MAXIMUM LIKELIHOOD ESTIMATION OF A
UNIMODAL DENSITY FUNCTION

by

EDWARD J. WEGMAN
Department of Statistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No.608

January 1969

This research arises in part from the author's doctoral dissertation submitted to the Department of Statistics, The Univeristy of Iowa. Supported in part by PHS Research Grant No. GM-14448-01 from the Div. of General Medical Sciences and in part by N. S. F. Development Grant GU-2059.

Maximum Likelihood Estimation of a Unimodal Density Function

Edward J. Wegman

1. Introduction. Robertson [4] has described a maximum likelihood estimate of a unimodal density when the mode is known. This estimate is represented as a conditional expectation given a σ-lattice. Discussions of such conditional expectations are given by Brunk [1] and [2].

A σ-lattice, L, of subsets of a measure space (Ω, A, μ) is a collection of subsets of Ω closed under countable unions and countable intersections and containing both ϕ and Ω. A function f is measurable with respect to a σ-lattice, L, if the set, $[f > a]$, is in L for each real a. If Ω is the real line, A is the collection of Borel sets, and $\mu = \lambda$ is Lebesgue measure, let L_2 be the set of square-integrable functions and $L_2(L)$ be those members of L_2 which are measurable with respect to L. Brunk [1] shows the following definition of conditional expectation with respect to L is suitable.

Definition. If $f \in L_2$, then $g \in L_2(L)$ is equal to $E(f|L)$, the conditional expectation of f given L, if and only if

(1.1) $$\int f \cdot \theta(g)d\lambda = \int g \cdot \theta(g)d\lambda$$

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

(1.2) $$\int_A (f-g)d\lambda \leq 0$$

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

The collection of intervals about a fixed point, m, together with ϕ is a σ-lattice which we denote as $L(m)$. A function, f, is unimodal at M

This research arises in part from the author's doctoral dissertation submitted to the Department of Statistics, The 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 N. S. F. Development Grant GU-2059.

by definition if f is measurable with respect to $L(M)$. It is not difficult to see that this is equivalent to f nondecreasing at $x < M$ and f nonincreasing at $x > M$. If f is unimodal at every point of an interval, I, then we call I the modal interval of f and we shall write $L(I)$ for the lattice of intervals containing I. Clearly, f has modal interval I if and only if f is measurable with respect to $L(I)$. The center of I will be called the center mode.

Robertson's estimate of the unimodal density is maximum likelihood where the mode is known. The present solution is a maximum likelihood estimate when the mode is unknown. A peculiar characteristic of Robertson's estimate (and also a related estimate found in Wegman [5]) is a "peaking" near the mode. Figure 1 is based on a Monte Carlo sample of size 90 from a triangular density. This is Robertson's maximum likelihood estimate with mode known to be $\frac{1}{2}$. To eliminate the peaking, at least partially, we shall require our estimate to have a modal interval of length ϵ, where ϵ is some fixed positive number.

This also will have the effect of uniformly bounding the estimate for all sample sizes by $\frac{1}{\epsilon}$. Figure 2 is the estimate as described in this paper. The modal interval extends from .493 to .518. In this estimate, .493 is an observation.

Table 1 is the tabulation of the density drawn in Figure 1. Table 2 is the tabulation for Figure 2. The reader should be warned that these computations were carried out on a computer and rounded. Thus, in the case of Figure 2, ϵ is actually .0234375 . Both estimates are based on a sample of size 90.

TABLE 1

Interval, I	Value of Estimate on I
(− ∞ , .011)	.0
[.011, .135)	.360
[.135, .248)	.490
[.248, .264)	.674
[.264, .345)	1.37
[.345, .414)	1.59
[.414, .432)	2.04
[.432, .448)	2.17
[.448, .493)	2.34
[.493, .500)	4.02
[.500, .505]	4.54
(.505, .510]	2.26
(.510, .553]	1.74
(.553, .583]	1.48
(.583, .834]	.973
(.834, .890]	.605
(.890, .915]	.427
(.915, .977]	.357
(.977, ∞)	.0

TABLE 2

Interval, I	Value of Estimate on I
(- ∞ , .011)	.0
[.011, .135)	.360
[.135, .248)	.490
[.248, .264)	.674
[.264, .345)	1.37
[.345, .414)	1.59
[.414, .432)	2.04
[.432, .448)	2.17
[.448, .493)	2.34
[.493, .518]	2.37
(.518, .553]	2.16
(.553, .583]	1.48
(.583, .834]	.973
(.834, .890]	.605
(.890, .915]	.427
(.915, .977]	.357
(.977, ∞)	.0

