The Annals of Statistics
1985, Vol. 13, No. 3, 932–946

STRONG CONSISTENCY OF APPROXIMATE MAXIMUM LIKELIHOOD ESTIMATORS WITH APPLICATIONS IN NONPARAMETRICS[1]

By Jane-Ling Wang

University of California, Davis and The University of Iowa

Wald's general analytic conditions that imply strong consistency of the approximate maximum likelihood estimators (AMLEs) have been extended by Le Cam, Kiefer and Wolfowitz, Huber, Bahadur, and Perlman. All these conditions use the log likelihood ratio of the type $\log[f(x, \theta)/f(x, \theta_0)]$, where θ_0 is the true value of the parameter. However these methods usually fail in the nonparametric case. Thus, in this paper, for each $\theta \neq \theta_0$, we look at the log likelihood ratio of the type $\log[f(x, \theta)/f(x, \theta_r(\theta))]$, where $\theta_r(\theta)$ is a certain parameter selected in a neighborhood V_r of θ_0. Some general analytic conditions that imply strong consistency of the AMLE are given. The results are shown to be applicable to several nonparametric families having densities, e.g., concave distributions functions, and increasing failure rate distributions. In particular, they can be applied to several censored data cases.

1. Introduction. The strong consistency of approximate maximum likelihood estimators (AMLEs), under some regularity conditions, has been investigated by many statisticians, notably Wald (1949), Le Cam (1953), Kiefer and Wolfowitz (1956), Bahadur (1967), Huber (1967), and Perlman (1972). Each of the other papers (except for Le Cam, 1953) uses conditions which are stronger than Perlman's (1972) sufficient conditions based on dominance or semidominance by zero of the log likelihood ratio (LLR) of a distribution to the true one, while Le Cam's conditions are equivalent to those based on dominance. They share the common assumption that this LLR is locally dominated (see Perlman, 1972, for definitions).

More specifically, let $f(x, \theta_0)$ be the true density function and let θ^* be any point in the parameter space. Local dominance requires the existence of a neighborhood V of θ^* such that $\log[f(x, \theta)/f(x, \theta_0)]$ is dominated for θ in V. However, it is easy to give examples (cf. next paragraph) where the AMLE is consistent even though this local dominance assumption is violated. This is especially true in certain nonparametric families where the density functions exist. While local dominance usually fails for such families, in many instances the consistency of AMLEs still can be proved.

One such example is provided by the class $\mathscr{P}$ of all distributions F with decreasing density function f on $[0, \infty)$. Using f itself as the parameter, we can still define the maximum likelihood estimator (MLE) $\hat{\theta}$ to be that decreasing

Received March 1983; revised January 1985.

[1] This research was supported in part by National Science Foundation Grant MCS78-25301 and Army Research Office Contract DAAG 29 79 C0093.

AMS 1980 *subject classifications*. Primary 62F12, 62G05; secondary 60F15.

Key words and phrases. Maximum likelihood estimator, consistency, dominance by zero, concave distribution, increasing failure rate, decreasing failure rate.

density which maximizes the likelihood function. The corresponding $\hat{F}$ was found by Grenander (1956) to be the least concave majorant to the empirical distribution function. The strong consistency of this MLE easily follows from the continuity of the concave majorant functional or from Marshall's lemma (1970). However, $\mathscr{P}$ is not locally dominated; hence all the previous methods fail to apply. It is natural to consider a more general analytic sufficient condition for the strong consistency of AMLEs. Examining Perlman's paper (1972), we notice that his condition (1.7) is stronger than is needed. The family of all concave distribution functions on $[0, \infty)$ is a convex set. Thus, if we consider a log likelihood ratio of the type $\log\{f/[(1-\varepsilon)f_0 + \varepsilon f]\}$, we easily obtain an upper bound on this ratio. It only remains to check some of the regularity conditions and an information inequality, and these do still hold on $\mathscr{P}$, implying the strong consistency of the AMLEs in $\mathscr{P}$. These considerations motivate the basic idea of this paper.

We provide some new general techniques, like those in Wald (1949) and Perlman (1972), which prove the strong consistency of AMLE (defined in Section 2). We let Θ be the parameter space (possibly infinite dimensional) and let θ_0 denote the true parameter. For any θ in Θ and any neighborhood $V_r(\theta_0)$ of θ_0, we work with LLRs of the type $\log[f(x, \theta)/f(x, \theta_r(\theta))]$, where $\theta_r(\theta)$ is selected in $V_r(\theta_0)$ so that for any θ^* not in $V_r(\theta_0)$ there exists a neighborhood V of θ^* on which $\log[f(x, \theta)/f(x, \theta_r(\theta))]$ is dominated. In the previous example, we take $\theta_r(f) = (1-\varepsilon)f_0 + \varepsilon f$ for the family $\mathscr{P}$. Notice that all the previous authors used $\theta_r(\theta)$ identically equal to θ_0 and this restricts the applicability of their techniques. A sufficient condition (Lemma 2.1) for the strong consistency of AMLE is based on this new technique. Theorem 2.1 shows that these sufficient conditions are very close to being necessary as well. Several other sufficient conditions are given in Theorem 2.2 and Theorem 3.1. In Sections 4 and 5 the sufficient conditions in Theorem 3.1 are applied to several nonparametric families to show the strong consistency of AMLEs. While the technique of this paper is applicable to parametric families, it seems necessary only for nonparametric problems.

2. Sufficient conditions based on dominance and semidominance. To conserve space, we adopt the notation and some of the definitions in Perlman (1972), hereafter abbreviated as [P](1972). Let $\mathscr{P}$, Θ, θ, θ_0, $P_0, {}_*P_0$, $\{V_r(\theta_0)\}$, $r \geq 1$ and $\Omega_r(\theta_0)$ be defined as in [P](1972), pages 264–265; $\mathscr{P}$ is a set of distinct probability measures with parameter space Θ (possibly infinite dimensional), θ_0 and P_0 denote the true paramerer and probability measure, ${}_*P_0$ denotes the inner measure induced by P_0 on the product spaces of all sequences $(X_1, X_2, \cdots)$, $\{V_r(\theta_0)\}$, $r \geq 1$ is a decreasing sequence of basic neighborhoods of θ_0, and $\Omega_r(\theta_0) = \Theta - V_r(\theta_0)$. For a sequence $\{X_1, X_2, \cdots\}$ of i.i.d. random variables with probability measure P_0, let $\{T_n\} = \{T_n(x_1, \cdots, x_n)\}$ be any estimating sequence such that T_n is Θ-valued and is not necessarily measurable. Refer to page 265 of [P](1972) for the definition of a strongly consistent estimating sequence.

