ASYMPTOTIC INFERENCE WHEN THE AMOUNT OF INFORMATION IS RANDOM

ROBERT B. DAVIES Department of Scientific and Industrial Research

ABSTRACT: Some results of asymptotic statistical decision theory are extended to allow for the situation in which the Fisher information should be treated as random. They are applied to parameter estimation and hypothesis testing for the supercritical Galton-Watson process and to sequential analysis.

Key Words and Phrases: asymptotic inference, conditional inference, contiguity, curved exponential family, Galton-Watson process, information matrix, sequential estimation, stochastic process estimation.

AMS 1980 Subject Classifications: primary 62F12; secondary 62F05 62L12, 62M05, 62M09.

Part of this project was carried out using the facilities of the Statistical Laboratory, University of California, Berkeley, with the partial support of the National Institutes of Health Grant USPHS-ES01299-17 and Office of Naval Research Contract N000-14-75-C-0159.

From *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, Volume II, Lucien M. Le Cam and Richard A. Olshen, eds.,

1. INTRODUCTION

This work was suggested by a series of papers by Heyde and others on parameter estimation and hypothesis testing for certain "nonergodic" or "evolving" stochastic processes, especially the supercritical Galton-Watson process. See, for example, Heyde (1975, 1978), Basawa, Feigin, and Heyde (1976), Basawa and Scott (1976), Basawa (1977), and Dion and Keiding (1978). An extensive list of references is given by Basawa and Scott (1983). For these processes it was shown that the maximum likelihood estimator has asymptotic properties that must be regarded as extensions of the usual ones. In particular, the Fisher information matrix must be replaced by a random matrix. However, questions about optimality are difficult to answer, particularly when one needs to consider limits as n tends to infinity.

Rather than proving asymptotic properties directly, we show that the real problem can be approximated by what we refer to as the *limit experiment*. Questions of statistical inference and optimality can then be discussed for the limit experiment independently from any asymptotic arguments.

The appropriate limit experiment is introduced in Section 2, and in Section 3 the asymptotic approximation argument is given. The approach and technology is that of Le Cam (1969), pp. 57-87. Our work closely parallels work by Jeganathan (1980, 1981, 1982) and Basawa and Scott (1983). However, we try to be rather more explicit in our separation of optimality and asymptotic arguments.

The application to the supercritical Galton-Watson process is considered in Section 5. It turns out that our framework is also appropriate for some sequential analysis problems, and this is considered in Section 6. Rather than proving directly that the conditions given in Section 3 are satisfied for these two situations, we find it convenient to imbed them in an i.i.d. situation and deduce the necessary results from the corresponding results for this i.i.d. situation. The theory for doing this is given in Section 4.

Finally, in Section 7 we briefly look at optimal inference methods in the limit experiment and the transferring of these to the real problem.

We use $\| \cdot \|$ to denote the L_2 norm of a vector or the norm $\sup\{f(x)dM(x) :\ f| < 1\}$ of a signed measure M. Transposes of vectors are indicated by $'$; $1(A)$ denotes the indicator function of a set A; pr means probability; $\mathscr{L}(X)$ denotes the distribution of a random variable X; $\mathscr{N}(\mu, \Sigma)$ denotes the normal distribution with mean μ and variance/covariance matrix Σ. A sequence of random variables X_n in R^r, or their distributions, will be said to be relatively compact if $\lim_{c \to \infty} \limsup_{n \to \infty} pr(\|X_n\| > c) = 0$.

We shall have to use the following lemma several times; so it is convenient to state it here.

Lemma 1.1: Let P_0 and P_1 denote two probability measures on a measure space $(\mathscr{X}, \mathscr{A})$ and Q_0 and Q_1 denote the probability measures they induce on a sigma field $\mathscr{B} \subset \mathscr{A}$. Suppose ϕ is a bounded $\mathscr{A}$-measurable function such that $\phi = 0$ when P_1 is not absolutely continuous with respect to P_0. Then

$$\frac{dQ_1}{dQ_0} = E_1^{\mathscr{B}}(1 - \phi)\frac{dQ_1}{dQ_0} + E_0^{\mathscr{B}}\left(\phi \frac{dP_1}{dP_0}\right),$$

where $E_0^{\mathscr{B}}$ denotes conditional expectation under P_0 given $\mathscr{B}$. In particular, if $P_1 \ll P_0$, then

$$\frac{dQ_1}{dQ_0} = E_0^{\mathscr{B}}\left(\frac{dP_1}{dP_0}\right)$$

2. THE LIMIT EXPERIMENT

The *limit experiment*, that is, the hypothetical experiment used to approximate an actual experiment, is defined as follows. We observe Δ, an r-dimensional random vector, and τ, an $r \times r$ random positive definite matrix, where Δ and τ have joint density with respect to some measure of the form

$$p(\Delta, \tau; t) = \exp\{t'\Delta - t'\tau t/2\}p(\Delta, \tau; 0), \tag{2.1}$$

where t is an r-dimensional vector of unknown parameters. Note that if X represents the outcome of an experiment and has density of the form

$$p(X;\ t) = \exp\{t'\Delta(X) - t'\tau(X)t/2\}p(X;\ 0) \tag{2.2}$$

with respect to some measure, where $\Delta(X)$ and $\tau(X)$ are a vector and a positive definite matrix function of X, then $\Delta(X)$ and $\tau(X)$ are sufficient for t and have joint density (2.1). Thus (2.2) can be regarded as an equivalent form of the limit experiment.

Examples

(i) Suppose X represents a Wiener process with drift t observed over a time $[0,\ \tau]$, where τ is a stopping time (possibly randomized but almost surely finite for all t) and $\Delta = X(\tau)$ represents the value of the process at the stopping time. Then, according to Freedman (1971, p. 98), X has a density of the form (2.2) with dimension $r = 1$. However, not all one-dimensional models of the form (2.1) can be represented in this way.

(ii) Observe a random positive definite $r \times r$ matrix whose distribution is independent of t and then a random r-dimensional vector Δ such that

$$\mathscr{L}(\Delta|\tau;\ t) = \mathscr{N}(\tau t,\ \tau). \tag{2.3}$$

Then Δ and τ have a distribution of the form (2.1). In fact, this simple version of the model turns out to be the one that is usually appropriate for estimation problems.

We examine some simple properties of (2.1). Let q_t denote the marginal probability measure of τ. Using Lemma 1.1, we have

$$\frac{dq_t}{dq_0} = E_0\left\{\frac{p(\Delta,\ \tau;\ t)}{p(\Delta,\ \tau;\ 0)}\middle|\tau\right\} = e^{-t'\tau t/2}E_0(e^{t'\Delta}|\tau), \tag{2.4}$$

where E_0 denotes expectation under $p(\Delta,\ \tau;\ 0)$. Using moment generating functions, we can then show the following proposition.

Proposition 2.1: If Δ and τ have joint density given by (2.1), then the following three conditions are equivalent:

(i) The marginal distribution of τ is independent of t.

