D.S. Newman

D1-82-0308

BOEING SCIENTIFIC RESEARCH LABORATORIES

Non-Parametric Maximum Likelihood Estimation

G. B. Crawford

S. C. Saunders

Mathematics Research

October 1963

D1-82-0308

NON-PARAMETRIC MAXIMUM LIKELIHOOD ESTIMATION

by

G. B. Crawford and S. C. Saunders

Mathematical Note No. 326

Mathematics Research Laboratory

BOEING SCIENTIFIC RESEARCH LABORATORIES

October 1963

Summary

Given that a distribution function is a member of a subclass of absolutely continuous measures, we consider the problem of non-parametric estimation, with the method of maximum likelihood, of the underlying density function of a given sample of independent identically distributed random variables. Sufficient conditions on the space of probability densities and its topology are given for the consistency of such an estimate.

1. A Topology of Pointwise Convergence. Let our random variables take values in some σ-finite measure space $(\Omega,\mathcal{B},\mu)$. Hereafter we refer to the Radon-Nikodym derivative of a measure with respect to μ as its density. By a probability measure (denoted F, G, H with or without subscripts) we mean a nonnegative Borel measure on $(\Omega,\mathcal{B})$ with $F(\Omega) \leq 1$. (Note that $F(\Omega)$ may be < 1.) We shall consider a subclass $\mathcal{E}$ of densities (denoted f, g, h, $f = dF/d\mu$, etc.) of absolutely continuous probability measures, and give $\mathcal{E}$ a topology of pointwise convergence. Specifically, given an element $f \in \mathcal{E}$, we assume there exists a fixed set $B_f \in \mathcal{B}$, called the exceptional set of f, such that $\mu(B_f) = 0$ and a net $\{f_\alpha, \alpha \in A\}$ in $\mathcal{E}$ converges to f whenever

$$\underline{\mathrm{Lim}}_\alpha B_{f_\alpha} = \bigcup_{\alpha\in A} \bigcap_{\beta>\alpha} B_{f_\beta} \subset B_f$$

and

$$\mathrm{Lim}_\alpha f_\alpha(x) = f(x) \text{ all } x \text{ not in } B_f.$$

It is shown that the collection of all pairs $(\mathcal{J},f)$, where $\mathcal{J} = \{f_\alpha\}$ is a net converging to f, both f and f_α elements of $\mathcal{E}$, satisfy the conditions for a convergence class (c.f. Section 4). Hence there is a unique topology on $\mathcal{E}$ such that we have precisely the convergence indicated. (The collection of pairs $(\mathcal{J},f)$ will define a closure operator, which in turn defines a topology.)

To exemplify the idea of exceptional sets, we mention briefly an application. Let $(\Omega,\mathcal{B},\mu)$ be the right half line $[0,+\infty)$, the Borel field, and Lebesgue measure. Let $\mathcal{E}$ be the class of densities which are non-increasing on Ω, with total mass one. Then any element of $\mathcal{E}$ has at most countably many discontinuities. Given an element $f \in \mathcal{E}$, let the exceptional set of f be the collection of its discontinuities. Then our topology on $\mathcal{E}$ is precisely the same as the topology of convergence at points of continuity.

2. Nonparametric Maximum Likelihood Estimation. We shall be concerned with estimating by the method of maximum likelihood the underlying probability density f given a sample $\bar{x}_n$ of size n of independent identically distributed random variables with observations $x_1, x_2, \ldots, x_n$, and given that f is a member of some class $\mathcal{E}$.

That is, given a sample $\bar{x}_n$, we define the "likelihood" functional L on $\mathcal{E}$:

$$L(g, \bar{x}_n) = \sum_{i=1}^{n} \log g(x_i),$$

and for a fixed sample $\bar{x}_n$, we define our "maximum likelihood estimate" of the density of $\bar{x}_n$ to be a point (if one exists) of $\mathcal{E}$ which maximizes $L(\cdot, \bar{x}_n)$.

It is our intent to show that with certain restrictions on the class of distributions and its topology the maximum likelihood estimate is consistent, that is, as our sample size gets large, the estimate of the density almost surely converges to the true density.

