The Annals of Mathematical Statistics
1969, Vol. 40, No. 5, 1661–1667

A NOTE ON ESTIMATING A UNIMODAL DENSITY[1]

BY EDWARD J. WEGMAN

University of North Carolina

1. Introduction. This paper is concerned with the problem of estimating a unimodal density with unknown mode. Robertson [5] has shown in the case that the mode is known a solution can be represented as a conditional expectation given a σ-lattice. Brunk [2] discusses such conditional expectations as well as other problems.

A σ-lattice, $\mathfrak{L}$, of subsets of a measure space $(\Omega, \mathfrak{A}, \mu)$ is a collection of subsets of Ω closed under countable unions and countable intersections and containing both ϕ and Ω. If Ω is the real line, then the collection of intervals containing a fixed point, m, is a σ-lattice which we shall denote as $\mathfrak{L}(m)$. A function, f, is measurable with respect to a σ-lattice, $\mathfrak{L}$, if the set, $[f > a]$, is in $\mathfrak{L}$ for each real a. In this paper, we shall say a function f is unimodal at M when f is measurable with respect to $\mathfrak{L}(M)$. This definition is equivalent to a more usual definition as seen in the following easily verified remark.

REMARK. A function f is unimodal at M if and only if f is non-decreasing at x for $x < M$ and f is nonincreasing at x for $x > M$.

Let Ω be the real line and let $\mu = \lambda$ be Lebesgue measure. Let L_2 be the set of square integrable functions and $L_2(\mathfrak{L})$ be those members of L_2 which are also measurable with respect to $\mathfrak{L}$. We shall adopt the following definition of conditional expectation with respect to a σ-lattice.

DEFINITION 1.1. If $f \in L_2$, then $g \in L_2(\mathfrak{L})$ is equal to $E(f \mid \mathfrak{L})$, the conditional expectation of f give $\mathfrak{L}$ if and only if

$$\int f\cdot\theta(g)\,d\lambda = \int g\cdot\theta(g)\,d\lambda \tag{1.1}$$

for every θ, a real-valued function such that $\theta(g) \in L_2$ and $\theta(0) = 0$, and

$$\int_A (f - g)\,d\lambda \leqq 0 \tag{1.2}$$

for each $A \in \mathfrak{L}$ with $0 < \lambda(A) < \infty$.

(Brunk [1] shows such a function, g, exists and is unique up to a set of Lebesgue measure 0). In this paper, the estimation of the density will be based upon a strongly consistent estimate of the mode such as those of Nadaraya [3] or of Venter [6]. (The author is aware of a third estimate, as of now unpublished, given by Robertson and Cryer).

2. An estimate of the density. The density, f, to be estimated has mode M which is unknown. Let us assume $y_1 \leqq y_2 \leqq \cdots \leqq y_n$ is the ordered sample of

Received 18 November 1968.

[1] This research arises in part from the author's doctoral dissertation submitted to the Department of Statistics, University of Iowa. It was supported in part by PHS Research Grant No. GM-14448-01 from the Division of General Medical Sciences and in part by U.S. Air Force Grant AFOSR-68-1415.

size n chosen according to the density f. Let $\{M_n\}$ be a sequence, (we shall not yet require $\{M_n\}$ to be a consistent estimate of M. In fact, our only restriction now is $y_1 \leqq M_n \leqq y_n$ for $n \geqq 2$.) Let $y_{q(n)}$ be the largest observation less than or equal to M_n. Let $A_1 = [y_1, y_2)$, $A_2 = [y_2, y_3)$, $\cdots$, $A_{q(n)} = [y_{q(n)}, M_n]$, $A_{q(n)+1} = [M_n, y_{q(n)+1}]$, $\cdots$, $A_n = (y_{n-1}, y_n]$. Notice, if M_n is an observation, $A_{q(n)} = \phi$ and will be deleted from this collection of intervals. With this convention in mind, let us define

