ON ESTIMATING A DENSITY WHICH IS MEASURABLE WITH RESPECT TO A σ-LATTICE[1]

By Tim Robertson

University of Iowa

1. Introduction and summary. This paper is concerned with the problem of estimating a probability density which is known to be measurable with respect to a σ-lattice of subsets of the space on which it is defined. Our solution is represented as a conditional expectation. (Generally in the literature "conditional expectation" refers to "conditional expectation given a σ-field." However since we shall be concerned exclusively with "conditional expectation given a σ-lattice" we shall use this abbreviated terminology for the latter, more general concept.) Brunk [2] discusses conditional expectations and many of the extremum problems for which they provide solutions.

We shall consider the case where the measure space, $(\Omega, \mathcal{A}, \mu)$, on which the density is defined is totally finite. Let $\mathcal{L}$ denote a σ-lattice of subsets of $\Omega(\mathcal{L} \subset \mathcal{A})$. A σ-lattice, by definition, is closed under countable unions and intersections and contains both Ω and the null set $\emptyset$. Let $\omega_1, \omega_2, \cdots, \omega_n$ be a sample of independent observations chosen in Ω according to the unknown, $\mathcal{L}$-measurable density f. We say that a point is chosen in Ω according to f if the probability that it will lie in any set A in $\mathcal{A}$ is given by $\int_A f\, d\mu$. The function f is $\mathcal{L}$-measurable if the set $[f > a]$ is in $\mathcal{L}$ for each real number a.

We shall use the maximum likelihood criterion for choosing an estimate. In other words we wish to find an $\mathcal{L}$-measurable density $\hat{f}$ such that the product of the values of $\hat{f}$ at the observed points is at least as large as the product of the values of any other $\mathcal{L}$-measurable density at those points. Such a function will be called a maximizing function.

Clearly the σ-lattice $\mathcal{L}$ must satisfy some restrictions in order for the problem to be of any interest at all. For example if $(\Omega, \mathcal{A}, \mu)$ is a finite subinterval of the real line together with Borel subsets and Lebesgue measure and if $\mathcal{L} = \mathcal{A}$ then there are many obvious solutions if the density is bounded, and none at all if it is not bounded. The second section of this paper is devoted to the restrictions that we impose and to showing that these restrictions are satisfied in some problems which are of interest.

The fourth section of this paper is devoted to some results on the asymptotic properties of our estimates in three special cases. The methods used are similar to those used by Marshall and Proschan [4]. The final section contains some observations on the problem of estimating a density on a non-totally finite measure space.

Received 18 December 1965.

[1] This research forms part of the author's doctoral dissertation submitted in the Department of Statistics, University of Missouri. It was sponsored by the U. S. Army Research Office-Durham under contract number DA-31-124-ARO(D)-158.

Let L_2 denote the set of square integrable random variables and $L_2(\mathcal{L})$ the collection of all those members of L_2 which are $\mathcal{L}$-measurable. Let $R(\mathcal{L})$ denote the collection of all $\mathcal{L}$-measurable random variables. Let $\mathcal{B}$ denote the collection of Borel subsets of the real line. We shall adopt the following definition for the conditional expectation, $E_\mu(f \mid \mathcal{L})$, of a random variable given a σ-lattice.

Definition 1.1. If $f \varepsilon L_2$ then $g \varepsilon L_2(\mathcal{L})$ is equal to $E_\mu(f \mid \mathcal{L})$ if and only if g has both of the following properties:

$$\int (f - g)h \, d\mu \leqq 0 \quad \text{for all} \quad h \,\varepsilon\, L_2(\mathcal{L}) \tag{1.1}$$

and

$$\int_B (f - g) \, d\mu = 0 \quad \text{for all} \quad B \,\varepsilon\, g^{-1}(\mathcal{B}). \tag{1.2}$$

(Brunk [1] shows that there is such a random variable g associated with each $f \varepsilon L_2$ and that g is unique in the sense that if g' is any other member of $L_2(\mathcal{L})$ having these properties then $g = g'[\mu]$.)

In order to motivate the consideration of a problem such as the one we take up in this paper we introduce the following examples. The first example is discussed in [5].

Example 1.1. Suppose Ω is a finite set and we denote its elements by 1, 2, $\cdots$, k. Let $\mathcal{A}$ be the collection of all subsets of Ω and suppose μ assigns positive mass to each point in Ω. Suppose $\mathcal{L}$ is an arbitrary σ-lattice of subsets of Ω. Let n_i denote the number of times the point i is observed. We wish to find an $\mathcal{L}$-measurable density $\hat{f}$ on Ω such that

$$\prod_{i=1}^{k} \hat{f}(i)^{n_i} \geqq \prod_{i=1}^{k} h(i)^{n_i}$$

for every other $\mathcal{L}$-measurable density h. It is shown in [5] that a solution is given by $\hat{f} = E_\mu(g \mid \mathcal{L})$ where $g(i) = n_i[n \cdot \mu(i)]^{-1}$.

The problem posed in the next example was solved by Pyke (personal communication with Professor Brunk) and for a monotone density by Grenander [3].