Often our estimate f_n will not be a proper member of $\mathcal{E}$, but instead the density f_n will be a limit of measures in $\mathcal{E}$. Thus we are led to consider a compactification $\mathcal{L}$ of $\mathcal{E}$.

We assume the elements of $\mathcal{L}$ are densities of probability measures, and corresponding to every element in $\mathcal{L}$ there is an exceptional set of μ measure zero such that f_n converges to f if and only if $f_n(x)$ converges to $f(x)$ for all x not in the exceptional set and $\underline{\mathrm{Lim}}_n B_{f_n} \subset B_f$.

Since the correspondence between measures and densities is not one to one, we find it convenient to work in quotient spaces $\mathcal{E}/\mathfrak{R}$, $\mathcal{L}/\mathfrak{R}$, where $\mathfrak{R}$ is the equivalence relation

$$f \mathfrak{R} g \quad \text{in case} \quad f(x) = g(x)$$

except on a set of μ measure zero. (c.f. Section 5.)

The added complexity of considering the quotient space with the quotient topology is bothersome but unavoidable. If $\mathcal{E}$ and $\mathcal{L}$ are such that (as is the case in the classical k-dimensional cases) f, g in $\mathcal{L}$ and $f \mathfrak{R} g$ if and only if $f = g$, then the quotient spaces $\mathcal{E}/\mathfrak{R}$ and $\mathcal{L}/\mathfrak{R}$ are precisely the same as $\mathcal{E}$ and $\mathcal{L}$.

Classically the method of maximum likelihood has been restricted to families of distribution functions having some k dimensional parameterization. The distributions are then given the metric topology induced by the usual metric on the parameter space and the estimates are shown to be consistent subject to certain regularity conditions.

Consistency, however, is essentially a topological property, and it is unduly restrictive to establish it only on the basis of a parameterization.

The problem of estimating a measure instead of its density is in many ways more appealing. However, with the method of maximum likelihood it is the density which is important in picking an estimate. The transformation between $\mathcal{L}/\mathfrak{R}$ and the corresponding class of probability measures is, of course, one to one, but the corresponding estimate of the measure will in general be consistent if and only if the transformation from densities to measures is continuous. That is, our corresponding estimate of the measure will be consistent when we give the measures a topology which is contained in the topology induced by the transformation from elements of $\mathcal{L}/\mathfrak{R}$ to the corresponding measures. It is interesting to note (c.f. Section 6) that in many cases the topology induced on the corresponding measures is precisely the topology of convergence in distribution. For instance, this is the case if we take $\mathcal{E}$ to be the class of all unimodal densities uniformly bounded by M, (c.f. Section 3) or if we take $\mathcal{E}$ to be the class of densities with increasing hazard rates, (c.f. [4].)

Note that in the classical cases of k dimensional parameterization there is a continuous one to one transformation from the parameter space to the densities (with our topology) and similarly to the class of measures with the topology of convergence in distribution. Since Euclidean k-space has the property of invariance of domain it follows that these transformations are bicontinuous, and hence the transformation from densities to measures is continuous. Moreover, it is clear that the classical classes of densities are locally compact and locally separable, since they are homeomorphic to a subspace of k-space.

Following A. Wald (c.f. [5]), we intend to give a proof of consistency assuming that we have a locally separable, locally compact quotient space $\mathcal{E}/\mathfrak{R}$ of probability densities, which together with a suitable compactification $\mathcal{L}/\mathfrak{R}$ satisfies Conditions 1 - 3. By locally separable, we mean the neighborhood system at a point has a countable base.

Section 6 is a topological discussion which is intended to make these hypotheses more amenable.

Condition 1. $\mathcal{E}/\mathfrak{R}$ with the above mentioned topology of pointwise convergence, is a locally separable, locally compact Hausdorff space.

We assume that there exists some Hausdorff compactification $\mathcal{L}$ of $\mathcal{E}$ such that $\mathcal{L}/\mathfrak{R}$ and $\mathcal{E}/\mathfrak{R}$ satisfy the following conditions.

The following is of paramount importance to the method of maximum likelihood and is known to be true under a wide range of conditions. Let P be the projection of $\mathcal{L}$ into $\mathcal{L}/\mathfrak{R}$. (c.f. Theorem 6.1, pages 5, 6, 14 of [3].)