$$\hat{g}_n = \sum_{i=1}^{n} n_i \cdot [n\lambda(A_i)]^{-1} \cdot I_{A_i} \tag{2.1}$$

where n_i is the number of observations in A_i and I_{A_i} is the indicator of A_i. In particular, $n_i = 1$ except when M_n is an observation, in which case $n_{q(n)} = 0$ and $n_{q(n)+1} = 2$. Finally let $\mathfrak{L}_n$ be the lattice of intervals containing M_n. Then $\hat{f}_n = E(\hat{g}_n \mid \mathfrak{L}_n)$, the conditional expectation of $\hat{g}_n$ given the σ-lattice of intervals containing M_n, is an estimate of the density. Robertson [5] was able to show that $\hat{f}_n$ is an unimodal maximum likelihood estimate with mode M_n when M_n is not an observation. If M_n is an observation, we could just as well have written $A_{q(n)-1} = [y_{q(n)-1}, M_n]$ and $A_{q(n)+1} = (M_n, y_{q(n)+1}]$ instead of $A_{q(n)-1} = [y_{q(n)-1}, M_n)$ and $A_{q(n)+1} = [M_n, y_{q(n)+1}]$. This shift may affect the likelihood product, whereas in case M_n is not an observation, there is no effect. See [7] for a construction of the maximum likelihood estimate when the mode is an observation.

3. Conditional expectations of the true density. In this section we shall represent the conditional expectation of the true density with respect to the lattice of intervals containing a point in $\{y : f(y) > 0\}$, the support of f. Since the definition of conditional expectation requires that f be in L_2, we shall make that assumption here.

THEOREM 3.1. *If f is a density function unimodal at M and m is a point in the support of f, then*

(i) *If $m > M$, there is an interval $[a, m]$ such that*
$$E(f \mid \mathfrak{L}(m)) = f \text{ on } [a, m]^c \quad \text{and}$$
$$E(f \mid \mathfrak{L}(m)) = (m - a)^{-1} \cdot \textstyle\int_{[a,m]} f\, d\lambda \text{ on } [a, m].$$
(ii) *If $m < M$, there is an interval $[m, b]$ such that*
$$E(f \mid \mathfrak{L}(m)) = f \text{ on } [m, b]^c \quad \text{and}$$
$$E(f \mid \mathfrak{L}(m)) = (b - m)^{-1} \cdot \textstyle\int_{[m,b]} f\, d\lambda \text{ on } [m, b].$$
(iii) If $m = M$, $E(f \mid \mathfrak{L}(m)) = f$ *everywhere.*

PROOF. Case (iii) is easily verified and case (ii) can be shown in a manner similar to case (i). We shall show case (i).

Let $a = \sup \{y \leqq M : (m - y)^{-1} \cdot \int_{[y,m]} f\, d\lambda \geqq f(y)\}$. To observe that the set is not empty, notice for any $y \leqq M$, for which $f(y) \leqq f(m)$, y is an element of the set. Let us define $f^*(y) = f(y)$ for y belonging to $[a, m]^c$ and $f^*(y) = (m - a)^{-1} \cdot \int_{[a,m]} f\, d\lambda$ for y belonging to $[a, m]$. We want to show $f^* = E(f \mid \mathfrak{L}(m))$. Clearly $f^*(z) \leqq f^*(m)$ for each z, so it is not difficult to verify f^* is unimodal at m. Hence $f^* \varepsilon L_2(\mathfrak{L}(m))$. We need only show (1.1) and (1.2) hold.