(ii) $\mathscr{L}(\Delta|\tau;\ 0) = \mathscr{N}(0,\ \tau)$.

(iii) $\mathscr{L}(\Delta|\tau;\ t) = \mathscr{N}(\tau t,\ \tau)$ for all t.

In the special case (2.3), τ is an ancillary statistic, and the obvious estimator of t is

$$T = \tau^{-1}\Delta. \quad (2.5)$$

This estimator is also the maximum likelihood estimator of t in the general case and approximately Bayes if t is assigned a diffuse normal prior. However, the family of densities described by (2.1) is an example of a curved exponential family as defined by Efron (1975), and so entirely satisfactory hypothesis tests will not be available. Questions of optimal inference are discussed further in Section 7.

3. THE ASYMPTOTIC THEORY

This section contains the main approximation theorem. Parts of this section represent an extension of results given by Le Cam (1969, pp. 57-85), and are closely related to results given by Jeganathan (1980, 1982) and Basawa and Scott (1983). Other very relevant work appears in Le Cam (1974, particularly pp. 195-224).

In the tradition of statistical asymptotic theory, we imagine that we have a sequence of experiments. Let X_n denote the totality of observations from the n^{th} experiment, $\mathscr{X}_n$ is the set of all possible values of X_n, $\mathscr{A}_n$ is a σ-algebra defined on $\mathscr{X}_n$, and $P_n(\theta)$ is a family of probability measures defined on $\mathscr{A}_n$ which, for a given value of θ, describes the distribution of X_n. Let Θ denote the set of possible values of θ, the unknown parameter about which we want to make inferences. We suppose Θ is a subset of r-dimensional Euclidean space R^r. We also suppose that for each θ there is a sequence of positive numbers $\delta_n(\theta)$ tending to zero as n tends to infinity. These are intended to show the general size of the error made by a good estimator of θ. Thus in the regular i.i.d. situation, one could take $\delta_n(\theta) = n^{-1/2}$.

We want to prove that for each $\theta \in \Theta$ there is a family of distributions of the form (2.2) with the following properties.

B1. If $S_n(X_n)$ is a sequence of random variables in R^k, relatively compact under $P_n(\theta)$, then there is a random variable $S(\Delta, \tau)$ with Δ, τ distributed as in (2.2) and a subsequence n' such that for each K

$$\mathscr{L}\{S_{n'}(X_{n'});\ P_{n'}(\theta + \delta_{n'}(\theta)t\} \to \mathscr{L}_t\{S(\Delta, \tau)\} \tag{3.1}$$

uniformly for $\|t\| < K$.

B2. If $S(\Delta, \tau)$ is a function of Δ, τ that is continuous almost everywhere under (2.1) with $t = 0$, then the sequence

$$S_n = S\{\tau_n(T_n - \theta)/\delta_n(\tau),\ \tau_n\} \tag{3.2}$$

satisfies (3.1). T_n and τ_n are functions of X_n to be defined later in this section.

These results make it easy to relate asymptotic properties of tests and estimators to corresponding (nonasymptotic) properties of tests and estimators for the limit experiment described in Section 2. This is considered briefly in Section 7.

Now consider the conditions under which B1 and B2 hold. Define the log-likelihood ratio (taking values in $[-\infty, \infty]$),

$$\Lambda_n(\theta_1, \theta_2) - \log\{dP_n(\theta_1)/dP_n(\theta_2)\}. \tag{3.3}$$

Then the conditions we require are as follows:

A0. Θ is an open set in R^r.

A1. For each $\theta \in \Theta$, $\{\delta_n(\theta)\}_{n=1,2,\dots}$ is an infinite sequence of real numbers such that

(i) $\delta_n(\theta) \downarrow 0$ as $n \to \infty$;

(ii) for each $K < 0$, $\delta_n\{\theta + \delta_n(\theta)t\}/\delta_n(\theta) \to 1$ uniformly in t for t an r-dimensional vector with $\|t\| < K$.

A2. For every vector t and $\theta \in \Theta$, the two sequences of probability measures $\{P_n(\theta + \delta_n(\theta)t)\}$ and $\{P_n(\theta)\}$ are contiguous [see Le Cam (1960, p. 40)].

A3. There exist sequences of random vectors $\{\Delta_n(\theta)\}$ and symmetric random matrices $\{\tau_n(\theta)\}$ such that for each r-vector t and $\theta \in \Theta$,

$$\Lambda_n\{\theta + \delta_n(\theta)t,\ \theta\} = t'\Delta_n(\theta) - t'\tau_n(\theta)t/2 + \varepsilon_{n,\theta,t},$$

where $\varepsilon_{n,\theta,t} \to 0$ in $P_n(\theta)$ probability.

A4. If t_n is a sequence of r-vectors with $t_n \to t$ and $\theta \in \Theta$, then

$$\Lambda_n\{\theta + \delta_n(\theta)t_n,\ \theta\} - \Lambda_n\{\theta + \delta_n(\theta)t,\ \theta\} \to 0$$

in $P_n(\theta)$ probability.

A5. *Either* A5.1: For each $\theta \in \Theta$,

$$\mathscr{L}\{\Delta_n(\theta), \tau_n(\theta)\} \to \mathscr{L}_\theta\{\Delta(\theta), \tau(\theta)\}$$

under $P_n(\theta)$, where $\Delta(\theta)$ is a random vector, $\tau(\theta)$ is an almost surely positive definite random matrix.

Or A5.2: For each $\theta \in \Theta$ and each r-vector t,

$$\mathscr{L}\{\tau_n(\theta)\} \to \mathscr{L}_{\theta, t}\{\tau(\theta)\}$$

under $P_n(\theta + \delta_n(\theta)t)$, where $\tau(\theta)$ is a random matrix, almost surely positive definite under $\mathscr{L}_{\theta, 0}$.

A6. For each $\theta \in \Theta$ and each $K > 0$,

$$\tau_n\{\theta + \delta_n(\theta)t\} - \tau_n(\theta) \to 0$$

in $P_n(\theta)$ probability, uniformly in t, an r-dimensional vector with $\|t\| < K$.

A7. For each n there exists an estimator $\tilde{\theta}_n$ (i.e., $\tilde{\theta}_n$ is a function of X_n) such that for each $\theta \in \Theta$ the sequence $\{\|\tilde{\theta}_n - \theta\|/\delta_n(\theta)\}$ is relatively compact under $P_n(\theta)$.

This completes the statements of the conditions. One can also consider the conditions applied for a single value of θ, in which case A1(ii) and A7 can be omitted. Our conditions represent an extension of the conditions of Le Cam (1969, pp. 61, 79). In contrast to Le Cam's condition 2, we have included a random quadratic term in A3 in place of a fixed term of unspecified form. Le Cam shows that his fixed term must, in fact, be a quadratic term, and Jeganathan (1980) shows that, in our case, the random quadratic term is the only form possible.

Our first theorem follows from Theorem 2.1(6) of Le Cam (1960, p. 40).

