The Annals of Mathematical Statistics
1971, Vol. 42, No. 6, 1977–1991

ASYMPTOTIC PROPERTIES OF MAXIMUM LIKELIHOOD ESTIMATORS FOR THE INDEPENDENT NOT IDENTICALLY DISTRIBUTED CASE

BY BRUCE HOADLEY
Bell Laboratories, Inc.

Conditions are established under which maximum likelihood estimators are consistent and asymptotically normal in the case where the observations are independent but not identically distributed. The key concept employed is uniform integrability; and the required convergence theorems which involve uniform integrability, and are of independent interest, appear in the appendix.

A motivational example involving estimation under variable censoring is presented. This example invokes the full generality of the theorems with regard to lack of i.i.d. and lack of densities wrt Lebesgue or counting measure.

1. Introduction. The asymptotic properties of maximum likelihood estimates (MLE's) have been studied by many people under a variety of conditions. Usually it is assumed that the observations, on which the MLE's are based, are independent identically distributed (i.i.d.) [see, for example, Chanda (1954), Cramér (1946), Daniels (1961), Doob (1934), Doss (1962), (1963), Huber (1967), Kulldorff (1957), LeCam (1953), (1966), Wald (1949), and Wolfowitz (1949)]. Some results have been obtained for models in which the observations are not i.i.d. For example, Bradley and Gart (1962) generalized the work of Chanda to the case where the observations are independent but not identically distributed (i.n.i.d.). Halperin (1952) considered the case where only the r smallest order statistics are observable; Billingsley (1961) and Roussas (1965), (1967) dealt with Markov processes, which are stationary and ergodic; and Silvey (1961) provided a very nice discussion of the problem for arbitrary stochastic processes, but his conditions are too restrictive and are not easily checked for the case considered in this paper.

The author was motivated to reconsider the i.n.i.d. case by an example which was not explicitly covered by the conditions of Bradley and Gart. The data arose in a study, which was designed to estimate, among other things, the cdf of nonservice time, which is the length of time that a dwelling, where telephone service was disconnected, is without service. The study was conducted during a fixed interval of time $[0, T]$. Throughout this interval, disconnections occurred at many dwellings and their nonservice times were observed. Of course, if the kth disconnection occurred at time s_k, and service was not reestablished by time T, then it was only

Received July 17, 1971.

Key words and phrases. Asymptotic theory, censoring, central limit theorem, convergence theorems, estimation, law of large numbers, maximum likelihood estimation, uniform integrability.

observed that the nonservice time was greater than $t_k = T-s_k$. To formalize this, let Z_k, $k = 1, 2, \cdots$, be the random nonservice time of the kth disconnected dwelling and assume that the Z_k's are i.i.d. If $Z_k \leqq t_k$, it is observed exactly; however, if $Z_k > t_k$, then this fact is all that is observed. This problem has been studied by Bartholomew (1957), (1963) and Bartlett (1953) under the assumption that the lifetime distribution is exponential. If it is assumed that the nonservice time cdf, $G(z \mid \boldsymbol{\theta})$, is absolutely continuous wrt Lebesgue measure λ with pdf $g(z \mid \boldsymbol{\theta})$, then it can be shown (see (1.4)) that the likelihood function is

$$L_n(\boldsymbol{\theta}) = \{\prod_{j=1}^{r} g(Z_{k_j} \mid \boldsymbol{\theta})\}\{\prod_{j=r+1}^{n} [1-G(t_{k_j} \mid \boldsymbol{\theta})]\}, \tag{1.1}$$

where $Z_{k_1}, \cdots, Z_{k_r}$ are those nonservice times which are observed exactly. At first sight, it is not clear how to apply the standard asymptotic theory of MLE's to this likelihood function; so it is now shown how this problem can be formulated as a standard one involving i.n.i.d. observations, which have a continuous and discrete part (a case not explicitly covered by Bradley and Gart).

Define the random variables

$$\begin{aligned} Y_k &= Z_k \quad \text{if} \quad Z_k \leqq t_k, \\ &= t_k \quad \text{if} \quad Z_k > t_k, \qquad k = 1, 2, \cdots; \end{aligned} \tag{1.2}$$

let ν be the σ-finite measure on the Borel real line which assigns measure 1 to each point in $\mathscr{T} = \{t_k \colon k = 1, 2, \cdots\}$; and let $D_k = \{y \colon 0 \leqq y < t_k, y \notin \mathscr{T}\}$

$$\begin{aligned} f_k(y \mid \boldsymbol{\theta}) &= g(y \mid \boldsymbol{\theta}) \quad &&\text{if} \quad y \in D_k; \\ &= 1-G(t_k \mid \boldsymbol{\theta}) \quad &&\text{if} \quad y = t_k; \\ &= 0 \quad &&\text{otherwise.} \end{aligned} \tag{1.3}$$

Then the observations, $\{Y_k, k = 1, 2, \cdots\}$, are i.n.i.d.; and $f_k(y \mid \boldsymbol{\theta})$ is a version of the density of Y_k wrt $\mu = \nu+\lambda$. So the likelihood function is

$$L_n(\boldsymbol{\theta}) = \prod_{k=1}^{n} f_k(Y_k \mid \boldsymbol{\theta}), \tag{1.4}$$

the form usually dealt with in asymptotic theory. Note that if $Y_k = t_{k'} < t_k$, then $L_n(\boldsymbol{\theta}) \equiv 0$; however, since this happens with probability zero, no difficulty arises in the distributional results.

Another example in which i.n.i.d. observations arise is the reliability growth model discussed by Dubman and Sherman (1969). If Y_k, $k = 1, 2, \cdots$, is the waiting time between failures $(k-1)$ and k, then the Y_k's are independent and

$$\begin{aligned} f_k(y \mid \boldsymbol{\theta}) &= P\{Y_k = y \mid \boldsymbol{\theta}\} \\ &= p_k(\boldsymbol{\theta})[1-p_k(\boldsymbol{\theta})]^{y-1}, \qquad y = 1, 2, \cdots, \end{aligned} \tag{1.5}$$

where $\boldsymbol{\theta} = (p, \beta)$ and $p_k(\boldsymbol{\theta}) = p\beta^{k-1}$ is the conditional probability of failure on the next trial, given that $k-1$ failures have occurred.

Observations which are i.n.i.d. also occur in any kind of regression model where the distribution of Y_k depends on the value of some concomitant vector x_k. Further discussion of these examples appears in Section 5.

The purpose of this paper is to present for the i.n.i.d. case an alternative set of conditions which implies consistency of the MLE and another set which implies asymptotic normality. This extends the asymptotic theory to many additional interesting examples. As a sequel to this paper, Chao (1970) has developed a somewhat different set of conditions which imply strong consistency. (See Section 3 for further remarks on Chao's conditions.)

Various convergence theorems which are needed in the proofs, and are of independent interest, appear in the appendix. For example, a useful uniform convergence theorem which gives conditions under which $\lim E[X_k(s)] = E[\lim X_k(s)]$ as $s \to s_0$ uniformly in k, where $\{X_k(s): k = 1, 2, \cdots; s \text{ in some set}\}$ is a collection of random variables, is presented.

