The Annals of Statistics
1980, Vol. 8, No. 6, 1375–1381

UNIFORM ASYMPTOTIC NORMALITY OF THE MAXIMUM LIKELIHOOD ESTIMATOR

BY T. J. SWEETING

University of Surrey

A very general result concerning the weak consistency and uniform asymptotic normality of the maximum likelihood estimator is presented. The result proves to be of particular value in establishing uniform asymptotic normality of randomly normalized maximum likelihood estimators of parameters in stochastic processes. The only conditions imposed are certain regularity conditions on the (random) information function, easily verified in practice. Application of the result is briefly considered.

1. Introduction. In this paper we present a very general result on the weak consistency and asymptotic normality (a.n.) of maximum likelihood (m.l.) estimators, which proves to be of particular value in inference for stochastic processes. Recently there has been much interest in large sample inference for stochastic processes—in particular, for the "nonergodic" case where the conditional information function does not behave asymptotically like a constant (for example, branching processes, the pure birth process, some diffusion processes). The standard approach is to prove the asymptotic equivalence of the m.l. estimator and the first derivative $U_t(\theta)$ of log likelihood (both suitably normalized) and then to establish a.n. for $U_t(\theta)$. In the independent case this follows from the ordinary central limit theorem when the Lindeberg condition holds, and in the "ergodic" dependent case from a martingale central limit theorem. (See for example [5], [11], [14], [3], [2], and [7].) In certain nonergodic cases it is possible to express $U_t(\theta)$ as the random time change of a process with stationary independent increments and derive a.n. from this fact (cf. Keiding (1975), Feigin (1976)).

It is shown here that under some reasonable stipulations on the (random) information function, a.n. of m.l. estimators results. Moreover, the convergence to normality is shown to be uniform in compact subsets of the parameter space, a statistically essential requirement for constructing approximate confidence regions or assessing the power of tests. The conditions imposed are briefly discussed and some applications given in Section 5.

2. Regularity assumptions and conditions on the information function. Let $\{(\Omega_t, \mathscr{A}_t)$ be a family of measurable spaces, where t is a discrete or continuous parameter, and let P_θ^t be a probability measure defined on $(\Omega_t, \mathscr{A}_t)$ depending on the parameter $\theta \in \Theta$, an open subset of $\mathbb{R}^k$. Assume that, for each t and $\theta \in \Theta$, P_θ^t is absolutely continuous with respect to a σ-finite measure λ_t; let $p_t(\theta)$ be the density of P_θ^t with respect to λ_t. Then the function $l_t(\theta) = \log p_t(\theta)$ exists a.e. (λ_t). We assume that the second-order partial derivatives of $p_t(\theta)$ exist and are continuous a.e. for all $\theta \in \Theta$. Let $U_t(\theta) = l_t'(\theta)$, the vector of first-order derivatives of $l_t(\theta)$, and $l_t''(\theta)$ be the matrix of second-order derivatives. Define the (random) information matrix $\mathscr{I}_t(\theta)$ to be $\mathscr{I}_t(\theta) = -l_t''(\theta)$.

The symbol $\to_u$ will mean uniform convergence in compact subsets of Θ, and $\Rightarrow_u$ will mean uniform weak convergence (see Section 4 for definitions and properties). Let M_k be the space of all $k \times k$ matrices. The norm $|A|$ of the matrix A is $|A| = (\text{tr } A^T A)^{1/2}$; a sequence (A_n) of matrices converges to the limit A iff $|A_n - A| \to 0$. If the matrix A is p.d. we write $A > 0$; if $A > 0$ then $A^{1/2}$ denotes the (symmetric) positive square root of A. The identity matrix in M_k will be I_k.

Received March 1979; revised September 1979.
AMS 1970 *subject classifications.* Primary 62E20; secondary 62M99.
Key words and phrases. Uniform asymptotic normality, maximum likelihood estimation, inference from stochastic processes.

If Γ is the matrix $(\theta_1, \cdots, \theta_k)$ where $\theta_i \in \Theta$, $i = 1, \cdots, k$, define $\mathscr{I}_t(\Gamma)$ to be $\mathscr{I}_t$ with row i evaluated at θ_i. We assume that $\mathscr{I}_t(\theta)$ satisfies the following conditions:

C1 (*Growth and convergence*). There exist nonrandom square matrices $A_t(\theta)$, continuous in θ, satisfying $\{A_t(\theta)\}^{-1} \to_u 0$ such that

$$W_t(\theta) \equiv \{A_t(\theta)\}^{-1}\mathscr{I}_t(\theta)[\{A_t(\theta)\}^{-1}]^T \qquad \Rightarrow_u W(\theta)$$

where $\Pr(W(\theta) > 0) = 1$.

