BOEING SCIENTIFIC RESEARCH LABORATORIES DOCUMENT D1-82-0599

SCIENTIFIC RESEARCH LABORATORIES

Maximum Likelihood Estimation: A Practical Theorem on Consistency of the Nonparametric Maximum Likelihood Estimates with Applications

Gordon B. Crawford

MATHEMATICS RESEARCH FEBRUARY 1967

0 9
180

D1-82-0599

MAXIMUM LIKELIHOOD ESTIMATION: A PRACTICAL THEOREM ON CONSISTENCY OF THE NONPARAMETRIC MAXIMUM LIKELIHOOD ESTIMATES WITH APPLICATIONS

by

Gordon B. Crawford

Mathematical Note No. 503

Mathematics Research Laboratory

BOEING SCIENTIFIC RESEARCH LABORATORIES

February 1967

Summary

Sufficient conditions for consistency of a nonparametric maximum likelihood estimate are given which are applicable to those problems where a class of distribution functions is specified only in terms of its graphs.

Consistency is proven and applications are given.

Introduction

In many statistical analyses there may be prior considerations which give information about the shape of a distribution function (c.f. examples 1-3) whereas there may not be sufficient information to consign the distribution to some class having a finite dimensional parameterization.

Given a class of densities and a sample of independent identically distributed random variables we may pick the most likely density without any reference to a parameterization. We intend to give conditions on the class in question which will insure that the corresponding distribution function converges to the true distribution function at points of continuity of the latter.

In addition to being a more general consideration, consistency of the estimate of the distribution function is in many ways a more natural problem than consistency of the estimates of parameters.

In the paper [7] by J. Kiefer and J. Wolfowitz there is a far-reaching paragraph on page 893 which indicates that their results can be extended to a general case of nonparametric maximum likelihood estimation of distribution functions.

In most cases of nonparametric maximum likelihood estimation that have come to our attention, (for instance [8]), application of [7] or [3] is hindered by the fact that justification of the hypothesis is more difficult than a direct proof of consistency based on the form of the estimates.

Conditions which seem to be amenable to a wide class of nonparametric estimation problems are given which are sufficient for consistency of the M.L.E.

In view of the examples, it is felt that these conditions may be more easily investigated and satisfied than perhaps even Wald's hypothesis [10] for the classical parametric estimation problem.

Under our hypothesis the correspondence between the class of densities from which we pick the most likely and the corresponding class of distribution functions may be many to one. By avoiding the requirement that this relation be one to one we have included (in example 2) consistency in the class of distribution functions having unimodal densities.

Since the M.L.E. in this class is, at some observation, neither right nor left continuous, there will almost surely be no M.L.E. if we consider only a class of densities which are in one-to-one correspondence with their distribution functions. In this context, our result is slightly more general than [7] but in that paper Kiefer and Wolfowitz do not hypothesize that the distribution functions must be absolutely continuous with respect to some fixed underlying sigma-finite measure as is hypothesized here. (Condition 1)

<u>Notation</u>

In the sequel, a distribution function $F(\cdot)$ is a monotone non-decreasing function with range in $[0,1]$.

A proper distribution function $F(\cdot)$ is a right continuous distribution function with range 1.

Given a distribution function $F(\cdot)$, there is, of course, the corresponding measure $F\{\cdot\}$ on R with the property that $F\{x : a < x \leq b\} = F(b) - F(a)$ whenever a,b are continuity points of $F(\cdot)$. We make no real distinction here, except to reserve the notation $(\cdot)$ for point functions and $\{\cdot\}$ for set functions.

We assume that we are given a class $\mathcal{F}$ of proper distribution functions, and we have the *a priori* knowledge that the distribution function of a random variable X to be observed is in $\mathcal{F}$. We give $\mathcal{F}$ the topology of weak convergence: a sequence $F_n(\cdot)$, $n=1,2,\ldots$ converges to $F(\cdot)$ whenever $F_n(x)$ converges to $F(x)$ for all x in the continuity set of $F(\cdot)$.

Let $\mathcal{G}$ denote a compactification of $\mathcal{F}$ whose elements are distribution functions, again we give $\mathcal{G}$ the topology of weak convergence.