Example 1.2. Suppose Ω is a closed subinterval of the real line ($\Omega = [c, d]$), $\mathcal{A}$ is the collection of Borel subsets of Ω and μ is Lebesque measure. We wish to estimate the density f which is known to be unimodal at some unknown point in Ω. Suppose that our observations are ordered: $\omega_1 < \omega_2 < \cdots < \omega_n$.

If h is any unimodal density with mode at a and $\omega_j < a < \omega_{j+1}$ then define the function g by:

$$\begin{aligned} g(x) &= h(\omega_j) && (\omega_j \leqq x < a) \\ &= h(\omega_{j+1}) && (a \leqq x \leqq \omega_{j+1}) \\ &= h(x) && \text{otherwise.} \end{aligned}$$

It is easily seen that the density $\hat{f} = [\int g \, d\mu]^{-1} \cdot g$ is unimodal with mode at ω_j or ω_{j+1} and that the product of the values of $\hat{f}$ at the observed points is at least as large as the product of the corresponding values of h. Hence our problem reduces to finding an estimate which has mode at one of the observed points.

Similarly we can show that any maximizing estimate must be constant on every open interval joining two consecutive observed values. The next remark is easy to verify.

REMARK 1.1. Let $\mathcal{L}$ be the σ-lattice of subsets of Ω consisting of all those intervals containing the point a. A function f on Ω is unimodal at a if and only if f is $\mathcal{L}$-measurable.

If we can estimate the density subject to the restriction that it is unimodal at a fixed point then we can select an estimate by comparing the ones we get by assuming the mode is at particular observations. We see thus that the problem of estimating a unimodal density reduces to estimating a density which is measurable with respect to a σ-lattice and constant on intervals joining consecutive observations. We shall see that these results for a unimodal density are typical of a larger class of problems.

Now consider the problem of estimating a unimodal density on the reals together with Lebesgue measure. Clearly any estimate must be zero outside of the smallest closed interval containing all the observed points. Hence this problem reduces to the one discussed above.

2. σ-lattices and partial orderings. Recall that we are given a totally finite measure space $(\Omega, \mathcal{A}, \mu)$ and a σ-lattice $\mathcal{L}$ of subsets of Ω. Define the relation $\ll$ on Ω by:

$$\omega \ll \omega' \quad \text{if and only if} \quad \omega \,\varepsilon\, L \,\varepsilon\, \mathcal{L} \quad \text{imply that} \quad \omega' \,\varepsilon\, L.$$

This relation gives an "almost" partial ordering of Ω in the sense that it is transitive and reflexive but not necessarily symmetric. To see that it is not necessarily symmetric consider: $\Omega = \{x, y\}$ and $\mathcal{L} = \{\varnothing, \Omega\}$.

We say that a σ-lattice is complete if it is closed under arbitrary unions and intersections. A function F on Ω is isotone with respect to $\ll$ provided that $\omega_1 \ll \omega_2$ implies that $F(\omega_1) \leqq F(\omega_2)$.

THEOREM 2.1. *If the σ-lattice $\mathcal{L}$ is complete then a necessary and sufficient condition for a function f on Ω to be $\mathcal{L}$-measurable is that it be isotone with respect to the almost partial ordering induced by $\mathcal{L}$.*

PROOF. The necessity is easy to show. To prove the sufficiency suppose f is isotone with respect to the partial ordering and a is any real number. Then for each ω in the set $[f > a]$ and ω' not in that set there is a set $L(\omega, \omega')$ in $\mathcal{L}$ which contains ω and does not contain ω'. Since $\mathcal{L}$ is complete the set $[f > a]$ is in $\mathcal{L}$ because

$$[f > a] = \bigcup_{\omega \varepsilon [f>a]} \bigcap_{\omega' \varepsilon [f>a]^c} L(\omega, \omega').$$

We now make the following assumptions about the σ-lattice $\mathcal{L}$:

(i) $\mathcal{L}$ is complete and (ii) corresponding to each pair ω and ω' of elements of Ω there is a unique element v of Ω such that $\omega \ll v$, $\omega' \ll v$, and if $\omega \ll y$ and $\omega' \ll y$ then $v \ll y$. (We denote v by $\max(\omega, \omega')$.)

In many situations we begin with a partial ordering of the space Ω and wish to estimate a density which is known to be isotone with respect to this partial ordering. Suppose the relation which gives this partial ordering is denoted by $\leqq$. Let the σ-lattice $\mathcal{L}$ be the collection of all subsets L of Ω with the property that whenever L contains a point ω then it contains all points ω' such that $\omega \leqq \omega'$. This σ-lattice $\mathcal{L}$ in turn induces the relation $\ll$. It is easy to show that this σ-lattice is complete and that the two relations, $\leqq$ and $\ll$, are equivalent. Hence if $\leqq$ has property (ii) then so does $\ll$. Using assumptions (i) and (ii) the following remarks are easy to verify.

REMARK 2.1. Corresponding to each point ω in Ω there is a unique member of $\mathcal{L}$, denoted by $L(\omega)$, with the property that $L(\omega)$ is a subset of each member of $\mathcal{L}$ which contains ω.