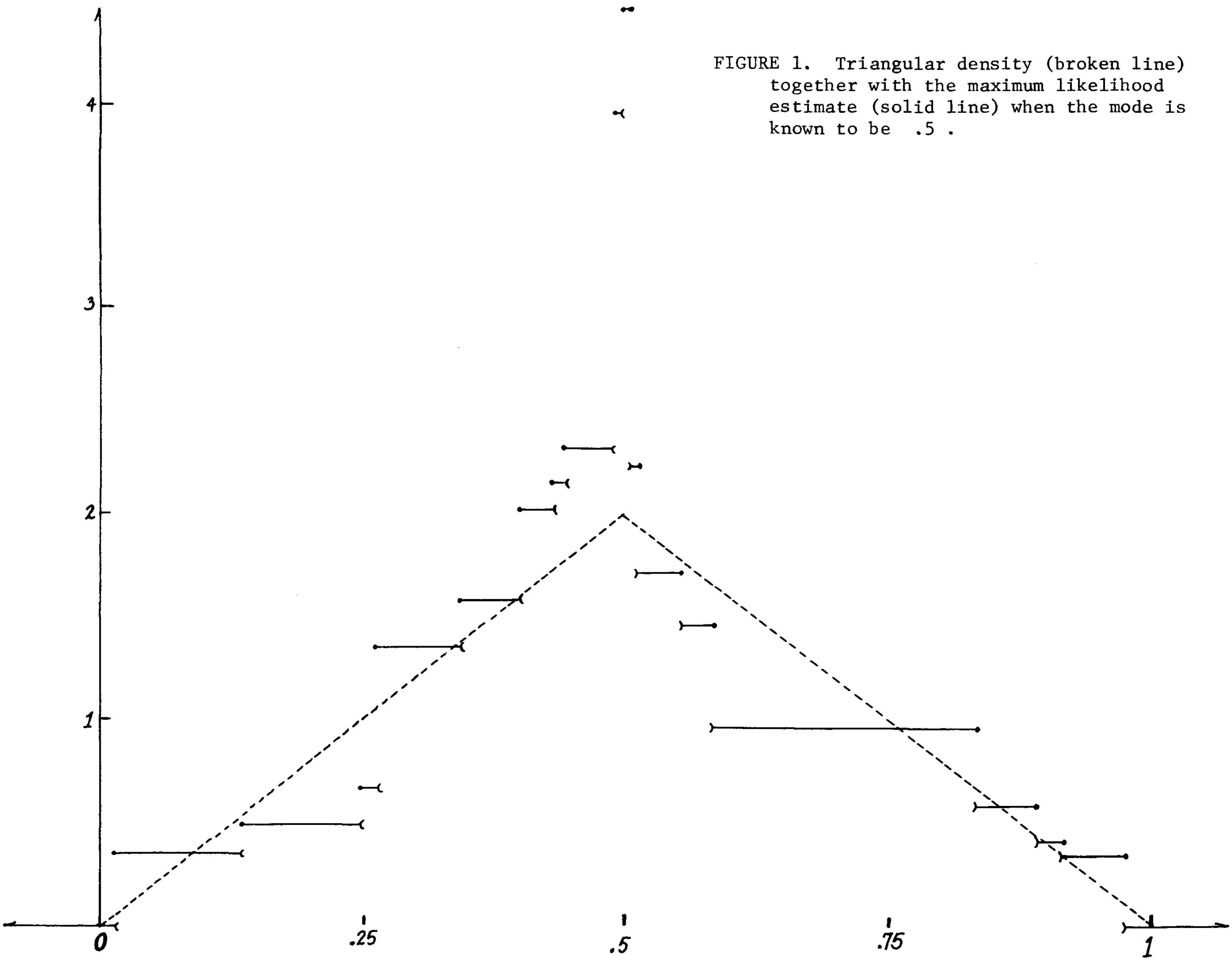
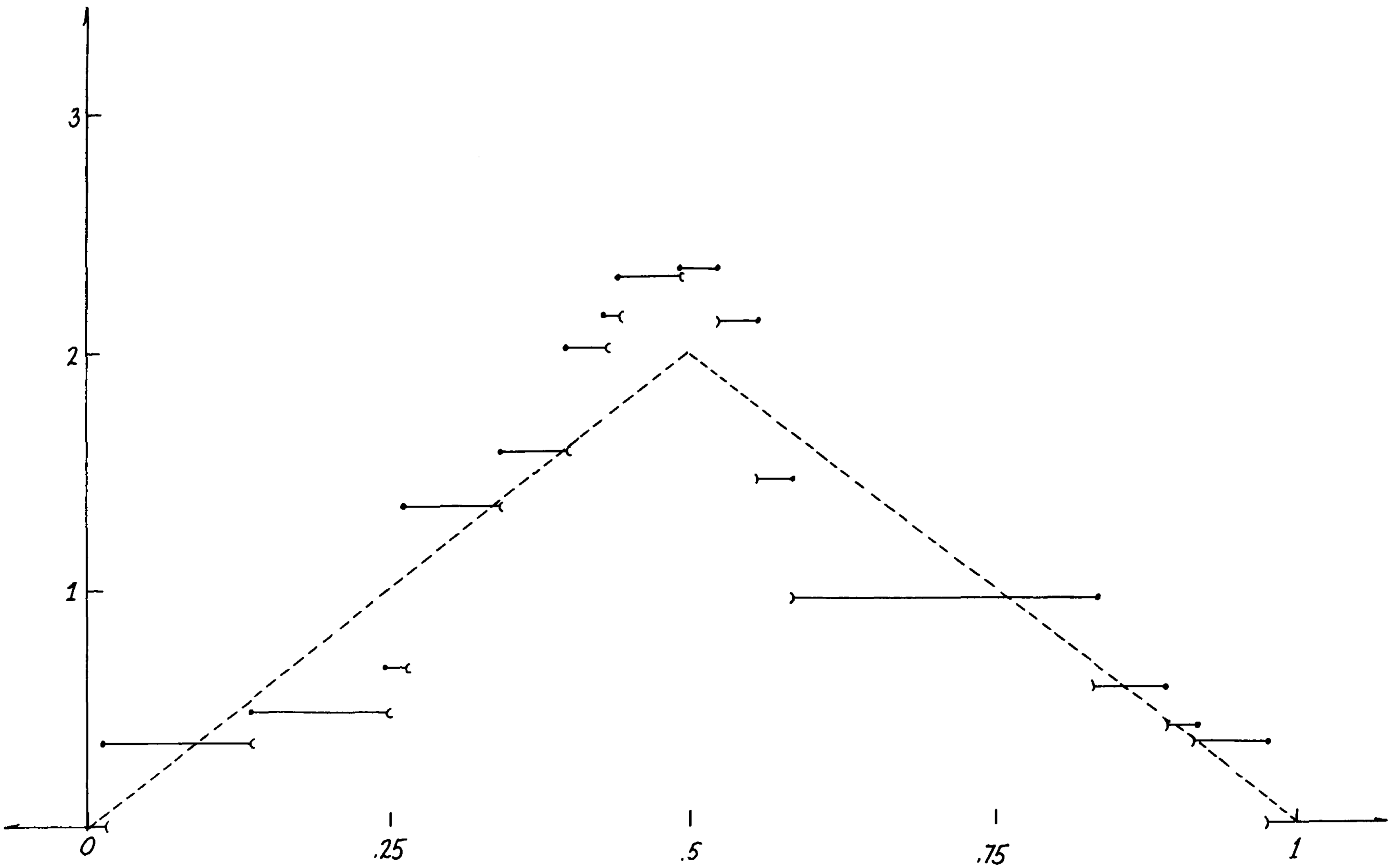


FIGURE 1. Triangular density (broken line) together with the maximum likelihood estimate (solid line) when the mode is known to be .5 .