2. Notation and Preliminaries. Let $Y_1, Y_2, \cdots$ be a sequence of independent random variables, which are defined on the probability space $(\Omega, \mathscr{F}, P_\theta)$, and take values in a measure space $(\mathscr{Y}, \mathscr{A}, \mu)$. $\mathscr{Y}$ could be $\mathscr{R}^m$ (Euclidean m-space); and $\boldsymbol{\theta} \in \Theta \subset \mathscr{R}^p$. Let $\|\cdot\|$ be the ordinary Euclidean norm on $\mathscr{R}^p$.

Assume that Y_k has density $f_k(y \mid \boldsymbol{\theta})$ wrt μ, the σ-finite measure on $(\mathscr{Y}, \mathscr{A})$; so for $A \in \mathscr{A}$,

$$P\{Y_k \in A \mid \boldsymbol{\theta}\} = \int_A f_k(y \mid \boldsymbol{\theta})\, d\mu(y). \tag{2.1}$$

The likelihood function suitable to the above structure is given by

$$L_n(\boldsymbol{\theta}) = \prod_{k=1}^n f_k(Y_k \mid \boldsymbol{\theta}). \tag{2.2}$$

The MLE of the true parameter value $(\boldsymbol{\theta}_0)$ is denoted by $\hat{\boldsymbol{\theta}}_n$, and is defined to be any point in Θ satisfying

$$L_n(\hat{\boldsymbol{\theta}}_n) \geqq L_n(\boldsymbol{\theta}) \qquad \text{for all } \boldsymbol{\theta} \in \Theta. \tag{2.3}$$

It is possible that for some values of n no such point exists, in which case $\hat{\boldsymbol{\theta}}_n$ is set equal to an arbitrary point in Θ (say $\boldsymbol{\theta}_1$). Of course we assume that $\boldsymbol{\theta}_0 \in \Theta$.

The extended random variables which are convenient to work with while proving consistency are

$$\begin{aligned} R_k(\boldsymbol{\theta}) &= \ln[f_k(Y_k \mid \boldsymbol{\theta})/f_k(Y_k \mid \boldsymbol{\theta}_0)] && \text{if } f_k(Y_k \mid \boldsymbol{\theta}_0) > 0 \\ &= 0 && \text{otherwise.} \\ R_k(\boldsymbol{\theta}, \rho) &= \sup\{R_k(\mathbf{t}): \|\mathbf{t} - \boldsymbol{\theta}\| \leqq \rho\} \\ V_k(r) &= \sup\{R_k(\boldsymbol{\theta}): \|\boldsymbol{\theta}\| > r\}. \end{aligned} \tag{2.4}$$

For any random variable X, let

$$\begin{aligned} X^{(B)} &= X && \text{if } X \geqq -B \\ &= -B && \text{otherwise,} \end{aligned} \tag{2.5}$$

where $B \geqq 0$. The expectations, when $\boldsymbol{\theta}_0$ obtains, of $R_k(\boldsymbol{\theta})$, $R_k(\boldsymbol{\theta}, \rho)$, $V_k(r)$, $R_k^{(B)}(\boldsymbol{\theta})$, $R_k^{(B)}(\boldsymbol{\theta}, \rho)$, and $V_k^{(B)}(r)$ shall be denoted by $r_k(\boldsymbol{\theta})$, $r_k(\boldsymbol{\theta}, \rho)$, $v_k(r)$, $r_k^{(B)}(\boldsymbol{\theta})$, $r_k^{(B)}(\boldsymbol{\theta}, \rho)$, and $v_k^{(B)}(r)$, respectively.

For the asymptotic normality section define

$$\Phi_k(y, \boldsymbol{\theta}) = \ln f_k(y \mid \boldsymbol{\theta}); \tag{2.6}$$

let $\dot{\boldsymbol{\Phi}}_k(y, \boldsymbol{\theta})$ be the $p \times 1$ vector whose ith component is

$$\dot{\boldsymbol{\Phi}}_{k,i}(y, \boldsymbol{\theta}) = \frac{\partial}{\partial \theta_i} \Phi_k(y, \boldsymbol{\theta}); \tag{2.7}$$

and let $\ddot{\boldsymbol{\Phi}}_k(y, \boldsymbol{\theta})$ be the $p \times p$ matrix whose (i, j)th component is

$$\ddot{\boldsymbol{\Phi}}_{k,ij}(y, \boldsymbol{\theta}) = \frac{\partial^2}{\partial \theta_i \, \partial \theta_j} \Phi_k(y, \boldsymbol{\theta}). \tag{2.8}$$

To simplify notation, let $\sum_k a_k$ denote $\sum_{k=1}^n a_k$; let $\bar{a}_n$ denote $[\sum_k a_k]/n$; let $E(\cdot)$ and $P\{\cdot\}$ denote expectation and probability wrt $\boldsymbol{\theta}_0$; and for all limits as $n \to \infty$, the notation $n \to \infty$ will be suppressed. Also, $\to_P$ and $\to_L$ denote convergence in probability and law, respectively. The open sphere with center $\boldsymbol{\theta}$ and radius ρ will be denoted by $S(\boldsymbol{\theta}, \rho)$, and the closed sphere by $\bar{S}(\boldsymbol{\theta}, \rho)$. Positive constants K and δ are generic so that, e.g.,

$$E|X_k|^{1+\delta} \leqq K \tag{2.9}$$

shall mean that there exist positive constants K and δ so that (2.9) holds for $k = 1, 2, \cdots$.

Some of the assumptions made by Bradley and Gart which are not made for both the consistency and asymptotic normality part of this paper are:

1. The measure μ is either Lebesgue measure or counting measure.
2. For k fixed, the support set of Y_k (i.e $\{y : f_k(y \mid \boldsymbol{\theta}) > 0\}$) is the same for all $\boldsymbol{\theta} \in \Theta$ (this assumption is almost implicitly made in the asymptotic normality section of this paper because Assumption N3 implies that $P\{f_k(Y_k \mid \boldsymbol{\theta}) > 0\} = 1$ for all $\boldsymbol{\theta}$).
3. $\dfrac{\partial^3}{\partial \theta_r \, \partial \theta_s \, \partial \theta_t} \Phi_k(Y_k, \boldsymbol{\theta})$ exists, a.s. $[P]$.
4. $\left| \dfrac{\partial}{\partial \theta_r} f_k(Y_k \mid \boldsymbol{\theta}) \right| < F_{kr}(Y_k)$ for all $\boldsymbol{\theta} \in \Theta$ a.s. $[P]$, where $\int_{R_k} F_{kr}(y) \, d\mu(y) < \infty$.

For the nonservice time example in Section 5, Assumptions 1 and 4 are not satisfied. If Y_k has a uniform distribution on $(0, \theta)$, then Assumption 2 is not satisfied (note that asymptotic normality does not hold for this example, but consistency does).

Proofs of the main results rely heavily on the theorems in the appendix which are numbered A.x, $x = 1, \cdots, 6$.

3. Consistency. The approach to consistency will be similar to that taken by Wald (1949); however, in order to handle the lack of i.i.d., the conditions will be somewhat different. The conditions are:

C1. Θ is a closed subset of $\mathscr{R}^p$.