REMARK 2.2. Suppose L is a member of $\mathcal{L}$ which contains the point ω. L is equal to $L(\omega)$ if and only if $\omega \ll \omega'$ for each $\omega' \varepsilon L$.

REMARK 2.3. For any two points ω and ω' in Ω we have

$$L(\max(\omega, \omega')) = L(\omega) \cap L(\omega').$$

3. Estimation of an $\mathcal{L}$-measurable density. In this section we derive an expression for a maximizing function. It can be shown that any maximizing function is simple and assumes at most n values different from zero. However, if we can find any collection of sets on which a maximizing function must be constant then our problem can be reduced to the one discussed in Example 1.1. Recall that $L(\omega_i)$ denotes the smallest member of $\mathcal{L}$ containing the ith observed point. Form the sets B_i ($i = 0, 1, 2, \cdots, H$) by taking all possible intersections of the sets $L(\omega_1), L(\omega_2), \cdots, L(\omega_n)$ in the following way:

$$B_0 = \textstyle\bigcap_{i=1}^n L(\omega_i), \qquad B_1 = \bigcap_{i\neq n} L(\omega_i), \qquad B_2 = \bigcap_{i\neq n-1} L(\omega_i), \cdots,$$

$$B_n = \textstyle\bigcap_{i\neq 1} L(\omega_i), \qquad B_{n+1} = \bigcap_{i\neq n-1,n} L(\omega_i), \cdots, B_H = L(\omega_n).$$

Also let $A_0 = B_0$ and $A_i = B_i - \bigcup_{j<i} B_j = B_i - \sum_{j<i} A_j$ for $i = 1, 2, \cdots, H$.

Since each B_i is a finite intersection of the sets $L(\omega_i)$ it follows from Remark 2.3 and our second assumption that there is a point σ_i in B_i such that $B_i = L(\sigma_i)$. Further from Remark 2.2 it follows that if any A_i is non-null then it must contain the corresponding point σ_i. Suppose that k of the sets $A_0, A_1, \cdots, A_H$ are non-null and that we label them $A_1, A_2, \cdots, A_k$. Also relabel the corresponding sets B_i and points σ_i so that we have $\sigma_i \varepsilon A_i \subset B_i = L(\sigma_i)$. Note that the sets $A_1, A_2, \cdots, A_k$ are pairwise disjoint and that $\sum_{i=1}^k A_i = \bigcup_{i=1}^n L(\omega_i)$.

In many interesting cases we will, with probability one, draw a sample such that the numbers $\mu(A_1), \mu(A_2), \cdots, \mu(A_k)$ are all positive. We shall assume that this is the case for the remainder of this paper.

THEOREM 3.1. *If there is a maximizing function then there is a maximizing function which is constant on each of the sets* $A_1, A_2, \cdots, A_k$.

PROOF. Suppose $\hat{f}$ is a maximizing function and let $h = \sum_{i=1}^k \hat{f}(\sigma_i) \cdot I_{A_i}$. We first show that h is $\mathcal{L}$-measurable. By Theorem 2.1 it suffices to show that h is isotone with respect to $\ll$. Suppose $\omega \ll \omega'$. We may assume that ω is in $\sum_{i=1}^k A_i$ for if not then $0 = h(\omega) \leqq h(\omega')$.

Suppose the subscripts m_1 and m_2 are chosen so that $\omega \varepsilon A_{m_1}$ and $\omega' \varepsilon A_{m_2}$. Let $B_{m_1} = \bigcap_{i \varepsilon T_1} L(\omega_i)$, and $B_{m_2} = \bigcap_{i \varepsilon T_2} L(\omega_i)$. Now $\omega \varepsilon B_{m_1}$ and $\omega \not\varepsilon B_i$ for $i < m_1$. Similarly $\omega' \not\varepsilon B_{m_2}$ and $w' \not\varepsilon B_i$ for $i < m_2$. Since $\omega \ll \omega'$ it follows from the order in which we defined the B's that $T_2 \supset T_1$ and hence that $\sigma_{m_1} \ll \sigma_{m_2}$. Since $\hat{f}$ is $\mathfrak{L}$-measurable we conclude that

$$h(\omega) = \hat{f}(\sigma_{m_1}) \leqq \hat{f}(\sigma_{m_2}) = h(\omega').$$

Now $\sigma_i \ll \omega$ for all ω in A_i so that $0 < \int h \, d\mu \leqq \int \hat{f} \, d\mu = 1$. Further if any observation ω_j is in A_i then $\hat{f}(\omega_j) = \hat{f}(\sigma_i)$ since $\hat{f}$ is $\mathfrak{L}$-measurable and in this case $A_i \subset B_i \subset L(\omega_j)$. It follows easily from these remarks that the function $\hat{h} = [\int h \, d\mu]^{-1} \cdot h$ which is constant on $A_1, A_2, \cdots, A_k$ is a maximizing function.

Let n_i denote the number of observations in A_i, $(i = 1, 2, \cdots, k)$. Further, let $\Omega^* = \{1, 2, \cdots, k\}$ and form the σ-lattice $\mathfrak{L}^*$ of subsets of Ω^* as follows: $\varnothing \varepsilon \mathfrak{L}^*$ and $T \varepsilon \mathfrak{L}^*$ if and only if $\sum_{i \varepsilon T} A_i \varepsilon \mathfrak{L}$.