Let $a' = \sup \{y < m : (m - a)^{-1} \cdot \int_{[a,m]} f\, d\lambda \leqq f(y)\}$. For y such that $a \leqq y < a'$,

$(m - a)^{-1} \cdot \int_{[a,m]} f \, d\lambda = f^*(y) \leqq f(y)$ and for y such that $a' < y \leqq m, f^*(y) \geqq f(y)$. Now let A be any element of $\mathfrak{L}(m)$ with $0 < \lambda(A) < \infty$. Since we are dealing with Lebesgue measure, we may assume A is a closed interval, say $[b_1, b_2]$. Since $b_2 \geqq m$, we may write,

$$\int_{[b_1,b_2]} (f - f^*) \, d\lambda = \int_{[b_1,m]} (f - f^*) \, d\lambda. \tag{3.1}$$

Now, if $a' \leqq b_1 \leqq m$, since we have $f^* - f \geqq 0$ on $(a', m]$

$$\int_{[b_1,m]} (f - f^*) \, d\lambda \leqq 0. \tag{3.2}$$

If $a \leqq b_1 < a'$, since $f - f^* \geqq 0$ on (a, a'),

$$\int_{[b_1,a']} (f - f^*) \, d\lambda \leqq \int_{[a,a']} (f - f^*) \, d\lambda. \tag{3.3}$$

But $\int_{[a,m]} (f - f^*) \, d\lambda = 0$, so that

$$\int_{[a,a']} (f - f^*) \, d\lambda = -\int_{[a',m]} (f - f^*) \, d\lambda.$$

This together with (3.3) gives us

$$\int_{[b_1,m]} (f - f^*) \, d\lambda \leqq 0. \tag{3.4}$$

Finally, if $b_1 < a$

$$\int_{[b_1,m]} (f - f^*) \, d\lambda = 0. \tag{3.5}$$

Hence, combining (3.1) with either (3.2), (3.4), or (3.5), we have for A in $\mathfrak{L}(m)$ such that $0 < \lambda(A) < \infty$, $\int_A (f - f^*) \, d\lambda \leqq 0$.

This shows condition (1.2) holds. To show (1.1) holds, let θ be any Borel function with $\theta(f^*) \varepsilon L_2$ and $\theta(0) = 0$. Then $\int (f - f^*) \cdot \theta(f^*) \, d\lambda = \int_{[a,m]} (f - f^*) \cdot \theta(f^*(m)) \, d\lambda = \theta(f^*(m)) \cdot \int_{[a,m]} (f - f^*) \, d\lambda$. But $\int_{[a,m]} (f - f^*) \, d\lambda = 0$, so $\int (f - f^*) \cdot \theta(f^*) \, d\lambda = 0$. Thus $f^* = E(f \mid \mathfrak{L}(m))$ and the proof is complete.

4. Consistency. Let F be the distribution corresponding to f, and F_n the empirical distribution. Let $\Omega' = [\lim_{n \to \infty} \sup_y |F_n(y) - F(y)| = 0]$. It is well known that this set has probability one. We shall assume henceforth that all observations arise from points in Ω'. If F is the distribution function corresponding to a random variable Y with density f, it is known that $Z = k \cdot F(Y)$ is distributed uniformly over the set $[0, k]$. Moreover, if $Y_1 \leqq Y_2 \leqq \cdots \leqq Y_n$ are the order statistics corresponding to a sample from F, then $Z_1 = k \cdot F(Y_1), \cdots, Z_n = k \cdot F(Y_n)$ are distributed like the order statistics from a uniform distribution with range 0 to k. We shall denote this uniform distribution by H^k. If $y_1 \leqq y_2 \leqq \cdots \leqq y_n$ is an ordered sample and $z_1 = k \cdot F(y_1), \cdots, z_n = k \cdot F(y_n)$ are the transformed points, we denote the empirical distribution based on $z_1, \cdots, z_n$ by H_n^k. Since

$$\sup_z |H_n^k(z) - H^k(z)| = \sup_y |F_n(y) - F(y)|,$$

$$\Omega' = [\lim_{n \to \infty} \sup_z |H_n^k(z) - H^k(z)| = 0],$$

for each $k > 0$. The reader may satisfy himself that for points in Ω', the following statements are true.

1. The largest observation less than and the smallest observation greater than a number in the support of f converge to that number.