(If $\mathcal{G}$ contains two or more distribution functions which are identical except on the discontinuity set of one, then $\mathcal{G}$ is not a Hausdorff space, but by identifying all such distribution functions we may form a quotient space which is Hausdorff, and in this quotient space the topology of weak convergence is the same as the metric topology given by:

$$\rho(F,G) = \int e^{-|x|}|F(x)-G(x)|dx.$$

It follows that the decreasing sequence of neighborhoods used in Lemma 1 and the sequel always exists. In the following it is tacitly assumed that when necessary we have in mind this quotient space; that "distinct" distribution functions do not differ only on their discontinuity sets; and $\bigcap_n \vartheta_n = G$ means that $\bigcap_n \vartheta_n$ does not contain any distribution function which is distinct from G.)

Condition 1. All elements of $\mathcal{G}$ are absolutely continuous with respect to some fixed σ-finite measure $\mu\{\cdot\}$ on R.

Let $\mathcal{E}$ be a set of densities with the property that every element of $\mathcal{E}$ is the density of an element of $\mathcal{G}$, and every element of $\mathcal{G}$ has at least one density in $\mathcal{E}$. Our maximum likelihood estimate will be determined by finding the density $f \in \mathcal{E}$ that maximizes the likelihood of a given sample $X_1,\ldots,X_n = \underline{X}_n$. To facilitate discussion of the estimate we assume that $\mathcal{E}$ and $\mathcal{G}$ are so chosen that $\mathcal{E}$ will almost surely contain a $g(\cdot)$ which maximizes the likelihood

$$L(g,\underline{X}_n) = \frac{1}{n} \sum_{i=1}^{n} \ell g\ g(X_i)$$

where the X_i are independent identically distributed random variables whose cumulative distribution function is an element of $\mathcal{G}$.

Condition 2. There exists a countable subset g_m, $m=1,2,\ldots,$ of $\mathcal{E}$ with the property that for all small neighborhoods ϑ of G:

$$\sup_{h:H\in\vartheta} h(x) = \sup_{g_m:G_m\in\vartheta} g_m(x) \qquad \text{a.e. } \mu\{\cdot\}.$$

Let $s(\cdot,\vartheta)$ denote the supremum function:

$$s(x,\vartheta) = \sup_{h:H\in\vartheta} h(x).$$

Then Condition 2 assures the measurability of $s(x,\vartheta)$. We also require

Condition 3. Given F in $\mathcal{F}$, G in $\mathcal{G}$, there exists a small neighborhood ϑ of G such that the function

$$\ell g\ s(x,\vartheta)$$

is bounded above by a function which is integrable with respect to $F\{\cdot\}$.

Main Theorem and Lemmas

Lemma 1. Let ϑ_n be a decreasing sequence of neighborhoods of G in $\mathcal{G}$ such that $\cap\vartheta_n = G$. Then, if $g \in \mathcal{E}$ is any density of G, we have

$$\text{(i)} \qquad \text{Lim}_n E_F[\ell g\ s(x,\vartheta_n) - \ell g\ g(x)] = 0.$$

Proof. Using Condition 3 we have:

$$\text{Lim}_n E_F[\ell g\ s(x,\vartheta_n) - \ell g\ g(x)] =$$

$$E_F[\text{lim}_n(\ell g\ s(x,\vartheta_n) - \ell g\ g(x))].$$

We intend to show that the function in round brackets above converges in F measure to zero. It follows (Halmos, p. 91) that it is convergent a.e. (with respect to F) to the zero function, since it is decreasing in n; hence the limit (i) is zero, as was to be shown.

To show

$$\text{Lim}_n F\{x : \ell g\ s(x,\vartheta_n) - \ell g\ g(x) \geq \varepsilon\} = 0,$$

define (with reference to the countable subset $g_m(\cdot)$, m=1,2,..., of $\mathcal{E}$),

$$B_{m_n} = \begin{cases} \{x : \ell g\ g_m(x) - \ell g\ g(x) > \varepsilon/2\} & \text{if } G_m(\cdot) \in \vartheta_n \\ \phi & \text{otherwise.} \end{cases}$$

Then

$$\text{Lim}_n F\{x : \ell g\ s(x,\vartheta_n) - \ell g\ g(x) \geq \varepsilon\} \leq \text{Lim}_n F\{\bigcup_m B_{m_n}\}\ .$$