Remark 3.1. A non-negative function $f^* = (x_1, x_2, \cdots, x_k)$ on Ω^* is $\mathfrak{L}^*$-measurable if and only if $\sum_{i=1}^k x_i \cdot I_{A_i}$ is $\mathfrak{L}$-measurable.

Our problem is then to find an $\mathfrak{L}^*$-measurable, non-negative function $\hat{f}^* = (y_1, y_2, \cdots, y_k)$ on Ω^* such that $\sum_{i=1}^k y_i \cdot \mu(A_i) = 1$ and

$$\prod_{i=1}^k y_i^{n_i} \geqq \prod_{i=1}^k z_i^{n_i}$$

for every other function $h^* = (z_1, z_2, \cdots, z_k)$ having these properties. The solution of this problem is given in Example 1.1. We state the solution in the following theorem.

Theorem 3.2. *A maximizing function is given by*

$$\hat{f} = \sum_{i=1}^k y_i \cdot I_{A_i}$$

where $\hat{f}^* = (y_1, y_2, \cdots, y_k)$ *is equal to* $E_{\mu^*}(g^* \mid \mathfrak{L}^*)$, $\mu^*(i) = \mu(A_i)$, *and* $g^*(i) = n_i \cdot [n \cdot \mu(A_i)]^{-1}$, $(i = 1, 2, \cdots, k)$.

We shall for the remainder of this paper denote this maximizing function by $\hat{f}$ or by $\hat{f}_n$ in the section on consistency. Let $\mathfrak{L}_n$ be the σ-lattice of subsets consisting of Ω, $\varnothing$ and all sets of the form $A_{i_1} + A_{i_2} + \cdots + A_{i_m}$ which are in $\mathfrak{L}$. It is shown in [5] that an equivalent definition for $E_\mu(f \mid \mathfrak{L}) = g$ can be obtained by substituting

$$(3.1) \qquad \int_A (f - g) \, d\mu \leqq 0 \quad \text{for each} \quad A \varepsilon \mathfrak{L}$$

for (1.1) in the definition. Using this result the following theorem is easy to verify.

Theorem 3.3. $\hat{f} = E_\mu(\hat{g} \mid \mathfrak{L}_n)$ *where*

$$\hat{g} = \sum_{i=1}^k n_i \cdot [n \cdot \mu(A_i)]^{-1} \cdot I_{A_i}.$$

In certain special cases we have been able to show that $\hat{f} = E_\mu(\hat{g} \mid \mathfrak{L})$. One of these cases was introduced in Example 1.2.

Theorem 3.4. *In the special case described in Example* 1.2 *we have* $\hat{f} = E_\mu(\hat{g} \mid \mathfrak{L})$.

Proof. The only difficulty involved is in showing that if $h \varepsilon L_2(\mathfrak{L})$ then $\int (\hat{g} - \hat{f}) h \, d\mu \leqq 0$. Since both $\hat{f}$ and $\hat{g}$ are constant on each of the sets $A_1, A_2,$

$\cdots, A_k$ and are both zero on the complement of $\sum_{i=1}^k A_i$ we can write

$$\int (\hat{g} - \hat{f}) \cdot h \, d\mu = \int (\hat{g} - \hat{f}) \cdot h' \, d\mu$$

where h' is constant on the sets $A_1, A_2, \cdots, A_k$, and $[\sum_{i=1}^k A_i]^c$, its value on A_i being $[\mu(A_i)]^{-1} \cdot \int_{A_i} h \, d\mu$ and its value on $[\sum_{i=1}^k A_i]^c$ being less than the minimum of these values. It is now easy to complete the argument by showing that h' is $\mathcal{L}_n$-measurable.

4. Consistency. In this section we present some results on the asymptotic properties of our estimate. We consider Example 1.1, a special case of Example 1.2 and a third example which we introduce in this section.

EXAMPLE 1.1. Recall that $\Omega = \{1, 2, \cdots, k\}$, $\mathfrak{A}$ is the collection of all subsets of Ω and μ is a measure on $\mathfrak{A}$ which assigns positive mass to each point in Ω. $\mathcal{L}$ is an arbitrary σ-lattice of subsets of Ω and n_i denotes the number of times i was observed ($i = 1, 2, \cdots, k$). Our estimate is given by $\hat{f}_n = E_\mu(g_n \mid \mathcal{L})$ where $g_n(i) = n_i \cdot [n \cdot \mu(i)]^{-1}$. It follows from the Borel strong law of large numbers that $\lim_n g_n(i) = f(i)$ with probability one. Brunk [1] observes that

$$\int (g_n - f)^2 \, d\mu \geqq \int (\hat{f}_n - f)^2 \, d\mu.$$

It follows from these observations that for each i, $\lim_n \hat{f}_n(i) = f(i)$ with probability one.

We now consider the problem of estimating a density which is known to be unimodal at a predetermined point in a finite subinterval of the real line. In this discussion and the discussion of Example 4.1 we shall use a representation for the conditional expectation which is given in [6].