C2 (*Continuity*). For all $c > 0$

(i) $\sup |\{A_t(\theta)\}^{-1}A_t(\theta') - I_k| \quad \to_u 0$

where the sup is over the set $|\{A_t(\theta)\}^T(\theta' - \theta)| \leq c$, and

(ii) $\sup |\{A_t(\theta)\}^{-1}[\mathscr{I}_t(\Gamma) - \mathscr{I}_t(\theta)][\{A_t(\theta)\}^{-1}]^T| \quad \to_u 0$

in probability, where the sup is over the set $|\{A_t(\theta)\}^T(\theta_i - \theta)| \leq c$, $1 \leq i \leq k$.

It is easily seen that C1 and C2 are equivalent to the single condition

$$(1) \qquad \{A_t(\theta')\}^{-1}\mathscr{I}_t(\Gamma)[\{A_t(\theta')\}^{-1}]^T \Rightarrow_u W(\theta)$$

where $|\{A_t(\theta)\}^T(\theta' - \theta)| \leq c$, $\theta_1, \cdots, \theta_k$ are random variables satisfying $|\{A_t(\theta)\}^T(\theta_i - \theta)| \leq c$ and $\Pr(W(\theta) > 0) = 1$. Verification of C1 and C2 is briefly discussed in Section 5. In most applications, $A_t(\theta)\{A_t(\theta)\}^T = E_\theta\mathscr{I}_t(\theta)$; however we do not require here that $E_\theta\mathscr{I}_t(\theta)$ even exist. When it does exist, $A_t(\theta)$ can often be taken as $\{E_\theta\mathscr{I}_t(\theta)\}^{1/2}$, but this is not always so—for example, when information about different parameters tend to infinity at different rates (cf. [8]). If t is the index parameter of a stochastic process one can often prove uniform convergence in probability of $W_t(\theta)$ to some random variable $W(\theta) > 0$, which will then imply C1 (see Lemma 2). However, C1 allows the random variables to be defined on different spaces.

3. Statement of main results. We assume throughout this section that the regularity assumptions and conditions C1, C2 in Section 2 hold. Normalize $U_t(\theta)$ by defining

$$X_t(\theta) = \{A_t(\theta)\}^{-1}U_t(\theta).$$

The main result concerns the asymptotic joint distribution of $X_t(\theta)$ and the "normalized information" $W_t(\theta)$.

THEOREM 1. $(X_t(\theta), W_t(\theta)) \Rightarrow_u (\{W(\theta)\}^{1/2}Z, W(\theta))$ *where* Z *is a standard normal random vector in* $\mathbb{R}^k$, *independent of* $W(\theta)$.

One can deduce uniform convergence of certain probabilities from Theorem 1 since the distribution of $W(\theta)$ is continuous in θ—see Lemmas 1 and 3. To deduce the asymptotic joint distribution of the m.l. estimator $\hat{\theta}_t$ and $W_t(\theta)$, we need a link between $\hat{\theta}_t$ and $X_t(\theta)$; this is given in Theorem 2. Write

$$Y_t(\theta) = \{A_t(\theta)\}^T(\hat{\theta}_t - \theta).$$

THEOREM 2. *There exists a local maximum* $\hat{\theta}_t$ *of* $l_t(\theta)$ *with probability tending to one satisfying*

$$X_t(\theta) - W_t(\theta)Y_t(\theta) \to_u 0$$

in probability.

It follows from Theorems 1 and 2 and the continuous mapping theorem that

$$(\{W_t(\theta)\}^{1/2}Y_t(\theta), W_t(\theta)) \quad \Rightarrow_u (Z, W(\theta)).$$

For statistical application, one can deduce the following result:

COROLLARY 1. *Conditional on* $W_t(\hat{\theta}_t)$, $Y_t(\theta)$ *is asymptotically normally distributed with zero mean and covariance matrix* $\{W_t(\hat{\theta}_t)\}^{-1}$.