But $F\{\cdot\}$ is finite, and B_{m_n}, $n=1,2,...$ is a decreasing nest of sets with

$$\text{Lim}_n F\{B_{m_n}\} = 0,$$

therefore

$$\text{Lim}_n F\{\bigcup_m B_{m_n}\} = 0. \quad \blacksquare$$

Lemma 2. If $F \varepsilon \mathcal{G}$ and $G \varepsilon \mathcal{G}$ are distinct elements of $\mathcal{G}$ and f and g are any two corresponding densities in $\mathcal{E}$, then

$$E_F[\ell g\ g(x) - \ell g\ f(x)] < 0.$$

Proof. Let f have support S, then

$$E_F\ \ell g[g(x)/f(x)] \leq \ell g\ E_F g(x)/f(x) = \ell g \int_S g(x)\mu(dx) \leq 0.$$

Moreover, $E_F\ \ell g[g(x)/f(x)] < \ell g\ E_F g(x)/f(x)$ unless on S the real valued random variable $g(x)/f(x)$ is almost surely equal to some constant c; but $c \leq 1$ since $\int f(x)\mu(dx) = 1$ and $\int g(x)\mu(dx) \leq 1$; and $c \neq 1$ since F and G are distinct; hence either $E_F\ \ell g\ g(x)/f(x) < \ell g\ E_F g(x)/f(x)$ or $\ell g \int_S g(x)\mu(dx) = \ell g\ c < 0$. $\blacksquare$

Theorem 1. Let G in $\mathcal{G}$, $F \varepsilon \mathcal{G}$, distinct, and let $f \varepsilon \mathcal{E}$ be a density of F. If $X_1, X_2, ..., X_n = \underline{X}_n$ is a sample with distribution function F, then there exists a neighborhood ϑ of G such that

$$\text{Lim sup}_n \left[\sup_{h:H\varepsilon\vartheta} L(h,\underline{X}_n) - L(f,\underline{X}_n)\right] < 0$$

with F probability 1.

Proof. Given G not equal to F there exists (by Lemma 1 and 2) a small neighborhood ϑ of G such that

$$(ii) \quad \mathcal{E}_F[\ell g\ s(x,\vartheta) - \ell g\ f(x)] =$$
$$\mathcal{E}_F[(\ell g\ s(x,\vartheta) - \ell g\ g(x)) + (\ell g\ g(x) - \ell g\ f(x))] < 0.$$

But

$$\sup_{h:H\epsilon\vartheta} L(h,\underline{X}_n) - L(f,\underline{X}_n) \leq \frac{1}{n}\Sigma(\ell g\ s(x_i,\vartheta) - \ell g\ f(x_i)),$$

and by the law of large numbers the latter quantity converges to the negative expression (ii).

Corollary to Theorem 1. Let $X_1,X_2,\ldots,$ and f be as in Theorem 1. Let D be any closed set not containing F. Then

$$\text{Lim sup}_n\left[\sup_{h:H\epsilon D} L(h,\underline{X}_n) - L(f,\underline{X}_n)\right] < 0$$

with F probability 1.

Proof. By Theorem 1, any G in D can be covered by an open neighborhood U_G with the property that

$$\text{Lim sup}_n\left[\sup_{h:H\epsilon U_G} L(h,\underline{X}_n) - L(f,\underline{X}_n)\right] < 0.$$

From the open cover $\{U_G,\ G\ \epsilon\ D\}$ of D, let $U_1,\ldots,U_m$ be a finite subcover. Then with F probability 1,

$$\text{Lim sup}_n\left[\sup_{h:H\epsilon D} L(h,\underline{X}_n) - L(f,X_n)\right] \leq$$

$$\text{Lim sup}_n\left\{\max_{1\leq i\leq m}\left[\sup_{h:H\epsilon U_i} L(h,\underline{X}_n)\right]\right\} = \max_{1\leq i\leq m}\ \text{Lim sup}_n\left[\sup_{h:H\epsilon U_i} L(h,\underline{X}_n) - L(f,\underline{X}_n)\right] > 0$$

as was to be shown.