Theorem 3.1: Under A0–A4 and A5.1,

$$\mathscr{L}\{\Delta_n(\theta), \tau_n(\theta); P_n(\theta + \delta_n(\theta)t)\} \to \mathscr{L}_{\theta, t}(\Delta, \tau)$$

where Δ and τ have a joint density of the form

$$p_\theta(\Delta, \tau; t) = \exp(t'\Delta - t'\tau t/2)p_\theta(\Delta, \tau; 0). \tag{3.4}$$

Two corollaries follow from formula (2.4) and Proposition 2.1.

Corollary 3.2: Under A0–A4, the conditions A5.1 and A5.2 are equivalent.

Corollary 3.3: Under A0-A5 with the notation of Theorem 3.1, the following three conditions are equivalent:

(i) $\mathscr{L}_{\theta,0}(\Delta|\tau) = \mathscr{N}(0, \tau)$.
(ii) $\mathscr{L}_{\theta,t}(\Delta|\tau) = \mathscr{N}(\tau t, \tau)$ for all t.
(iii) $\mathscr{L}_{\theta,t}(\tau)$ is independent of t.

Note that if $\mathscr{L}\{\tau_n(\theta); P_n(\theta)\}$ converges to its limiting distribution uniformly in θ, and if this limiting distribution is a continuous function of θ, then the third condition of Corollary 3.3 must hold. So for most estimation problems where the conditions hold for a range θ, the appropriate limit experiment is the mixed normal case (2.3).

We now state the main approximation theorem, which is a direct adaptation of a corresponding theorem of Le Cam.

Theorem 3.4: Suppose A0-A5 are satisfied. Then for each θ there exists a family of probability measures $\{Q_{n,\theta,t}: t \in R^r\}$ of the form

$$dQ_{n,\theta,t} = \exp\{t'\Delta_n(\theta) - t'\tau_n(\theta)t/2\}c_n(\theta, t)dQ_{n,\theta,0}, \tag{3.5}$$

where $c_n(\theta, t)$ is nonrandom such that for each K

$$\sup_{\|t\| < K} \|Q_{n,\theta,t} - P_n(\theta + \delta_n(\theta)t)\| \to 0$$

as $n \to \infty$, and

$$\sup_{\|t\| < K} |c_n(\theta, t) - 1| \to 0$$

as $n \to \infty$.

Proofs of this theorem have been given by Jeganathan (1980, 1982) and Basawa and Scott (1983). Nevertheless, this theorem is central to our discussion; so a proof is given in the appendix. The family $Q_{n,\theta,t}$ is not exactly of the form (2.2) owing to the presence of $c_n(\theta, t)$. It would be possible to get rid of $c_n(\theta, t)$ by applying small corrections to $\Delta_n(\theta)$ and $\tau_n(\theta)$. However, (3.5) is all we need to show $\{\Delta_n(\theta), \tau_n(\theta)\}$ to be sufficient for t.

This establishes that $\{\Delta_n(\theta_0), \tau_n(\theta_0)$ is asymptotically sufficient for θ in neighborhoods of θ_0 of radius of order $\delta_n(\theta_0)$. To get a statistic that does not depend on θ_0, we need to replace θ_0 by an estimate. First, we need a lemma that is proved in the same way as the corresponding lemma given by Le Cam (1960, p. 55).

Lemma 3.5: Suppose conditions A0-A6 hold. Then $\Delta_n(\theta)$ can be chosen so that for all $\theta \in \Theta$ and bounded sequences t_n,

$$\Delta_n\{\theta + \delta_n(\theta)t_n\} - \{\Delta_n(\theta) - \tau_n(\theta)t_n\} \to 0 \tag{3.6}$$

in $P_n(\theta)$ probability.

To avoid extra continuity conditions of $\Delta_n(\theta)$ when we replace t_n with $(\tilde{\theta}_n - \theta)/\delta_n(\theta)$, we follow the approach of Le Cam and modify $\tilde{\theta}_n$ so it can take on only a bounded number of values. This is slightly more complicated to do when δ_n depends on θ, but can be done by letting $\hat{\theta}_n$ be the integer multiple of

$$2^{\mathrm{Int}[\log_2\{\delta_n(\tilde{\theta}_n)\}]}$$

nearest to $\tilde{\theta}_n$, where $\mathrm{Int}(x)$ denotes the integer part of x. If $\tilde{\theta}_n$ satisfies A7, so does $\hat{\theta}_n$; but for each θ, with a high probability, $\hat{\theta}_n$ has only a bounded number of possible values. With Δ_n satisfying (3.6) one can show that

$$\{\Delta_n(\hat{\theta}_n) + \tau_n(\hat{\theta}_n - \theta)/\delta_n(\theta)\} - \Delta_n(\theta) \to 0$$

and also that

$$\tau_n(\hat{\theta}_n) - \tau_n(\theta) \to 0,$$

both in $P_n(\theta)$ probability. Let

$$\tau_n = \tau_n(\hat{\theta}_n); \tag{3.7a}$$

$$T_n = \hat{\theta}_n + \delta_n(\hat{\theta}_n)\tau_n^{-1}\Delta_n(\hat{\theta}_n). \tag{3.7b}$$

Then $\tau_n(T_n - \theta)/\delta_n(\theta)$ and τ_n are candidates for $\Delta_n(\theta)$ and $\tau_n(0)$ in A3, A5, and A6. Thus we can state Theorem 3.6.

Theorem 3.6: Suppose A0-A7 are satisfied, $\Delta_n(\theta)$ satisfies (3.6), and T_n and τ_n are as defined in (3.7). Then for any θ there is a family of probability measures $Q_{n,\theta,t}$ such that for any K

$$\lim_{n\to\infty} \sup_{\|t\| < K} \|Q_{n,\theta,t} - P_n(\theta + \delta_n(\theta)t)\| = 0, \tag{3.8}$$

and in which (T_n, τ_n) is sufficient for t.

Moreover,

$$\mathscr{L}\{(T_n - \theta)/\delta_n(\theta),\ \tau_n;\ Q_{n,\theta,t}\} \to \mathscr{L}_{\theta,t}(T,\ \tau), \tag{3.9}$$

where Δ and τ are distributed as in (3.4) and

$$T = \tau^{-1}\Delta. \tag{3.10}$$

In fact, for a given value of θ it is possible to perturb T_n and τ_n slightly, call the perturbed versions $\tilde{T}_n$ and $\tilde{\tau}_n$, so that (3.8) still holds and (3.9) can be replaced by convergence of probability measures in norm rather than in law. For further details see Le Cam (1969, p. 86).