It may be shown from the uniform stochastic boundedness (u.s.b.) of $Y_t(\theta)$ (see Lemma 4) and C2(i) that $\{A_t(\theta)\}^{-1}A_t(\hat{\theta}_t) \to_u I_k$ in probability and so $Y_t(\theta)$ in Corollary 1 may be replaced by $\{A_t(\hat{\theta})\}^T(\hat{\theta}_t - \theta)$. Thus confidence regions for θ are based on the approximate normality of $\hat{\theta}_t$, mean θ, covariance matrix $\{\mathscr{I}_t(\hat{\theta}_t)\}^{-1}$

The justification for considering the sampling distribution of $\hat{\theta}_t$ conditional on $W_t(\hat{\theta}_t)$ is provided by the following conditionality argument. From (1) and $Y_t(\theta)$ u.s.b. it follows that $\{\mathscr{I}_t(\hat{\theta}_t)\}^{-1} \to_u 0$ in probability. Since the asymptotic distribution of $W_t(\hat{\theta}_t)$ is continuous in θ (from Lemma 3 and $W_t(\hat{\theta}_t) - W_t(\theta) \to_u 0$), the distribution of $W_t(\hat{\theta}_t)$ as a function of θ is effectively constant over the main range of variation of the distribution of $\hat{\theta}_t$. Thus $W_t(\hat{\theta}_t)$ behaves like an ancillary statistic for θ, which suggests basing inferences about θ on the distribution of $\hat{\theta}_t$ conditional on $W_t(\hat{\theta}_t)$.

4. Proof of results. We first define the notion of uniform weak convergence and give some simple properties. Let g_n, $n \geqslant 1$ be arbitrary real functions and g a real continuous function on a metric space X. Several times we shall make use of the fact that

$$\begin{aligned} &g_n(s) \to g(s) \qquad &&\text{uniformly in } s \\ \text{iff } &g_n(s_n) \to g(s) &&\text{for every sequence } s_n \to s. \end{aligned} \tag{2}$$

Let P_s, $P_{n,s}$, $n \geqslant 1$ be probability measures defined on the Borel subsets of a metric space depending on the arbitrary parameter s, and let C be the space of real bounded uniformly continuous functions. We shall say that $P_{n,s} \Rightarrow P_s$ uniformly in s iff

$$\int u\, dP_{n,s} \to \int u\, dP_s \qquad \text{uniformly in } s, \text{ for all } u \in C. \tag{3}$$

If $s \in X$, another metric space, then the family (P_s) of probability measures is *continuous* if $P_{s_n} \Rightarrow P_s$ whenever $s_n \to s$. The set B will be called a *(P_s)-continuity set* if $P_s(\partial B) = 0$ for all $s \in X$, where ∂B is the boundary of B. The following lemma gives some simple consequences of uniform weak convergence when the limit family (P_s) is continuous.

LEMMA 1. *Suppose that $P_{n,s} \Rightarrow P_s$ uniformly in s and (P_s) is continuous. Then*
(i) *The convergence in* (3) *holds for all bounded continuous u.*
(ii) *$P_{n,s}(B) \to P_s(B)$ uniformly in s for every (P_s)-continuity set B.*

PROOF. Let s_n be an arbitrary sequence converging to s.

(i) From (2), $\int u\, dP_{n,s_n} \to \int u\, dP_s$ for all $u \in C$, and hence for all bounded continuous u (Theorem 2.1 in [4]); (i) follows on applying (2) once more.

(ii) Since $P_{n,s_n} \Rightarrow P_s$ from (2), it follows from Theorem 2.1 in [4] that $P_{n,s_n}(B) \to P_s(B)$; (ii) now follows from (2) since $P_s(B)$ is a continuous function of s.

Let X_s, $X_{n,s}$, $n \geqslant 1$ be random k-vectors on a probability space $(\Omega, \mathscr{T}, P_s)$.

LEMMA 2. *If $X_{n,s} \to_p X_s$ uniformly in s then $X_{n,s} \Rightarrow X_s$ uniformly in s.*

The proof is elementary and therefore omitted.

Assume throughout the remainder of this section that the regularity assumptions and conditions C1, C2 of Section 2 hold.