Suppose $\mathscr{P}$ is dominated by some σ-finite measure μ, that is, $P \ll \mu$ for each P in $\mathscr{P}$. Let $f(x; \theta) = (dP/d\mu)(x)$ be a version of the Radon-Nykodym density function of P with respect to μ, where $\theta = \theta(P)$. Then the likelihood function is $L(x_1, \cdots, x_n; \theta) = \Pi_1^n f(x_i; \theta)$.

We shall define MLE and AMLE slightly different than [P](1972).

DEFINITION 2.1. An estimating sequence $\{T_n\}$ is called an MLE of θ if for all P_0 in $\mathscr{P}$,

$$_*P_0\{\sup_\Theta \log(L(x_1, \cdots, x_n; \theta)/L(x_1, \cdots, x_n; T_n)) = 0 \text{ a.e.n}\} = 1,$$

where if $\{A_n\}$ is any sequence of sets, $\{A_n \text{ a.e.n.}\}$ is the set $\lim_n \inf A_n = \bigcup_{n=1}^\infty \bigcap_{k\geq n} A_k$. (All suprema in this paper are taken with respect to θ over the indicated set.)

$\{T_n\}$ is called an AMLE of θ if

$$_*P_0\{\sup_\Theta \log(L(x_1, \cdots, x_n; \theta)/L(x_1, \cdots, x_n; T_n)) \to 0\} = 1.$$

Using the above definition, a sufficient condition for the AMLE to be strongly consistent follows immediately in the following lemma. (Compare to Lemma 1.1 in [P](1972) and note that the present sufficient condition is more general.)

LEMMA 2.1. *Let $\{T_n\}$ by any AMLE of θ. If for each $P_0 \in \mathscr{P}$, $r \geq 1$ and each θ in $\Omega_r(\theta_0)$ there exists $\theta_r(\theta)$ in $V_r(\theta_0)$ such that*

$$(2.1) \quad _*P_0\{\limsup_n \sup_{\Omega_r(\theta_0)} \log(L(x_1, \cdots, x_n; \theta)/L(x_1, \cdots, x_n; \theta_r(\theta))) < 0\} = 1,$$

then $\{T_n\}$ is strongly consistent.

PROOF. For $\omega = (x_1, x_2, \cdots)$ in the $_*P_0 = 1$ set of (2.1) there exists an $\varepsilon > 0$ sufficiently small such that for n sufficiently large,

$$\sup_{\Omega_r(\theta_0)} \log(L(x_1, \cdots, x_n; \theta)/L(x_1, \cdots, x_n; \theta_r(\theta))) \leq -\varepsilon.$$

This implies, for n sufficiently large, T_n must be in $V_r(\theta_0)$. □

From the following theorem, we see that the conditions in Lemma 2.1 are very close to being necessary.

THEOREM 2.1. *If the parameter space Θ is locally compact and $f(x; \theta)$ is an upper semicontinuous function at θ for almost all x (the exceptional set may possibly depend on θ) and all θ, then (2.1) is also necessary for any AMLE to be strongly consistent.*

PROOF. The proof follows from the facts that we can choose $V_r(\theta_0)$ to be compact neighborhoods and, of course, that an upper semicontinuous function attains its supremum on a compact subset. □

Let dominance and dominance by zero be defined as in [P](1972), page 266, Definition 1.

We are now ready to present a sufficient condition for the strong consistency of all AMLEs.

THEOREM 2.2. *If for each P_0 in $\mathscr{P}$, $r \geq 1$, and θ in $\Omega_r(\theta_0)$ there exists $\theta_r(\theta)$ in $V_r(\theta_0)$ such that $g_r(x, \theta) = \log[f(x, \theta)/f(x, \theta_r(\theta))]$ is dominated by zero on $\Omega_r(\theta_0)$ with respect to P_0, then any AMLE is consistent.*

PROOF. The proof follows from our Lemma 2.1 and Theorem 2.1 of [P](1972). □

One way to verify the dominance by zero condition (for the special case of $k = 1$ as defined in Definition 1 of [P](1972)) in Theorem 2.2 will be given in the next section. Notice that Theorem 2.2 can easily be generalized to the semidominance case as Theorem 2.1 in [P](1972). Since our examples in Sections 4 and 5 use only dominance by zero, we shall restrict Theorem 2.2 only to the case of dominance by zero and refer the reader to [P](1972) or Wang (1983) for the more general treatment.

Since Perlman's conditions imply our sufficient conditions in Theorem 2.2 where $\theta_r(\theta)$ is taken to be θ_0 for all θ in $\Omega_r(\theta_0)$ and $r \geq 1$, all the previous results by the authors mentioned in Section 1 also do.

3. Some regularity conditions. We shall present a series of assumptions that will imply the dominance by zero condition in Theorem 2.2.

DEFINITION 3.1. A first countable Hausdorff space $\overline{\Theta}$ is a compactification of Θ if $\overline{\Theta}$ is compact and Θ is a topological subspace of $\overline{\Theta}$.

In many nonparametric cases (cf. Sections 4 and 5), for θ in $\overline{\Theta}$, the definition of P_θ can be extended naturally. Note that Definition 3.1 is weaker than the compactification defined by Kiefer and Wolfowitz (1956) or Bahadur (1967). However, in many cases (e.g., in all the examples of Sections 4 and 5), if there exists a compactification as in Definition 3.1, there also exist compactifications of the type defined by the above authors.

We shall impose the following assumptions on $\mathscr{P}$:

ASSUMPTION 1. There exists a compactification of Θ, say $\overline{\Theta}$, which is separable.

Under Assumption 1, for any θ in $\overline{\Theta}$ and $r \geq 1$, let $V_r(\theta)$ be a sequence of decreasing open neighborhoods of θ in $\overline{\Theta}$, and let $\Omega_r(\theta)$ be its complement in $\overline{\Theta}$ (that is, $\Omega_r(\theta) = \overline{\Theta} - V_r(\theta)$). Therefore $\Omega_r(\theta)$ is a compact subset of $\overline{\Theta}$.

For definition of local dominance, refer to [P](1972) page 271.

ASSUMPTION 2. For any θ_0 in Θ and $r \geq 1$, there exists a function θ_r: $\overline{\Theta} \to V_r(\theta_0)$ such that (a) $\log[f_\theta(x)/f_{\theta_r(\theta)}(x)]$ is locally dominated on $\overline{\Theta}$ with respect to P_{θ_0} and (b) $\theta_r(\theta)$ is in Θ if θ is in Θ.

Note that it follows from Theorem 2.3(ii) of [P](1972) that Assumptions 1 and 2 imply that $\log[f_\theta(x)/f_{\theta_r(\theta)}(x)]$ is dominated on $\overline{\Theta}$ with respect to P_{θ_0}.