Condition 2. If $Pg \in \mathcal{L}/\mathfrak{R}$, $Pf \in \mathcal{E}/\mathfrak{R}$, are distinct elements of $\mathcal{L}/\mathfrak{R}$, then

$$\mathcal{E}_F \log g(x) < \mathcal{E}_F \log f(x)$$

and

$$-\infty < \mathcal{E}_F \log f(x) < +\infty.$$

(Here $\mathcal{E}_F$ denotes integration over Ω with respect to F.)

For a subset B of $\mathcal{L}/\mathfrak{R}$ let

$$s(x,B) = \sup_{Pf \in B} f(x).$$

Condition 3. If $Pg \in \mathcal{L}/\mathfrak{R}$, $Pf \in \mathcal{E}/\mathfrak{R}$ are distinct points of $\mathcal{L}/\mathfrak{R}$, then for sufficiently small neighborhoods B of Pg, the function

$$\log s(\cdot,B)$$

is measurable and bounded above by some function which is integrable with respect to F.

Lemma 1. Let $\{U_n\}$ be a decreasing sequence of neighborhoods of Pg in $\mathcal{L}/\mathcal{R}$ such that $\cap U_n = \{Pg\}$.

Then

$$\mathrm{Lim}_n\, E_F \log s(x,U_n) = E_F \log g(x) \quad \text{for any } F \text{ in } \mathcal{E}.$$

Proof. In view of Condition 2 we have only to show that

$$(1.1) \qquad \mathrm{Lim}_n\, s(x,U_n) = g(x) \quad \text{a.s. } F;$$

for the asserted result then follows from the fact that $s(x,U_n)$ is decreasing and the bounded convergence theorem. To show (1.1) we begin by using Condition 1 and throwing out a set B of μ measure zero (and hence of F measure 0) so that on $\Omega \sim B$, g_n converging to g implies $g_n(x)$ converges to $g(x)$. Now, for fixed x not in B and any ϵ we can pick g_n in U_n so that $s(x,U_n) - g_n(x) < \epsilon$. It follows that $\mathrm{Lim}\, s(x,U_n) - g_n(x) = \mathrm{Lim}\, s(x,U_n) - g(x) < \epsilon$, for any ϵ, thus $s(x,U_n) \downarrow g(x)$, as was to be shown.

Given a sample $x_1,\ldots,x_n$ of independent random variables with density f, let f_n be a point in $\mathcal{L}$ which maximizes the likelihood functional $L(\cdot,\bar{x}_n)$. We say that the sequence of estimates $\{f_n\}$ is consistent in case Pf_n converges to Pf in the topology of $\mathcal{L}/\mathcal{R}$ with F probability one. Note that this is in general stronger than saying that $f_n(x)$ converges almost surely to $f(x)$ except on a set of μ measure 0.

It is not obvious that the event $\{Pf_n$ converges to Pf in the topology of $\mathcal{L}/\mathcal{R}\}$ is always measurable. It is our intent to show that the complement of this event is a subset of a measurable set of F measure zero, and hence measurable. Similarly, in the sequel, measurability is implicit in the statements that events have F measure one.

The essence of the fact that maximum likelihood estimates are consistent is contained in the following theorem.

Theorem 1. Let Pg in $\mathcal{L}/\mathcal{R}$ and Pf in $\mathcal{E}/\mathcal{R}$ be two distinct elements of $\mathcal{L}/\mathcal{R}$ and let $X_1, X_2 \ldots$ be a sample with density function f. Then for sufficiently small neighborhoods U of Pg:

$$\text{Lim sup}_n [\sup_{Ph \in U} L(h, \overline{X}_n) - L(f, \overline{X}_n)] < 0$$

with F probability 1.

Proof. Given Pg not equal to Pf there exists (by Condition 2 and Lemma 1) a small neighborhood U of Pg such that

$$(1.2) \qquad E_F \log s(X,U) < E_F \log f(X) .$$

Now

$$\sup_{h \in U} \frac{1}{n} \Sigma \log h(X_i) \leq \frac{1}{n} \Sigma \log s(X_i, U).$$

Taking limits we see by the strong law of large numbers that with F probability one

$$\frac{1}{n} \Sigma \log s(x_i, U) \rightarrow E_F \log s(X,U)$$

and

$$\frac{1}{n} \Sigma \log f(X_i) \rightarrow E_F \log f(X)$$

and the result follows by (1.2).