<u>Theorem 2</u>. (Consistency of Maximum Likelihood Estimate). Let $f(\cdot;\underline{X}_n)$ be a point of $\mathcal{E}$ depending on the random variable $X_1,\ldots,X_n = \underline{X}_n$ such that for all f in $\mathcal{E}$ and some fixed $c > 0$,

$$\frac{\prod_1^n f(X_i;\underline{X}_n)}{\prod_1^n f(X_i)} \geq c.$$

Then $F(\cdot,\underline{X}_n)$ converges to the distribution function $F(\cdot)$ of $X_1,\ldots,X_n$ with F probability 1.

<u>Proof</u>. For notational simplicity, write $F_n(\cdot) = F(\cdot,\underline{X}_n)$, $f_n(\cdot) = f(\cdot,\underline{X}_n)$; if

$$\frac{\prod_1^n f_n(x_i)}{\prod_1^n f(x_i)} \geq c > 0,$$

then

$$\text{(2.1)} \qquad \operatorname{Lim\,sup} \left[\prod_1^n \frac{f_n(x_i)}{f(x_i)}\right]^{-1/n} \geq 1,$$

and therefore

$$\operatorname{Lim\,sup}_n [L(f_n,\underline{X}_n) - L(f,\underline{X}_n)] \geq 0.$$

Let U be any open neighborhood of F. If F_n is outside of U infinitely often, then

$$\operatorname{Lim\,sup}_n \left[\sup_{h:H\in\mathcal{G}\sim U} L(h,\bar{x}_n) - L(f,\bar{x}_n)\right] \geq 0.$$

By the corollary to Theorem 1, this can occur only on a set of measure zero. Now, let U_i be a decreasing sequence of neighborhoods of F whose intersection is F. Corresponding to each U_i there is, by the corollary, an event S_i of F measure zero such that on the complement of S_i, F_n is eventually in U_i. It follows that on the complement of $\bigcup_{i=1}^{\infty} S_i$, F_n is eventually in every neighborhood of F. Thus F_n coverges to F with F probability 1, as was to be shown. ▌

Examples

Example 1. (c.f. [9]). Let $\mathcal{G}^m$ be the class of distribution functions on $[0,+\infty)$ which passes a Radon-Nykodymn derivative which is nonincreasing and bounded above by m, and let $\mathcal{G} = \mathcal{G}^{+\infty}$. It follows from the form of the estimate and considerations of truncated data that if $G \varepsilon \mathcal{G}$, and t^* is fixed, $m > g(t^*)$, then consistency of the estimate in $\mathcal{G}^m$ implies consistency of the estimate in $\mathcal{G}$ at the point t^*; but t^* is arbitrary.

We will show how to compute the estimate in $\mathcal{G}$ and prove consistency in $\mathcal{G}^m$; thus establishing consistency in $\mathcal{G}$.

Computation of the M.L.E. in $\mathcal{G}$. Given a sample $X_1 \leq \cdots \leq X_n$, it is clear that the M.L.E. $\hat{f}_n(\cdot)$ of the density will put as little mass between observations as is consistent with the hypothesis that $\hat{f}_n(\cdot)$ be decreasing. Thus $\hat{f}_n(\cdot)$ will be a left continuous step function with heights a_i, subject to:

(i) $a_1 \geq a_2 \geq \cdots \geq a_n$,

(ii) $\sum_{1}^{n} a_i(X_i - X_{i-1}) \leq 1$,

(letting $X_0 = 0$).

If we consider maximizing the likelihood $\prod_1^n \tilde{a}_i$ subject only to the requirement (ii) we see that the maximum occurs when

$$\tilde{a}_i(X_i - X_{i-1}) = \tilde{a}_n(X_n - X_{n-1}) \quad \text{all} \quad i,$$

hence

$$\tilde{a}_i(X_i - X_{i-1}) = 1/n,$$

$$\tilde{a}_i = 1/n(X_i - X_{i-1}).$$

The collection of a_i's which maximizes $\prod_1^n a_i$ subject to the constraints (i) and (ii) may be obtained from the sequence $\tilde{a}_i$ by a direct application of [1]:

$$a_i = n^{-1}\left\{\max_{v \geq i} \min_{u \leq i-1} (v-u)[X_v - X_u]^{-1}\right\}.$$