2. Corresponding to every pair, $r_1 < r_2$, of numbers in the support of f, there is eventually a pair of observations $y_{r_1(n)}$ and $y_{r_2(n)}$ satisfying

$$r_1 < y_{r_1(n)} < y_{r_2(n)} < r_2 .$$

We shall implicitly use these statements in the proof of the consistency theorem.

Our estimate is given by $\hat{f}_n = E(\mathfrak{g}_n \mid \mathfrak{L}_n)$. Suppose y_0 is an arbitrary point in the support of f and let $t = \hat{f}_n(y_0)$. Let $P_t = [\hat{f}_n > t]$ and $T_t = [\hat{f}_n \geqq t]$. If $\mathcal{K}_1(T_t) = \{L \,\varepsilon\, \mathfrak{L}_n : \lambda(T_t - L) > 0\}$ and $\mathcal{K}_2(P_t) = \{L \,\varepsilon\, \mathfrak{L}_n : \lambda(L - P_t) > 0\}$, the result of Robertson [4] gives us

$$\hat{f}_n(y_0) = \inf_{L\varepsilon\mathcal{K}_1(T_t)} [\lambda(T_t - L)]^{-1} \cdot \int_{T_t - L} \mathfrak{g}_n \, d\lambda \tag{4.1}$$

and

$$\hat{f}_n(y_0) = \sup_{L\varepsilon\mathcal{K}_2(P_t)} [\lambda(L - P_t)]^{-1} \cdot \int_{L - P_t} \mathfrak{g}_n \, d\lambda. \tag{4.2}$$

(Notice that Robertson's theorem is stated for finite measure spaces. Once the sample in question has been chosen we may restrict our attention to $[y_1, y_n]$ for which $\lambda[y_1, y_n] < \infty$.) The proof of the next theorem uses methods similar to those of Robertson [5].

THEOREM 4.1. *If f is a unimodal density with mode M and $\{M_n\}$ is a sequence converging to m in the support of f, then let $f_m = E(f \mid \mathfrak{L}(m))$ and let $\hat{f}_n$ be the estimate described in Section* 2. *For every $y_0 < m$*

$$f_m(y_0^-) \leqq \liminf \hat{f}_n(y_0) \leqq \limsup \hat{f}_n(y_0) \leqq f_m(y_0^+) \tag{4.3}$$

and for every $y_0 > m$

$$f_m(y_0^-) \geqq \limsup \hat{f}_n(y_0) \geqq \liminf \hat{f}_n(y_0) \geqq f_m(y_0^+) \tag{4.4}$$

Inequalities (4.3) *and* (4.4) *hold for points in Ω', hence with probability one.*

PROOF. Inequalities (4.3) and (4.4) are proven in a similar manner although they will require different forms of the representation theorem given in [4]. We shall restrict our attention to (4.3). Let us suppose $m > M$ and let a and a' be defined as in Section 3. Let $y_1 \leqq y_2 \leqq \cdots \leqq y_n$ be the ordered sample based on a point from Ω'. Let $x < m$ and x a number in the interior of the support of f. Let $y_{s(n)}$ be the largest observation less than x. There is a number $x_1 < x$ in the support of f since the interior is an open interval (including possibly $(-\infty, \infty)$). Let $y_{u(n)}$ be the smallest observation greater than x_1. Let n be sufficiently large so that $y_{s(n)} > y_{u(n)}$. Let $t = \hat{f}_n(x)$ and $P_t = [\hat{f}_n > t]$. Since $\hat{f}_n$ is constant on intervals between observations (a fact demonstrated by Robertson in [5]) and unimodal with mode $M_n > x$ for sufficiently large n, $P_t = [y_{j(n)}, y_{i(n)}]$. Notice for sufficiently large n, $[y_{u(n)}, y_{i(n)}] \,\varepsilon\, \mathcal{K}_2(P_t)$ so that by (4.2),

$$\hat{f}_n(x) \geqq (y_{j(n)} - y_{u(n)})^{-1} \cdot \int_{[y_{u(n)}, y_{j(n)})} \mathfrak{g}_n \, d\lambda$$

for sufficiently large n. But $\int_{[y_{u(n)}, y_{j(n)})} \hat{g}_n \, d\lambda = n^{-1} \cdot (j(n) - u(n))$, so that for sufficiently large n

$$\hat{f}_n(x) \geqq n^{-1} \cdot (j(n) - u(n)) \cdot (y_{j(n)} - y_{u(n)})^{-1}.$$