<u>Corollary to Theorem 1</u>. Let $X_1, X_2, \ldots,$ and Pf be as in Theorem 1. Let D be any closed set not containing Pf. Then

$$\text{Lim sup}_n [\sup_{Ph \in D} L(h, \bar{X}_n) - L(f, \bar{X}_n)] < 0$$

with F probability 1.

<u>Proof</u>. By Theorem 1, any Pg in D can be covered by an open neighborhood U_g with the property that

$$\text{Lim sup}_n [\sup_{h \in U_g} L(h, \bar{X}_n) - L(f, \bar{X}_n)] < 0.$$

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

$$\text{Lim sup}_n [\sup_{Ph \in D} L(h, \bar{X}_n) - L(f, \bar{X}_n)] \leq$$

$$\text{Lim sup}_n \{ \max_{1 \leq i \leq m} [\sup_{Ph \in D} L(h, \bar{X}_n) - L(f, \bar{X}_n)]\} = \max_{1 \leq i \leq m} \text{Lim sup}_n [\sup_{Ph \in U_i} L(h, \bar{X}_n) - L(f, \bar{X}_n)] < 0$$

as was to be shown.

<u>Theorem 2</u>. (Consistency of Maximum Likelihood Estimate) Let $f(\cdot; \bar{X}_n)$ be a point of $\mathcal{L}$ depending on the random variable $X_1, \ldots, X_n = \bar{X}_n$ such that

$$\frac{\prod_1^n f(X_i; \bar{X}_n)}{\prod_1^n f(X_i)} \geq c \qquad \text{where } c > 0.$$

Then $Pf(\cdot, \bar{X}_n)$ converges to Pf in the topology of $\mathcal{L}/\mathcal{R}$ with F probability one.

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

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

then

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

and therefore

$$\text{Lim sup}_n [L(f_n, \bar{x}_n) - L(f, \bar{x}_n)] \geq 0.$$

Let U be any open neighborhood of Pf. If Pf_n is outside of U infinitely often, then

$$\text{Lim sup}_n [\sup_{Ph \in \mathcal{L} \sim U} L(h, \bar{x}_n) - L(f, \bar{x}_n)] \geq 0 \, .$$

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 Pf whose intersection is Pf. Corresponding to each U_i there is an event S_i of F measure zero such that on the complement of S_i, Pf_n is eventually in U_i. It follows that on the complement of $\bigcup_{i=1}^{\infty} S_i$, Pf_n is eventually in every neighborhood of Pf. Thus Pf_n converges to Pf in the topology of $\mathcal{L}/\mathcal{R}$ with F probability 1, as was to be shown.

3. Application to Estimation of a Unimodal Density. Let our sample space be the real line with the usual Borel field and Lebesgue measure. Take $\mathcal{E}$ to be the class of unimodal densities uniformly bounded by some constant M and such that

$$\int_{-\infty}^{\infty} f(x)\,dx = 1.$$

That is, the class of densities for which there exists some x such that $f(x) \leq M$, f is nondecreasing on $(-\infty, x]$, and nonincreasing on $[x, +\infty)$. Any such x is called a mode of f. Note that if x and y are modes of f, then $f(x) = f(y)$. If x is a mode of f, we define the height of f to be the value $f(x)$.

Any element of $\mathcal{E}$ has at most countably many points of discontinuity. Give $\mathcal{E}$ the topology of pointwise convergence on points of continuity. That is, given an element f in $\mathcal{E}$, we take the exceptional set B_f to be the points of discontinuity of f.

Add to $\mathcal{E}$ all densities of the form $p\,f$, $0 \leq p < 1$, and f in $\mathcal{E}$, and give $\mathcal{L}$ the topology of convergence on points of continuity. Again in this case the exceptional set of $p\,f$ will consist of the points of discontinuity of f.