ASSUMPTION 3. For any θ_0 in Θ, if θ in $\overline{\Theta}$ is different from θ_0 and $r \geq 1$, then the functions θ_r obtained in Assumption 2 satisfy $E_{\theta_0}\log[f_\theta(x)/f_{\theta_r(\theta)}(x)] < 0$.

ASSUMPTION 4. For any θ_0 in Θ, θ in $\overline{\Theta}$, and $r \geq 1$, $\log[f_\theta(x)/f_{\theta_r(\theta)}(x)]$ is lower semicontinuous at θ except for x in a μ-null set which is indepenent of θ.

ASSUMPTION 5. For any θ_0 in Θ, θ in $\overline{\Theta}$, and $r \geq 1$, $\log[f_\theta(x)/f_{\theta_r(\theta)}(x)]$ is upper semicontinuous at θ except for x in a μ-null set possibly depending on θ.

THEOREM 3.1. *Under Assumptions* 1 *through* 5, *any AMLE in* Θ *is strongly consistent.*

PROOF. Let θ_0 be any point in Θ and $r \geq 1$. Assumption 4 and the separability of $\overline{\Theta}$ imply that, for any open subset U of $\overline{\Theta}$,

$$\sup_{\theta \in U}\log[f_\theta(x)/f_{\theta_r(\theta)}(x)]$$

is a measurable function. Assumption 5 now implies that for any θ^* in $\overline{\Theta}$,

$$\sup_{V_h(\theta^*)}\log[f_\theta(x)/f_{\theta_r(\theta)}(x)] \downarrow \log[f_{\theta^*}(x)/f_{\theta_r(\theta^*)}(x)] \quad \text{as} \quad h \to \infty,$$

where $\downarrow$ means "decreasingly converges to." It then follows from Assumptions 2 and 3, that for any θ^* in $\Omega_r(\theta_0)$, there exists $N \geq 1$ such that

$$E_{\theta_0}\sup_{V_N(\theta^*)}\log[f_\theta(x)/f_{\theta_r(\theta)}(x)] < 0.$$

Therefore $\log[f_\theta(x)/f_{\theta_r(\theta)}(x)]$ is dominated by zero on $V_N(\theta^*)$. Theorem 2.3 of [P](1972) and the fact that $\Omega_r(\theta_0)$ is compact imply that $\log[f_\theta(x)/f_{\theta_r(\theta)}(x)]$ is dominated by zero on $\Omega_r(\theta_0)$ and hence dominated by zero on $\Omega_r(\theta_0) \cap \Theta$. Note that Assumption 2 implies that $\theta_r(\theta)$ is in $V_r(\theta_0) \cap \Theta$ for θ in Θ. Theorem 2.2 now implies any AMLE in Θ is strongly consistent. □

4. Applications to some nonparametric families.

4.1 *Strong consistency of AMLEs for concave distributions.* For any distribution function F on the real line, let $\alpha_0(F)$, $\alpha_1(F)$ denote the left- and right-hand endpoint of its interval of support. That is, $\alpha_0(F) = \inf\{x: F(x) > 0\}$, $\alpha_1(F) = \sup\{x: F(x) < 1\}$.

A distribution function F is said to be concave (convex) if it is concave (convex) on its interval of support $[\alpha_0(F), \alpha_1(F)]$. Let $\mathscr{P}_1$ be the family of all concave continuous distributions on $[0, \infty)$. The MLE of this family exists and was first found by Grenander (1956) to be the least concave majorant C_n of the empirical distribution function F_n of n independent observations according to some F in $\mathscr{P}_1$. As mentioned in Section 1, while the consistency of the MLE in $\mathscr{P}_1$ has been proved by Robertson (1967), Prakasa Rao (1969) and Marshall (1970), it cannot be deduced from any of the general analytic approaches mentioned in Section 1 because $\log[f_\theta(x)/f_{\theta_0}(x)]$ is not locally dominated. Here we show the strong consistency of the MLE by the method of Sections 2 and 3. The new proof does not require the knowledge of either the existence or the shape of the MLE.

Let $\mathscr{P}_1 = \{F: \alpha_0(F) = 0, F$ is continuous and concave on $[0, \infty)\}$. Any F in $\mathscr{P}_1$ is absolutely continuous; denote its density function by f. We can choose f to be right continuous and nonincreasing on $[0, \alpha)$; let Θ_1 be the set of all such f, so that $\theta: \mathscr{P}_1 \to \Theta_1$ defined by $\theta(F) = f$ is a parameterization of $\mathscr{P}_1$. On Θ_1 define the metric $d(f, g) = \inf\{h: f(x + h) - h < g(x)$ for all $x \geq 0$, and $g(x) < f(x - h) + h$ for all $x \geq h\}$; d is a metric similar to the Levy distance for distribution functions (see Gnedenko and Kolmogorov, 1954, pages 33–37).

We shall now enlarge the parameter space to $\overline{\Theta}_1$, where $\overline{\Theta}_1$ is the set of all nonincreasing, right continuous subdensity functions on $[0, \infty)$. Note that f is a subdensity function on $[0, \infty)$, if and only if $f(x) \geq 0$ for all x in $[0, \infty)$ and $\int_0^\infty f(x)\,dx \leq 1$. Then $F(x) = \int_0^x f(y)\,dy$ is the corresponding subdistribution function. For any f in $\overline{\Theta}_1$, the support of f is the set on which f is nonzero. Let $\alpha_0(f)$ and $\alpha_1(f)$ be the left- and right-hand endpoint of its interval of support. If f is identically zero, take $\alpha_0(f) = \alpha_1(f) = 0$. Note that $\alpha_0(f) = 0$ for all f in Θ_1 and $\alpha_0(f) = \alpha_0(F)$ for all f in Θ_1. It is obvious that both Θ_1 and $\overline{\Theta}_1$ are convex sets; that is, $\lambda f + (1 - \lambda)g$ is in Θ_1 ($\overline{\Theta}_1$) for any f, g in Θ_1 ($\overline{\Theta}_1$) and $0 \leq \lambda \leq 1$. The definition of d can be extended to $\overline{\Theta}_1$ and convergence on this metric space $(\overline{\Theta}_1, d)$ is similar to weak convergence of distribution functions. More specifically, we have

LEMMA 4.1. *$(\overline{\Theta}_1, d)$ is a metric space. For any f and sequence $\{f_n\}$ in $\overline{\Theta}_1$, $d(f_n, f) \to 0$, if and only if $f_n(x) \to f(x)$ at all nonzero continuity points x of f.*

PROOF. Similar to the proof in Gnedenko and Kolmogorov (1954, pages 33–37). □

Notice the difference at zero. Convergence of $\{f_n\}$ in $(\overline{\Theta}_1, d)$ to f does not imply convergence of $f_n(0)$ to $f(0)$ even though 0 may be a continuity point of f. For example, $f_n(x) = nI_{[0,1/n]}(x) \to 0$ but $f_n(0) \to \infty$.