If $Z_i = f(x_1)^{-1} \cdot F(y_i)$, $i = 1, 2, \cdots, n$, then since f is nondecreasing on $(-\infty, m)$ $Z_{j(n)} - Z_{u(n)} \geqq y_{j(n)} - y_{u(n)}$, so that for sufficiently large n

$$\hat{f}_n(x) \geqq n^{-1} \cdot (j(n) - u(n)) \cdot (Z_{j(n)} - Z_{u(n)})^{-1}.$$

If $k = f(x_1)^{-1}$, then $n^{-1} \cdot (j(n) - u(n)) = H_n^k(Z_{j(n)}) - H_n^k(Z_{u(n)})$. Finally, for sufficiently large n

$$\hat{f}_n(x) \geqq [H_n^k(Z_{j(n)}) - H_n^k(Z_{u(n)})] \cdot (Z_{j(n)} - Z_{u(n)})^{-1}.$$

Since $\lim_{n \to \infty} Z_{u(n)} = [f(x_1)]^{-1} \cdot F(x_1)$, $\liminf [Z_{j(n)} - Z_{u(n)}] > 0$. Thus because of the uniform convergence of H_n^k to H^k,

$$\lim_{n \to \infty} [H_n^k(Z_{j(n)}) - H_n^k(Z_{u(n)})] \cdot (Z_{j(n)} - Z_{u(n)})^{-1} = 1/k = f(x_1),$$

so that $\liminf \hat{f}_n(x) \geqq f(x_1)$. In the discussion of Section 3, we saw for $x \leqq a'$, $f_m(x) \leqq f(x)$. Hence we have

$$f_m(x^-) \leqq \liminf \hat{f}_n(x) \quad \text{for} \quad x \leqq a'. \tag{4.5}$$

Using the representation given by (4.1), we can show in a similar manner that

$$f(x^+) \geqq \limsup \hat{f}_n(x) \quad \textit{for} \quad x < m.$$

Again by the discussion of Section 3, for $x < a, f_m(x) = f(x)$ and for $a' < x < m$, $f_m(x) \geqq f(x)$. Clearly then for $y_0 < a$

$$f_m(y_0^-) \leqq \liminf \hat{f}_n(y_0) \leqq \limsup \hat{f}_n(y_0) \leqq f_m(y_0^+).$$

Moreover, for $a' < x < m$,

$$f_m(x^+) \geqq \limsup \hat{f}_n(x) \tag{4.6}$$

Now let y_0 be chosen in (a, m). Then there is a pair (x_1, x_2) such that $x_1 < y_0 < x_2$ and $a < x_1 < a'$ and $a' < x_2 < m$. Clearly by (4.5) and (4.6)

$$f_m(x_1^-) \leqq \liminf \hat{f}_n(y_0) \leqq \limsup \hat{f}_n(y_0) \leqq f_m(x_2^+).$$

Since f_m is constant on $[a, m]$, we have

$$\lim_{n \to \infty} \hat{f}_n(y_0) = f_m(y_0).$$

This provides the desired conclusion, (4.3) if y_0 is in the interior of the support of f. If $f(y_0^-) = 0$, the result is clear. The remaining cases may be proven in a similar manner.

If f is continuous, then $\lim_{n \to \infty} \hat{f}_n(y_0) = f_m(y_0)$ for $y_0 \neq m$. Using this and applying methods similar to those of the Glivenko-Cantelli Theorem, we have the following corollary.