Now putting these results together: Suppose $S_n(X_n)$ is the sequence of random variables defined in B1. For a given value of θ, we can find a family of probability measures $Q_{n,\theta,t}$ that approximates $P_n(\theta + \delta_n(\theta),\ t)$ and with $(\tilde{T}_n,\ \tilde{\tau}_n)$ sufficient for t. Thus under $Q_{n,\theta,t}$ we can replace $S_n(X_n)$ by $\tilde{S}_n(\tilde{T}_n,\ \tilde{\tau}_n)$, where $\tilde{S}_n$ can involve an additional randomizing variable. Further, the distribution of $\{(\tilde{T}_n - \theta)/\delta_n(\theta),\ \tilde{\tau}_n\}$ tends, in norm, to that of $(T,\ \tau)$ with T as in (3.10) and Δ and τ distributed as in (3.4); so the distribution of $\tilde{S}_n(\tilde{T}_n,\ \tilde{\tau}_n)$ and hence that of $S_n(X_n)$ tends to that of $\tilde{S}_n(\theta + T\delta_n(\theta),\ \tau)$ as n tends to infinity. That is,

$$pr\{S_n(X_n) \in A;\ P_n(\theta + \delta_n(\theta)t)\} - pr\{\tilde{S}_n(\theta + T\delta_n(\theta),\ \tau) \in A;\ t\} \to 0 \tag{3.11}$$

uniformly in A and $\|t\| < K$, say.

To prove statement B1 we need to show that there exists $S(T,\ \tau)$ and a subsequence n' of n such that

$$\mathscr{L}\{\tilde{S}_{n'}(\theta + T\delta_n(\theta),\ \tau);\ t\} \to \mathscr{L}_t\{S(T,\ \tau)\},$$

and this follows from the following lemma which is proved in the appendix.

Lemma 3.7: Let $P_\theta : \theta \in \Theta$ be a family of probability measures on $\mathscr{R}^j$. Suppose for some particular member of Θ, denoted by 0, and for each $\theta \in \Theta$, $P_\theta << P_0$ and also dP_θ/dP_0 is a continuous function on $\mathscr{R}^j$. If $X \in \mathscr{R}^j$ is distributed according to P_θ, and $S_n(X) \in \mathscr{R}^k$ is a sequence of random variables, relatively compact under P_0, then there is a random variable $S(X)$, a function of X and some randomizing variable, and a subsequence n' such that for each θ,

$$\mathscr{L}\{S_{n'}(X), X; P_\theta\} \to \mathscr{L}\{S(X), X; P_\theta\}.$$

Theorem 3.8: If A0-A7 and (2.6) hold, then B1 and B2 follow with T_n and τ_n defined by (3.7). If $(T_n - \theta)/\delta_n(\theta)$ and τ_n are components of S_n, then these components can be identified with T and τ in the limit experiment.

4. EXPERIMENTS USING ONLY PART OF THE DATA

In the next two sections we consider problems that can (almost) be considered as being an experiment that consists of making inferences from a sequence of independent, identically distributed random variables, but in which only part of the data is used. The following theorem enables one to deduce A2-A4 for an experiment, given only part of the data, from the corresponding results for the experiment with the complete data.

Theorem 4.1: Suppose conditions A0-A4 are satisfied for a sequence of experiments $\{(\mathscr{X}_n, \mathscr{A}_n, \tilde{P}_n(\theta) : \theta \in \Theta)\}$ with the $\Delta_n(\theta)$ and $\tau_n(\theta)$ being denoted by $\tilde{\Delta}_n(\theta)$, $\tilde{\tau}_n(\theta)$, and suppose $\tilde{\tau}_n(\theta) = \tilde{\tau}(\theta)$ is nonrandom and does not depend on n (the usual i.i.d. situation). Suppose $\{\mathscr{B}_n\}$ is a sequence of sigma fields with $\mathscr{B}_n \subset \mathscr{A}_n$ and

$$E_\theta^{\mathscr{B}_n}[\exp\{iu'\tilde{\Delta}_n(\theta)\}] - \exp\{iu'\mu_n(\theta) - u'\Sigma_n(\theta)u/2\} \to 0 \qquad (4.1)$$

for all u in $\tilde{P}_n(\theta)$ probability for some $\mathscr{B}$-measurable vector $\mu_n(\theta)$ and $\mathscr{B}_n$-measurable symmetric matrix $\Sigma_n(\theta)$. Then A2, A3, and A4 are satisfied for the experiments $\{\mathscr{X}_n, \mathscr{B}_n, P_n(\theta) : \theta \in \Theta\}$, where $P_n(0)$ is the measure induced on $\mathscr{B}_n$ by $\tilde{P}_n(\theta)$ with

$$\Delta_n(\theta) = \mu_n(\theta), \quad \tau_n(\theta) = \tau(\theta) - \Sigma_n(\theta)$$

Proof: A2 follows immediately. To show A3 and A4, it will be sufficient to show for any bounded sequence of r-dimensional vectors t_n

$$dP_n(\theta + \delta_n(\theta)t_n)/dP_n(\theta) - \exp\{t'\Delta_n(\theta) - t'\tau_n(\theta)t\} \to 0$$

in $P_n(\theta)$ probability, since, by contiguity, the likelihood ratio is bounded away from zero, in probability. Let ϕ_n be an $\mathscr{A}_n$-measurable function, 0 if $\tilde{P}_n(\theta + \delta_n(\theta)t_n)$ is not absolutely continuous with respect to $\tilde{P}_n(\theta)$, otherwise a nonincreasing, nonnegative continuous function ϕ of $d\tilde{P}_n(\theta + \delta_n(\theta)t_n)/d\tilde{P}_n(\theta)$, where $\phi(x) = 1$ if $x < a$, 0 if $x > a + 1$. Then by Lemma 1.1,

$$\frac{dP_n(\theta + \delta_n(\theta)t_n)}{dP_n(\theta)} = E^{\mathscr{B}_n}\{(1 - \phi_n);\ \tilde{P}_n(\theta + \delta_n(\theta)t_n\}\frac{dP_n(\theta + \delta_n(\theta)t_n)}{dP_n(\theta)} + E^{\mathscr{B}_n}\left\{\phi_n \frac{d\tilde{P}_n(\theta + \delta_n(\theta)t_n)}{d\tilde{P}_n(\theta)};\ \tilde{P}_n(\theta)\right\}.$$

The expectation of the first term on the right under $P_n(\theta)$ is less than or equal to $E\{(1 - \phi_n);\ \tilde{P}_n(\theta + \delta_n(\theta)t_n)\}$, which can be made arbitrarily small by choosing a and n large enough. Hence the first term itself can be made arbitrarily small (in probability) by choosing a and n large enough.