LEMMA 4.2. *$(\overline{\Theta}_1, d)$ is a separable compact metric space.*

PROOF. Since $(\overline{\Theta}_1, d)$ is a metric space, compactness is equivalent to sequential compactness. Using the fact that $f(x) < x^{-1}$ for any f in $\overline{\Theta}_1$, the proof of sequential compactness is similar to that of Helly's extraction principle. For example, Chung (1974, page 83) provides such a proof. Therefore, $(\overline{\Theta}_1, d)$ is compact. Since a compact metric space is separable, the lemma is proved. □

Lemma 4.2 shows that $\overline{\Theta}_1$ is a compactification of Θ_1. For any f in $\overline{\Theta}_1$ and $r \geq 1$, let $V_r(f)$ denote the open ball with center f and radius r^{-1}, and $\Omega_r(f) = \overline{\Theta}_1 - V_r(f)$.

LEMMA 4.3. *For any f_0 in $\overline{\Theta}_1$ and $r \geq 1$, there exists $0 < \varepsilon < 1$ such that $(1 - \varepsilon)f_0 + \varepsilon f$ is in $V_r(f_0)$ for all f in $\overline{\Theta}_1$.*

PROOF. Choose $0 < \varepsilon < \min\{r^{-2}, 1\}$. By convexity of $\overline{\Theta}_1$, $(1 - \varepsilon)f_0 + \varepsilon f$ is in

$\overline{\Theta}_1$. For $x \geq r^{-1}$,

$$(1-\varepsilon)f_0(x) + \varepsilon f(x) \leq (1-\varepsilon)f_0(x) + \varepsilon/x < f_0(x) + \varepsilon r \leq f_0(x - r^{-1}) + r^{-1}.$$

For $x \geq 0$,

$$(1-\varepsilon)f_0(x) + \varepsilon f(x) \geq (1-\varepsilon)f_0(x + r^{-1}) \geq f_0(x + r^{-1}) - \varepsilon f_0(r^{-1})$$
$$\geq f_0(x + r^{-1}) - \varepsilon r > f_0(x + r^{-1}) - r^{-1}.$$

Hence $(1-\varepsilon)f_0 + \varepsilon f$ is in $V_r(f_0)$. □

The following lemma is an extension of the information inequality.

LEMMA 4.4. *Let f_0 be any density function. For any subdensity function f different from f_0, we have*

$$E_{f_0}\log\{f/[(1-\varepsilon)f_0 + \varepsilon f]\} < 0 \quad \textit{for all} \quad 0 \leq \varepsilon < 1.$$

PROOF. $E_{f_0}|f/f_0| \leq \int_{-\infty}^{\infty} f(x)\,dx \leq 1$.

CASE 1. $P_{f_0}(f = 0) > 0$. Then

$$E_{f_0}\log(f/[(1-\varepsilon)f_0 + \varepsilon f]) = -\infty.$$

Hence the assertion follows.

CASE 2. $P_{f_0}(f = 0) = 0$. By Jensen's inequality,

$$E_{f_0}\log(f/f_0) \leq \log E_{f_0}(f/f_0) \leq 0,$$

and the first equality sign holds only if f/f_0 is equal to a constant c with (f_0) probability one. Since f, f_0 are subdensity and density functions, respectively and, $f \neq f_0$, c must be less than 1, hence $E_{f_0}\log(f/f_0) < 0$. Because

$$E_{f_0}\log\{[(1-\varepsilon)f_0 + \varepsilon f]/f\} \geq E_{f_0}[(1-\varepsilon)\log(f_0/f) + \varepsilon\log(f/f)]$$
$$= (1-\varepsilon)E_{f_0}\log(f_0/f) > 0$$

for $0 \leq \varepsilon < 1$, the lemma follows. □

THEOREM 4.1. *Any AMLE with values in Θ_1 is strongly consistent.*

PROOF. We shall show that $\mathscr{P}_1$ satisfies Assumptions 1 through 5 in Section 3 and hence the result follows from Theorem 3.1. It may be seen that Assumption 1 results from Lemma 4.2, Lemma 4.3 and the convexity of Θ_1 gives Assumption 2 by taking $\theta_r(f) = (1-\varepsilon)f_0 + \varepsilon f$, while Assumption 3 follows from Lemma 4.4.

To show that Assumption 4 holds, let $\{f_n\}$ be any sequence in $(\overline{\Theta}_1, d)$ that converges to f. Using the right continuity property of f and Lemma 4.1, we have

$$\liminf_n f_n(x) \geq f(x) \quad \text{for all} \quad x > 0. \tag{4.1}$$

This implies

$$\liminf_n f_n(x)/[(1-\varepsilon)f_0(x)+\varepsilon f_n(x)] \geq f(x)/[(1-\varepsilon)f_0(x)+\varepsilon f(x)],$$

for all $x > 0$. Assumption 4 must hold since $\log\{f(x)/[(1-\varepsilon)f_0(x)+\varepsilon f(x)]\}$ is a lower semicontinuous function of f for all $x > 0$. Finally, Assumption 5 follows from Lemma 4.1. □

Note that the consistency of an AMLE of f_0 in Θ_1 is uniform on any compact interval on which f_0 is continuous. If $\hat{f}_n$ is an AMLE of f_0 in Θ_1, its corresponding distribution function $\bar{F}_n$ is an AMLE of F_0 in $\mathscr{P}_1$, where F_0 is the distribution function with density f_0. We have the following corollary.

COROLLARY 4.1. (i) $\sup_{-\infty<x<+\infty}|F_n(x) - F_0(x)| \to 0$ *with probability one.*

(ii) *If* f_0 *is continuous on* $[a, b]$, *then* $\sup_{x\in[a,b]}|\hat{f}_n(x) - f_0(x)| \to 0$ *with probability one.*

PROOF. (i) follows from Scheffé's Theorem; see Billingsley (1968, page 224). (ii) follows immediately. □

REMARK. If the left endpoint of support of functions in Θ_1 is not 0 but any fixed known α, i.e., $\alpha_0(F) = \alpha$ for all F in $\mathscr{P}_1$, Theorem 4.1 is still true. Now consider α fixed and known, and let $\mathscr{P}_{2,\alpha}$ denote the set of all continuous distribution functions F with $\alpha_1(F) = \alpha$, such that F is convex on $(-\infty, \alpha]$. Using a reflection argument, we have the following:

COROLLARY 4.2. *For* α *fixed and known, any AMLE in* $\mathscr{P}_{2,\alpha}$ *is strongly consistent.*

Another direct consequence of Theorem 4.1 is the consistency of AMLE in the family of distributions with decreasing failure rate, as is now shown.

