&\quad + y(\theta^2 - 1)^{-1/2} I_n^{-1/2}(\theta) \textstyle\sum_1^n \theta^{j-1}(W_j - \theta W_{j-1}) + o(1)
\end{aligned}
$$

where W_n was defined by (4.4). The first two terms on the right are easily seen to converge to zero in probability. The third term is a linear combination of normal variates. Using the result

$$\mathrm{Cov}(W_j, W_k \mid Y = y) = \frac{\theta^{-|j-k|} - \theta^{-|j+k|}}{\theta^2 - 1}$$

we can verify that

$$\mathrm{Var}\{I_n^{-1/2}(\theta) \textstyle\sum_1^n \theta^{j-1}(W_j - \theta W_{j-1}) \mid Y = y\} \to \theta^2 - 1 \quad \text{as} \quad n \to \infty.$$

Finally we obtain the desired result that under $P_\theta(\cdot \mid Y = y)$

$$I_n^{-1/2}(\theta) \frac{\partial}{\partial\theta} \log p_n(X(n); \theta) \to_d N(0, y^2).$$

It is fortunate in the autoregressive model we have considered that the conditional process has such a simple structure. Other examples for which this occurs may be found in Basawa (1981a, b). It should be noted however that explicit knowledge of the conditional distributions is not essential for the derivation of conditional optimality results using Theorem 2.1.

Acknowledgments. We are indebted to Richard Davis for valuable suggestions, particularly with regard to Section 4, and to the referees for many helpful comments and suggested improvements.

REFERENCES

ANDERSON, T. W. (1959). On asymptotic distributions of estimates of parameters of stochastic difference equations. *Ann. Math. Statist.* **30** 676–687.

BARNDORFF-NIELSEN, O. (1978). *Information and Exponential Families in Statistical Theory.* Wiley, New York.

BASAWA, I. V. (1981a). Efficient conditional tests for mixture experiments with applications to the birth and branching processes. *Biometrika* **68** 153–165.

BASAWA, I. V. (1981b). Efficiency of conditional maximum likelihood estimators and confidence limits for mixtures of exponential families. *Biometrika* **68** 515–523.

BASAWA, I. V. and KOUL, H. L. (1979). Asymptotic tests of composite hypotheses for non-ergodic type stochastic processes. *Stochastic. Process. Appl.* **9** 291–305.

BASAWA, I. V. and KOUL, H. L. (1983). Asymptotically minimax tests of composite hypotheses for non-ergodic type stochastic processes. *Stochastic Process. Appl.* **14** 41–54.

BASAWA, I. V. and PRAKASA RAO, B. L. S. (1980). Asymptotic inference for stochastic processes. *Stochastic Process. Appl.* **10** 221–254.

BASAWA, I. V. and SCOTT, D. J. (1977). Efficient tests for stochastic processes. *Sankhyā Ser. A* **39** 21–31.

BASAWA, I. V. and SCOTT, D. J. (1983). *Asymptotic optimal inference for nonergodic models. Springer Lecture Notes in Statist.* **17** Springer, New York.

BILLINGSLEY, PATRICK (1971). Weak convergence of measures: applications in probability. Regional Conference Series in Applied Mathematics, SIAM, Philadelphia.

COX, D. R. and HINKLEY, D. V. (1974). *Theoretical Statistics.* Chapman, London.

CRAMER, H. and LEADBETTER, M. R. (1967). *Stationary and Related Stochastic Processes.* Wiley, New York.

FEIGIN, P. D. (1976). Maximum likelihood estimation for continuous time stochastic processes. *Adv. Appl. Probab.* **8** 712–736.

FEIGIN, P. D. (1981). Conditional exponential families and a representation theorem for asymptotic inference. *Ann. Statist.* **9** 597–603.

HEYDE, C. C. and FEIGIN, P. D. (1975). On efficiency and exponential families in stochastic process estimation. *Statistical Distributions in Scientific Work* **1** 227–240. (ed. G. P. Patil, S. Kotz, J. K. Ord). Reidel, Bordrecht.

HEYDE, C. C. (1978). On an optimal property of maximum likelihood estimator of a parameter from a stochastic process. *Stochastic. Process. Appl.* **8** 1–9.

KEIDING, N. (1974). Estimation in the birth process. *Biometrika* **61** 71–80 and 647.

LAURITZEN, S. L. (1982). *Statistical Models as Extremal Families.* Aalborg University Press, Aalborg.

LE CAM, L. (1960). Locally asymptotically normal families of distributions. *Univ. Calif. Pub. Probab. Statist.* **3** 27–98.

MATHEMATICAL STATISTICS DEPARTMENT
LA TROBE UNIVERSITY
BUNDOORA, VICTORIA
AUSTRALIA, 3803

STATISTICS DEPARTMENT
COLORADO STATE UNIVERSITY
FORT COLLINS, COLORADO 80521