Consider the second term: For arbitrary u, letting $\tilde{\Lambda}_n$ denote the log-likelihood ratio, it follows from (4.1) and A3 and A4 for the parent problem that

$$E_\theta^{\mathscr{B}_n}[\exp\{iu\tilde{\Lambda}_n(\theta + \delta_n(\theta)t_n,\ \theta)\}] - E_\theta^{\mathscr{B}_n}[\exp\{iu(t'\mu_n(\theta) + Z_n - t'\tilde{\tau}(\theta)t/2)\}] \to 0, \tag{4.2}$$

where Z_n is normally distributed with zero mean and variance $t'\Sigma_n(\theta)t$. Take expectations. Since $\tilde{\Lambda}_n(\theta + \delta_n(\theta)t_n,\ \theta)$ converges in law along suitable subsequences, so do the expected values of the terms in (4.2), and so $t'\mu_n(\theta) - Z_n - t'\tilde{\tau}(\theta)t/2$ is relatively compact. Also, $\phi(e^x)e^x$ can be arbitrarily accurately approximated over an arbitrary range by a finite number of terms of the form e^{iux}. Hence, from (4.2),

$$E_\theta^{\mathscr{B}_n}\{\phi\ d\tilde{P}_n(\theta + \delta_n(\theta)t_n)/d\tilde{P}_n(\theta)\} - E_\theta^{\mathscr{B}_n}\left\{\phi(e^{t'\mu_n(\theta) + Z_n - t'\tilde{\tau}(\theta)t/2})e^{t'\mu_n(\theta) + Z_n - t'\tilde{\tau}(\theta)t/2}\right\} \to 0.$$

The second term approaches

$$e^{t'\mu_n(\theta) + t'\Sigma_n(\theta)t/2 - t'\tilde{\tau}(\theta)t/2}$$

as $a \to \infty$, and this is what we needed to prove.

In effect (4.1) is requiring

$$\mathscr{L}(\tilde{\Delta}_n(\theta) \mid \mathscr{B}_n) \sim \mathscr{N}(\mu_n(\theta), \Sigma_n(\theta)).$$

Theorem 4.1 can also be useful for showing that the standard Le Cam conditions hold, that is, where $\Sigma_n(\theta)$ is nonrandom [see Davies (1984)].

5. APPLICATION TO THE SUPERCRITICAL GALTON-WATSON PROCESS

We show that under suitable conditions a supercritical Galton-Watson branching process satisfies A0-A7. Suppose the n^{th} experiment consists of the observation of the first n generation sizes of the process

$$X_n = \{x_1, x_2, \ldots, x_n\}.$$

Let the offspring distribution be given by

$$pr(x_1 = x; \theta) = p(x; \theta),$$

where $\theta \in \Theta = (\theta_1, \theta_2)$ is a single unknown real parameter. For simplicity we suppose $p(0; \theta) = 0$ so that extinction of the process is impossible. Also, suppose that θ denotes the expected offspring number

$$E_\theta(x_1) = \Sigma x p(x; \theta) = \theta, \tag{5.1}$$

and that the variance of the offspring number is finite and a continuous function of θ,

$$\text{var}_\theta(x_1) = \delta^2(\theta) < \infty.$$

We require $p(x; \theta)$ to be sufficiently regular for the discrete analogue of Hájek's (1972, p. 189) L.A.N. conditions to be satisfied. Then our conditions A0-A6 are satisfied for the situation in which the n^{th} experiment consists of observing a random sample $(y_1, y_2, \ldots, y_n)$ from $p(\cdot; \theta)$ with δ_n, Δ_n, and τ_n given by

$$n^{-1/2}, \quad n^{-1/2} \sum_1^n \frac{d}{d\theta} \log p(y_i; \theta), \quad E_\theta\left\{\frac{d}{d\theta} \log p(y_i; \theta)\right\}^2, \tag{5.2}$$

respectively. Now define

$$\delta_n(\theta) = \theta^{-n/2} \tag{5.3a}$$

$$\Delta_n(\theta) = \theta^{-n/2} \sum_{k=0}^{n-1} (x_{k+1} - \theta x_k)/\sigma^2(\theta), \tag{5.3b}$$

$$\tau_n(\theta) = \theta^{-n} \sum_{k=0}^{n-1} x_k/\sigma^2(\theta), \tag{5.3c}$$

where $x_0 = 1$.

Theorem 5.1: Under the conditions and definitions given above, the family of probability measures $\{P_n(\theta)\}$ describing the distributions of the X_n satisfies A0-A7. In addition, the special condition of Corollary 3.3 is satisfied, as is condition (3.6).

Proof: A. R. Swensen (1980) proves this result directly. Our proof is based on the results of Section 4. A0, A1, A6, and (3.6) follow directly. Now

$$E_\theta \sum_0^{n-1} x_k = (\theta^n - 1)/(\theta - 1);$$

so for a given value of θ and arbitrary $\eta > 0$, we can choose c so that

$$P_\theta\left(\sum_0^{n-1} x_k > c^2\theta^n\right) < \eta.$$

Let $m = m(n)$ be the first integer above $c^2\theta^n$. Then we could generate the branching process X_n from a random sample $Y_m = \{y_1, \ldots, y_m\}$ from $p(\cdot;\ \theta)$ with a probability of less than η of "running out" of y_i. Since η is arbitrary, it is sufficient to prove A2-A4 for the part of the sample space where we do not "run out" of y_i. We define our n^{th} "parent" experiment to be the observation of Y_m and let

$$\tilde{\Lambda}_n(\theta) = \{m(n)\}^{-1/2} \sum_0^{m(n)} \frac{d}{d\theta} \log p(y_i;\ \theta)/c, \tag{5.4a}$$

$$\tilde{\tau}(\theta) = E_\theta\left\{\frac{d}{d\theta} \log p(y_i;\ \theta)\right\}^2/c^2. \tag{5.4b}$$

Then A2-A4 are satisfied for this sequence of parent experiments with δ_n, Δ_n, and τ given by (5.3a), (5.4a), and (5.4b). We now need to show that (4.1) holds. Let

$$\phi_{i,j} = \frac{d}{d\theta} \log p(y_{i,j}),$$

where $y_{i,j}$ is the number of offspring in the j^{th} family of the i^{th} generation, $i = 0, \ldots, n-1$; $j = 1, \ldots, x_i$. Let

$$\phi_j;\ j = 1, \ldots, m - \sum_0^{n-1} x_i$$

similarly be related to the "unused" y_i. Note that $E(\phi_{i,j}) = 0$, $\text{var}(\phi_{i,j}) = \tilde{\tau}(\theta)$, and $\text{cov}(y_{i,j}, \phi_{i,j}) = 1$. We need to find the limiting conditional distribution of $\tilde{\Delta}_n(\theta)$ given X_n. To do this, we use a conditional central limit theorem which can be obtained as a generalization of the local limit theorem of Gnedenko and Kolmogorov (1954, pp. 233-235); see also Holst (1979). It follows that

$$E\left\{\exp\ ium^{-1/2} \sum_{j=1}^{x_{n-k}} \phi_{n-k,j} \,\middle|\, X_n\right\} - E\{\exp(iuZ_{n,k})\} \to 0$$

in $P_n(\theta)$ probability, where $Z_{n,k}$ denotes a normal random variable with expectation $m^{-1/2}(x_{n-k+1} - \theta x_{n-k})/\sigma^2(\theta)$ and variance $m^{-1}x_{n-k}(\tilde{\tau}(\theta) - 1/\sigma^2(\theta))$. Hence