LEMMA 3. *The distribution G_θ of $W(\theta)$ is continuous in θ.*

PROOF. Let $G_{t,\theta}$ be the distribution of $W_t(\theta)$. Since, from C1, $\int u\, dG_{t,\theta} \to \int u\, dG_\theta$ uniformly in compact subsets of Θ for every $u \in C$, it suffices to show that $G_{t,\theta}$ is continuous in θ for each t. But if $u \in C$, θ, $\theta_m \in \Theta$, $m \geqslant 1$, $\theta_m \to \theta$,

$$\left| \int u\, dG_{t,\theta_m} - \int u\, dG_{t,\theta} \right|$$

$$(3) \qquad = \left| \int u(W_t(\theta_m))\, dP_{t,\theta_m} - \int u(W_t(\theta))\, dP_{t,\theta} \right| \leqslant \|u\| \int |p_t(\theta_m) - p_t(\theta)|\, d\lambda_t + \int |u(W_t(\theta_m)) - u(W_t(\theta))|\, dP_{t,\theta}.$$

Since $p_t(\theta)$ is continuous a.e., the first integral tends to zero as $m \to \infty$ by Scheffé's Theorem. Since $W_t(\theta)$ is continuous a.e. (λ_t) and $u \in C$, the second integral tends to zero by Dominated Convergence.

In the proofs of Theorem 1 and Lemma 4 below, θ_t, $\theta \in \Theta$ and $\theta_t \to \theta$ as $t \to \infty$. Write $W_t = W_t(\theta_t)$, $A_t = A_t(\theta_t)$, $X_t = A_t^{-1} l_t'(\theta_t) = X_t(\theta_t)$. Repeated use is made of (2) in the proofs. Note in particular that from C1 and (2) we have

$$(4) \qquad A_t^{-1} \to 0.$$

PROOF OF THEOREM 1. For brevity we omit the fixed argument θ in $W(\theta)$ and elsewhere in the proof. Let $s \in \mathbb{R}^k$ and define $\psi_t = \theta_t + \{A_t^{-1}\}^T s$; from (4), $\psi_t \in \Theta$ for $t > t_0$ and $\psi_t \to \theta$. Assume $t > t_0$; from the regularity assumptions we can write a.e. (λ_t)

$$l_t(\psi_t) = l_t(\theta_t) + (\psi_t - \theta_t)^T l_t'(\theta_t) + \tfrac{1}{2}(\psi_t - \theta_t)^T l_t''(\phi_t)(\psi_t - \theta_t)$$

where $\phi_t = \alpha_t\theta_t + (1 - \alpha_t)\psi_t$, $0 < \alpha_t < 1$ (α_t is random). Write $V_t = A_t^{-1}\mathscr{I}_t(\phi_t)\{A_t^{-1}\}^T$; taking exponentials and rearranging gives

$$(5) \qquad e^{s^T X_t} p_t(\theta_t) = e^{1/2 s^T V_t s} p_t(\psi_t).$$

It follows from (1), (2) and Lemma 3 that $V_t \Rightarrow W$ under both the (θ_t) and (ψ_t) families of distributions. Let $0 < \epsilon < 1$ and choose K such that $\Pr(|W| \geqslant K) \leqslant \epsilon$ and $\Pr(|W| = K) = 0$ (possible since the distribution of $|W|$ has a denumerable number of atoms). Since $V_t \Rightarrow W$ and $\{|x| < K\}$ is a G_θ-continuity set, it follows that

$$(6) \qquad P_{\theta_t}^t(|V_t| < K) \to \Pr(|W| < K).$$

Let (Q^t) be the family of distributions $(P_{\theta_t}^t)$ conditional on $(|V_t| < K)$; that is, Q^t has density

$$q_t = \begin{cases} p_t(\theta_t)/P_{\theta_t}^t(|V_t| < K), & |V_t| < K \\ 0, & \text{otherwise.} \end{cases}$$

Let u be a bounded function on M_k, continuous on $|A| < K$, with $u(A) = 0$ for $|A| \geqslant K$. Let E_t^* denote expectation under Q^t. Multiplying (5) through by $u(V_t)$ and integrating with respect to λ_t over the set $(|V_t| < K)$ yields

$$\begin{aligned} E_t^*\{u(V_t)e^{s^T X_t}\} &= E\{u(V_t)e^{1/2 s^T V_t s}\}/P_{\theta_t}^t(|V_t| < K) \\ &\to E\{u(W)e^{1/2 s^T W s}\}/\Pr(|W| < K) \\ &= E^*\{u(W)e^{1/2 s^T W s}\} \end{aligned}$$