DEFINITION 4.1. For any distribution function F, $H_F(x) = -\log(1 - F(x))$ is called the hazard function of F. If F has density f with respect to some σ-finite measure μ, the failure rate γ of F is defined to be $\gamma(x) = f(x)/[1 - F(x)]$ for $F(x) < 1$.

DEFINITION 4.2. A distribution function F is said to have decreasing failure rate (DFR) if the support of F is of the form $[\alpha, \infty)$, $\alpha < -\infty$, and if $H_F(x)$ is concave on $[\alpha, \infty)$.

From Definition 4.1 the failure rate is the derivative of the hazard function. Marshall and Proschan (1965) showed that a distribution function F with DFR on $[\alpha, \infty)$ is absolutely continuous except for the possibility of a discontinuity at the left end α. We shall only consider the case where F is continuous at α, hence absolutely continuous on its support.

Let $\mathscr{P}_3$ be the set of all continuous distribution functions F with DFR on $[0, \infty)$. It is obvious that $\mathscr{P}_3$ is a subset of $\mathscr{P}_1$ so we can use the same parameterization of $\mathscr{P}_1$. Let Θ_3 be the set of all densities in Θ_1 with DFR. Barlow, Marshall and Proschan (1963) showed that Θ_3 is a convex set.

THEOREM 4.2. *Any AMLE of Θ_3 is strongly consistent.*

PROOF. Θ_3 is a convex subset of Θ_1. Theorem 3.1 and the proof of Theorem 4.1 together imply the result. □

REMARKS. (1) The MLE for $\mathscr{P}_3$ was found by Grenander (1956) to be the distribution corresponding to the least concave majorant of the empirical hazard function and proved to be consistent by Marshall and Proschan (1965). Theorem 4.2 provides another way of proving the consistency of MLE without knowing its explicit form.

(2) If $\hat{f}_n$ is an AMLE of f_0 in Θ_3, let $\hat{\gamma}_n$ be the corresponding failure rate function for $\hat{f}_n$; then $\hat{\gamma}_n$ is an AMLE of γ_0, the true failure rate. Since Corollary 4.1 is still true, $\hat{\gamma}_n$ is also a strongly consistent estimator of γ_0 for the topology induced by the Lévy distance on the space of decreasing functions on $[0, \infty)$.

The method used in this section to prove consistency of AMLEs can also be applied to any convex parameter space which satisfies the regularity assumptions in Section 3.

4.2 *Strong consistency of AMLE for IFR distributions.* In this section, we shall prove the strong consistency of AMLE for the family $\mathscr{P}_4^M$ which consists of all IFR continuous distributions F with failure rate uniformly bounded by M on $[0, \infty)$ and $F(0-) = 0$. The MLE in $\mathscr{P}_4^M$ exists and was proved by Marshall and Proschan (1965) to be consistent. We shall give another proof of its consistency without knowing either its form or existence. Let us first define IFR distributions.

DEFINITION 4.3. A distribution function F is said to have increasing failure rate (IFR), if the support of F is an interval $[\alpha_0(F), \alpha_1(F)]$ and its hazard function (cf. Definition 4.1) is convex on this interval.

REMARK. Contrary to the DFR distributions, an IFR distribution is absolutely continuous except possibly for a discontinuity at the right endpoint $\alpha_1(F)$.

Let F be a continuous distribution function with IFR on $[0, \infty)$. The above remark implies that F is absolutely continuous. Let f be its density function, and let γ be its failure rate as given in Definition 4.2. Then γ is nondecreasing and can be taken to be left continuous. We shall also assume that f is left continuous. Hence $\mathscr{P}_4^M = \{F: F \text{ has IFR on } [\alpha_0(F), \infty)\ \alpha_0(F) \geq 0 \text{ and } \gamma \leq M\}$. Let Θ_4^M be the set of all nondecreasing left continuous functions γ on $[0, \infty)$ such that $\gamma \leq M$ and γ is not identically zero. There is a one-to-one correspondence between $\mathscr{P}_4^M$ and Θ_4^M. For any F in $\mathscr{P}_4^M$, we parameterize it by its failure rate γ, and the corresponding density function will be denoted by f_γ. That is,

$$f_\gamma(x) = \gamma(x) \cdot \exp\left\{-\int_0^x \gamma(t)\, dt\right\}.$$

Let $\overline{\Theta}_4^M = \Theta_4^M \cup \{0\}$ where 0 is the zero function. For convenience of exposition, from now on we shall use $\overline{\Theta}_4$, Θ_4 and $\mathscr{P}_4$ to denote $\overline{\Theta}_4^M$, Θ_4^M and $\mathscr{P}_4^M$, respectively. It is obvious that both $\overline{\Theta}_4$ and Θ_4 are convex sets. For any γ_1, γ_2 in $\overline{\Theta}_4$, we define $d(\gamma_1, \gamma_2) = \inf\{h\colon \gamma_2(x - h) - h < \gamma_1(x)$ for all $x \geq h$, and $\gamma_1(x) < \gamma_2(x + h) + h$ for all $x \geq 0\}$. Just as in Section 4.1, d is similar to the Lévy distance and the weak convergence property also carries over to $\overline{\Theta}_4$ due to the fact that $0 \leq \gamma \leq M$ for any γ in $\overline{\Theta}_4$. Therefore Lemmas 4.1 and 4.2 also apply to $(\overline{\Theta}_4, d)$ and we have

LEMMA 4.5. *$(\overline{\Theta}_4, d)$ is a separable compact metric space. For any γ and sequence $\{\gamma_n\}$ in $\overline{\Theta}_4$, $d(\gamma_n, \gamma) \to 0$ if and only if $\gamma_n(x) \to \gamma(x)$ at all nonzero continuity points x of γ.*

For any γ in $\overline{\Theta}_4$ and $h \geq 1$, let $V_h(\gamma)$ denote the open ball with center γ and radius h^{-1}, and $\Omega_h(\gamma) = \overline{\Theta}_4 - V_h(\gamma)$ be its complement in $\overline{\Theta}_4$. For fixed γ_0 in Θ_4, for any γ in $\overline{\Theta}_4$ and any $0 \leq \varepsilon \leq 1$, we shall denote $f_{(\gamma,\varepsilon)}(x)$ and $F_{(r,\,\varepsilon)}(x)$ to be the density and distribution functions corresponding to the failure rate $(1 - \varepsilon)\gamma_0(x) + \varepsilon\gamma(x)$.