C2. $f_k(Y_k \mid \boldsymbol{\theta})$ is an upper semicontinuous (u.s.c.) function of $\boldsymbol{\theta}$, uniformly in k, a.s. $[P]$.

C3. There exists $\rho^* = \rho^*(\boldsymbol{\theta}) > 0$ and $r > 0$ for which
 (i) $E[R_k^{(0)}(\boldsymbol{\theta}, \rho)]^{1+\delta} \leqq K, \quad 0 \leqq \rho \leqq \rho^*$;
 (ii) $E[V_k^{(0)}(r)]^{1+\delta} \leqq K$.

C4. There exists $B > 0$ for which
 (i) $\bar{r}^{(B)}(\boldsymbol{\theta}) = \limsup \bar{r}_n^{(B)}(\boldsymbol{\theta}) < 0, \quad \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$;
 (ii) $\limsup \bar{v}_n^{(B)}(r) < 0$.

C5. $R_k(\boldsymbol{\theta}, \rho)$ and $V_k(r)$ are measurable functions of Y_k.

Note that if the domain of $f_k(y \mid \boldsymbol{\theta})$, viewed as a function of y, depends on $\boldsymbol{\theta}$, then there may be a $\boldsymbol{\theta}$ for which $P\{f_k(Y_k \mid \boldsymbol{\theta}) = 0 \mid \boldsymbol{\theta}_0\} > 0$, i.e., $P\{R_k(\boldsymbol{\theta}) = -\infty \mid \boldsymbol{\theta}_0\} > 0$. But this should not affect the consistency of the MLE; so, for example, it should not be required that $E|R_k(\boldsymbol{\theta})|^{1+\delta} < \infty$. However, an assumption about the right tail of the distribution of $R_k(\boldsymbol{\theta}, \rho)$ is necessary. This explains why C3(i) is stated in terms of $R_k^{(0)}(\boldsymbol{\theta}, \rho)$ rather than $R_k(\boldsymbol{\theta}, \rho)$. To prove strong consistency, Chao (1970) assumed that for $0 < s, \rho < \varepsilon(\boldsymbol{\theta})$, $E[\exp\{sR_k(\boldsymbol{\theta}, \rho)\}] \leqq K$, which is stronger than C3(i). He also replaced C3(ii) and C4(ii) by: $f_k(Y_k \mid \boldsymbol{\theta})/f_k(Y_k \mid \boldsymbol{\theta}_0) \to 0$ uniformly in k, a.s. $[P]$ as $\|\boldsymbol{\theta}\| \to \infty$; which is not satisfied by the example given in Section 5.

Stronger but more easily applicable replacements for C3 and C4 are:

C3′. There exists $\rho^* = \rho^*(\boldsymbol{\theta}) > 0$ and $r > 0$ for which
 (i) $E[R_k(\boldsymbol{\theta}, \rho)]^2 \leqq K, \quad 0 \leqq \rho \leqq \rho^*$;
 (ii) $E[V_k(r)]^2 \leqq K$.

C4′. (i) $\lim \bar{r}_n(\boldsymbol{\theta}) < 0, \quad \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$;
 (ii) $\lim \bar{v}_n(r) < 0$.

A comment is in order for C4′(i). If for $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, the distribution of Y_k when $\boldsymbol{\theta}$ obtains is not the same as the distribution of Y_k when $\boldsymbol{\theta}_0$ obtains, then Wald (1949) showed that $r_k(\boldsymbol{\theta}) < 0$. Condition C4′(i) states that this is true on the average.

THEOREM 1. *If conditions* C1–C5 *are satisfied, then*

$$\hat{\boldsymbol{\theta}}_n \to_p \boldsymbol{\theta}_0.$$

PROOF. For $\eta > 0$, define $\Theta(\eta) = \Theta - S(\boldsymbol{\theta}_0, \eta)$. It then suffices to show that

$$P\{\hat{\boldsymbol{\theta}}_n \in \Theta(\eta)\} \to 0. \tag{3.1}$$

Let

$$R_n{}^* = \sup\{\ln \textstyle\prod_{k=1}^n [f_k(Y_k \mid \boldsymbol{\theta})/f_k(Y_k \mid \boldsymbol{\theta}_0)] : \boldsymbol{\theta} \in \Theta(\eta)\} \\ = \sup\{\textstyle\sum_k R_k(\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta(\eta)\}. \tag{3.2}$$

Since $\{\hat{\boldsymbol{\theta}}_n \in \Theta(\eta)\} \subset \{R_n{}^* \geqq 0\}$, it suffices to show that

$$R_n{}^* \to_P -\infty. \tag{3.3}$$

For the r in the conditions, let

$$\omega = \Theta(\eta) \cap \bar{S}(\mathbf{0}, r)$$
$$R_{n,1}^* = \sup\{\textstyle\sum_k R_k(\boldsymbol{\theta}) : \boldsymbol{\theta} \in \omega\} \tag{3.4}$$
$$R_{n,2}^* = \sup\{\textstyle\sum_k R_k(\boldsymbol{\theta}) : \|\boldsymbol{\theta}\| > r\}.$$

It now suffices to show that $R_{n,1}^* \to_P -\infty$ and $R_{n,2}^* \to_P -\infty$. First consider $R_{n,1}$. C2 implies that as $\rho \downarrow 0$, $R_k{}^{(B)}(\boldsymbol{\theta}, \rho) \downarrow R_k{}^{(B)}(\boldsymbol{\theta})$, uniformly in k, a.s. $[P]$; and C3(i) implies that $\{R_k{}^{(B)}(\boldsymbol{\theta}, \rho) : k = 1, 2, \cdots; 0 \leqq \rho \leqq \rho^*\}$ is uniformly integrable (u.i.); hence, by Theorem A.3(ii), $r_k{}^{(B)}(\boldsymbol{\theta}, \rho) \downarrow r_k{}^{(B)}(\boldsymbol{\theta})$ as $\rho \downarrow 0$, uniformly in k. So for each $\boldsymbol{\theta} \in \omega$, there exists a $\rho(\boldsymbol{\theta}) \leqq \rho^*$ for which

$$r_k{}^{(B)}(\boldsymbol{\theta}, \rho(\boldsymbol{\theta})) < r_k{}^{(B)}(\boldsymbol{\theta}) - \bar{r}^{(B)}(\boldsymbol{\theta})/2. \tag{3.5}$$

Now $\{S(\boldsymbol{\theta}, \rho(\boldsymbol{\theta}))\}$ forms an open cover of the compact set ω; so there exist $\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_g \in \omega$ for which

$$\omega \subset \textstyle\bigcup_{i=1}^g S(\boldsymbol{\theta}_i, \rho(\boldsymbol{\theta}_i)). \tag{3.6}$$

By (3.5) and C4(i), $\limsup \bar{r}_n{}^{(B)}(\boldsymbol{\theta}_i, \rho(\boldsymbol{\theta}_i)) < 0$. Also it follows from C3(i) that $E|R_k{}^{(B)}(\boldsymbol{\theta}_i, \rho(\boldsymbol{\theta}_i))|^{1+\delta} \leqq K$ (remember that K is generic); so Theorem A.4 applies to give

$$\textstyle\sum_k R_k(\boldsymbol{\theta}_i, \rho(\boldsymbol{\theta}_i)) \to_P -\infty. \tag{3.7}$$

That $R_{n,1}^* \to_P -\infty$ follows from (3.7) and the fact that

$$R_{n,1}^* \leqq \max\{\textstyle\sum_k R_k(\boldsymbol{\theta}_i, \rho(\boldsymbol{\theta}_i)) : 1 \leqq i \leqq g\}. \tag{3.8}$$

Conditions C3(ii) and C4(ii) along with Theorem A.4 insure that

$$\textstyle\sum_k V_k(r) \to_P -\infty,$$

which implies $R_{n,2}^* \to_P -\infty$. □

4. Asymptotic normality. The approach to asymptotic normality will be related to that taken by Roussas (1968). The conditions are:

N1. Θ is an open subset of $\mathscr{R}^p$.