Thus

$$\hat{f}_n(y) = \frac{1}{n}\left\{\max_{v \geq i} \min_{u \leq i-1} \frac{v-u}{X_v - X_u}\right\}, \quad X_{i-1} < y \leq X_i.$$

<u>Consistency of the M.L.E. in $\mathcal{G}^m$.</u> If $\{f_n(\cdot)\}$ is a sequence of densities in $\mathcal{G}^m$, then by the Helly weak compactness theorem there exists a convergent subsequence, call it $f_{m_n}(\cdot)$ which converges to a density $f(\cdot)$ in $\mathcal{G}^m$. If we could show that this implies convergence of the subsequence $F_{m_n}(\cdot)$ of $\mathcal{G}^m$ to the point $F(\cdot)$ of $\mathcal{G}^m$, then we would have shown that $\mathcal{G}^m$ is compact. Take x fixed; it suffices to show

$$\lim_n \int_0^x f_{m_n}(y)dy = \int_0^x \lim_n f_{m_n}(y)dy.$$

This interchange follows from the bounded convergence theorem, since any $f(\cdot)$ in $\mathcal{G}^m$ is less than or equal to

$$h(y) = \begin{cases} m & 0 \leq y \leq 1/m \\ 1/y & y \geq 1/m \end{cases}$$

and $h(\cdot)$ is integrable on $(0,x)$.

Condition 1 is automatic; μ = Lebesgue measure.

Condition 2. Let $\{g_m(\cdot)\}$ be the densities in $\mathcal{G}^m$ which are "rational step functions", i.e., step functions with jumps at the rationals and rational values.

Condition 3. $\ell g\ s(x,v) \leq \ell g\ m$, which is integrable with respect to $F(\cdot)$.

Thus, consistency follows from the theorem on consistency of the M.L.E.

Example 2. (c.f. [9]). Let $\mathcal{H}^m$ be the class of distribution functions which passes a Radon-Nikodymn derivative $f(\cdot)$ which is bounded above by m and is unimodal, i.e., there exists some fixed t_f such that if $t_1 > t_2 > t_f$ or $t_f > t_2 > t_1$, then $f(t_2) \geq f(t_1)$. (Any such t_f will be referred to as a turning point.)

As in Example 1, consistency of the M.L.E. in the class $\mathcal{H} = \bigcup_{m<+\infty} \mathcal{H}^m$ follows from consistency of the M.L.E. in $\mathcal{H}^m$.

Computing the M.L.E. in $\mathcal{H}^m$. As before, to maximize the likelihood at the observations it is necessary to make the M.L.E. of the density as small as possible between observations subject to the condition that it be unimodal. Thus we can make the following statements about the M.L.E.:

Some observation X_j will be a turning point. At this observation the M.L.E. $\hat{f}_n(\cdot)$ will have the maximum permissible value m, since we may take $\hat{f}_n(\cdot)$ to be discontinuous at this X_j.

To the right of X_j, $\hat{f}_n(\cdot)$ will be a left continuous step function and will be zero to the right of X_n.

To the left of X_j, $\hat{f}_n(\cdot)$ will be a right continuous step function and will be zero to the left of x_1.

If we knew the turning point X_j and knew that $\hat{f}_n(\cdot)$ had mass δ, $0 \leq \delta \leq 1$, to the left of X_j, then we could proceed as before:

$$y < X_j: \quad f_{n_j}(y) = \frac{\delta}{j-1}\left\{\min_{v \geq i+1}\ \max_{u \leq i} \frac{v-u}{X_j - X_u}\right\}, \quad X_i \leq y < X_{i+1}$$

$$y > X_j: \quad f_{n_j}(y) = \frac{1-\delta}{n-j}\left\{\max_{v \geq i+1}\ \min_{u \leq i} \frac{v-u}{X_v - X_u}\right\}; \quad X_i < y \leq X_{i+1}.$$

Keeping X_j fixed and writing the likelihood as a function δ:

$$\prod_{i=1}^{n} f_{n_j}(X_i) = \delta^{j-1}(1-\delta)^{n-j} \cdot \psi(X_1,\ldots,X_n)$$

where $\psi(\cdot)$ is independent of δ. Thus the choice of δ which maximizes the likelihood based on the assumption that X_j is the turning point is $\delta = j-1/n-1$. Hence the M.L.E. $\hat{f}_n(\cdot)$ is equal to that $f_{n_j}(\cdot)$ for which the likelihood (using the optimal choice of δ) is a maximum over all choices of j.