Our estimate in both cases is given by $\hat{f}_n = E_\mu(g_n \mid \mathcal{L}_n)$. Suppose ν_0 is an arbitrary point in Ω and $t = \hat{f}_n(\nu_0)$. Let $P_t = [\hat{f}_n > t]$ and $\mathfrak{N}_2(P_t) = \{L; L \in \mathcal{L}_n, \mu(L - P_t) > 0\}$. The result in [6] then gives us

$$\hat{f}_n(\nu_0) = \sup_{L \in \mathfrak{N}_2(P_t)} [\mu(L - P_t)]^{-1} \cdot \int_{L - P_t} g_n \, d\mu.$$

Let $\mathfrak{A}$ be the collection of all Borel subsets of $\Omega = [c, d]$ and let μ be Lebesgue measure. The σ-lattice $\mathcal{L}$ is the collection of all subintervals of Ω containing the point $a (c \leqq a \leqq d)$.

We wish to estimate the $\mathcal{L}$-measurable density f on the basis of the ordered sample $\omega_1 < \omega_2 < \cdots < \omega_n$. Suppose that $q(n)$ is the subscript of the largest observation less than a. The sets $A_1, A_2, \cdots, A_k$ on which the estimate must be constant are given by:

$$A_1 = [\omega_1, \omega_2), \quad A_2 = [\omega_2, \omega_3), \cdots, \quad A_{q(n)} = [\omega_{q(n)}, a),$$
$$A_{q(n)+1} = [a, \omega_{q(n)+1}], \cdots, \quad A_k = A_n = (\omega_{n-1}, \omega_n].$$

Our estimate is given by:

$$\hat{f}_n = E_\mu(g_n \mid \mathcal{L}_n) = E_\mu(g_n \mid \mathcal{L})$$

where $g_n = \sum_{i=1}^n n_i \cdot [n \cdot \mu(A_i)]^{-1} \cdot I_{A_i}$.

The proof of the next theorem uses methods similar to those used by Marshall and Proschan [4].

THEOREM 4.1. *If f is $\mathfrak{L}$-measurable then for every $\nu_0 < a$,*

$$f(\nu_0-) \leqq \lim_n \inf \hat{f}_n(\nu_0) \leqq \lim_n \sup \hat{f}_n(\nu_0) \leqq f(\nu_0+)$$

with probability one. Furthermore for every $\nu_0 > a$

$$f(\nu_0-) \geqq \lim_n \sup \hat{f}_n(\nu_0) \geqq \lim_n \inf \hat{f}_n(\nu_0) \geqq f(\nu_0+)$$

with probability one.

PROOF. The proofs of the various parts are similar although they may use different forms of the representation theorem given in [6]. We shall prove that if $\nu_0 < a$, then

$$f(\nu_0-) \leqq \lim_n \inf \hat{f}_n(\nu_0)$$

with probability one. If $f(\nu_0-) = 0$ then the desired result is obvious since $\hat{f}_n$ is a non-negative function. Suppose $\nu_1 < \nu_0$ and $f(\nu_1) > 0$. Let $s(n)$ be the subscript of the largest observed value less than or equal to ν_0 and let $r(n)$ be the subscript of the smallest observed value greater than or equal to ν_1.

Consider the above mentioned representation for $\hat{f}_n = E_\mu(g_n \mid \mathfrak{L}_n)$. Let $t = f_n(\nu_0)$ and $P_t = [\omega_{j(n)}, \omega_{i(n)}]$. Then $j(n) > s(n)$ so that $[\omega_{r(n)}, \omega_{i(n)}] \in \mathfrak{N}_2(P_t)$ for sufficiently large n with probability one. Hence

$$\hat{f}_n(\nu_0) \geqq [\omega_{j(n)} - \omega_{r(n)}]^{-1} \cdot \int_{[\omega_{r(n)}, \omega_{j(n)}]} g_n \, d\mu$$

with probability one for sufficiently large n. However

$$\int_{[\omega_{r(n)}, \omega_{j(n)}]} g_n \, d\mu = (1/n) \cdot [j(n) - r(n)]$$

so that

$$\hat{f}_n(\nu_0) \geqq (1/n) \cdot [j(n) - r(n)](\omega_{j(n)} - \omega_{r(n)})^{-1}.$$

If we let $F(x) = \int_{[c,x]} f \, d\mu$ and $Z_i = [f(\nu_1)]^{-1} \cdot F(\omega_i)$ $(i = 1, 2, \cdots, n)$ then since f is non-decreasing on $[c, a]$,

$$Z_{j(n)} - Z_{r(n)} \geqq \omega_{j(n)} - \omega_{r(n)}$$

so that

$$\hat{f}_n(\nu_0) \geqq (1/n) \cdot [j(n) - r(n)](Z_{j(n)} - Z_{r(n)})^{-1}.$$

It is well known that $Z_1, Z_2, \cdots, Z_n$ represents an ordered sample from the uniform $[0, 1/f(\nu_1)]$ distribution. Let G denote the distribution function associated with this distribution and G_n the empirical distribution function determined by $Z_1, Z_2, \cdots, Z_n$. Then

$$(1/n)[j(n) - r(n)] = G_n(Z_{j(n)}) - G_n(Z_{r(n)})$$

so that

$$\hat{f}_n(\nu_0) \geqq [G_n(Z_{j(n)}) - G_n(Z_{r(n)})](Z_{j(n)} - Z_{r(n)})^{-1}.$$