FIGURE 2. Triangular density (broken line) and the maximum likelihood estimate (solid line) with modal interval of length .0234375 .



2. <u>The estimate</u>. Let us assume $y_1 \le y_2 \le \dots \le y_n$ is the ordered sample selected according to a unimodal density f. Let L and R be points such that $\lambda[L, R] = R - L = \epsilon$. Let $\ell(n)$ and $r(n)$ be the subscripts of the largest observation less than or equal to L and the smallest observation greater than or equal to R respectively. If h is any estimate which has $[L, R]$ for its modal interval, define $g(y)$ by

$$g(y) = \begin{cases} h(y_{\ell(n)}) & y_{\ell(n)} \le y < L \\ h(y_{r(n)}) & R < y \le y_{r(n)} \\ h(y) & \text{otherwise} . \end{cases}$$

It follows that $\hat{f} = [\int g d\lambda]^{-1}$. g has modal interval $[L, R]$ and has a likelihood product no smaller than that of h. Similarly, any estimate, which is to maximize the likelihood product, must be constant on every open interval joining consecutive observations in the complement of $[L, R]$, $[L, R]^c$. The problem then is to find an estimate which is constant on

$$A_1 = [y_1, y_2), \quad A_2 = [y_2, y_3), \dots, A_{\ell(n)} = [y_{\ell(n)}, L) ,$$

$$A_{\ell(n)+1} = [L, R], \quad A_{\ell(n)+2} = (R, y_{r(n)}], \dots, A_k = (y_{n-1}, y_n].$$

Recall $m = \frac{1}{2}(R + L)$ is the center mode and $L([L, R])$ is the lattice of intervals containing $[L, R]$. Let $\mathcal{A}_k$ denote the σ-field whose atoms are $A_1, A_2, \dots, A_k$ and $(\bigcup_{j=1}^{k} A_j)^c$. Finally, let $L_n([L, R])$ be the intersection of $L([L, R])$ with $\mathcal{A}_k$. An application of a theorem of Robertson [4] shows that the maximum likelihood estimate is given by

$$\hat{f}_{nm} = E(\hat{g}_{nm} | L_n([L, R]))$$

where

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

Here n_i is the number of observations in A_i and I_{A_i} is the indicator of A_i. Hence we know the form of the maximum likelihood estimate given the location of the modal interval. We wish to determine the location of the modal interval. The following lemma shows that we need only consider a finite number of locations for the modal interval.

<u>Lemma 2.1</u>. If $\hat{f}_{nm}$ is the maximum likelihood estimate, we may assume that one of the endpoints of the modal interval lies in the set $\{y_1, \ldots, y_n\}$.

<u>Proof</u>: Recall $L = m - \frac{1}{2}\epsilon$ and $R = m + \frac{1}{2}\epsilon$.

Suppose neither L nor R belongs to $\{y_1, \ldots, y_n\}$, but there are observations in $(-\infty, L)$, in $[L, R]$, and in (R, ∞). Let

$$\eta = \min\{L - y_{\ell(n)},\ y_{\ell(n)+1} - L,\ R - y_{r(n)-1},\ y_{r(n)} - R\}.$$

We may assume $\hat{f}_{nm}(y_{r(n)}) > \hat{f}_{nm}(y_{\ell(n)})$. (A similar argument holds for $\hat{f}_{nm}(y_{r(n)}) < \hat{f}_{nm}(y_{\ell(n)})$. Let $R' = R + \frac{1}{2}\eta$ and $L' = L + \frac{1}{2}\eta$. Define $g(y)$ by

$$g(y) = \begin{cases} \hat{f}_{nm}(y) & \text{for } y < L \text{ or } y > R' \\ \hat{f}_{nm}(y_{\ell(n)}) & \text{for } L \le y < L' \\ \epsilon^{-1} \cdot \{\int_{[L',R']} \hat{f}_{nm} d\lambda + \int_{[L,L']} (\hat{f}_{nm} - \hat{f}_{nm}(y_{\ell(n)}))d\lambda\} & \\ \qquad\qquad \text{for } L' \le y \le R'. & \end{cases}$$

Clearly g has modal interval $[L', R']$. $g = \hat{f}_{nm}$ on $(y_{\ell(n)}, y_{r(n)})^c$, so to show g has a larger likelihood product than $\hat{f}_{nm}$, we need only show $g > \hat{f}_{nm}$ on $[y_{\ell(n)+1}, y_{r(n)-1}]$.

Noting that $\hat{f}_{nm} = \hat{f}_{nm}(L)$ on $[L, R]$ and that $\hat{f}_{nm} = \hat{f}_{nm}(y_{r(n)})$ on $(R, R']$, we have on $[L', R']$

$$g(y) = \epsilon^{-1} \cdot \{\int_{[L',R]} \hat{f}_{nm}(L)\,d\lambda + \int_{[R,R']} \hat{f}_{nm}(y_{r(n)})\,d\lambda + \int_{[L,L']} \hat{f}_{nm}(L)\,d\lambda - \int_{[L,L']} \hat{f}_{nm}(y_{\ell(n)})\,d\lambda\} .$$

Combining the first and third terms of the right-hand side and observing $\lambda[L, R] = \epsilon$, we have for $y \in [L', R']$

$$g(y) = \hat{f}_{nm}(L) + \epsilon^{-1} \cdot \{\int_{[R,R']} \hat{f}_{nm}(y_{r(n)})\,d\lambda - \int_{[L,L']} \hat{f}_{nm}(y_{\ell(n)})\,d\lambda\} .$$