N2. $\hat{\boldsymbol{\theta}}_n \to_P \boldsymbol{\theta}_0$.

N3. $\dot{\boldsymbol{\Phi}}_k(Y_k, \boldsymbol{\theta})$ and $\ddot{\boldsymbol{\Phi}}_k(Y_k, \boldsymbol{\theta})$ exist, a.s. $[P]$.

N4. $\ddot{\boldsymbol{\Phi}}_k(Y_k, \boldsymbol{\theta})$ is a continuous function of $\boldsymbol{\theta}$, uniformly in k, a.s. $[P]$, and is a measurable function of Y_k.

N5. $E[\dot{\boldsymbol{\Phi}}_k(Y_k, \boldsymbol{\theta}) \mid \boldsymbol{\theta}] = 0 \quad k = 1, 2, \cdots$.

N6. $\boldsymbol{\Gamma}_k(\boldsymbol{\theta}) = E[\dot{\boldsymbol{\Phi}}_k(Y_k, \boldsymbol{\theta})\dot{\boldsymbol{\Phi}}_k(Y_k, \boldsymbol{\theta})' \mid \boldsymbol{\theta}] = -E[\ddot{\boldsymbol{\Phi}}_k(Y_k, \boldsymbol{\theta}) \mid \boldsymbol{\theta}]$.

N7. $\bar{\boldsymbol{\Gamma}}_n(\boldsymbol{\theta}) \to \bar{\boldsymbol{\Gamma}}(\boldsymbol{\theta})$, and $\bar{\boldsymbol{\Gamma}}(\boldsymbol{\theta})$ is positive definite.

N8. For some $\delta > 0$, $\sum_k E|\boldsymbol{\lambda}'\dot{\boldsymbol{\Phi}}_k(Y_k, \boldsymbol{\theta}_0)|^{2+\delta}/n^{(2+\delta)/2} \to 0$ for all $\boldsymbol{\lambda} \in \mathscr{R}^p$.

N9. There exist $\varepsilon > 0$ and random variables $B_{k,ij}(Y_k)$ such that
   (i) $\sup\{|\ddot{\boldsymbol{\Phi}}_{k,ij}(Y_k,\mathbf{t})| : \|\mathbf{t}-\boldsymbol{\theta}_0\| \leqq \varepsilon\} \leqq B_{k,ij}(Y_k)$.
   (ii) $E|B_{k,ij}(Y_k)|^{1+\delta} \leqq K$.

Some comments on these conditions are in order. Conditions N5 and N6 are standard conditions for the asymptotic normality of MLE's, and they are implied by:

$$N5'. \quad \frac{\partial}{\partial\theta_i}\int f_k(y \mid \boldsymbol{\theta})\, d\mu(y) = \int \frac{\partial}{\partial\theta_i} f_k(y \mid \boldsymbol{\theta})\, d\mu(y)$$

$$N6'. \quad \frac{\partial}{\partial\theta_i \partial\theta_j}\int f_k(y \mid \boldsymbol{\theta})\, d\mu(y) = \int \frac{\partial^2}{\partial\theta_i \partial\theta_j} f_k(y \mid \boldsymbol{\theta})\, d\mu(y).$$

A stronger, but more easily applicable, replacement for N8 is:

N8′. $E|\dot{\boldsymbol{\Phi}}_{k,i}(Y_k, \boldsymbol{\theta}_0)|^3 \leqq K$.

THEOREM 2. *If conditions* N1 *to* N9 *are satisfied, then*

$$n^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \to_L N(\mathbf{0}, \bar{\boldsymbol{\Gamma}}^{-1}(\boldsymbol{\theta}_0)).$$

PROOF. By N1 and N2, there exists $\eta > 0$ such that $S(\boldsymbol{\theta}_0, \eta) \subset \Theta$, and $P\{\hat{\boldsymbol{\theta}}_n \in S(\boldsymbol{\theta}_0, \eta)\} = 1 - \varepsilon_n$, where $\varepsilon_n \to 0$. So with probability $1 - \varepsilon_n$,

$$\mathbf{0} = \begin{bmatrix} \frac{\partial}{\partial\theta_1} \ln L_n(\boldsymbol{\theta}) \\ \cdot \\ \cdot \\ \cdot \\ \frac{\partial}{\partial\theta_p} \ln L_n(\boldsymbol{\theta}) \end{bmatrix}_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_n} = \sum_k \dot{\boldsymbol{\Phi}}_k(Y_k, \hat{\boldsymbol{\theta}}_n). \tag{4.1}$$

Define

$$\psi_k(\gamma) = \dot{\boldsymbol{\Phi}}_k(y, \boldsymbol{\theta} + \gamma(\mathbf{t} - \boldsymbol{\theta})). \tag{4.2}$$

Then, by the fundamental theorem of calculus,

$$\psi_k(1)-\psi_k(0) = \textstyle\int_0^1 \psi'(\xi)\,d\xi; \tag{4.3}$$

or

$$\begin{aligned}\dot{\mathbf{\Phi}}_k(y,\mathbf{t})-\dot{\mathbf{\Phi}}_k(y,\boldsymbol{\theta}) &= \int_0^1 \frac{d}{d\gamma}\dot{\mathbf{\Phi}}_k(y,\boldsymbol{\theta}+\gamma(\mathbf{t}-\boldsymbol{\theta}))\,\big|_{\gamma=\xi}\,d\xi \\ &= [\textstyle\int_0^1 \ddot{\mathbf{\Phi}}_k(y,\boldsymbol{\theta}+\xi(\mathbf{t}-\boldsymbol{\theta}))\,d\xi](\mathbf{t}-\boldsymbol{\theta}).\end{aligned} \tag{4.4}$$

Now by setting $y = Y_k$, $t = \hat{\boldsymbol{\theta}}_n$, $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, summing over k from 1 to n, and recalling (4.1), one gets

$$n^{-\frac{1}{2}}\textstyle\sum_k \dot{\mathbf{\Phi}}_k(Y_k,\boldsymbol{\theta}_0) = I_n[n^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0)], \tag{4.5}$$

where

$$I_n = \textstyle\int_0^1 n^{-1}\sum_k[-\ddot{\mathbf{\Phi}}_k(Y_k,\boldsymbol{\theta}_0+\xi(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0))]\,d\xi. \tag{4.6}$$