Note that the class of unimodal densities which are finite linear combinations of characteristic functions is dense in $\mathcal{L}$. Moreover, the characteristic functions may be taken to be 1 on intervals with rational end points, and the multiplicative coefficients may be taken to be rational. Thus unimodal densities of this type form a countable dense subset of $\mathcal{L}$. Let C denote this countable dense subset of $\mathcal{L}$.

We will now construct a base for the neighborhood system at a point f of $\mathcal{L}$. Let D be a countable dense subset of the points of continuity of f. For every positive integer n and every finite subset $d_1, \ldots, d_m$ of D, we construct a neighborhood U of f as follows.

$U = \{g :$ i) $|g(d_i) - f(d_i)| < 1/n$ for $i = 1, \ldots, m$

ii) there is a mode m_1 of g and a mode m_2 of f such that $|m_1 - m_2| < 1/n$

iii) If f has a mode which is a point of continuity, then $|\text{height of } g - \text{height of } f| < 1/n$, otherwise there is no restraint on the height of g.}

It follows that $\mathcal{L}$ has a countable dense subset and is locally separable, thus there exists countable base for the topology of $\mathcal{L}$. Hence, it suffices to verify that $\mathcal{L}$ is sequentially compact.

Suppose $\{f_n\}$ is a sequence in $\mathcal{L}$. Let m_n be a mode of f_n. Then, if $\limsup m_n = +\infty$, there is a subsequence of $\{f_n\}$ which converges to the density $z(\cdot)$ which is identically zero. Similarly if $\liminf m_n = -\infty$, there is a subsequence of $\{f_n\}$ convering to z.

If, on the other hand, $-\infty < a \leq m_n \leq b < +\infty$ for all large n, then pick a subsequence such that m_n converges to some finite value m. Now, we have a sequence of uniformly bounded functions which are nondecreasing on $(-\infty, m_n]$, and m_n converges to m. Hence by considerations similar to the Helly Weak Compactness theorem we may pick a subsequence which is convergent on $(-\infty, m]$. Similarly, from this we may pick a subsequence which is convergent on $[m, +\infty)$, and this sequence is convergent in $\mathcal{L}$, as was to be shown.

It can be seen that $s(x,U)$ is in fact equal to the supremum of $g(x)$ where g is in the intersection of U and the countable dense set C. Thus $s(x,U)$ is the supremum of countably many measurable functions and as such is measurable. The function $\log s(x,U)$ is F integrable since it is bounded by $\log M$.

For Condition 2, recall that $f \in \mathcal{E}$ implies that

$$\int_{-\infty}^{\infty} f(x)dx = 1,$$

and note Theorem 6.1.

Thus the class of bounded unimodal densities satisfies all the regularity conditions, and it follows that the maximum likelihood estimate of the density, given a sample of observations, is consistent.

R. Pyke (in a private conversation) has suggested an algorithm for computing this estimate. This and other applications are to appear in a forthcoming paper.

In another paper [4] to appear, A. W. Marshall and Frank Proschan consider the maximum likelihood estimate of a distribution with monotone hazard rate. If we take Ω to be the half line $[0,+\infty)$, these distributions satisfy our requirements for consistency of the estimate. However, they consider Ω to be the whole real line and give a direct method of proving consistency. An easy algorithm is given for computing this estimate.

4. Appendix, Convergence Classes. Let $\mathfrak{X}$ be a space of densities of probability measures on $(\Omega, \mathcal{B}, \mu)$, and assume that with each $f \in \mathfrak{X}$ we associate a set $B_f \in \mathcal{B}$ of μ measure zero. Then form the class $\mathcal{C}$ of pairs $(\mathcal{F}, f)$, where $f \in \mathfrak{X}$ and $\mathcal{F} = \{f_\alpha, \alpha \in A\}$ is a net in $\mathfrak{X}$ such that $f_\alpha(x)$ converges to $f(x)$ for all x not in B_f and $\underline{\mathrm{Lim}}_\alpha B_{f_\alpha} \subset B_f$. Then, in the notation of Kelley, [2], we have:

Theorem 4.1 $\mathcal{C}$ is a convergence class.

Proof First note that if $f_\alpha = f$ for all $\alpha \in A$, then f_α, $\alpha \in A$ converges $\mathcal{C}$ to f.