where E^* denotes expectation conditional on $|W| < K$. This follows from (6) and because the function $u(w)e^{1/2 s^T w s}$ is a bounded G_θ-continuous function ([4], Theorem 5.2). But the right-hand side is equal to $E^*\{u(W)e^{s^T W^{1/2} Z}\}$ where Z is a standard normal random vector in $\mathbb{R}^k$ independent of W. By the uniqueness of bilateral Laplace transforms and the weak compactness theorem, it follows that $(X_t, V_t) \Rightarrow (W^{1/2}Z, W) \mid |W| < K$ with respect to the family (Q^t) of distributions. Since ϵ was arbitrary, it follows easily by standard computations that $(X_t, V_t) \Rightarrow (W^{1/2}Z, W)$ and hence

$$(7) \qquad (X_t, W_t) \Rightarrow (W^{1/2}Z, W),$$

since $V_t - W_t \to_p 0$ with respect to $(P_{\theta_t}^t)$. Finally, since (7) holds for arbitrary $\theta_t \to \theta$ and since the distribution of $(\{W(\theta)\}^{1/2}Z, W(\theta))$ is continuous in θ from Lemma 3, the theorem follows on application of (2).

A family $(T_t(\theta))$ of (possibly extended) $\mathscr{A}_t$-measurable functions is *uniformly stochastically bounded* (u.s.b.) if for each $\epsilon > 0$ and compact set K in Θ there exists c and t_0 such that

$$P_\theta(|T_t(\theta)| > c) < \epsilon$$

for all $t > t_0$ and $\theta \in K$. Let $\hat{\theta}_t$ be any local maximum of $l_t(\theta)$; if none exists put $\hat{\theta}_t = +\infty$.
Recall

$$Y_t(\theta) = \{A_t(\theta)\}^T(\hat{\theta}_t - \theta).$$

LEMMA 4. *There exist local maxima $\hat{\theta}_t$ with $(Y_t(\theta))$ u.s.b.*

PROOF. Write $P = P^t_{\theta_t}$ for brevity. Let $S_t = \{\phi \in \mathbb{R}^k: |A_t^T(\phi - \theta_t)| = c\}$. If $(\phi - \theta_t)^T l_t'(\phi) < 0$ for all $\phi \in S_t$, then there exists a local maximum $\hat{\theta}_t$ of $l_t(\theta)$ satisfying $|Y_t(\theta_t)| \leq c$ (cf. Aitchison and Silvey (1958)). Let $\pi_t = P(\sup_{\phi \in S_t} (\phi - \theta_t)^T l_t'(\phi) \geq 0)$. We shall show that

$$\limsup_{t\to\infty} \pi_t \to 0 \tag{8}$$

as $c \to \infty$, from which the lemma will follow. If $\phi \in S_t$ we have a.e.

$$(\phi - \theta_t)^T l_t'(\phi) = (\phi - \theta_t)^T l_t'(\theta_t) - (\phi - \theta_t)^T \mathscr{I}_t(\Gamma)(\phi - \theta_t)$$

where $\Gamma = (\theta_1, \cdots, \theta_k)$ and $|A_t^T(\theta_i - \theta_t)| \leq c$, $1 \leq i \leq k$. Taking the sup and inf over the set $\{x \in \mathbb{R}^k: |x| = 1\}$ we therefore have

$$\pi_t \leq P(\sup x^T X_t \geq c \inf x^T U_t x)$$

where $U_t = A_t^{-1}\mathscr{I}_t(\Gamma)\{A_t^{-1}\}^T$. But $\inf x^T U_t x = \mu_t$, the smallest eigenvalue of U_t, and from (1) and (2) it follows that $\mu_t \Rightarrow \mu(\theta)$, the smallest eigenvalue of $W(\theta)$. But for all $\epsilon > 0$

$$\pi_t \leq P(|X_t| \geq c\mu_t) \leq P(|X_t| \geq c\epsilon) + P(\mu_t \leq \epsilon).$$

Since from Theorem 1 $X_t \Rightarrow \{W(\theta)\}^{1/2}Z$, and $\mu_t \Rightarrow \mu(\theta)$ it follows from (iii) of Theorem 2.1 in [4] that

$$\limsup_{t\to\infty} \pi_t \leq \Pr(|\{W(\theta)\}^{1/2}Z| \geq c\epsilon) + \Pr(\mu(\theta) \leq \epsilon)$$
$$\to \Pr(\mu(\theta) \leq \epsilon)$$

as $c \to \infty$ and (8) follows since $\Pr(\mu(\theta) \leq \epsilon) \to 0$ as $\epsilon \to 0$.