Finally, observing that both $\lambda[R, R']$ and $\lambda[L, L']$ are $\frac{1}{2}\eta$, that $\hat{f}_{nm}(y_{r(n)}) > \hat{f}_{nm}(y_{\ell(n)})$, and that $\hat{f}_{nm}(y) = \hat{f}_{nm}(L)$ on $[L', R]$,

$$g(y) > \hat{f}_{nm}(y) \quad \text{for all } y \text{ in } [L', R].$$

Thus it follows that

$$g(y) > \hat{f}_{nm}(y) \quad \text{for all } y \text{ in } [y_{\ell(n)+1}, y_{r(n)-1}] .$$

This implies that g has a larger likelihood product than $\hat{f}_{nm}$, which is contrary to choice of $\hat{f}_{nm}$. Hence either L or R must lie in $\{y_1, \ldots, y_n\}$ or $\hat{f}_{nm}(y_{\ell(n)}) = \hat{f}_{nm}(y_{r(n)})$. If the latter case holds, an estimate equally likely as $\hat{f}_{nm}$ may clearly be defined with modal interval having proper end points. If one or more of the sets $(-\infty, L)$, $[L, R]$, or (R, ∞) contain no observations, we may define an alternate function, g , in a similar manner to the above. This completes the proof.

There are at most $2n$ intervals of the form $[L, R]$ where one of L or R belongs to $\{y_1, \ldots, y_n\}$. If the modal interval is unspecified, compute $2n$ estimates $\hat{f}_{nm_i}$, $i = 1, 2, \ldots, 2n$, corresponding to these $2n$ intervals. Calculate the $2n$ likelihood products,

$$L_{nm_i} = \prod_{j=1}^{n} f_{nm_i}(y_j).$$

Select the maximum of these 2n numbers and let $\hat{f}_n$ be the corresponding estimate. Let m_n be the center mode. (Notice that if 2 or more of the L_{nm_i} are equal, we could have some ambiguity of definition of $\hat{f}_n$. Let us adopt the convention of choosing $\hat{f}_n$ to be the estimate with the smallest center mode). Clearly the likelihood product of $\hat{f}_n$ is as large as that of any other estimate.

Before closing this section, we note the following lemma that $L_n([L, R])$ may be replaced by $L([L, R])$.

<u>Lemma 2.2</u>. The maximum likelihood estimate $\hat{f}_{nm}$ which is measurable with respect to $L([L, R])$ is given by $E(\hat{g}_{nm}|L([L, R]))$, where $\hat{g}_{nm}$ is as before.

<u>Proof</u>: Let us abbreviate for purpose of this lemma $E(\hat{g}_{nm}|L_n([L, R]))$ by $\hat{f}$ and $\hat{g}_{nm}$ by $\hat{g}$. We want to show $\hat{f} = E(\hat{g}_{nm}|L([L, R]))$. Clearly, $\hat{f}$ is measurable with respect to $L([L, R])$. Property (1.1) is easy to verify, so let us show

$$\int_A (\hat{g} - \hat{f})d\lambda \le 0 \qquad \text{for } A \in L([L, R]) , \quad 0 < \lambda(A) < \infty .$$

We may assume A is a closed interval, [a, b] and suppose

$$y_i \le a < y_{i+1} \le y_{\ell(n)} \qquad \text{and} \qquad y_{r(n)} \le y_j < b \le y_{j+1} .$$

Then

$$\int_A (\hat{g}-\hat{f})d\lambda = \int_{[a,y_{i+1}]} (\hat{g}-\hat{f})d\lambda + \int_{[y_{i+1},y_j]} (\hat{g}-\hat{f})d\lambda + \int_{[y_j,b]} (\hat{g}-\hat{f})d\lambda.$$

The center term of the right-hand side must be non-positive by (1.2), since $[y_{i+1}, y_j] \in L_n([L, R])$. Thus, if $\int_A (\hat{g} - \hat{f})d\lambda > 0$, either

$$\int_{[a,y_{i+1}]} (\hat{g} - \hat{f})d\lambda > 0 \quad \text{or} \quad \int_{[y_j,b]} (\hat{g} - \hat{f})d\lambda > 0 .$$

Say the latter is the case. Then since $\hat{f}$ and $\hat{g}$ are both constant on $(y_j, y_{j+1}]$, $\hat{g}(y_{j+1}) > \hat{f}(y_{j+1})$. From this, it follows that

$$0 < \int_{[y_j,b]} (\hat{g} - \hat{f})d\lambda < \int_{[y_j,y_{j+1}]} (\hat{g} - \hat{f})d\lambda.$$

Thus

$$\int_{[b,y_{j+1}]} (\hat{g} - \hat{f})d\lambda > 0.$$

But

$$\int_{[a,y_{j+1}]} (\hat{g} - \hat{f})d\lambda = \int_{[a,b]} (\hat{g} - \hat{f})d\lambda + \int_{[b,y_{j+1}]} (\hat{g} - \hat{f})d\lambda .$$

Since both terms on the right-hand side are positive, we have

$$\int_{[a,y_{j+1}]} (\hat{g} - \hat{f})d\lambda > 0.$$

Rewriting

$$\int_{[a,y_{i+1}]} (\hat{g} - \hat{f})d\lambda + \int_{[y_{i+1},y_{j+1}]} (\hat{g} - \hat{f})d\lambda > 0 .$$

The last term on the left-hand side is nonpositive by (1.2), since $[y_{i+1}, y_{j+1}]$ belongs to $L_n([L, R])$. From this, we have

$$\int_{[a,y_{i+1}]} (\hat{g} - \hat{f})d\lambda > 0.$$

Since $\hat{f}$ and $\hat{g}$ are also constant on $[y_i, y_{i+1})$, we have $\hat{g}(y_i) > \hat{f}(y_i)$. Thus

$$0 < \int_{[a,y_{i+1})} (\hat{g} - \hat{f})d\lambda < \int_{[y_i,y_{i+1}]} (\hat{g} - \hat{f})d\lambda .$$

Hence

$$\int_{[y_i, a]} (\hat{g} - \hat{f}) d\lambda > 0,$$

so that

$$\int_{[y_i, y_{j+1}]} (\hat{g} - \hat{f}) d\lambda > 0 .$$

But $[y_i, y_{j+1}] \in L_n([L, R])$, which means by (1.2)

$$\int_{[y_i, y_{j+1}]} (\hat{g} - \hat{f}) d\lambda \leq 0 .$$

This is a contradiction. Our assumption that $\int_{[a,b]} (\hat{g} - \hat{f}) d\lambda > 0$ is false. A similar argument follows if either $y_{\ell(n)} \leq a < L$ or $R < b \leq y_{r(n)}$ or both. This concludes the proof.