It follows from the Borel strong law of large numbers that $\lim_n Z_{r(n)} = [f(\nu_1)]^{-1} \cdot F(\nu_1)$ with the probability one so that

$$\lim_n \inf [Z_{j(n)} - Z_{r(n)}] > 0$$

with probability one. Using this and the Glivenko-Cantelli theorem we infer that

$$\lim_n [G_n(Z_{j(n)}) - G_n(Z_{r(n)})](Z_{j(n)} - Z_{r(n)})^{-1} = f(\nu_1)$$

with probability one. We can now draw the desired conclusion.

Clearly if f is continuous at ν_0 then this theorem implies that $\lim_n \hat{f}_n(\nu_0) = f(\nu_0)$ with probability one. Using this observation and methods similar to those used in proving the Glivenko-Cantelli theorem we can prove the following corollary.

COROLLARY 4.1. *If f is continuous on $[a_1, a_2] \subset [c, d]$ then*

$$\lim_n \sup_{x \varepsilon [a_1, a_2]} |\hat{f}_n(x) - f(x)| = 0$$

with probability one.

If we let $\hat{F}_n(x) = \int_{[c,x]} \hat{f}_n \, d\mu$ then we get the following corollary.

COROLLARY 4.2. *If f is continuous on $[c, d]$ then*

$$\lim_n \sup_x |\hat{F}_n(x) - F(x)| = 0$$

with probability one.

EXAMPLE 4.1. Let $\Omega = [0, 1) \times [0, 1)$, $\mathfrak{A}$ be the collection of Borel subsets of Ω and μ be Lebesgue measure. Define a partial ordering of Ω by: $(x_1, y_1) \leqq (x_2, y_2)$ if and only if $x_1 \leqq x_2$ and $y_1 \leqq y_2$. Suppose we wish to estimate the density f which is known to be isotone with respect to this partial ordering.

Denote our observations by $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$. Recall that when we derived our estimate of f all that was required was to have a finite number of sets of non-zero measure on which the estimate is known to be constant. If we draw the lines $x = x_i$ and $y = y_i$ $(i = 1, 2, \cdots, n)$ then these lines partition Ω into $(n + 1)^2$ sets $B_1, B_2, \cdots, B_J$ which give a refinement of the partition $A_1, A_2, \cdots, A_k, (\sum_{i=1}^k A_i)^c$ on which the estimate must be constant. Clearly the estimate must also be constant on these sets.

Now let n_i denote the number of observations in B_i and

$$g_n = \sum_{i=1}^J n_i \cdot [n \cdot \mu(B_i)]^{-1} \cdot I_{B_i} .$$

If we let $\mathfrak{L}$ denote the σ-lattice induced by the partial ordering then our estimate of f is given by $\hat{f}_n = E_\mu(g_n \mid \mathfrak{L}_n)$ where $\mathfrak{L}_n$ consists of Ω and $\varnothing$ together with all those sets of the form $B_{i_1} + B_{i_2} + \cdots + B_{i_m}$ which are in $\mathfrak{L}$.

THEOREM 4.2. *If g_n and $\hat{f}_n$ are the functions described above then*

$$\hat{f}_n = E_\mu(g_n \mid \mathfrak{L}).$$

PROOF. As in the proof of Theorem 3.4 the only difficulty involved is in showing that if h is $\mathfrak{L}$-measurable then the function h' defined by

$$h' = \sum_{i=1}^J [\mu(B_i)]^{-1} \cdot [\int_{B_i} h \, d\mu] \cdot I_{B_i} ,$$

is $\mathfrak{L}_n$-measurable. However a function is $\mathfrak{L}_n$-measurable if and only if it is constant on each of the sets $B_1, B_2, \cdots, B_J$ and is isotone with respect to the partial ordering. To see that it is isotone with respect to the partial ordering we compare its value on adjacent sets B_i and B_j. First consider the case where B_i and B_j have exactly one boundary point ν_0 in common and $\nu_1 \leqq \nu_0 \leqq \nu_2$ for $\nu_1 \varepsilon B_i$ and $\nu_2 \varepsilon B_j$. Next consider the case where B_i and B_j are separated by one of the lines $x = x_i$. Partition B_i and B_j by equally spaced horizontal lines, use the first result and take the limit as the distance between these lines goes to zero. The third case follows in the same way.

For each positive integer m partition Ω into 2^{2m} squares each having the width 2^{-m}. Let $\mathfrak{F}_m$ be the σ-field of subsets of Ω generated by these 2^{2m} squares.

Lemma 4.1. *For any $L \varepsilon \mathfrak{L}$ there are sets E_0 and E_1 in $\mathfrak{F}_m$ such that $E_0 \subset L \subset E_1$ and both $L - E_0$ and $E_1 - L$ have μ measure less than $2/2^m$.*

Proof. Let E_0 be the union of all those members of the partition generating $\mathfrak{F}_m$ which are entirely contained in L. Let E_1 be the union of all those members of the partition which have a non-null intersection with L. Note that whenever L contains a point in Ω it must contain all points greater than or equal to that point so that the set $E_1 - E_0$ can contain at most $2 \cdot 2^m - 1$ members of the partition. The desired conclusion follows from this observation.