$$E\left\{\exp\ ium^{-1/2} \sum_{i=0}^{n-1} \sum_{j=1}^{x_i} \phi_{i,j} + \sum_k \phi_k \,\middle|\, X_n\right\} - E\{\exp(iuZ_n)\} \to 0 \tag{5.4}$$

in $P_n(\theta)$ probability, where Z_n denotes a normal random variable with expectation

$$m^{-1/2} \sum_0^{n-1} (x_{i+1} - \theta x_i)/\sigma^2(\theta),$$

and variance

$$\tilde{\tau}(\theta) - m^{-1} \sum_{i=0}^{n-1} x_i/\sigma^2(\theta),$$

since the sum over i in (5.4) is dominated by a finite number of terms. The conditions of Theorem 4.1 and hence A2-A4 follow. To prove A5.2, note that

$$\theta^{-n} \sum_{k=0}^{n-1} X_k \to W$$

in $P_n(\theta)$ probability for some random variable W. Hence

$$\lim_{m\to\infty} \limsup_{n\to\infty} pr\left\{\left|\theta^{-n}\sum_{k=0}^{n-1}X_k - \theta^{-m}\sum_{k=0}^{n-1}X_k\right| > \varepsilon\right\} = 0$$

under $P_n(\theta)$ probability, and hence also under $P_n(\theta + \delta_n(\theta)t)$ probability. Also by continuity, for fixed m,

$$\mathscr{L}\left\{\theta^{-m}\sum_{k=0}^{m-1}X_k;\ P_n(\theta + \delta_n(\theta)t)\right\} \to \mathscr{L}\left\{\theta^{-m}\sum_{k=0}^{m-1}X_k;\ \theta\right\}$$

as $n \to \infty$. Combining these equations, we get

$$\mathscr{L}\{\tau_n(\theta);\ P_n(\theta + \delta_n(\theta)t)\} \to \mathscr{L}\{W/\sigma^2(\theta)\};$$

so A5.2 is satisfied, and also the situation described by Corollary 3.3 applies.

Candidates for $\tilde{\theta}_n$ in A7 include

$$\tilde{\theta}_n = X_n/X_{n-1}$$

and

$$\tilde{\theta}_n = \sum_1^n X_k \Big/ \sum_0^{n-1} X_k.$$

This completes the proof.

Note that if we want to estimate a second parameter, for example, the variance of the family size, the rate of convergence is $O(n^{-1/2})$; so one has either to treat its estimation as a separate problem with another limit experiment or to follow Jeganathan (1980) with matrix valued δ_n.

6. SEQUENTIAL INFERENCE

This section shows that our results are also relevant to sequential estimation and testing. But see also Le Cam (1979) for related results obtained directly. Suppose our n^{th} experiment consists of observing N_n i.i.d. observations

$$X_n = \{x_1, \ldots, x_{N_n}\}$$

from some distribution with density $p(x;\ \theta)$, where N_n is a stopping rule. Let $P_n(\theta)$ denote the probability measure of X_n. Suppose also that the fixed sample size problem with n observations satisfies conditions A0–A7 with

$$\delta_n(\theta) = n^{-1/2}; \quad \tilde{\Delta}_n(\theta) = n^{-1/2} \sum_1^n \phi_i(\theta); \quad \tilde{\tau}(\theta) = \text{var}\{\phi_i(\theta)\}.$$

See Le Cam (1966) and Hájek (1972) for appropriate conditions.

Now suppose that

$$\mathscr{L}\{\delta_n^2(\theta)N_n;\ P_n(\theta + \delta_n(\theta)t)\} \to \mathscr{L}_{\theta,\,t}(N(\theta)),$$

say, where under $\mathscr{L}_{\theta,\,0}$ the random variable $N(\theta)$ is almost surely positive. Then Theorem 6.1 follows.

Theorem 6.1: Under the conditions just stated, A0–A7 are satisfied with

$$\Delta_n(\theta) = \delta_n(\theta) \sum_1^{N_n} \phi_i(\theta),$$

$$\tau_r(\theta) = \delta_n(\theta)^2 N_n \tilde{\tau}(\theta).$$

The proof is based on Theorem 4.1 and will be omitted.

In the sequential estimation situation, one would expect the remarks following Corollary 3.3 to apply and hence the limit experiment to be of a mixed normal form rather than of a sequential form. On the other hand, when one is considering tests such as the sequential likelihood ratio test, the $\delta_n(\theta)$ are likely to vary rapidly with θ so that condition A2(ii) is not satisfied. Nevertheless, this condition is not required for Theorem 3.4, and so this theorem, applied to values of θ under the hypothesis is useful for investigating properties of a test under the hypothesis or contiguous alternatives.

7. INFERENCE FROM THE LIMIT EXPERIMENT

In this section we look at the special case of the limit experiment when $r = 1$ and

$$\mathscr{L}(T|\tau;\ t) = \mathscr{N}(t,\ 1/\tau), \tag{7.1}$$

with $\mathscr{L}(\tau)$ being independent of t, and we show how some of the asymptotic results obtained by other authors can be derived as analogues of optimality results on the limit experiment.

Theorem 7.1: Suppose T and τ are distributed as in (7.1), and S is a function of T and τ and a randomizing variable such that $\mathscr{L}_t(S - t, \tau)$ does not depend on t. Then $S = T + Z$, where $\mathscr{L}(Z)$ does not depend on t, and where Z and T are conditionally independent given τ.

Proof (essentially Bickel's see Roussas, 1972, p. 136): Since $\mathscr{L}_t(S - t|\tau)$ is independent of t, the whole argument is conditional, given τ, and for simplicity we omit the references to the conditioning variable τ.

$$E_t(e^{iu(S-t)}) = E_w(e^{iuS - iut - Tw + w^2/2 + Tt - t^2/2}).$$

This expression is invariant in t, and the right-hand side is analytic in t. Substitute $t = w - iu$ on the right-hand side to get

$$E_0 e^{iuS} = E_w\{e^{iu(S-T)}\}e^{-u^2/2}$$

Thus $\mathscr{L}_t(S - T)$ is independent of t, and hence $S - T$ is independent of T; the result follows.

The following corollary gives the corresponding asymptotic result. It is essentially a one-dimensional version of a theorem of Jeganathan (1980). See also Basawa and Scott (1983, p. 46). In effect, it shows that among estimators of θ that satisfy (7.2), T_n has the most concentrated distribution.

Corollary 7.2: Suppose A0-A7 are satisfied, with limit experiment as in (7.1), τ_n is as in (3.7a), and $S_n(X_n)$ is such that for each t and some $\theta \in \Theta$,

$$\mathscr{L}\{(S_n - \theta)/\delta_n(\theta),\ \tau_n;\ P_n(\theta + \delta_n(\theta)t)\} \to \mathscr{L}(S_0 + t,\ \tau_0) \qquad (7.2)$$

for some S_0 and τ_0 whose joint distribution does not depend on t. Then the conditional distribution of S_0 given τ_0 is equal to the convolution of a centered normal random variable with variance $1/\tau_0$ and some other distribution (which may depend on τ_0). If $S_n = T_n$ defined in (3.7b), then (7.2) is satisfied, and the second distribution in the convolution is degenerate.