3. Conditional Expectations of the true density. In this section, we investigate the conditional expectation, $E(f | L([L, R]))$, of the true density f with respect to the σ-lattice of intervals containing the interval, $[L, R]$. Notice $\int f^2 d\lambda < \infty$ is a condition necessary to the definition of the conditional expectations. We shall need to require this condition in the remainder of this paper. We also will require continuity of f. It is not difficult to see that continuity together with unimodality of the density f implies

$$\int f^2 \, d\lambda < \infty.$$

The main theorem of this section is an analogue of Theorem 3.1 in [5].

Theorem 3.1. If f is a continuous unimodal density with a unique mode M, and $[L, R]$ is an interval with center m, let

$$K = (R - L)^{-1} \cdot \int_{[L,R]} f \, d\lambda.$$

i. If $f(L) \leq K > f(R)$ or $f(L) < K \geq f(R)$,

$E(f|L([L, R])) = f$ on $[L, R]^c$ and

$E(f|L([L, R])) = (R - L)^{-1} \cdot \int_{[L,R]} f \, d\lambda$ on $[L, R]$.

ii. If $f(L) \geq K > f(R)$ or if $f(L) = K = f(R)$ and $L \geq M$ there is an interval $[a, R]$ such that

$E(f|L([L, R])) = f$ on $[a, R]^c$ and

$E(f|L([L, R])) = (R - a)^{-1} \cdot \int_{[a,R]} f \, d\lambda$ on $[a, R]$

iii. If $f(L) < K \leq f(R)$ or if $f(L) = K = f(R)$ and $R \leq M$ there is an interval $[L, b]$ such that

$E(f|L([L, R])) = f$ on $[L, b]^c$ and

$E(f|L([1, R])) = (b - L)^{-1} \cdot \int_{[L,b]} f \, d\lambda$ on $[L, b]$

Proof: Notice that $f(L) \geq K < f(R)$ and $f(L) > K \leq f(R)$ are impossible. Thus cases i, ii, and iii exhaust all possibilities. Case i is easily verified and case iii follows in a manner similar to case ii. Let $a = \sup\{y \leq L': (R - y)^{-1} \cdot \int_{[y,R]} f \, d\lambda \geq f(y)\}$ with $L' = \min\{L, M\}$. If $y \leq L'$ and $f(y) \leq K$, it is clear that y belongs to the set defining a. Hence the set is not empty, and a, indeed, exists. Let $f^*(y)$ be given by $f(y)$ for y in $[a, R]^c$ and by $(R - a)^{-1} \cdot \int_{[a,R]} f \, d\lambda$ for y in $[a, R]$. We wish to show $f^* = E(f|L([L, R]))$. Clearly $f^* \in L([L, R])$. Hence it remains that

$$\int_A (f - f^*) d\lambda \leq 0 \quad \text{for every} \quad A \in L([L, R]) \quad \text{with} \quad 0 < \lambda(A) < \infty$$

and

$$\int (f - f^*)\theta(f^*) d\lambda = 0 \quad \text{for every Borel Function } \theta \text{ such that } \theta(0) = 0.$$

Let $a' = \sup\{y < R: (R - a)^{-1} \cdot \int_{[a,R]} f \, d\lambda \le f(y)\}$. Thus for $a < x < a'$, $f(x) \ge f^*(x)$ and for $a' < x < R$, $f(x) \le f^*(x)$. Now let $A \in L([L, R])$ such that $0 < \lambda(A) < \infty$. We may assume A is a closed interval, say $[b_1, b_2]$. If $b_2 > R$,

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

If $a' \le b_1 < R$, we have $f - f^* \le 0$ on (b_1, R), so that

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

If $a \le b_1 < a'$, $f - f^* \ge 0$ on (b_1, a'). Using this

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

Now since $\int_{[a,R]} (f - f^*)d\lambda = 0$, we have

$$\int_{[a,a']} (f - f^*)d\lambda = -\int_{(a',R]} (f - f^*)d\lambda . \tag{3.4}$$

Combining (3.3) and (3.4), for $a \le b_1 < a'$

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

Finally if $b_1 < a$,

$$\int_{[b_1,R]} (f - f^*)d\lambda = \int_{[a,R]} (f - f^*)d\lambda = 0 . \tag{3.6}$$

Combining (3.1) with either (3.2), (3.5), or (3.6) we have,

$$\int_A (f - f^*)d\lambda \le 0 \quad \text{for} \quad 0 < \lambda(A) < \infty .$$

Now let θ be any Borel function for which $\theta(0) = 0$. Then

$$\int(f - f^*) \cdot \theta(f^*)d\lambda = \int_{[a,R]} (f - f^*) \cdot \theta(f^*)d\lambda = \theta(f^*) \cdot \int_{[a,R]} (f - f^*)d\lambda.$$

But $\int_{[a,R]} (f - f^*)d\lambda = 0$. Thus $f^* = E(f|\mathcal{L}([L, R]))$. This completes the proof of case ii.

Notice that if $f(L) = K > f(R)$ case i and ii appear to overlap. In this case, $a = L$ so that there is no contradiction.

4. Convergence of $\hat{f}_{nm_n}$. We shall show the consistency of the density estimate based on a certain consistency of the center mode. Let

$$\Omega' = [\lim_{n \to \infty} \sup_y |F_n(y) - F(y)| = 0] ,$$

where F is the distribution corresponding to density f , and F_n is the empirical distribution based on a sample of size n from F . It is well known that this set has probability one. In addition, for points in Ω'

1. The largest observation less than and the smallest observation greater than a number converge to that number.
2. Corresponding to every pair of numbers, $r_1 < r_2$, in the support of f , $\{x: f(x) > 0\}$, there is eventually a pair of observations $y_{r_1(n)}$ and $y_{r_2(n)}$ satisfying $r_1 < y_{r_1(n)} < y_{r_2(n)} < r_2$.