Consistency of the M.L.E. in $\mathcal{H}^m$. Again, by considerations similar to the Helly weak compactness theorem, we may show that the collection of unimodal densities of $\mathcal{H}^m$ is weakly compact. (Recall that $\mathcal{H}^m$ will contain other than proper distribution functions.) What remains is to show that if $F_n(\cdot)$, $F(\cdot)$ are in $\mathcal{H}^m$, and $f_n(x) \to f(x)$ at all points of continuity of $f(\cdot)$, then $F_n(x) \to F(x)$.

Let t_n be a turning point of $f_n(\cdot)$, then if necessary we may take a subsequence t_{n_m} which converges to some point t_0, $-\infty \leq t_0 \leq +\infty$. For all large n, $t_n \varepsilon (t_0-1, t_0+1)$, thus as before, for sufficiently large n, $f_n(y) \leq h(y)$, where

$$
\begin{aligned}
h(y) &= m, \quad t_0 - 1 - 1/m \leq y \leq t_0 + 1 + 1/m, \\
&= 1/y-(t_0+1), \quad t_0 + 1 + 1/m \leq y \\
&= 1/(t_0-1)-y, \quad y \leq t_0 - 1 - 1/m.
\end{aligned}
$$

Hence for fixed, finite a, b:

$$\lim_n \int_a^b f_n(y)dy = \int_a^b \lim_n f_n(y)dy,$$

$$\lim_n F_n(b) - F_n(a) = F(b) - F(a).$$

But every element of $\mathcal{H}^m$ is the integral of a unimodal density and therefore $F_n(-\infty) = 0$, $F(-\infty) = 0$, hence $F_n(b) \to F(b)$ for all b, as was needed to show $\mathcal{H}^m$ compact.

The conditions 1, 2 and 3 are clearly satisfied; μ = Lebesque measure; $\{g_m(\cdot)\}$ may be taken as "rational step functions", and $\ell g\ s(x,\mathcal{U}) \leq \ell g\ m$, hence consistency follows from Theorem 1.

Example 3. Let $\mathcal{G}^m$ $(m < +\infty)$ denote the class of distribution functions $F(\cdot)$ with the property that $-\ell g(1-F(\cdot))$ has a Radon-Nykodymn derivative $r(\cdot)$ which is bounded above by m, and $-r(\cdot)$ is unimodal; i.e., there exists t_0, $0 \leq t_0 \leq +\infty$, such that if $t_0 \leq t_1 \leq t_2$ or if $t_2 \leq t_1 \leq t_0$, then $r(t_2) \geq r(t_1)$. (Any such t_0 will be referred to as a turning point.)

Let the class $\mathcal{G} = \mathcal{G}^{+\infty}$ be the class of distribution functions having unimodal hazard rates.

Computing the M.L.E. in $\mathcal{G}^m$. (c.f. [2]).

As in other examples of hazard rate estimation, the unrestrained M.L.E. has a hazard rate which is a step function, in this case it is of the form

$$r_n^*(\cdot) = (n-j)(x_{j+1}-x_j)^{-1}, \quad t \in (x_j, x_{j+1}).$$

The likelihood is a monotone function of these values and subject to the bounds m, ε, it is maximizing by setting

$$\tilde{r}_n(t) = (n-j)(x_{j+1}-x_j)^{-1} \wedge m, \quad t \in (x_j, x_{j+1}).$$

As before, if we knew the turning point t_0, it would follow that the M.L.E. having unimodal hazard rate is just the decreasing "Brunkized" version of the $\tilde{r}_n(t)$ to the left of t_0 and the increasing "Brunkized" version of the $\tilde{r}_n(t)$ to the right of t_0. Hence the M.L.E. in $\mathcal{G}^{m,\varepsilon}$ can be obtained by maximizing this expression over all choices of t_0.

It follows from the form of the expression (c.f. [8]) that the estimate in the class $\mathcal{G}^m$ is equal to $(\hat{r}_n(t) \wedge m)$ where $\hat{r}_n(t)$ is the hazard rate of the estimate in $\mathcal{G}$.