Recall that the probability that any one observation will be in a set A in $\mathfrak{A}$ is given by $\int_A f\,d\mu$. Define the probability measure γ on $\mathfrak{A}$ by $\gamma(A) = \int_A f\,d\mu$. For each positive number δ let $\sum_n(\delta)$ be the collection of all those sets in $\mathfrak{A}$ of the form $L - L'$ where both L and L' are in $\mathfrak{L}$ and $\gamma(L - L') \geqq \delta$. For a given sequence of observations let $N_n(A)$ denote the number of the first n observations which are in A.

Theorem 4.3. *If f is bounded then for each $\delta > 0$*

$$\lim_n \sup_{A \varepsilon \Sigma_n(\delta)} |(1/n) \cdot N_n(A)(\gamma(A))^{-1} - 1| = 0,$$

with probability one.

Proof. Let R be the set of all sample sequences for which the above limit is zero. Let T be the set of all sample sequences for which

$$\lim_n (1/n) \cdot N_n(S)(\gamma(S))^{-1} = 1$$

for every set S which is a member of some $\mathfrak{F}_m$ and has positive γ measure. It can be shown, using the Borel strong law of large numbers, that T has probability one. We shall show that T is a subset of R.

Let ℓ be an upper bound on the values of f. Suppose that we are dealing with a sample sequence in T and that η is an arbitrary positive number. We may, without loss of generality, assume that $\eta < \frac{1}{6}$.

Choose m so that $(\frac{1}{2})^m < (\eta \cdot \delta)/(48 \cdot \ell)$ and $(\frac{1}{2})^m < \delta(/8 \cdot \ell.)$ Since $\mathfrak{F}_m$ is finite we can choose N such that $n \geqq N$ implies that

$$|(1/n) \cdot N_n(S)(\gamma(S))^{-1} - 1| < \eta/6$$

for all S in $\mathfrak{F}_m$ such that $\gamma(S) > 0$.

Now suppose $A \varepsilon \sum_n(\delta)$. Using Lemma 3.1 we can choose S_0 and S_1 in $\mathfrak{F}_m$ such that $S_0 \subset A \subset S_1$ and such that both $\mu(S_1 - A)$ and $\mu(A - S_0)$ are less than or equal to $4/2^m$. It follows that $\gamma(S_1 - S_0) \leqq (8\cdot\ell)/2^m < (\eta\cdot\delta)/6$. Further it follows from the second restriction on $(\frac{1}{2})^m$ that both $\gamma(S_1)$ and $\gamma(S_0)$ are greater than $\delta/2$. Now

$$(1/n)\cdot N_n(S_0)(\gamma(S_0))^{-1}\cdot\gamma(S_0)(\gamma(S_1))^{-1} \leqq (1/n)\cdot N_n(A)(\gamma(A))^{-1}$$
$$\leqq (1/n)\cdot N_n(S_1)(\gamma(S_1))^{-1}\cdot\gamma(S_1)(\gamma(S_0))^{-1}.$$

Hence

$$(1 - \eta/6)\cdot(1 - \eta/3) \leqq (1/n)\cdot N_n(A)(\gamma(A))^{-1} \leqq (1 + \eta/6)(1 + \eta/3).$$

The desired conclusion follows from this result.

Let $(x_1, y_1) < (x_2, y_2)$ if and only if $x_1 < x_2$ and $y_1 < y_2$. Since f is isotone with respect to the partial ordering we can define

$$f(\omega_0-) = \sup_{\omega<\omega_0} f(\omega)$$

and

$$f(\omega_0+) = \inf_{\omega>\omega_0} f(\omega).$$

THEOREM 4.4. *If the density f is $\mathfrak{L}$-measurable and bounded above then for every $\nu_0 \varepsilon \Omega$ we have*

$$f(\nu_0-) \leqq \lim_n \inf \hat{f}_n(\nu_0) \leqq \lim_n \sup \hat{f}_n(\nu_0) \leqq f(\nu_0+)$$

with probability one.

PROOF. We shall show that

$$f(\nu_0-) \leqq \lim_n \inf \hat{f}_n(\nu_0)$$

with probability one. If $f(\nu_0-) = 0$ the conclusion is obvious so suppose $\nu_1 < \nu_0$ and $f(\nu_1) > 0$. For each n let $\omega_{r(n)}$ be one of the first n observations and suppose that it is chosen so that it is greater than ν_1 and so that there are no observations greater than ν_1 and less than $\omega_{r(n)}$. Since $f(\nu_1) > 0$ it follows from the Borel strong law of large numbers that $\lim_n \omega_{r(n)} = \nu_1$ with probability one.