Recall that the maximum likelihood estimate with center mode m is given by $\hat{f}_{nm}=E(\hat{g}_{nm}|\mathcal{L}_n([L, R]))$, where $\hat{g}_{nm}$, $\mathcal{L}_n([L, R])$, L and R are as in Section 2. Suppose y_0 is a point in the support of f and let $t = \hat{f}_{nm}(y_0)$. Let $P_t = [\hat{f}_{nm} > t]$ and $T_t = [\hat{f}_{nm} \geq t]$. If

$$H_1(T_t) = \{L \in \mathcal{L}([L, R]): \lambda(T_t - L) > 0\}$$

and

$$H_2(P_t) = \{L \in \mathcal{L}([L, R]): \lambda(L - P_t) > 0\} ,$$

a result of Robertson [3] gives us

(4.1) $$\hat{f}_{nm}(y_0) = \inf_{L \in H_1(T_t)} [\lambda(T_t - L)]^{-1} \cdot \int_{T_t - L} \hat{g}_{nm}\, d\lambda$$

and

(4.2) $$\hat{f}_{nm}(y_0) = \sup_{L \in H_2(P_t)} [\lambda(L - P_t)]^{-1} \cdot \int_{L - P_t} \hat{g}_{nm}\, d\lambda .$$

Robertson's theorem is stated for finite measure spaces. We restrict our space to $[y_1, y_n]$ where $y_1 \le \ldots \le y_n$ is the ordered sample.

Theorem 4.1. If f is a continuous unimodal density with unique mode M and m_n is a sequence converging to m not necessarily M, let

$$R_n = m_n + \frac{1}{2}\epsilon \quad \text{and} \quad L_n = m_n - \frac{1}{2}\epsilon , \quad \epsilon > 0 .$$

Let $\hat{f}_{nm_n}$ be the maximum likelihood estimate with center mode m_n. Let $f_m = E(f | L([L, R]))$, with $R = m + \frac{1}{2}\epsilon$ and $L = m - \frac{1}{2}\epsilon$. Then for points in Ω', $\hat{f}_{nm_n}$ converges pointwise to f_m except possibly at L or R or both.

Proof: The proof varies depending upon which representation in Theorem 3.1 holds. If the first characterization holds, we have $f_m = f$ on $[L, R]^c$ and $f_m = (R - L)^{-1} \cdot \int_{[L,R]} f\, d\lambda$ on $[L, R]$. For $y_0 < L$ or $y_0 > R$, we use methods quite analogous to those found in Robertson [4] or Wegman [5]. Let us turn our attention to $L < y_0 < R$.

For sufficiently large n, $y_0 \in [L_n, R_n]$. Thus if $t = \hat{f}_{nm_n}(y_0)$, $P_t = [\hat{f}_{nm_n} > t] = \emptyset$ eventually. In this case, any element of $L([L_n, R_n])$ belongs to $H_2(P_t)$. Let η be any positive number. For sufficiently large n, $[L - \eta, R + \eta] \in L([L_n, R_n])$, so that for sufficiently large n,

$$\hat{f}_{nm_n}(y_0) \ge (R - L + 2\eta)^{-1} \cdot \int_{[L-\eta, R+\eta]} \hat{g}_{nm_n}\, d\lambda .$$

If n^* is the number of observations in $[L - \eta, R + \eta]$, it is not difficult to see

$$(n^* + 2)/n \geq \int_{[L-\eta,R+\eta]} \hat{g}_{nm_n}\, d\lambda \geq (n^* - 1)/n .$$

Let us write $F_n(x-)$ for $\lim_{y\uparrow x} F_n(y)$. Then we may rewrite the above inequality as

$$F_n(R + \eta) - F_n(L - \eta-) + \frac{2}{n} \geq \int_{[L-\eta,R+\eta]} g_{nm_n}\, d\lambda \geq F_n(R + \eta) - F_n(L - \eta-) - \frac{1}{n} .$$

Hence we have

$$\lim_{n\to\infty} \int_{[L-\eta,R+\eta]} \hat{g}_{nm_n}\, d\lambda = F(R + \eta) - F(L - \eta-) = \int_{[L-\eta,R+\eta]} f\, d\lambda.$$

From this,

$$\liminf \hat{f}_{nm_n}(y_0) \geq [R - L + 2\eta]^{-1} \cdot \int_{[L-\eta,R+\eta]} f\, d\lambda.$$

But this is true for all $\eta > 0$, so

$$\liminf \hat{f}_{nm}(y_0) \geq (R - L)^{-1} \cdot \int_{[L,R]} f\, d\lambda = f_m(y_0).$$

On the other hand, if $T_t = [\hat{f}_{nm_n} \geq t]$, then $T_t = [y_{i(n)}, y_{j(n)}]$ is an interval. Clearly $y_{i(n)} \leq L_n$ and $y_{j(n)} \geq R_n$, so that $\limsup y_{i(n)} \leq L$ and $\liminf y_{j(n)} \geq R$. We wish to show $\lim_{n\to\infty} y_{i(n)} = L$. To see this, it is necessary to show $\liminf y_{i(n)} \geq L$. If this is not true, there is a subsequence which we shall again denote $y_{i(n)}$ for simplicity, and a positive constant, η such that $y_{i(n)} \leq L - \eta$. For the moment, $\hat{f}_{nm_n}$ will mean the subsequence corresponding to $y_{i(n)}$. Since $\hat{f}_{nm_n}$ is constant on T_t eventually,

$$t = \hat{f}_{nm_n}(y_0) = \hat{f}_{nm_n}(L - \eta) = \hat{f}_{nm_n} \quad \text{eventually on} \quad [L - \eta, R).$$