LEMMA 4.6. *For any γ_0 in Θ_4 and $h \geq 1$, there exists $0 < \varepsilon < 1$ such that*

(a) *$(1 - \varepsilon)\gamma_0 + \varepsilon\gamma$ is in $V_h(\gamma_0) \cap \Theta_4$ for all γ in $\overline{\Theta}_4$, and*
(b) *$f_\gamma(x)/f_{(\gamma,\varepsilon)}(x)$ is dominated on $\overline{\Theta}_4$ with respect to P_{γ_0}.*

PROOF. Choose $0 < \varepsilon < \min\{(hM)^{-1}, 1\}$. Mimicking the proof of Lemma 4.3, it can be shown that $(1 - \varepsilon)\gamma_0 + \varepsilon\gamma(x)$ is in $V_h(\gamma_0)$. Also $(1 - \varepsilon)\gamma_0 + \varepsilon\gamma$ is not identically zero since γ_0 is not. This proves (a). Next,

$$\begin{aligned} f_{(\gamma,\varepsilon)}(x) &= [(1 - \varepsilon)\gamma_0(x) + \varepsilon\gamma(x)]\exp\left\{-\int_0^x [(1 - \varepsilon)\gamma_0(t) + \varepsilon\gamma(t)\, dt\right\} \\ &\geq \varepsilon\gamma(x)\exp\left\{-\int_0^x [\gamma_0(t) + \gamma(t)]\, dt\right\} \\ &= \varepsilon f_\gamma(x)\exp\left\{-\int_0^x \gamma_0(t)]\, dt\right\}. \end{aligned}$$

Hence

$$\log[f_\gamma(x)/f_{(\gamma,\varepsilon)}(x)] \leq \int_0^x \gamma_0(t)\, dt - \log \varepsilon.$$

Notice that

$$\int_0^x \gamma_0(t)\, dt = -\log[1 - F_{\gamma_0}(x)]$$

is the hazard function corresponding to the failure rate γ_0, and

$$E_{\gamma_0}\{-\log[1 - F_{\gamma_0}(x)]\}$$
$$= \int_{-\infty}^{\infty} -\log[1 - F_{\gamma_0}(x)]\, dF_{\gamma_0}(x) = \int_0^1 -\log(1 - u)\, du = 1,$$

which proves (b). □

LEMMA 4.7. *For any two distinct γ_0 and γ in Θ_4, and any $0 \le \varepsilon < 1$,*

$$E_{\gamma_0}\log[f_\gamma(x)/f_{(\gamma,\varepsilon)}(x)] < 0.$$

PROOF. From the definition of $F_{(\gamma,\varepsilon)}(x)$, we have

$$[1 - F_{(\gamma,\varepsilon)}(x)]/[1 - F_\gamma(x)] = \exp\left\{-\int_0^x (1 - \varepsilon)(\gamma_0 - \gamma)(t)\, dt\right\}, \tag{4.2}$$

and

$$[1 - F_{(\gamma,\varepsilon)}(x)]/[1 - F_{\gamma_0}(x)] = \exp\left\{\int_0^x \varepsilon(\gamma_0 - \gamma)(t)\, dt\right\}. \tag{4.3}$$

If $P_{\gamma_0}(f_\gamma = 0) > 0$, $E_{\gamma_0}\log[f_\gamma(x)/f_{(\gamma,\varepsilon)}(x)] = -\infty$, hence the assertion follows. In the case $P_{\gamma_0}(f_\gamma = 0) = 0$, we have

$$\begin{aligned}
E_{\gamma_0}[\log[f_{(\gamma,\varepsilon)}/f_\gamma]] &= E_{\gamma_0}\log\{[(1-\varepsilon)\gamma_0 + \varepsilon\gamma][1 - F_{(\gamma,\varepsilon)}]/[\gamma(1 - F_\gamma)]\} \\
&\ge (1-\varepsilon)E_{\gamma_0}\log\{\gamma_0[1 - F_{(\gamma,\varepsilon)}]/[\gamma(1 - F_\gamma)]\} \\
&\quad + \varepsilon E_{\gamma_0}\log\{[1 - F_{(\gamma,\varepsilon)}]/(1 - F_\gamma)\} \\
&= (1-\varepsilon)E_{\gamma_0}\log\{\gamma_0(1 - F_{\gamma_0})/[\gamma(1 - F_\gamma)]\} \\
&\quad + (1-\varepsilon)E_{\gamma_0}\log\{[1 - F_{(\gamma,\varepsilon)}]/(1 - F_{\gamma_0})\} \\
&\quad + \varepsilon E_{\gamma_0}\log\{[1 - F_{(\gamma,\varepsilon)}]/(1 - F_\gamma)\} \\
&= (1-\varepsilon)E_{\gamma_0}\log(f_{\gamma_0}/f_\gamma) + (1-\varepsilon)E_{\gamma_0}\left[\int_0^x \varepsilon(\gamma_0 - \gamma)(t)\, dt\right] \\
&\quad - \varepsilon E_{\gamma_0}\left[\int_0^x (1-\varepsilon)(\gamma_0 - \gamma)(t)\, dt\right] \quad \text{(from (4.2) and (4.3))} \\
&= (1-\varepsilon)E_{\gamma_0}\log(f_{\gamma_0}/f_\gamma) > 0,
\end{aligned}$$

by the proof of Lemma 4.4. □

LEMMA 4.8. *For any fixed γ_0 in Θ_4 and $0 \le \varepsilon \le 1$, $\log\{f_\gamma(x)/f_{(\gamma,\varepsilon)}(x)\}$ is lower semicontinuous at γ for all $x > 0$ and γ in $\overline{\Theta}_4$.*

PROOF. Let $\gamma_n \to \gamma$ in $\overline{\Theta}_4$. The left continuity of γ, the fact that $\overline{\Theta}_4$ consists of nondecreasing functions, and Lemma 4.5 together imply that

$$\lim_n \inf \gamma_n(x) \ge \gamma(x) \quad \text{for all} \quad x > 0. \tag{4.4}$$

On the other hand,

$$\int_0^x (\gamma_0 - \gamma_n)(t)\, dt \to \int_0^x (\gamma_0 - \gamma)(t)\, dt \quad \text{for all } x. \tag{4.5}$$

If $\lim_n \inf f_{\gamma_n}(x) \ne 0$,

$$f_{(\gamma_n,\varepsilon)}(x)/f_{\gamma_n}(x) = [(1-\varepsilon)\gamma_0/\gamma_n + \varepsilon]\exp\left[-\int_0^x (1-\varepsilon)(\gamma_0 - \gamma_n)(t)\, dt\right].$$