Let $L(\omega_{r(n)})$ be the smallest member of $\mathfrak{L}$ containing $\omega_{r(n)}$ and let $P = \{\omega; \hat{f}_n(\omega) > \hat{f}_n(\omega_0)\}$. As in the proof of Theorem 4.1 we can use the representation theorem in [6] to conclude that

$$\hat{f}_n(\nu_0) \geqq (1/n)N_n[L(\omega_{r(n)}) - P]\{\nu[L(\omega_{r(n)}) - P]\}^{-1}f(\nu_1)$$

with probability one. Using Theorem 4.3 we infer that limit of the term on the right hand side of the above inequality is $f(\nu_1)$ with probability one. The desired conclusion follows easily.

COROLLARY 4.3. *If besides satisfying the hypotheses of Theorem 4.4, f is continuous at ν_0 then*

$$\lim_n \hat{f}_n(\nu_0) = f(\nu_0)$$

with probability one.

5. On estimating a density on a non-totally finite measure space. We now suppose that $(\Omega, \mathcal{A}, \mu)$ is an arbitrary measure space (not necessarily totally finite). We shall assume that the σ-lattice $\mathcal{L}$ satisfies Assumptions (i) and (ii) of Section 2 so that a function f is $\mathcal{L}$-measurable if and only if it is isotone with respect to the almost partial ordering induced by $\mathcal{L}$. Let L_0 denote the smallest member of $\mathcal{L}$ which contains all of the observed points ($L_0 = \bigcup_{i=1}^{n} L(\omega_i)$). We consider three possibilities.

First, if $\mu(L_0) = 0$ then given any set L in $\mathcal{L}$ of positive but finite measure we could define an $\mathcal{L}$-measurable density g which is as large as we want at the observed points. Define g as follows:

$$\begin{aligned} g(\omega) &= k > 1/\mu(L), && \omega \,\varepsilon\, L_0, \\ &= 1/\mu(L), && \omega \,\varepsilon\, L - L_0, \\ &= 0, && \text{otherwise.} \end{aligned}$$

Next, suppose $\mu(L_0) = \infty$. Then $\mu(L(\omega_i)) = \infty$ for some i. Now any function which is $\mathcal{L}$-measurable and positive at ω_i cannot be integrable since $\omega_i \ll \omega$ for each ω in $L(\omega_i)$. Hence any estimate must be zero at one of the observed points.

Finally if $0 < \mu(L_0) < \infty$ then it is easy to see that any estimate must be zero on $\Omega - L_0$. Using this observation it can be shown that our problem reduces to estimating an $\mathcal{L}^*$-measurable density on the totally finite measure space $(\Omega^*, \mathcal{A}^*, \mu^*)$ where $\Omega^* = L_0$ and μ^*, $\mathcal{A}^*$, and $\mathcal{L}^*$ are the restrictions of μ, $\mathcal{A}$, and $\mathcal{L}$ to Ω^*.

Our estimate $\hat{f}$ is then the extension to Ω of the function $\hat{f}^* = E_{\mu^*}(g^* \mid \mathcal{L}_n^*)$ obtained by defining it to be zero on $\Omega - L_0$. Now let $\mathcal{L}_n$ be the σ-lattice of subsets of Ω obtained by adding Ω to $\mathcal{L}_n^*$ and let g be the extension of g^* to Ω obtained by defining it to be zero on $\Omega - L_0$. It can then be shown that our estimate $\hat{f}$ is equal to $E_\mu(g \mid \mathcal{L}_n)$ by simply verifying that $\hat{f}$ has all the properties of such a conditional expectation. Note that the definition for a conditional expectation on a non-totally finite measure space is different from the definition given in this paper for a totally finite measure space (cf. [1]).

In the special case where Ω is the real line, $\mathcal{A}$ is the collection of Borel sets, μ is Lebesgue measure and $\mathcal{L}$ is such that $\mathcal{L}$-measurability is equivalent to being unimodal at the real number a, our consistency theorem still holds. For example we can say that if f is $\mathcal{L}$-measurable and if $\nu_0 < a$ then

$$f(\nu_0-) \leqq \lim_n \inf \hat{f}_n(\nu_0) \leqq \lim_n \sup \hat{f}_n(\nu_0) \leqq f(\nu_0+),$$

with probability one.

6. Acknowledgment. I wish to thank Professor H. D. Brunk for suggesting the above investigation and for his guidance as my thesis advisor at the University of Missouri. I also wish to thank Professor D. L. Hanson for several valuable suggestions.

REFERENCES

[1] Brunk, H. D. (1963). On an extension of the concept conditional expectation. *Proc. Amer. Math. Soc.* **14** 298–304.

[2] Brunk, H. D. (1965). Conditional expectation given a σ-lattice and applications. *Ann. Math. Statist.* **36** 1339–1350.

[3] Grenander, Ulf (1956). On the theory of mortality measurement. *Skand. Aktuarietidskr.* **39** 125–153.

[4] Marshall, Albert W. and Proschan, Frank (1965). Maximum likelihood estimation for distributions with monotone failure rate. *Ann. Math. Statist* **36** 69–77.

[5] Robertson, Tim (1965). A note on the reciprocal of the conditional expectation of a positive random variable. *Ann. Math. Statist.* **36** 1302–1305.

[6] Robertson, Tim (1966). A representation for conditional expectations given σ-lattices. *Ann. Math. Statist.* **37** 1279–1283.