Second, if f_α converges $\mathcal{C}$ to f, then off B_f every subset of $f_\alpha(x)$ converges to $f(x)$ and hence every subset of $\{f_\alpha, \alpha \in A\}$ converges to f.

Third, if f_α does not converge $\mathcal{C}$ to f, then there is some point x not in $_f$ such that $f_\alpha(x)$ does not converge to $f(x)$. Hence, there is a subnet $\{f_\alpha(x), \alpha \in C\}$, no subnet of which converges to $f(x)$. Thus no subnet of $\{f_\alpha, \alpha \in C\}$ converges to f.

Finally, we must show that $\mathcal{C}$ satisfies the theorem on iterated limits. We can do no better here than to remind the reader of our definition of convergence and refer him to Kelley [2], pages 69, 73, and 74.

This completes the proof of Theorem 4.1, and hence having determined the exceptional sets, there is precisely one topology on $\mathcal{E}$ and one extension to $\mathcal{L}$ which gives us the convergence asserted.

5. Appendix, Quotient Spaces and Projections. Let $\mathfrak{R}$ be the equivalence relation on $\mathcal{L}$:

$$f \mathfrak{R} g \quad \text{whenever} \quad f(x) = g(x)$$

except on a set of μ measure zero.

That is, two densities are $\mathfrak{R}$-equivalent when they are equivalent forms of the Radon-Nikodym derivative of the same probability measure.

Let P be the projection of $\mathcal{L}$ onto $\mathcal{L}/\mathfrak{R}$. Then, with the quotient topology P is continuous and hence we have

Lemma 5.1 The space $\mathcal{L}/\mathfrak{R}$ is compact when $\mathcal{L}$ is.

Corollary 5.2 If $\mathcal{L}$ is a compactification of $\mathcal{E}$ whose elements are densities of probability measures, then $\mathcal{L}/\mathfrak{R}$ is a compactification of $\mathcal{E}/\mathfrak{R}$.

Appendix 6. On the Conditions 1 - 3. The following theorem concerning Condition 2 is well known.

Theorem 6.1 Let f be a probability density such that

$$\int_{\Omega} f(x)\mu(dx) = 1, \quad \text{and} \quad \mathcal{E}_F \log f(x) < +\infty.$$

Let g be any probability density which is not R-equivalent to f, then

(i) $\mathcal{E}_F \log g(x) < \mathcal{E}_F \log f(x)$

Proof Let f have support S, then

(ii) $\mathcal{E}_F \log \frac{g(x)}{f(x)} \leq \log \mathcal{E}_F \frac{g(x)}{f(x)} = \log \int_S \frac{g(x)}{f(x)} f(x)\mu(dx) = 0.$

Moreover

(iii) $\mathcal{E}_F \log \frac{g(x)}{f(x)} < \log \mathcal{E}_F \frac{g(x)}{f(x)}$

unless 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 $\int g(x)\mu(dx) \leq 1$. Now, if c were equal to 1, then g is equal to f almost everywhere on the support of f, and therefore g must be R-equivalent to f, which is a contradiction. Hence $c < 1$, and

$$\mathcal{E}_F \log c < 0$$

as was to be shown.

Some remarks on Condition 1 may be helpful.

In the sequel, let $(\Omega, \mathcal{B}, \mu)$ be the real line, Borel field, and Lebesque measure. If $\mathcal{F}$ is a class of absolutely continuous distribution functions, then given F in $\mathcal{F}$, there exists a set B_F of μ measure zero such that off B_F the formal derivative exists and is unique. Thus, if $\mathcal{E}$ is a class of formal derivatives of elements of $\mathcal{F}$ and the exceptional set of an element f in $\mathcal{E}$ is the forementioned set B_F depending on the corresponding distribution function, we have a natural and pleasing topology on $\mathcal{E}$. It turns out (Theorem 6.6) that projection into the quotient space $\mathcal{E}/\mathcal{R}$ is aclosed map, and the natural map from $\mathcal{E}/R$ into the topological space $\mathcal{F}$, (with the topology of convergence in distribution) is a homeomorphism. Thus consistency of the derivative (in our topology) is equivalent to the consistency of the corresponding estimate of the distribution function (with the topology of convergence in distribution).