Proof: It follows from B1 that we can find S, a function of T and τ such that

$$\mathscr{L}\{(S_n - \theta)/\delta_n, \tau_n; P_n(\theta + \delta_n(\theta)t)\} \to \mathscr{L}_t\{S(T, \tau), \tau\}$$

along suitable subsequences with T and τ distributed as in (7.1). In view of (7.2) the conditions of Theorem 7.1 hold, and the convolution result follows. The result when $S_n = T_n$ follows from B2.

We now look at problems of hypothesis testing. The topic is extensively discussed (see Basawa and Scott, 1983), and our purpose is to illustrate the use of the limit experiment for investigating simple hypothesis tests. We want to test the hypothesis $\theta = 0$ against the alternative $\theta > 0$. Three possible tests for the limit experiment have customarily been considered.

The locally optimal test maximizes the derivative of the power function at $t = 0$ and has critical region

$$\{T > c\tau^{-1}\}. \tag{7.3}$$

The beta-optimal test (Davies, 1969) is the test whose power function reaches a preassigned level β as quickly as possible and for the present problem is simplest when $\beta = 1 - \alpha$, α being the significance level. It has critical region

$$\{T > c\}. \tag{7.4}$$

Since τ is an ancillary statistic, we could condition on it and so get critical region

$$\{T > c\tau^{-1/2}\}.$$

The powers of these tests, when τ has an exponential distribution have been graphed by Sweeting (1978, p. 126).

The tests have critical regions of the form $\{T > c\tau^{-\rho}\}$ with $\rho = 1, 0, 1/2$, where c is chosen to give the required significance level. Transferring these back to the sequence of experiments, we obtain critical regions of the form $\{T_n > \delta_n(0)c\tau^{-\rho}\}$. The actual optimality properties are a little complicated because our theory deals only with convergence of probabilities and not with conditional probabilities or derivatives of probabilities.

Consider first the locally optimal test; $\rho = 1$. Let $\beta_n(t)$ denote the power of a proposed test when $\theta = \delta_n(0)t$, and let $\beta_n^*(t)$ denote the corresponding power of the test based on T_n. It follows from B1 that we can find a test for the limit experiment with power function $\beta(t)$ so that $\beta_n(t) \to \beta(t)$ and $\beta_n^*(t) \to \beta^*(t)$. It is necessary to use a "subsequence of subsequence" argument if there is not a unique limit. Suppose $\beta(0) = \beta^*(0) = \alpha$. Then in view of the local optimality of the test based on T_n, for arbitrary $\varepsilon > 0$, there exists $\eta > 0$ such that if $0 < t < \eta$, then

$$\{\beta^*(t) - \beta^*(0)\}/t > \{\beta(t) - \beta(0)\}/t - \varepsilon.$$

Hence

$$\lim_{t\to\infty} \liminf_{n\to\infty} \{\beta_n(t) - \beta_n(0)\}/t$$

is maximized by the test based on T_n with $\rho = 1$.

Similarly, where $\rho = 0$, the test based on T_n minimizes

$$\liminf_{n\to\infty} \{t : \beta_n(t) \geqslant 1 - \alpha\}.$$

When $\rho = 1/2$, the test maximizes

$$\lim_{\varepsilon\to 0} \liminf_{n\to\infty} pr\{\text{reject hypothesis} | x - \varepsilon < \tau_n < x + \varepsilon;\ P_n(\delta_n(0)t)\},$$

for all t and x such that $P_0(x - \varepsilon < \tau < x + \varepsilon) \neq 0$ for all $\varepsilon > 0$, among tests that satisfy

$$\lim_{\varepsilon\to 0} \limsup_{n\to\infty} pr\{\text{reject hypothesis} | x - \varepsilon < \tau_n < x + \varepsilon;\ P_n(0)\} = \alpha$$

for all x as above.

ACKNOWLEDGMENTS

I have discussed this project with many of the authors listed in the references and am very grateful for their help, insight, and encouragement.

Finally, I must acknowledge my debt to Professor Neyman, particularly his introducing me to asymptotic methods, his emphasis on the use of mathematics in statistics, his insistence on developing

optimal methods tailored for the problem in hand, his interest in nonstandard statistical problems, and especially for creating similar interests in his student, Professor Le Cam, on whose work much of the present paper is based.

APPENDIX

Proof of Theorem 3.4: Drop references to θ and omit $\delta_n(\theta)$ so that $P_n\{\theta + \delta_n(\theta)t\}$ is written $P_{n,\ t}$; $Q_{n,\ \theta,\ t}$ is written $Q_{n,\ t}$; and so on. Let ε_n, b_n be sequences of numbers,

$$A_{n,\ t} = \{|\log(dP_{n,\ t}/dP_{n,\ 0}) - t'\Delta_n + \tfrac{1}{2}t'\tau_n t| < \varepsilon_n\},$$

$$B_n = \{\|\Delta_n\| < b_n \text{ and } \tau_n \text{ positive definite}\},$$

and let $Q_{n,\ t}$ be a measure defined by

$$dQ_{n,\ t} = \exp(t'\Delta_n - \tfrac{1}{2}t'\tau_n t)1(B_n)dP_{n,\ 0}.$$

Then

$$A_{n,\ t} = \left\{\exp(-\varepsilon_n) < \frac{dP_{n,\ 0}}{dP_{n,\ t}} \exp(t'\Delta_n - \tfrac{1}{2}t'\tau_n t) < \exp(\varepsilon_n)\right\}.$$

Hence

$$\begin{aligned} dQ_{n,\ t} - dP_{n,\ t} &= \exp(t'\Delta_n - \tfrac{1}{2}t'\tau_n t)1(B_n)dP_{n,\ 0} - dP_{n,\ t} \\ &\leqslant \exp(\varepsilon_n)1(A_{n,\ t} \cap B_n)dP_{n,\ t} - dP_{n,\ t} \\ &\quad + \exp(t'\Delta_n - \tfrac{1}{2}t'\tau_n t)1(A^c_{n,\ t} \cap B_n)dP_{n,\ 0} \\ &\leqslant \{\exp(\varepsilon_n) - 1\}dP_{n,\ t} + \exp(\|t\|b_n)1(A^c_{n,\ t})dP_{n,\ 0}. \end{aligned}$$

Also,

$$\begin{aligned} dQ_{n,\ t} - dP_{n,\ t} &\geqslant \exp(-\varepsilon_n)1(A_{n,\ t} \cap B_n)dP_{n,\ t} - dP_{n,\ t} \\ &\quad + \exp(t'\Delta_n - \tfrac{1}{2}t'\tau_n t)1(A^c_{n,\ t} \cap B_n)dP_{n,\ 0} \\ &\geqslant \{\exp(-\varepsilon_n) - 1\}dP_{n,\ t} - \exp(-\varepsilon_n)\{1(A^c_{n,\ t}) \\ &\qquad + 1(B^c_n)\}dP_{n,\ t}. \end{aligned}$$