PROOF OF THEOREM 2. Let F_t be the set $(\hat{\theta}_t < \infty)$. On the set F_t, $U_t(\theta) = \mathscr{I}_t(\Gamma)(\hat{\theta}_t - \theta)$ where $\Gamma = (\theta_1, \cdots, \theta_k)$ and so

$$X_t(\theta) = W_t(\theta, \hat{\theta}_t)Y_t(\theta)$$

where $W_t(\theta, \hat{\theta}_t) = \{A_t(\theta)\}^{-1}\mathscr{I}_t(\Gamma)[\{A_t(\theta)\}^{-1}]^T$. Since $\{A_t(\theta)\}^T(\theta_i - \theta)$, $1 \leq i \leq k$, are u.s.b. it easily follows from C2(ii) that $W_t(\theta, \hat{\theta}_t) - W_t(\theta) \to_u 0$ in probability. Since $Y_t(\theta)$ is u.s.b., the theorem therefore holds conditionally on F_t; the result follows since $P_\theta^t(F_t) \to_u 1$.

PROOF OF COROLLARY 1. This follows since $Y_t(\theta)$ is u.s.b. and $W_t(\hat{\theta}_t) - W_t(\theta) \to_u 0$, as in the proof of Theorem 2.

5. Application of results. In this section we consider briefly the application of the results of this paper. Sufficient conditions are easily constructed for conditions C1 and C2 of Section 3; obvious sufficient conditions for C2 in terms of the derivatives (when they exist) of the elements of $A_t(\theta)$ and $\mathscr{I}_t(\theta)$ may be constructed—we omit further consideration of C2 and concentrate on C1. The approach to verifying C1 largely depends on whether one is in the ergodic ($W(\theta)$ degenerate) or nonergodic case. These two cases are briefly discussed below, along with some simple examples.

In the ergodic case the simplest general sufficient condition arises when $\mathscr{I}_t(\theta)$ is square-integrable, since then C1 holds if $E\{\mathscr{I}_t(\theta)\} \to_u W(\theta)$, $\operatorname{Var}\{\mathscr{I}_t(\theta)\} \to_u 0$. This requirement can of course be weakened; in the case of independent observations, conditions such as those in [5], [11] or [14], strengthened to hold uniformly in compacts of Θ, can be given. Asymptotic normality in the ergodic case is derived in Weiss (1973) under a similar set of conditions to those imposed here. Our result may be considered an improvement of Weiss's result since *uniform* asymptotic normality is obtained. In addition, our conditions are weaker since in [17] it is assumed that there are diagonal matrices Λ_t such that $\Lambda_t^{-1}A_t(\theta) \to_u B(\theta)$. This condition fails to hold for a nonhomogeneous Poisson process with intensity $\theta e^{\theta t}$ for example, which satisfies our conditions with $A_t(\theta) = te^{\theta t/2}$.

The nonergodic case usually arises when the process (X_t) itself is nonergodic and $\mathscr{I}_t(\theta)$ approximates a linear combination of X_t (or integral, when the parameter t is continuous). For then (specializing to $k = 1$), if there are constants $b_t(\theta)$ and random variables $Y(\theta)$ such that $Y_t(\theta) \equiv \{b_t(\theta)\}^{-1}X_t(\theta) \to_u Y(\theta)$ in first mean, an application of the Toeplitz lemma, or integral version, will usually give C1. A simple discrete example is the estimation of the mean θ of a power series offspring distribution in a supercritical Galton-Watson process. The likelihood function and m.l. estimator $\hat{\theta}_n$ were derived by Heyde (1975) and asymptotic normality of $\hat{\theta}_n$ (randomly normalized) discussed. Here $\mathscr{I}_n(\theta) = \{\sigma(\theta)\}^{-2}\sum_{i=0}^{n-1} X_i$ where X_i is the ith generation size and $\sigma^2(\theta)$ the offspring distribution variance (Sweeting (1978)). It is straightforward to show using the above argument that C1 holds with $\{A_n(\theta)\}^2 = \theta^n$; C2 is readily seen to hold, so that Corollary 1 applies. The uniform asymptotic normality of $\hat{\theta}_n$ (obtained by a different method) was used in [15] to make power calculations. The pure birth process provides a simple continuous-time example. The likelihood function and m.l. estimator $\hat{\lambda}_t$ of the birth parameter λ are derived by Keiding (1974), who obtains the asymptotic distribution of $\hat{\lambda}_t$. Here $\mathscr{I}_t(\lambda) = \lambda^{-2}(X_t - X_0)$ where X_t is the population size at time t, and it is easily shown that C1 and C2 hold with $\{A_t(\lambda)\}^2 = e^{\lambda t}$. The asymptotic results in [12] therefore follow with the addition of uniform convergence.