The next step in the proof is to show that

$$I_n \to_P \bar{\mathbf{\Gamma}}(\boldsymbol{\theta}_0). \tag{4.7}$$

To do this, it suffices to proceed with one component at a time. Conditions N4, N6, and N9 allow the application of Theorem A.5(ii) to $-\ddot{\mathbf{\Phi}}_{k,ij}(Y_k, \mathbf{s})$ with $\mathbf{s} \in S = \bar{S}(\boldsymbol{\theta}_0, \varepsilon)$. The result is

$$\sup\{|n^{-1}\textstyle\sum_k[-\ddot{\mathbf{\Phi}}_{k,ij}(Y_k,\mathbf{s})]-\bar{\Gamma}_{n,ij}(\mathbf{s})| : \|\mathbf{s}-\boldsymbol{\theta}_0\| \leqq \varepsilon\} \to_P 0, \tag{4.8}$$

where $\bar{\Gamma}_{n,ij}(\mathbf{s})$ is the (ij)th component of $\bar{\Gamma}_n(\mathbf{s})$. Letting $s = \boldsymbol{\theta}_0+\xi(\mathbf{t}-\boldsymbol{\theta}_0)$, where $0 \leqq \xi \leqq 1$, it is clear that $\|\mathbf{s}-\boldsymbol{\theta}_0\| \leqq \|\mathbf{t}-\boldsymbol{\theta}_0\|$; hence, with the aid of N7, (4.8) becomes

$$n^{-1}\textstyle\sum_k[-\ddot{\mathbf{\Phi}}_{k,ij}(Y_k,\boldsymbol{\theta}_0+\xi(\mathbf{t}-\boldsymbol{\theta}_0))] \to_P \bar{\Gamma}_{ij}(\boldsymbol{\theta}_0+\xi(\mathbf{t}-\boldsymbol{\theta}_0)), \tag{4.9}$$

uniformly for $0 \leqq \xi \leqq 1$ and $\|\mathbf{t}-\boldsymbol{\theta}_0\| \leqq \varepsilon$.

Now since N4 and N9 hold, Theorem A.5(i) can be applied to give $\lim \Gamma_{k,ij}(\boldsymbol{\theta}_0+\xi(\mathbf{t}-\boldsymbol{\theta}_0)) = \Gamma_{k,ij}(\boldsymbol{\theta}_0)$ as $\mathbf{t} \to \boldsymbol{\theta}_0$, uniformly in k. This, along with N7 is enough to insure that

$$\lim \bar{\Gamma}_{ij}(\boldsymbol{\theta}_0+\xi(\mathbf{t}-\boldsymbol{\theta}_0)) = \bar{\Gamma}_{ij}(\boldsymbol{\theta}_0) \qquad \text{as } \mathbf{t} \to \boldsymbol{\theta}_0, \tag{4.10}$$

uniformly for $0 \leqq \xi \leqq 1$.

Combining (4.6), (4.9), and (4.10) with the fact that $\hat{\boldsymbol{\theta}} \to_P \boldsymbol{\theta}_0$, we obtain the desired result stated in (4.7).

The next step in the proof is to show that

$$n^{-\frac{1}{2}}\textstyle\sum_k \dot{\mathbf{\Phi}}_k(Y_k,\boldsymbol{\theta}_0) \to_L N(\mathbf{0}, \bar{\mathbf{\Gamma}}(\boldsymbol{\theta}_0)). \tag{4.11}$$

But this follows from the multivariate form of Liapounov's theorem (Theorem A.6), because the required conditions are granted by N5, N6, N7, and N8.

Finally, it is clear that (4.7), (4.11), and the fact that (4.5) holds with probability $1-\varepsilon_n$, where $\varepsilon_n \to 0$, imply the conclusion of Theorem 2. □

5. Examples. Consider the nonservice time example mentioned in the introduction, and assume that $t_k \geqq M > 0 \quad k = 1, 2, \cdots,$

$$g(z \mid \theta) = \frac{\theta}{(1+\theta z)^2} \, z \geqq 0, \theta \geqq 0; \tag{5.1}$$

so that

$$\begin{aligned} f_k(y \mid \theta) = L_1(\theta \mid y) &= \frac{\theta}{(1+\theta y)^2} \qquad \text{if } \; y \in D_k; \\ &= L_2(\theta \mid t_k) = \frac{1}{(1+\theta t_k)} \qquad \text{if } \; y = t_k. \end{aligned} \tag{5.2}$$

Ordinarily one would assume that $\theta > 0$, but then Θ would not be a closed set; so we interpret $\theta = 0$ to mean that $P\{Y_k = t_k \mid \theta = 0\} = 1$ and (5.2) still holds. The assumption that $t_k \geqq M > 0$ insures that the amount of information in the kth observation does not tend to zero. Clearly this could be relaxed for finitely many k's and probably even for infinitely many. For example, it might be possible to construct a decreasing sequence of t_k's whose limit is 0, but for which the results are applicable.

First the conditions for consistency are checked. Condition C1 holds, because $\theta \geqq 0$. To establish C2, it suffices to show that $L_1(\theta \mid y)$ and $L_2(\theta \mid t)$ are u.s.c. at θ uniformly in y and t respectively. For $\theta > 0$, it is clear that continuity holds uniformly in y and t; but for $\theta = 0$, it does not. However, u.s.c. does hold uniformly at $\theta = 0$, because: $L_2(\varepsilon \mid t) \leqq L_2(0 \mid t) = 1$; and for $y \geqq 1/4\varepsilon$, $L_1(\varepsilon \mid y) \leqq \varepsilon = L_1(0 \mid y)+\varepsilon$; and for $y < 1/4\varepsilon$, $L_1(\varepsilon \mid y) = L_1(0 \mid y)+L_1{}'(0 \mid y)\varepsilon+L_1{}''(\theta^* \mid y)\varepsilon^2/2 \leqq L_1(0 \mid y)+\varepsilon$, because $\theta^* \leqq \varepsilon \leqq 1/4y \leqq 2/y$ and $L_1{}''(\theta^* \mid y) \leqq 0$ whenever $\theta^* \leqq 2/y$.

For C3 and C4 note that

$$\begin{aligned} R_k(\theta) &= 0 && \text{if } \; Y_k \in D_k, \quad \theta_0 = 0 \\ &= \ln\left(\frac{\theta}{\theta_0}\right)\left(\frac{1+\theta_0 Y_k}{1+\theta Y_k}\right)^2 && \text{if } \; Y_k \in D_k, \quad \theta_0 > 0 \\ &= \ln\left(\frac{1+\theta_0 t_k}{1+\theta t_k}\right) && \text{if } \; Y_k = t_k; \end{aligned} \tag{5.3}$$

$$\begin{aligned} R_k(\theta) &\leqq 0 && \text{if } \; \theta_0 = 0 \\ &\leqq \ln(\theta_0/\theta) && \text{if } \; 0 < \theta \leqq \theta_0 \\ &\leqq \ln(\theta/\theta_0) && \text{if } \; 0 < \theta_0 < \theta; \end{aligned} \tag{5.4}$$