COROLLARY 4.1. *If f is continuous and the conditions of Theorem* 4.1 *hold, then*

$\hat{f}_n$ converges to f_m uniformly except on an interval of arbitrarily small measure containing m. This holds for all points in Ω', hence with probability one.

If the limit of the sequence $\{M_n\}$ is M, which is the case if M_n is a strongly consistent estimate of the mode, then we may state the following corollary.

COROLLARY 4.2. *If $\{M_n\}$ converges to M, the mode of f, with probability one, then with probability one,*

$$f(x^-) \leqq \liminf \hat{f}_n(x) \leqq \limsup \hat{f}_n(x) \leqq f(x^+) \quad \textit{for} \quad x < M$$

and

$$f(x^-) \geqq \limsup \hat{f}_n(x) \geqq \liminf \hat{f}_n(x) \geqq f(x^+) \quad \textit{for} \quad x > M.$$

In addition, if f is continuous, with probability one $\hat{f}_n$ converges to f uniformly except on an interval of arbitrarily small measure containing M.

Of course, $\hat{f}_n$ also may have the maximum likelihood property mentioned in Section 2. The maximum likelihood property causes $\hat{f}_n$ to be a step function, so that if f is continuous we may be willing to sacrifice the maximum likelihood property in order to obtain a continuous estimate. If we assure our continuous estimate is sufficiently close to $\hat{f}_n$, then our continuous estimate will also be consistent. Clearly, there is no unique way to make $\hat{f}_n$ continuous. In the following, let f be a continuous density.

THEOREM 4.2. *If $\{f_n^*\}$ is a sequence of continuous density functions such that*

(1) f_n^* is unimodal with mode M_n,

(2) $f_n^*(M_n) = \hat{f}_n(M_n)$,

(3) If $\hat{f}_n$ is constant on $B_1, \cdots, B_k$, then $f_n^* = \hat{f}_n$ at least once on each $B_j, j = 1, \cdots, k$, then with probability one f_n^* converges to f uniformly except on an interval of arbitrarily small measure containing M.

PROOF. Let A be the interval of arbitrarily small measure. By Corollary 4.2, $\sup_{x \varepsilon A^c} |\hat{f}_n(x) - f(x)|$ converges to zero. Let $\epsilon > 0$ and choose n sufficiently large so that

$$\sup_{x \varepsilon A^c} |\hat{f}_n(x) - f(x)| < \epsilon/3.$$

It is easy to see that conditions (1), (2), and (3) are sufficient to imply that for any x in the support

$$|f_n^*(x) - \hat{f}_n(x)| \leqq \sup_{y_i} |\hat{f}_n(y_i^+) - \hat{f}_n(y_i^-)|,$$

where $y_i, i = 1, \cdots, n$ are the set of observations. But adding and subtracting $f(y_i)$ and using the triangle inequality,

$$\sup_{y_i \varepsilon A^c} |\hat{f}_n(y_i^+) - \hat{f}_n(y_i^-)| \leqq 2 \sup_{x \varepsilon A^c} |\hat{f}_n(x) - f(x)|.$$

Hence

$$\sup_{x \varepsilon A^c} |f_n^*(x) - \hat{f}_n(x)| \leqq (2/3)\epsilon,$$

or for sufficiently large n

$$\sup_{x \varepsilon A^c} |f_n^*(x) - f(x)| \leqq \epsilon.$$

This completes the proof.

5. Acknowledgments. I wish to thank Professor Tim Robertson for suggesting the topic from which this work arose and for his guidance as my thesis advisor at the University of Iowa.

REFERENCES

[1] BRUNK, H. D. (1963). On an extension of the concept of 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] NADARAYA, E. A. (1966). On nonparametric estimates of density functions and regression curves. *Theor. Probability Appl.* **10** 186–190.

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

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

[6] VENTER, J. H. (1967). On estimation of the mode. *Ann. Math. Statist.* **38** 1446–1455.

[7] WEGMAN, E. J. (1968). On Estimating a Unimodal Density. Ph.D. dissertation, Univ. Iowa.