By (4.4) and (4.5),

$$\limsup_n [f_{(\gamma_n,\varepsilon)}(x)/f_{\gamma_n}(x)] \le f_{(\gamma,\varepsilon)}(x)/f_\gamma(x)$$

for all x. Therefore

$$\liminf_n \log[f_{\gamma_n}(x)/f_{(\gamma_n,\varepsilon)}(x)] \ge \log[f_\gamma(x)/f_{(\gamma,\varepsilon)}(x)] \tag{4.6}$$

for all x. If $\lim_n \inf f_{\gamma_n}(x) = 0$, (4.4) and (4.5) imply that $f_\gamma(x) = 0$ and hence (4.6) follows. □

THEOREM 4.3. *Any AMLE in Θ_4 is strongly consistent.*

PROOF. We shall show that $\mathscr{P}_4$ satisfies Assumptions 1 through 5 in Section 3 and the assertion then follows from Theorem 3.1. Assumption 1 results from Lemma 4.5. Lemma 4.6 implies Assumption 2 by letting $\theta_h(\gamma) = (1-\varepsilon)\gamma_0 + \varepsilon\gamma$. Assumptions 3 and 4 follow from Lemmas 4.7 and 4.8, respectively. Lemma 4.5 and the Lebesgue Dominated Convergence Theorem imply Assumption 5. □

Let $\hat{\gamma}_n$ be an AMLE of γ_0 in Θ_4, and let $\hat{f}_n$, $\hat{F}_n$ be the corresponding density and distribution functions. Then $\hat{f}_n$ and $\hat{F}_n$ are AMLEs of the density and distribution functions, respectively. We have a result similar to Corollary 4.1.

COROLLARY 4.3. (i) $\sup_{-\infty<x<+\infty} |\hat{F}_n(x) - F_0(x)| \to 0$ *with probability one,*

(ii) *If γ_0 in Θ_4 is continuous on $[a, b]$, then*

$$\lim_n \sup_{x\in[a,b]} |\hat{\gamma}_n(x) - \gamma_0(x)| \to 0 \text{ with probability one, and}$$

$$\lim_n \sup_{x\in[a,b]} |\hat{f}_n(x) - f_0(x)| \to 0 \text{ with probability one.}$$

REMARKS. Note that we have assumed $\alpha_0(F) \ge 0$ for any F in $\mathscr{P}_4$. Such a restriction is unnecessary. However, we do need a lower bound for $\alpha_0(F)$. For $\alpha \ge -\infty$, let $\mathscr{P}_{4,\alpha}$ be the set of all continuous distribution functions F with IFR on

$[\alpha_0(F), \alpha_1(F)]$ where $\alpha_0(F) \geq \alpha$ and the failure rate γ of F is bounded above by M. Note that for F in $\mathscr{P}_{4,\alpha}$ with $\alpha > -\infty$, $\alpha_1(F)$ must be $+\infty$ and $\mathscr{P}_{4,0}$ is equal to $\mathscr{P}_4$. Using the same argument for $\mathscr{P}_4$, any AMLE in $\mathscr{P}_{4,\alpha}$ is consistent for $\alpha > -\infty$. If $\alpha = -\infty$, we can no longer use the Lebesgue Dominated Convergence Theorem, and Assumptions 4 and 5 may not be true.

5. Applications to estimators based on censored data. In the previous section, we showed the applicability of Theorem 3.1 to several nonparametric families for which the consistency of MLEs was already established. In this section, we shall show the applicability of Theorem 3.1 to some other nonparametric families for which, to our knowledge, the consistency of AMLEs has never been investigated.

5.1 *Concave lifetime distributions with censored data.* Suppose one has prior information that the lifetime distribution function F of certain items is concave, continuous and has zero as its left endpoint. That is, F is in $\mathscr{P}_1$, the family of concave distributions defined in Section 4. Let $X_1^0, \cdots, X_n^0$ be the true survival times of n such items which are censored from the right by a sequence of i.i.d. random variables $U_1, \cdots, U_n$. It is assumed that the censoring time of an item is independent of its survival time and the distribution function $U(x)$ of the censoring time is known. If we can only observe $X_i = \min(X_i^0, U_i)$, $i = 1, \cdots, n$, without knowing whether it is a censored observation or real death, then the distribution G of the observation X_i is given by $1 - G(x) = [1 - F(x)][1 - U(x)]$.

Our goal is to estimate the lifetime distribution F based on the observations $X_1, \cdots, X_n$. Since we are investigating AMLEs, let us also assume that the censoring distribution $U(x)$ is absolutely continuous with density $u(x)$. Hence $G(x)$ also has a density function, say $g(x)$, and the likelihood function is

$$L(x_1, \cdots, x_n) = \prod_{i=1}^n g(X_i).$$

Let $\mathscr{P} = \{G: (1 - G) = (1 - F)(1 - U), F \text{ in } \mathscr{P}_1\}$. We can parameterize $\mathscr{P}$ by the density of the lifetime distribution F. Let Θ_1 be defined as in Section 4. For any θ in Θ_1, we shall define F_θ to be its distribution function and G_θ to be the distribution function such that $1 - G_\theta = (1 - F_\theta)(1 - U)$. Then G_θ is in $\mathscr{P}$, and there is a one-to-one correspondence between $\mathscr{P}$ and Θ_1. Let g_θ denote the density of G_θ. We can use the same topology and compactification $\overline{\Theta}_1$ for Θ_1 as in Section 4. For fixed θ_0 and any $0 \leq \varepsilon \leq 1$, let $G_{(\theta,\varepsilon)} = G_{(1-\varepsilon)\theta_0+\varepsilon\theta}$, which has density $g_{(\theta,\varepsilon)} = (1 - \varepsilon)g_{\theta_0} + \varepsilon g_\theta$. Using this fact and Theorem 4.1, it is not hard to see that $\mathscr{P}$ satisfies all the assumptions in Section 3. We thus have

THEOREM 5.1. *Any AMLE of θ in Θ_1 is strongly consistent, and hence any AMLE of the lifetime distribution of F is also strongly consistent.*

5.2 *IFR lifetime distributions with censored data.* Let $\mathscr{P}_4^M$ and Θ_4^M be defined as in Section 4. Then any F in $\mathscr{P}_4^M$ has an increasing failure rate denoted by γ_F. Following the description in Section 5.1, we let $\mathscr{P}' = \{G: (1 - G) =$