if $\theta_0 > 0$,

$$
\begin{aligned}
R_k(0,\rho) &= \ln\left(\frac{\rho}{\theta_0}\right)\left(\frac{1+\theta_0 Y_k}{1+\rho Y_k}\right)^2 && \text{if} \quad Y_k \in D_k, && Y_k < 1/\rho \\
&= \ln\left(\frac{1}{4\theta_0 Y_k}\right)(1+\theta_0 Y_k)^2 && \text{if} \quad Y_k \in D_k, && Y_k \geqq 1/\rho \qquad (5.5)\\
&= \ln(1+\theta_0 t_k) && \text{if} \quad Y_k = t_k;
\end{aligned}
$$

and

$$
\begin{aligned}
V_k(r) &= 0 && \text{if} \quad Y_k \in D_k, && \theta_0 = 0 \\
&= \ln\left(\frac{1}{4\theta_0 Y_k}\right)(1+\theta_0 Y_k)^2 && \text{if} \quad Y_k \in D_k, && Y_k < 1/r, \theta_0 > 0 \qquad (5.6)\\
&= \ln\left(\frac{r}{\theta_0}\right)\left(\frac{1+\theta_0 Y_k}{1+r Y_k}\right)^2 && \text{if} \quad Y_k \in D_k, && Y_k \geqq 1/r, \theta_0 > 0 \\
&= \ln\left(\frac{1+\theta_0 t_k}{1+r t_k}\right) && \text{if} \quad Y_k = t_k.
\end{aligned}
$$

Now if $\theta > 0$, C3(i) follows from (5.4); and if $\theta = 0$, $0 < \rho < \rho^*(0) < \theta_0$, then manipulation of (5.5) yields

$$
\begin{aligned}
R_k^{(0)}(0,\rho) &\leqq 2\ln(1+\theta_0 Y_k) && \text{if} \quad Y_k \in D_k \qquad (5.7)\\
&\leqq \ln(1+\theta_0 t_k) && \text{if} \quad Y_k = t_k,
\end{aligned}
$$

and a direct integration shows that $E[R_k^{(0)}(0,\rho)]^2 \leqq K$. If $r > \theta_0$ and $1/r < M$, then

$$
\begin{aligned}
V_k^{(0)}(r) &\leqq \left[\ln\left(\frac{1}{4\theta_0 Y_k}\right)(1+\theta_0 Y_k)^2\right]^{(0)} && \text{if} \quad Y_k \in D_k, Y_k < 1/r, && \theta_0 > 0 \\
&\leqq \ln\left(\frac{r}{\theta_0}\right) && \text{if} \quad Y_k \in D_k, Y_k \geqq 1/r, && \theta_0 > 0 \qquad (5.8)\\
&\leqq 0 && \text{if} \quad Y_k \in D_k, \quad \theta_0 = 0 \quad \text{or} && Y_k = t_k;
\end{aligned}
$$

and a direct integration shows that $E[V_k^{(0)}(r)]^2 \leqq K$. So C3 holds.

Now if $\theta_0 \neq \theta > 0$, then by Lemma 1 of Wald (1949), $r_k(\theta) < 0$; so for θ fixed, $r_k(\theta)$ is a negative continuous function of t_k on $[M, \infty]$ for which it is easily shown that $\lim r_k(\theta) < 0$ as $t_k \to \infty$; therefore, $r_k(\theta) \leqq r(\theta) < 0$, $k = 1, 2, \cdots$, and C4(i) holds. It also holds for $\theta_0 = \theta = 0$, because $r_k(0) = -\infty$. Using (5.6), a direct integration shows that for r sufficiently large, $v_k(r) < v(r) < 0$; so C4(ii) holds.

As for the asymptotic normality, if one assumes that $\theta > 0$, and the t_k's are such that N7 is satisfied, then the conditions associated with Theorem 2 are easy to verify for this example.

For the reliability growth model discussed in the Introduction, Theorem 1 holds, but Theorem 2 does not, because $n^{\frac{1}{2}}(\hat{p}_n - p_0)$ and $n^{\frac{3}{2}}(\hat{\beta}_n - \beta_0)$ are asymptotically jointly normal. The reason is that as $k \to \infty$, $\Gamma_{k,22}(\theta) \to \infty$, so that N7 is violated. An intuitive way of looking at it is that as $k \to \infty$, $p_k(\theta) \to 0$, so that the waiting times to failure (the Y_k's) become stochastically longer and longer and hence contain more and more information. If one assumes that as $k \to \infty$, $p_k(\theta) \downarrow p(\theta)$, where $0 < p(\theta) < 1$, then it can be shown that Theorems 1 and 2 both apply.

APPENDIX

This appendix contains various convergence theorems which are used in the main body of the paper. The notational conventions adopted in Section 2 hold here. All random variables in the appendix are defined on the probability space $(\Omega, \mathscr{F}, P)$.

The concept of uniform integrability (ui) is fundamental to the approach adopted in this paper to the asymptotic theory of MLE's for i.n.i.d. observations. A nice discussion of ui can be found in Neveu (1965), pages 49–54. The results from that book needed in this paper are now presented.

DEFINITION. (Neveu, page 49). A family $\{X_i: i \in I\}$ of integrable random variables is said to be uniformly integrable (ui) if

$$\limsup \{\textstyle\int_{|X_i| > M} |X_i| \, dP : i \in I\} = 0 \quad \text{as} \quad M \to \infty.$$

THEOREM A.1 (Neveu, page 54). *A sufficient condition for* $\{X_i: i \in I\}$ *to be* ui *is that* $E|X_i|^{1+\delta} \leqq K$.

THEOREM A.2 (Neveu, page 52). *The following are equivalent*:

(i) $\{X_n: n = 1, 2, \cdots\}$ *is* ui *and* $X_n \to_P X$.

(ii) X *is integrable and* $E|X_n - X| \to 0$.

Theorem A.2 is now applied to obtain a uniform convergence theorem, which is used often in both the main body of the paper and the remainder of the appendix.

THEOREM A.3. *Let U be a subset of* $\mathscr{R}^p$.

If $\{X_k(u): k = 1, 2, \cdots, u \in U\}$ *is* ui, *and* $\lim X_k(u) = X_k$ *as* $u \to u_0$, a.s. $[P]$, *then*

(i) $\{X_k: k = 1, 2, \cdots\}$ *is* ui. *If in addition*, $\lim X_k(u) = X_k$ *as* $u \to u_0$, *uniformly in* k, a.s. $[P]$, *then*

(ii) *As* $u \to u_0$, $\lim E|X_k(u) - X_k| = 0$, *uniformly in k, so that* $\lim EX_k(u) = EX_k$, *uniformly in k.*

PROOF. (i) Choose M so large that

$$\int_{|X_k(u)| > M} |X_k(u)| \, dP < \varepsilon. \tag{A.1}$$

Holding k fixed, let

$$\begin{aligned} A &= \{\lim X_k(u) = X_k \quad \text{as} \quad u \to u_0\} \\ B(u) &= \{|X_k(u)| > M\} \\ B &= \{|X_k| > M\}. \end{aligned} \tag{A.2}$$