No attempt has been made here to work through details of more complex processes, but the results of this paper should apply to a very wide range of inference problems. Where asymptotic results are available in the literature (for example, the Ornstein-Uhlenbeck process with unknown drift parameter—see [6] and [9]), use of the general results proved here provides a simpler route to establishing asymptotic normality—and with the additional bonus of uniformity of convergence.

REFERENCES

[1] Aitchison, J. and Silvey, S. D. (1958). Maximum-likelihood estimation of parameters subject to restraints. *Ann. Math. Statist.* **29** 813–828.

[2] Bhat, B. R. (1974). On the method of maximum likelihood for dependent observations. *J. Roy. Statist. Soc. B.* **36** 48–53.

[3] Billingsley, P. (1961). *Statistical Inference for Markov Processes.* Univ. Chicago.

[4] Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

[5] Bradley, R. and Gart, J. (1962). The asymptotic properties of ML estimators when sampling from associated populations. *Biometrika* **49** 205–214.

[6] Brown, B. M. and Hewitt, J. I. (1975). Asymptotic likelihood theory for diffusion processes. *J. Appl. Probability* **12** 228–238.

[7] Crowder, M. J. (1976). Maximum likelihood estimation for dependent observations. *J. Roy. Statist. Soc. B* **38** 45–53.

[8] Dubman, M. and Sherman, B. (1969). Estimation of parameters in a transient Markov Chain arising in a reliability growth model. *Ann. Math. Statist.* **40** 1542–1556.

[9] Feigin, P. D. (1976). Maximum likelihood estimation for continuous-time stochastic processes. *Adv. Appl. Probability* **8** 712–736.

[10] Heyde, C. C. (1975). Remarks on efficiency in estimation for branching processes. *Biometrika* **63** 531–536.

[11] Hoadley, B. (1971). Asymptotic properties of maximum likelihood estimators for the independent not identically distributed case. *Ann. Math. Statist.* **42** 1977–1991.

[12] Keiding, N. (1974). Estimation in the birth process. *Biometrika* **61** 71–80.

[13] Keiding, N. (1975). Maximum likelihood estimation in the birth-and-death process. *Ann. Statist.* **3** 363–372.

[14] Philippou, A. N. and Roussas, G. G. (1973). Asymptotic normality of the maximum likelihood estimate in the independent not identically distributed case. *Ann. Inst. Statist. Math.* **27** 45–55.

[15] SWEETING, T. J. (1978). On efficient tests for branching processes. *Biometrika* **65** 123–127.

[16] WEISS, L. (1973). Asymptotic properties of maximum likelihood estimators in some nonstandard cases. *J. Amer. Statist. Assoc.* **63** 428–430.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF SURREY
GUILDFORD, SURREY GU2 5XH
ENGLAND

STATEMENT OF OWNERSHIP, MANAGEMENT AND CIRCULATION
(Required by 39 U.S.C. 3685)

1. TITLE OF PUBLICATION: Annals of Statistics (USPS 026-160)
A. PUBLICATION NO.: 0 0 9 0 5 3 6 4
2. DATE OF FILING: OCTOBER 2, 1980

3. FREQUENCY OF ISSUE: BIMONTHLY
A. NO. OF ISSUES PUBLISHED ANNUALLY: 6
B. ANNUAL SUBSCRIPTION PRICE: $ 45.00

4. LOCATION OF KNOWN OFFICE OF PUBLICATION *(Street, City, County, State and ZIP Code) (Not printers)*
3401 INVESTMENT BL. # 6 HAYWARD (ALAMEDA COUNTY) CA 94545

5. LOCATION OF THE HEADQUARTERS OR GENERAL BUSINESS OFFICES OF THE PUBLISHERS *(Not printers)*
3401 INVESTMENT BL. # 6 HAYWARD (ALAMEDA COUNTY) CA 94545