Since $\hat{f}_{nm_n}(L - \eta)$ converges to $f_m(L - \eta)$ we have

$$\limsup \hat{f}_{nm_n} = f_m(L - \eta) \quad \text{on} \quad [L - \eta, R) .$$

Eventually, $[L - \eta, R + \eta] \in L([L_n, R_n])$ so that by (1.2)

$$\int_{[L-\eta,R+\eta]} \hat{g}_{nm_n} \, d\lambda \le \int_{[L-\eta,R+\eta]} \hat{f}_{nm_n} \, d\lambda,$$

where $\hat{g}_{nm_n}$ is the subsequence corresponding to $y_{i(n)}$. By Fatou's lemma,

$$\limsup \int_{[L-\eta,R+\eta]} \hat{g}_{nm_n} \, d\lambda \le \int_{[L-\eta,R+\eta]} \limsup \hat{f}_{nm_n} \, d\lambda.$$

But $\limsup \hat{f}_{nm_n}(x) \le f_m(L-\eta)$ for all x so that

$$\limsup \int_{[L-\eta,R+\eta]} \hat{g}_{nm_n} \, d\lambda \le \int_{[L-\eta,R+\eta]} f_m(L - \eta) \, d\lambda .$$

We have just seen

$$\lim_{n \to \infty} \int_{[L-\eta,R+\eta]} \hat{g}_{nm_n} \, d\lambda = \int_{[L-\eta, R+\eta]} f \, d\lambda ,$$

so

$$\int_{[L-\eta,R+\eta]} f \, d\lambda \le \int_{[L-\eta,R+\eta]} f_m(L - \eta) \, d\lambda. \tag{4.3}$$

But $[L, R]$ must contain M since otherwise the characterization of case i in Theorem 3.1 could not hold. Since f has unique mode, $f > f_m(L - \eta)$ in some neighborhood of M. Hence (4.3) cannot hold. Thus $\lim_{n \to \infty} y_{i(n)} = L$. Similarly, $y_{j(n)}$ converges to R. Since ϕ belongs to $H_1(T_t)$, by (4.1)

$$\hat{f}_{nm_n}(y_0) \le (y_{j(n)} - y_{i(n)})^{-1} \cdot \int_{[y_{i(n)},y_{j(n)}]} \hat{g}_{nm_n} \, d\lambda.$$

Writing $\int_{[y_{i(n)},y_{j(n)}]} \hat{g}_{nm_n} \, d\lambda = F_n(y_{j(n)}) - F_n(y_{i(n)}) + (\frac{1}{n})$ it is clear that

$$\limsup \hat{f}_{nm_n}(y_0) \le (R - L)^{-1} \cdot (F(R) - F(L)) = f_m(y_0)$$

Thus $$\lim_{n \to \infty} \hat{f}_{nm_n}(y_0) = f_m(y_0) \text{ for } y_0 \text{ such that } L < y_0 < R.$$

This completes the proof in the case that i of Theorem 3.1 is the characterization of $E(f|\mathcal{L}([L, R]))$. The other cases may be proven in a similar manner.

We are easily able to upgrade the convergence in this case by applying methods similar to those of the Glivenko-Cantelli Theorem.

Corollary 4.1. If the conditions of 4.1. hold, then $\hat{f}_{nm_n}$ converges uniformly except on an interval of arbitrarily small measure containing L or R or the union of two such intervals. This convergence holds for all points in Ω', hence with probability one.

5. A convergence theorem for the center mode. Recall from Section 2, $\hat{f}_n$, the maximum likelihood estimate is given by a conditional expectation with respect to a lattice containing a modal interval, where m_n is the center mode. Let $L_n = m_n - \frac{1}{2}\epsilon$ and $R_n = m_n + \frac{1}{2}\epsilon$. It is desired to show that the sequence, $\{m_n\}$, has a limit m.

Roughly speaking, we shall accomplish this by showing that there is an m for which $\int \log f_m \cdot f \, d\lambda$ is maximized. If m_n fails to converge to that m, then we can pick another estimate which has a larger likelihood product eventually. This would be a contradiction to the choice of $\hat{f}_n$.

Lemma 5.1. Let $H(m) = \int \log f_m \cdot f \, d\lambda$, where $f_m = E(f|\mathcal{L}([L, R]))$ and $L = m - \frac{1}{2}\epsilon$ and $R = m + \frac{1}{2}\epsilon$. Then $H(m)$ is unimodal.

Proof: The real line may be divided into three regions, corresponding to the three different characterizations of f_m. Let

$$\mathcal{M} = \{m: f(L) \le (R - L)^{-1} \cdot \int_{[L,R]} f \, d\lambda > f(R) \quad \text{or}$$

$$f(L) < (R - L)^{-1} \cdot \int_{[L,R]} f \, d\lambda \ge f(R)\}.$$

Clearly $M - \frac{1}{2}\epsilon$ is a lower bound of $\mathcal{M}$ and $M + \frac{1}{2}\epsilon$ is an upper bound, where M is the mode of f. Let $m^- = \inf \mathcal{M}$ and $m^+ = \sup \mathcal{M}$. It is straightforward to show that f_m is characterized by i of Theorem 3.1 for each m such that $m^- \le m \le m^+$. For $m \le m^-$, f_m is characterized as in iii of Theorem 3.1 and for $m \ge m^+$, f_m is characterized as in ii of Theorem 3.1. If we can show H(m) is nondecreasing for $m \le m^-$ and nonincreasing for $m \ge m^+$ and unimodal for $m^- \le m \le m^+$, this will be sufficient.