For any $F \in \mathscr{F}$, let $I(F)$ denote the indicator random variable associated with F. It is then clear that as $u \to u_0$

$$\liminf I(A \cap B(u))|X_k(u)| \geqq I(A \cap B)|X_k|;$$

so by Fatou's lemma,

$$\begin{aligned} \int_{|X_k|>M} |X_k|\,dP &= E[I(A \cap B)|X_k|] \\ &\leqq E[\liminf I(A \cap B(u))|X_k(u)|] \\ &\leqq \liminf E[I(A \cap B(u))|X_k(u)|] \\ &= \liminf \int_{|X_k(u)|>M} |X_k(u)|\,dP \qquad \text{as } u \to u_0. \end{aligned} \tag{A.3}$$

But, by (A.1), the last expression in (A.3) is $\leqq \varepsilon$; thus, the result is established. □

(ii) If not, then there exist $\varepsilon > 0$ and sequences, $k_n \to \infty$, $s_n \to u_0$, for which

$$0 < \varepsilon < EZ_n, \tag{A.4}$$

where

$$Z_n = |X_{k_n}(s_n) - X_{k_n}|. \tag{A.5}$$

Now $\{X_{k_n}(s_n): n = 1, 2, \cdots\}$ is ui by assumption, and $\{X_{k_n}: n = 1, 2, \cdots\}$ is ui by part (i); hence,

$$\{Z_n: n = 1, 2, \cdots\} \quad \text{is ui.} \tag{A.6}$$

Also, since $X_k(u) \to X_k$ as $u \to u_0$, uniformly in k, a.s. $[P]$, it is clear that

$$Z_n \to 0 \text{ a.s. } [P]. \tag{A.7}$$

Now by Theorem A.2, (A.6) and (A.7) imply that $EZ_n \to 0$, which contradicts (A.4). □

Next, a sufficient condition for $\sum_k X_k \to_P -\infty$ is established. This is useful in the proof of consistency given in Section 3.

THEOREM A.4. *Let $\{X_k: 1, 2, \cdots\}$ be a sequence of independent random variables. Let $X_k^{(B)} = X_k$ when $X_k \geqq -B$, and $= -B$ otherwise; and let $\mu_k^{(B)} = EX_k^{(B)}$. If $E|X_k^{(0)}|^{1+\delta} \leqq K$ and $\mu^{(B)} = \limsup \bar{\mu}_n^{(B)} < 0$, then $\sum_k X_k \to_P -\infty$.*

PROOF. $\sum_k X_k \leqq \sum_k X_k^{(B)}$, and since $E|X_k^{(0)}|^{1+\delta} \leqq K$, $E|X_k^{(B)}|^{1+\delta} \leqq K$ (remember, K is generic). Hence, it follows from Markov's weak law of large numbers [see Loève (1960) page 275] that

$$\bar{X}_n^{(B)} - \bar{\mu}_n^{(B)} \to_P 0. \tag{A.8}$$

Now for sufficiently large n, $\bar{\mu}_n^{(B)} < \mu^{(B)}/2$; hence

$$\begin{aligned} P\{\textstyle\sum_k X_k^{(B)} \leqq n\mu^{(B)}/4\} &= P\{\bar{X}_n^{(B)} \leqq \mu^{(B)}/4\} \\ &\geqq P\{\bar{X}_n^{(B)} - \bar{\mu}_n^{(B)} \leqq -\mu^{(B)}/4\} \to 1 \qquad \text{by (A.8).} \end{aligned}$$

Therefore, $\sum_k X_k^{(B)} \to_P -\infty$. □

The next theorem is a weak law of large numbers which holds uniformly over a compact set. It is needed in the proof of asymptotic normality found in Section 4.

THEOREM A.5. *Let* $\{Y_k: k = 1, 2, \cdots\}$ *be independent random variables which assume values in some set* $\mathscr{Y}$ *endowed with the σ-field* $\mathscr{A}$. *Let* $H_k: \mathscr{Y} \times S \to \mathscr{R}^1$, *where* $S \subset \mathscr{R}^p$ *is compact; and let* $h_k(\mathbf{s}) = EH_k(Y_k, \mathbf{s})$. *Assume*:

(a) *For each* $\mathbf{s} \in S$, $H_k(y, \mathbf{s})$ *is* $\mathscr{A}$ *measurable.*

(b) $H_k(Y_k, \mathbf{s})$ *is continuous on* S, *uniformly in* k, a.s. $[P]$.

(c) *There exist measurable* $B_k: \mathscr{Y} \to \mathscr{R}^1$ *for which* $|H_k(y, \mathbf{s})| < B_k(y)$ *for all* $\mathbf{s} \in S$, *and* $E|B_k(Y_k)|^{1+\delta} \leqq K$.

Then

(i) $h_k(\mathbf{s})$ *is continuous on* S, *uniformly in* k.

(ii) $\sup\{|[\sum_k H_k(Y_k, \mathbf{s})]/n - \bar{h}_n(\mathbf{s})| : \mathbf{s} \in S\} \to_P 0$.

PROOF. (i) By assumption (b), for each $\mathbf{s}_0 \in S$, $\lim H_k(Y_k, \mathbf{s}) = H_k(Y_k, \mathbf{s}_0)$ as $\mathbf{s} \to \mathbf{s}_0$, uniformly in k, a.s. $[P]$. From assumption (c), it follows that $\{H_k(Y_k, \mathbf{s}): k = 1, 2, \cdots; \mathbf{s} \in S\}$ is ui. The result follows from Theorem A.3(ii).

(ii) Because of part (i), it can be assumed without loss of generality that $h_k(\mathbf{s}) = 0$. Let

$$H_k^*(y, \mathbf{s}, \rho) = \sup\{H_k(y, \mathbf{t}): \|\mathbf{t} - \mathbf{s}\| \leqq \rho\}$$
$$H_{*k}(y, \mathbf{s}, \rho) = \inf\{H_k(y, \mathbf{t}): \|\mathbf{t} - \mathbf{s}\| \leqq \rho\}. \tag{A.9}$$

These functions are $\mathscr{A}$ measurable, because S is separable, and $H_k(y, \mathbf{s})$ is continuous on S. From assumptions (b) and (c) it follows that

$$\lim H_k^*(Y_k, \mathbf{s}, \rho) = H_k(Y_k, \mathbf{s})$$
$$\lim H_{*k}(Y_k, \mathbf{s}, \rho) = H_k(Y_k, \mathbf{s}), \qquad \text{as } \rho \to 0, \tag{A.10}$$

uniformly in k, a.s. $[P]$; and

$$E|H_k^*(Y_k, \mathbf{s}, \rho)|^{1+\delta} \leqq K$$
$$E|H_{*k}(Y_k, \mathbf{s}, \rho)|^{1+\delta} \leqq K. \tag{A.11}$$

Theorem A.3(ii) now applies to give

$$\lim EH_k^*(Y_k, \mathbf{s}, \rho) = 0$$
$$\lim EH_{*k}(Y_k, \mathbf{s}, \rho) = 0, \qquad \text{as } \rho \to 0, \tag{A.12}$$

uniformly in k.

Equation (A.12) insures that for each $\mathbf{s} \in S$, there exists $\rho(\mathbf{s})$ so small that

$$-\varepsilon < EH_{*k}(Y_k, \mathbf{s}, \rho(\mathbf{s})) \leqq EH_k^*(Y_k, \mathbf{s}, \rho(\mathbf{s})) < \varepsilon. \tag{A.13}$$

The collection $\{S(\mathbf{s}, \rho(\mathbf{s}))\}$ forms an open cover of the compact set S; hence, there exist $\mathbf{s}_1, \cdots, \mathbf{s}_g \in S$ for which $S \subset \bigcup_{i=1}^g S(\mathbf{s}_i, \rho(\mathbf{s}_i))$.