6. NAMES AND COMPLETE ADDRESSES OF PUBLISHER, EDITOR, AND MANAGING EDITOR

PUBLISHER *(Name and Address)*
INSTITUTE OF MATHEMATICAL STATISTICS, 3401 INVESTMENT BLVD. # 6, HAYWARD, CA 94545

EDITOR *(Name and Address)*
DAVID V. HINKLEY DEPT. OF STATISTICS STANFORD UNIVERSITY STANFORD, CA 94305

MANAGING EDITOR *(Name and Address)*
DONALD TRUAX, MATHEMATICS DEPT UNIVERSITY OF OREGON EUGENE, OR 97403

7. OWNER *(If owned by a corporation, its name and address must be stated and also immediately thereunder the names and addresses of stockholders owning or holding 1 percent or more of total amount of stock. If not owned by a corporation, the names and addresses of the individual owners must be given. If owned by a partnership or other unincorporated firm, its name and address, as well as that of each individual must be giving. If the publication is published by a nonprofit organization, its name and address must be stated.)*

NAME	ADDRESS
INSTITUTE OF MATHEMATICAL STATISTICS	3401 INVESTMENT BLVD. # 6
(UNINCORPORATED NON-PROFIT SOCIETY)	HAYWARD, CA 94545

8. KNOWN BONDHOLDERS, MORTGAGEES, AND OTHER SECURITY HOLDERS OWNING OR HOLDING 1 PERCENT OR MORE OF TOTAL AMOUNT OF BONDS, MORTGAGES OR OTHER SECURITIES *(If there are none, so state)*

NAME	ADDRESS
NONE	

9. FOR COMPLETION BY NONPROFIT ORGANIZATIONS AUTHORIZED TO MAIL AT SPECIAL RATES *(Section 132.122, PSM)*
The purpose, function, and nonprofit status of this organization and the exempt status for Federal income tax purposes *(Check one)*

[XX] HAVE NOT CHANGED DURING PRECEDING 12 MONTHS [ ] HAVE CHANGED DURING PRECEDING 12 MONTHS *(If changed, publisher must submit explanation of change with this statement.)*

10. EXTENT AND NATURE OF CIRCULATION	AVERAGE NO. COPIES EACH ISSUE DURING PRECEDING 12 MONTHS	ACTUAL NO. COPIES OF SINGLE ISSUE PUBLISHED NEAREST TO FILING DATE
A. TOTAL NO. COPIES PRINTED *(Net Press Run)*	6708	6602
B. PAID CIRCULATION 1. SALES THROUGH DEALERS AND CARRIERS, STREET VENDORS AND COUNTER SALES	NONE	NONE
2. MAIL SUBSCRIPTIONS	4636	4506
C. TOTAL PAID CIRCULATION *(Sum of 10B1 and 10B2)*	4636	4506
D. FREE DISTRIBUTION BY MAIL, CARRIER OR OTHER MEANS SAMPLES, COMPLIMENTARY, AND OTHER FREE COPIES	12	12
E. TOTAL DISTRIBUTION *(Sum of C and D)*	4648	4518
F. COPIES NOT DISTRIBUTED 1. OFFICE USE, LEFT OVER, UNACCOUNTED, SPOILED AFTER PRINTING	2060	2084
2. RETURNS FROM NEWS AGENTS	NONE	NONE
G. TOTAL *(Sum of E, F1 and 2—should equal net press run shown in A)*	6708	6602

11. I certify that the statements made by me above are correct and complete.
SIGNATURE AND TITLE OF EDITOR, PUBLISHER, BUSINESS MANAGER, OR OWNER
[signature] TREASURER, IMS

12. FOR COMPLETION BY PUBLISHERS MAILING AT THE REGULAR RATES *(Section 132.121, Postal Service Manual)*

39 U. S. C. 3626 provides in pertinent part: "No person who would have been entitled to mail matter under former section 4359 of this title shall mail such matter at the rates provided under this subsection unless he files annually with the Postal Service a written request for permission to mail matter at such rates."

In accordance with the provisions of this statute, I hereby request permission to mail the publication named in Item 1 at the phased postage rates presently authorized by 39 U. S. C. 3626.

SIGNATURE AND TITLE OF EDITOR, PUBLISHER, BUSINESS MANAGER, OR OWNER
[signature] TREASURER, IMS