Consistency of the M.L.E. in $\mathcal{G}^m$

First we show that $\mathcal{G}^m$ is compact. Take $G_n(\cdot)$, $n=1,2,\ldots$ having hazard rates $\{r_n(\cdot)\}_{n=1}^{\infty}$ and turning points $\{t_n\}_{n=1}^{\infty}$. Then there is a subsequence, call it $\{t_m\}_{m=1}^{\infty}$ such that $t_m \to t_0$ $0 \leq t_0 \leq +\infty$. It follows from considerations similar to the Helly-Bray lemma that there is a subsequence of $\{r_m(\cdot)\}_{m=1}^{\infty}$, call it $\{r_\ell(\cdot)\}_{\ell=1}^{\infty}$, which converges to some nonincreasing function on $[0,t_0)$. Similarly, there is a subsequence (call it $\{r_k(\cdot)\}_{k=1}^{\infty}$) of $\{r_\ell(\cdot)\}_{\ell=1}^{\infty}$ which converges to some nondecreasing function on $[t_0,+\infty)$. It follows that $\{r_k(\cdot)\}_{k=1}^{\infty}$ converges to a unimodal hazard function which is bounded above by M; hence, (by the bounded convergence theorem) we have found a convergent subsequence of $G_n(\cdot)$ whose limit is in $\mathcal{G}^{m,\varepsilon}$, as was to be shown.

Again the conditions 1, 2, and 3 are immediate; μ : Lebesque measure; $\{g_m\}$ may be taken as densities with hazard rates which are unimodal "rational step functions", and $\ell g\ s(x,v) \leq \ell g\ m$.

Thus consistency follows from the theorem on consistency of the M.L.E.

Other examples. In the manner of examples 1-3 we could also show consistency for the class of distributions having increasing (or decreasing) hazard rates bounded above by m. As in example 1, consistency of these classes without the bound follows easily. (c.f. [8], [5]).

Following the lines of example 3 we could consistently estimate a bimodal density and compute the estimate in a similar manner. It should be mentioned that in [9] Rao gives (as the solution of a certain heat equation) the asymptotic distribution of the M.L.E. of a unimodal density with known mode.

Similarly one could prove the consistency of the M.L.E. of a distribution having convex hazard rate. An algorithm for computing such an estimate has been found in [4].

In the classical parametric estimation problems it suffices (in all those examples we have considered) to take a compactification of the parameter space and a countable dense subset of this space. The troublesome condition 3 is the same as one of Wald's hypotheses [10].

Acknowledgment

The author would like to thank Dr. Frank Proschan and Prof. Ronald Pyke for their helpful advice concerning this note and especially for their communications concerning computation of the estimates mentioned in examples 1-3.

Bibliography

[1] Brunk, H. D. On the estimation of parameters restricted by inequalities. *Ann. Math. Statist.*, *29*, pp. 437-454.

[2] Bray, T. A., Crawford, G. B. and Proschan, F. Maximum likelihood estimation of a failure rate initially decreasing and then increasing. To appear in *Proceedings of IEEE/ESD*, Spring Seminar, 1967.

[3] Crawford, G. B. and Saunders, S. C. Non-parametric maximum likelihood estimation. Boeing document D1-82-0308, October 1963.

[4] Crawford, G. B. and Saunders, S. C. Minimizing convex functions over a simplex. Boeing document D1-82-0428, April 1965.

[5] Grenander, V. On the theory of mortality measurements, Part II. *Skan. Oktuarietidskr*, *39*, pp. 125-153, 1956.

[6] Kelley, J. L. *General Topology*, Von Nostrand, 1955.

[7] Kiefer, J. and Wolfowitz, J. Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Ann. Math. Statist.*, *27*, pp. 887-906, 1956.

[8] Marshall, A. W., and Proschan, F. Maximum likelihood estimation for distributions with monotone failure rate. *Ann. Math. Statist.*, *36*, pp. 69-77, 1965.

[9] Rao, B. L. S. Prakasa. Asymptotic distributions in some non-regular statistical problems. Technical Report No. 9, Department of Statistics and Probability, Michigan State University, East Lansing, Michigan.

[10] Wald, A. Note on the consistency of the maximum likelihood estimate. *Ann. Math. Statist.*, *20*, pp. 595-601, 1949.