Now, it can be said that

$$\text{(A.14)}\quad \min\{[\sum_k H_{*k}(Y_k, \mathbf{s}_i, \rho(\mathbf{s}_i))]/n : 1 \leqq i \leqq g\}$$
$$\leqq [\sum_k H_k(Y_k, \mathbf{s})]/n \leqq \max\{[\sum_k H_k^*(Y_k, s_i, \rho(s_i))]/n : 1 \leqq i \leqq g\},$$

for all $\mathbf{s} \in S$. By (A.11), the Markov weak law of large numbers can be applied to each term in the brackets following min and max in (A.14). This, combined with (A.13), insures that with probability $1-\varepsilon_n$, $\varepsilon_n \to 0$, $[\sum_k H_k(Y_k, \mathbf{s})]/n$ lies between -2ε and 2ε for all $\mathbf{s} \in S$. The result follows. □

Finally, the Liapounov form of the multivariate central limit theorem is presented. Of course this plays the dominant role in proving that MLE's are asymptotically normal in the i.n.i.d. case.

THEOREM A.6. *Let* $\{\mathbf{X}_k : k = 1, 2, \cdots\}$ *be independent p-dimensional random vectors for which* $E\mathbf{X}_k = \mathbf{0}$, $\text{Cov}(\mathbf{X}_k) = \mathbf{\Gamma}_k$. *Assume*:

(a) $\bar{\mathbf{\Gamma}}_n \to \bar{\mathbf{\Gamma}}$; *and* $\bar{\mathbf{\Gamma}}$ *is positive definite.*

(b) *For some* $\delta > 0$, $\sum_k E|\boldsymbol{\lambda}'\mathbf{X}_k|^{2+\delta}/n^{(2+\delta)/2} \to 0$ *for all* $\boldsymbol{\lambda} \in \mathscr{R}^p$.

Then $n^{-\frac{1}{2}} \sum_k \mathbf{X}_k \to_L N(\mathbf{0}, \bar{\mathbf{\Gamma}})$.

PROOF. The assumptions allow the application of Liapounov's theorem [see Loève (1960) page 275] to $\sum_k \boldsymbol{\lambda}'\mathbf{X}_k$ for all $\boldsymbol{\lambda} \neq 0$.
The result is

$$\text{(A.15)}\qquad [\sum_k \boldsymbol{\lambda}' X_k]/[n\boldsymbol{\lambda}'\bar{\mathbf{\Gamma}}_n\boldsymbol{\lambda}]^{\frac{1}{2}} \to_L N(0, 1).$$

But $\boldsymbol{\lambda}'\bar{\mathbf{\Gamma}}_n\boldsymbol{\lambda} \to \boldsymbol{\lambda}'\bar{\mathbf{\Gamma}}\boldsymbol{\lambda} \neq 0$; therefore

$$\text{(A.16)}\qquad \boldsymbol{\lambda}'[n^{-\frac{1}{2}} \sum_k X_k] \to_L N(0, \boldsymbol{\lambda}'\bar{\mathbf{\Gamma}}\boldsymbol{\lambda}),$$

for all $\boldsymbol{\lambda} \neq 0$. This implies the desired result [see Rao (1965) page 109]. □

REFERENCES

BARTHOLOMEW, D. J. (1957). A problem in life testing. *J. Amer. Statist. Assoc.* **52** 350–355.

BARTHOLOMEW, D. J. (1963). The sampling distribution of an estimate arising in life testing. *Technometrics* **5** 361–374.

BARTLETT, M. S. (1953). On the statistical estimation of mean life-times. *Phil. Mag.* **44** 249–262.

BILLINGSLEY, P. (1961). *Statistical Inference for Markov Processes.* Univ. of Chicago Press.

BRADLEY, R. A. and GART, J. J. (1962). The asymptotic properties of ML estimators when sampling from associated populations. *Biometrika* **49** 205–214.

CHANDA, K. C. (1954). A note on the consistency and maxima of the roots of likelihood equations. *Biometrika* **41** 56–61.

CHAO, M. T. (1970). Strong consistency of maximum likelihood estimators when the observations are independent but not identically distributed. *Dr. Y. W. Chen's 60-year Memorial Volume.* Academia Sinica, Taipei.

CRAMÉR, H. (1946). *Mathematical Methods of Statistics.* Princeton Univ. Press.

DANIELS, H. E. (1961). The asymptotic efficiency of a maximum likelihood estimator. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* 151–163. Univ. of California Press.

DOOB, J. L. (1934). Probability and statistics. *Trans. Amer. Math. Soc.* **36** 759–775.

DOSS, S. A. D. C. (1962). A note on consistency and asymptotic efficiency of maximum likelihood estimates in multi-parametric problems. *Calcutta Statist. Assoc. Bull.* **11** 85–93.

DOSS, S. A. D. C. (1963). On consistency and asymptotic efficiency of maximum likelihood estimates. *J. Indian Soc. Agric. Statist.* **1** 232–241.

DUBMAN, M. and SHERMAN, B. (1969). Estimation of parameters in a transient Markov chain arising in a reliability growth model. *Ann. Math. Statist.* **40** 1542–56.

HALPERIN, M. (1952). Maximum likelihood estimation in truncated samples. *Ann. Math. Statist.* **23** 226–238.

HUBER, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* 221–233. Univ. of California Press.

KULLDORFF, G. (1957). On the conditions for consistency and asymptotic efficiency of maximum likelihood estimates. *Skand. Aktuarietidskr.* **40** 129–144.

LECAM, L. (1953). On some asymptotic properties of maximum likelihood estimates and related Bayes' estimates. *Univ. of California Publications in Statistics* **1** 277–330.

LECAM, L. (1966). Likelihood functions for large numbers of independent observations. *Research Papers in Statistics*, ed. F. N. David. Wiley, New York, 167–187.

LOÈVE, M. (1960). *Probability Theory*, 2nd ed. Van Nostrand, Princeton.

NEVEU, J. (1965). *Mathematical Foundations of the Calculus of Probability.* Holden-Day, San Francisco.

RAO, C. R. (1965). *Linear Statistical Inference and Its Applications.* Wiley, New York.

ROUSSAS, G. G. (1965). Extension to Markov processes of a result by A. Wald about the consistency of the maximum likelihood estimate. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **4** 69–73.

ROUSSAS, G. G. (1968). Asymptotic normality of the maximum likelihood estimate in Markov processes. *Metrika* **14** 62–70.

SILVEY, S. D. (1961). A note on maximum-likelihood in the case of dependent random variables. *J. Roy. Statist. Soc. Ser. B* **23** 444–452.

WALD, A. (1949). Note on the consistency of the maximum likelihood estimate. *Ann. Math. Statist.* **20** 595–601.

WOLFOWITZ, J. (1949). On Wald's proof of the consistency of the maximum likelihood estimate. *Ann. Math. Statist.* **20** 601–602.