$(1 - F)(1 - U)$, for F in $\mathscr{P}_4^M\}$, where U represents the known censoring distribution. We shall assume that U is absolutely continuous with failure rate γ_U such that $\gamma_U(x) \le c$ for all x. We can then parameterize $\mathscr{P}'$ by the failure rate of the lifetime distribution F, and for convenience of exposition, we denote the parameter space Θ_4^M by Θ_4. For any θ in Θ_4, let F_θ be the lifetime distribution function with failure rate θ, and G_θ be the distribution function in $\mathscr{P}'$ such that $1 - G_\theta = (1 - F_\theta)(1 - U)$. Any G_θ in $\mathscr{P}'$ is absolutely continuous; let g_θ denote its density. Note that G_θ has failure rate $\theta + \gamma_u$. For fixed θ_0 and any $0 \le \varepsilon \le 1$, let $F_{(\theta,\varepsilon)} = F_{(1-\varepsilon)\theta_0+\varepsilon\theta}$, $G_{(\theta,\varepsilon)} = G_{(1-\varepsilon)\theta_0+\varepsilon\theta} = (1 - F_{(\theta,\varepsilon)})(1 - U)$ and $g_{(\theta,\varepsilon)}$ be the density of $G_{(\theta,\varepsilon)}$. Then the failure rate of $G_{(\theta,\varepsilon)}$ is

$$(1 - \varepsilon)\theta_0 + \varepsilon\theta + \gamma_U = (1 - \varepsilon)(\theta_0 + \gamma_U) + \varepsilon(\theta + \gamma_U), \tag{5.1}$$

which is the convex combination of the failure rate of G_{θ_0} and G_θ.

THEOREM 5.2. *Any AMLE of θ in Θ_4 is strongly consistent, and any AMLE of the lifetime distribution F is also strongly consistent.*

PROOF. Assumptions 1, 2, 4, and 5 can be checked using arguments similar to those in the proof of Theorem 4.3. Assumption 3 can be proved as follows. By (5.1), we have

$$\begin{aligned}
\log[g_{(\theta,\varepsilon)}/g_\theta] &= \log\frac{[(1-\varepsilon)(\theta_0+\gamma_U)+\varepsilon(\theta+\gamma_U)](1-G_{(\theta,\varepsilon)})}{(\theta+\gamma_U)(1-G_\theta)}\\
&\ge (1-\varepsilon)\log\frac{(\theta_0+\gamma_U)(1-G_{(\theta,\varepsilon)})}{(\theta+\gamma_U)(1-G_\theta)}+\varepsilon\log\frac{1-G_{(\theta,\varepsilon)}}{1-G_\theta}\\
&= (1-\varepsilon)\log\frac{(\theta_0+\gamma_U)(1-G_{\theta_0})}{(\theta+\gamma_U)(1-G_\theta)}+(1-\varepsilon)\log\frac{1-G_{(\theta,\varepsilon)}}{1-G_{\theta_0}}\\
&\quad+\varepsilon\log\frac{1-G_{(\theta,\varepsilon)}}{1-G_\theta}\\
&= (1-\varepsilon)\log\frac{g_{\theta_0}}{g_\theta}+(1-\varepsilon)\log\frac{1-F_{(\theta,\varepsilon)}}{1-F_{\theta_0}}+\varepsilon\log\frac{1-F_{(\theta,\varepsilon)}}{1-F_\theta}\\
&= (1-\varepsilon)\log\frac{g_{\theta_0}}{g_\theta}+(1-\varepsilon)\varepsilon\log\frac{1-F_\theta}{1-F_{\theta_0}}+\varepsilon(1-\varepsilon)\log\frac{1-F_{\theta_0}}{1-F_\theta}\\
&= (1-\varepsilon)\log\frac{g_{\theta_0}}{g_\theta}.
\end{aligned}$$

Hence,

$$E_{\theta_0}\log[g_\theta/g_{(\theta,\varepsilon)}] \le (1 - \varepsilon)E_{\theta_0}\log[g_\theta/g_{\theta_0}] < 0.$$

The theorem now follows from Theorem 3.1. □

6. Conclusion. As mentioned in Section 4, the techniques we use in this paper may be applied to any convex parameter space with appropriate regularity assumptions. The main difficulty in applying those techniques to the nonparametric problems lies in the construction of a suitable topology on a parameter space having certain regularity properties. Once determined, however, convergence on this topological space is the same as (or implies) convergence of estimating sequences in the usual statistical sense.

While the techniques and theorems in this paper are only applied to approximate maximum likelihood estimates, they can be applied in general to approximate maximum ω estimators defined in [P](1972). Such a generalization is not repeated here.

Acknowledgements. I am grateful to Professor J. Kiefer and Professor L. Le Cam for helpful discussions at various stages of this work. I wish to thank the editor and referees for many suggestions which improved the quality of this paper.

REFERENCES

BAHADUR, R. (1967). Rates of convergence of estimates and test statistics. *Ann. Math. Statist.* **38** 303–324.

BARLOW, R. E., BARTHOLOMEW, P. J., BREMNER, J. N. and BRUNK, H. D. (1972). *Statistical Inference under Order Restrictions.* Wiley, New York.

BARLOW, R. E., MARSHALL, A. W. and PROSCHAN, F. (1963). Properties of probability distributions with monotone hazard rate. *Ann. Math. Statist.* **34** 375–389.

BILLINGSLEY, P. (1968) *Convergence of Probability Measures.* Wiley, New York.

CHUNG, K. L. (1974). *A Course in Probability Theory.* Academic, New York.

GNEDENKO, B. V. and KOLMOGOROV, A. M. (1954). *Limit Distributions for Sums of Independent Random Variables.* Addison-Wesley, Reading, Mass.

GRENANDER, U. (1956). On the theory of mortality measurement. Part II. *Skand. Akt.* **39** 125–153.

HUBER, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Proc. Fifth Berkeley Symp. Math. Statist. Probab.* **1** 221–233, Univ. California Press.

KIEFER, J. and WOLFOWITZ, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Ann. Math. Statist.* **27** 887–906.

LE CAM, L. (1953). On some asymptotic properties of maximum likelihood estimates and related Bayes estimates. *Univ. California Publ. Statist.* **1** 277–328.

MARSHALL, A. W. (1970). Discussion of Barlow and van Zwet's papers. In *Nonparametric Techniques in Statistical Inference.* (M. L. Puri, ed.), 175–176, Cambridge University Press.

MARSHALL, A. W. and PROSCHAN, F. (1965). Maximum likelihood estimation for distribution with monotone failure rate. *Ann. Math. Statist.* **36** 69–77.

PERLMAN, M. D. (1972). On the strong consistency of approximate maximum likelihood estimates. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **1** 263–281, Univ. California Press.

PRAKASA RAO, B. L. S. (1969). Estimation of a unimodal density. *Sankhyā Ser. A* **31** 23–36.

ROBERTSON, T. (1967). On estimating a density which is measurable with respect to a σ-lattice. *Ann. Math. Statist.* **38** 482–493.

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

WANG, J.-L. (1983). Strong consistency of approximate maximum likelihood estimators with applications in nonparametrics. Technical Report, University of Iowa.

DIVISION OF STATISTICS
UNIVERSITY OF CALIFORNIA, DAVIS
DAVIS, CALIFORNIA 95616