This leads us to the following definition: A topological space $\mathfrak{X}$ of densities is well defined in case

(i) The exceptional set of a density in $\mathfrak{X}$ depends only on the corresponding measure; that is, given a probability measure all versions of the Radon-Nikodym derivative which are in $\mathfrak{X}$ have the same exceptional set, call it B_F.

(ii) Given a probability measure F all versions of the Radon Nikodym derivative of F which are in $\mathfrak{X}$ are equal off the exceptional set B_F.

It should be pointed out that although the space $\mathcal{E}$ may be well defined, it is not in general true that an otherwise suitable compactification $\mathcal{L}$ of $\mathcal{E}$ will be well defined.

For example, if $\mathcal{E}$ is the class of unimodal derivatives of probability measures on $(-\infty, +\infty)$ with the topology of convergence at points of continuity, then any suitable compactification of $\mathcal{E}$ will contain densities which are not formal derivatives.

On the other hand, if $\mathcal{E}$ is the class of all nonincreasing derivatives of probability measures on $[0, \infty)$, then $\mathcal{E}$ is compact if we include measures with mass less than one. Hence $\mathcal{L}$ can be taken to be well defined, and $\mathcal{E}$ and $\mathcal{L}$ have the topology of convergence at points of continuity.

In the sequel, we assume that $\mathcal{L}$ is a compact well defined space of probability densities. We let φ be the natural map of an element f of $\mathcal{L}$ into $F(x) = \int_{-\infty}^{x} f(t)dt$, and let $\mathcal{G}$ be the image under φ of $\mathcal{L}$, with the topology of convergence in distribution. Since the continuous image of a compact space is compact, and a subset of a compact space is compact if and only if it is closed, we have:

Theorem 6.2 φ is a closed map.

Now, an element of $\mathcal{G}$ is uniquely determined if we know its value on some fixed countable dense subset, in fact we have the well known

Theorem 6.3 $\mathcal{G}$ is homeomorphic to a subspace of the cube of dimension ω (the first infinite ordinal).

Hence, $\mathcal{G}$ is locally separable (in fact has a countable base), and we need consider only sequences.

There is a natural one to one map (call it h) from $\mathcal{G}$ to $\mathcal{L}/\mathcal{R}$. We can give $\mathcal{L}/\mathcal{R}$ a topology (call it $\mathcal{C}_h$) such that h is a homeomorphism. It follows that the map $h(\varphi(\cdot))$ is a closed continuous map from the topological space $\mathcal{L}$ to the space $(\mathcal{L}/\mathcal{R}, \mathcal{C}_h)$, and therefore ([2], page 95) $\mathcal{C}_h$ is precisely the quotient topology on $\mathcal{L}/\mathcal{R}$. Hence

<u>Theorem 6.4</u> The projection of $\mathcal{L}$ onto $\mathcal{L}/\mathcal{R}$ is a closed map, and the natural map from $\mathcal{L}/\mathcal{R}$ to $\mathcal{G}$ is a homeomorphism.

Going back to the original space $\mathcal{E}$ we easily have

<u>Theorem 6.5</u> (Condition 1) $\mathcal{E}/\mathcal{R}$ is a locally separable,locally compact Hausdorff space, and is homeomorphic to $\varphi(\mathcal{E})$.

Condition 3 on the supremum function is not so easily analyzed. In all of the cases that have come to our attention, the supremum may be taken over a countable class of measurable functions, and as such is measurable. The integrability of this function may be assured by assuming that the class of densities is uniformly bounded above by some constant. If the resulting estimate does not depend on the constant, then we may conclude, by a limiting argument, that the estimate is consistent without the condition of boundedness.

References

[1] Kraft, C. and Le Cam, L., A remark on the roots of the maximum likelihood equation, Ann. Math. Stat., 27, 1174-1177.

[2] Kelley, J. L., General Topology, D. van Nostrand Company, 1955.

[3] Kullback, Solomon, Information Theory and Statistics, Wiley and Sons, pp. 5, 6, 14, 1959.

[4] Marshall, A. W. and Proschan, Frank, Maximum likelihood estimation for distributions with monotone hazard rates, to appear.

[5] Wald, Abraham, Note on the consistency of the maximum likelihood estimate, Ann. Math. Stat., 20, 595-601.