Thus

$$\begin{aligned} \|Q_{n,\ t} - P_{n,\ t}\| &\leqslant \{\exp(\varepsilon_n) - 1\} + P_{n,\ t}(A^c_{n,\ t}) + P_{n,\ t}(B^c_n) \\ &\quad + \exp(\|t\|b_n)P_{n,\ 0}(A^c_{n,\ t}). \end{aligned} \tag{a1}$$

We now show that for a given value of K the sequences $\{b_n\}$, $\{\varepsilon_n\}$ can be chosen so that

$$\sup_{\|t\| < K} \|Q_{n,\,t} - P_{n,\,t}\| \to 0. \tag{a2}$$

In view of a4, a5.1 we can extend a2, a3 to allow t to be replaced by a bounded sequence $\{t_n\}$. Then, for fixed $\varepsilon_n > 0$,

$$P_{n,\,t_n}(A^c_{n,\,t_n}) \to 0. \tag{a3}$$

Hence

$$\sup_{\|t\| < K} P_{n,\,t}(A^c_{n,\,t}) \to 0,$$

and so we can find $\varepsilon_n \to 0$ such that (a3) still holds. Hence we can find $b_n \to \infty$ such that the last term in (a1) tends to zero uniformly for $\| t \| < K$. Then

$$\sup_{\|t\| < K} P_{n,\,t}(B^c_n) \to 0.$$

Thus a2 holds. To show that (a2) holds for all K, we can think of ε_n and b_n as being functions of K and then use sequences $\varepsilon_n(K_n)$, $b_n(K_n)$ with K_n tending to infinity sufficiently slowly so as not to disturb the convergence to zero. Since $Q_{n,\,t}$ does not necessarily have total mass equal to 1, it must be rescaled, and the term $c_n(\theta,\, t)$ must be included. This completes the proof.

Proof of Lemma 3.7: We can think of $(S_n(X),\, X)$ being a random variable in R^{j+k}. Let $Q_{n,\,\theta}$ denote the joint probability measure corresponding to P_θ. Then there is a subsequence n' on which $Q_{n',\,0} \to Q_0$ weakly, for some measure Q_0. Now $dQ_{n,\,\theta}/dQ_{n,\,0} = dP_\theta/dP_0$ by Lemma 1.1, and so $dQ_{n',\,\theta} \to dP_\theta/dP_0 \cdot dQ_0 = dQ_\theta$, say. Then $dQ_\theta/dQ_0 = dP_\theta/dP_0$. Let $(S,\, X)$ be distributed according to Q_0. Then X is sufficient for θ (Halmos and Savage, 1949, p. 233), and so S can be considered as a possibly randomized function of X. This completes the proof.

REFERENCES

Basawa, I. V., Feigen, P. D., and Heyde, C. C. (1976). Asymptotic properties of maximum likelihood estimators for stochastic processes. *Sankhya A*, *38*, 259-270.

Basawa, I. V., and Scott, D. J. (1976). Efficient tests for branching processes. *Biometrika 63*, 531-536.

Basawa, I. V. (1977). Asymptotic statistical curvature for dependent data. *Bull. Int. Stat. Inst. 47*(4), 45-48.

Basawa, I. V., and Scott, D. J. (1983). *Asymptotic Optimal Inference for Nonergodic Models*. Lecture Notes in Statistics 17. New York: Springer-Verlag.

Davies, R. B. (1969). Beta-optimal tests and an application to the summary evaluation of experiments. *J. Roy. Statist. Soc. B*, *31*, 524-538.

Davies, R. B. (1984). Asymptotic inference in the frequency domain. In *Handbook of Statistics—3: Optimal Inference in the Frequency Domain Case*. Edited by D. R. Brillinger and P. R. Krishnaiah. Amsterdam: North-Holland, pp. 73-92.

Dion, J., and Keiding, N. (1978). Statistical Inference in Branching Processes. In *Advances in Probability and Related Topics, V: Branching Processes*, pp. 105-140. Edited by A. Joffe and P. Ney. New York: Dekker.

Efron, B. (1975). Defining the curvature of a statistical problem (with discussion). *Ann. Statist. 6*, 1189-1242.

Freedman, D. (1971). *Brownian Motion and Diffusion.* San Francisco: Holden-Day.

Gnedenko, B. V., and Kolmogorov, A. N. (1954). *Limit Theorems for Sums of Independent Random Variables.* Cambridge, Mass.: Addison-Wesley.

Hájek, J. (1972). Local asymptotic minimax and admissibility in estimation. In *Proc. 6th Berkeley Symposium Math. Stat. Prob. 1*, 175-194.

Halmos, P. R., and Savage, L. J. (1949). Application of the Radon Nikodym theorem to the theory of sufficient statistics. *Ann. Math. Statist. 20*, 225-241.

Heyde, C. C. (1975). Remarks on efficiency in estimation for branching processes. *Biometrika 62*, 49-55.

Heyde, C. C. (1978). On an optimal property of the maximum likelihood estimator of a parameter from a stochastic process. *Stochastic Processes Appl. 8*, 1-9.

Holst, L. (1979). Two conditional limit theorems with applications. *Ann. Statist. 7*, 351-357.

Jeganathan, P. (1980). Asymptotic Theory of Estimation when the Limit of the Log-Likelihood Ratios is Mixed Normal. Ph.D. dissertation, I.S.I., Calcutta.

Jeganathan, P. (1981). On a decomposition of the limit distribution of a sequence of estimators. *Sankhya A*, *43*, 26-36.

Jeganathan, P. (1982). On the asymptotic theory of estimation when the limit of the log-likelihood ratio is mixed normal. *Sankhya A*, *44*, 173-212.

Le Cam, L. (1960). Locally asymptotically normal families of distributions. *University of California Publications in Statistics 3*, 37-98.

Le Cam, L. (1966). Likelihood functions for large numbers of independent observations. In *Research Papers in Statistics*, pp. 167-188. Edited by F. N. David. New York: Wiley.

Le Cam, L. (1969). *Théorie Asymptotique de la Décision Statistique.* Montreal: University of Montreal Press.

Le Cam, L. (1974). *Notes on Asymptotic Methods in Statistical Decision Theory.* Montreal: University of Montreal Press.

Le Cam, L. (1979). A reduction theorem for certain sequential experiments, II. *Ann. Statist. 7*, 847-859.

Roussas, G. G. (1972). *Contiguous Probability Measures: Some Applications in Statistics.* Cambridge: Cambridge University Press.

Sweeting, T. J. (1978). On efficient tests for branching processes. *Biometrika 65*, 123-127.

Swensen, A. R. (1980). Asymptotic Inference for a Class of Stochastic Processes. Ph.D. dissertation, University of California, Berkeley.