Let us consider $m < m^* \le m^-$. $H(m^*) - H(m) = \int \log \frac{f_{m^*}}{f_m} \cdot f\, d\lambda$. Let L, L*, b and b* have their obvious meanings. Since

$$(y - L)^{-1} \cdot \int_{[L,y]} f\, d\lambda \le (y - L^*)^{-1} \cdot \int_{[L^*,y]} f\, d\lambda$$

it follows that if $(y - L)^{-1} \cdot \int_{[L,y]} f\, d\lambda - f(y) \ge 0$, so is $(y - L^*) \cdot \int_{[L^*,y]} f\, d\lambda - f(y) \ge 0$. From a consideration of the sets defining b and b*, we may conclude $b \ge b^*$. Hence both f_{m^*} and f_m are constant on $[L^*, b^*]$. From this, we have

$$\int \log \frac{f_{m^*}}{f_m} f\, d\lambda = \int_{[L^*,b^*]^c} \log \frac{f_{m^*}}{f_m} f\, d\lambda + \log \frac{f_{m^*}(L^*)}{f_m(L^*)} \int_{[L^*,b^*]} f\, d\lambda .$$

But

$$\int_{[L^*,b^*]} f\, d\lambda = \int_{[L^*,b^*]} f_{m^*}\, d\lambda$$

and since

$$f = f_{m^*} \quad \text{on} \quad [L^*, b^*]^c ,$$

$$\int \log\left(\frac{f_{m^*}}{f_m}\right) f\, d\lambda = \int \log\left(\frac{f_{m^*}}{f_m}\right) f_{m^*}\, d\lambda .$$

The latter quantity is the Kullback-Leibler information number and is non-negative. Hence

$$H(m^*) \geq H(m) .$$

A similar proof holds for $m \geq m^+$. Let us turn our attention to $m^- \leq m \leq m^+$. Let $\eta(t) = t \log t$ for $t > 0$. We may expand $H(m)$ as follows

$$H(m) = \log\left[\frac{\int_{[L,R]} f d\lambda}{\epsilon}\right] \cdot \int_{[L,R]} f d\lambda + \int_{(-\infty,L]} \log(f) \cdot f d\lambda + \int_{[R,\infty)} \log(f) \cdot f d\lambda.$$

Recalling $L = m - \frac{1}{2}\epsilon$ and $R = m + \frac{1}{2}\epsilon$ and differentiating with respect to m, we obtain

$$H'(m) = \eta'\left[\frac{\int_{[L,R]} f\ d\lambda}{\epsilon}\right] \cdot \left[f(R) - f(L)\right] + \eta(f(L)) - \eta(f(R)).$$

Rewriting this,

$$\frac{H'(m)}{f(R) - f(L)} = \eta'\left[\frac{\int_{[L,R]} f\ d\lambda}{\epsilon}\right] - \left[\frac{\eta(f(R)) - \eta(f(L))}{f(R) - f(L)}\right] .$$

Noticing that $f(R) \leq \epsilon^{-1} \cdot \int_{[L,R]} f\ d\lambda$ and that $\eta'(t)$ is an increasing function, we have

$$\frac{H'(m)}{f(R) - f(L)} \geq \eta'(f(R)) - \left[\frac{\eta(f(R)) - \eta(f)L))}{f(R) - f(L)}\right].$$

Let us assume $f(R) - f(L) \geq 0$. In this case, the right hand side represents the slope of a chord subtracted from a tangent. It is clear in this case, that

$$\frac{H'(m)}{f(R) - f(L)} \geq 0.$$

Hence, we have that if $f(R) - f(L) \geq 0$, $H'(m) \geq 0$. Correspondingly, if $f(R) - f(L) \leq 0$, then $H'(m) \leq 0$. Hence, $H(m)$ is unimodal. The mode of H is that m for which $f(R) = f(L)$. Clearly there need not be a unique mode, but if not $H(m)$ will have a modal interval, namely $\{m: f(R) = f(L)\}$. If f is symmetric, then it is clear that the mode of H is the mode of f, that is, the mode of H is M. In any case, since $M - \frac{1}{2}\epsilon$ and $M + \frac{1}{2}\epsilon$ are lower and upper bounds respectively on M, it is clear that the difference between the modes of H and of f is less than $\frac{1}{2}\epsilon$.

Let us assume, in general that m is the mode of H. In order to avoid cumbersome details, let us assume f is strictly increasing at x for $x < M$ and strictly decreasing at x for $x > M$. Thus H has a unique mode. An obvious modification exists if there are some "plateaus" in f.

If $\{m_n\}$ is the sequence of center modes of $\{\hat{f}_n\}$, we wish to show $\{m_n\}$ converges to m. The next lemma applies in the case that the support of f is $(-\infty, \infty)$.

Lemma 5.2. If the entropy, $-\int \log f \cdot f \, d\lambda$ of f is finite, with probability one

$$-\infty < \liminf m_n \leq \limsup m_n < \infty.$$

Proof: If $\limsup m_n = \infty$, there is a subsequence (which we will also label $\{m_n\}$) which diverges to ∞. Let γ be chosen in $(0, 1)$ and choose x_0 so that $1 - F(x_0) \leq \gamma$. Suppose $\eta > 0$ and $\hat{f}_n(x_0) > \eta$ infinitely often. Then $\hat{f}_n(x) > \eta$ for all $x \in (x_0, m_n)$ infinitely often. This is impossible as soon as $m_n - x_0 > (\frac{1}{\eta})$. Thus $f_n(x_0) \leq \eta$ for sufficiently large n. Hence for $x \leq x_0$, $\hat{f}_n$ converges uniformly to zero.

Let us write

$$L_n(\hat{f}_n) = \frac{1}{n} \cdot \sum_{i=1}^{n} \log(\hat{f}_n(y_i)).$$

We may write $L_n(\hat{f}_n)$ as follows

$$L_n(\hat{f}_n) = n^{-1} \left| \sum_{y_i \le x_0} \log(\hat{f}_n(y_i)) + \sum_{y_i > x_0} \log(\hat{f}_n(y_i)) \right| .$$

If we let n_0 be the number of observations less than or equal to x_0 since $\hat{f}_n$ is bounded by ϵ^{-1}, and since $n^{-1}(n - n_0)$ converges to $1 - F(x_0) \le \gamma$, we have

$$\lim\sup n^{-1} \sum_{y_i > x_0} \log(\hat{f}_n(y_i)) \le \log(\epsilon^{-1}) \cdot \gamma .$$

On the other hand, since $\hat{f}_n(x)$ converges uniformly to 0 for $x \le x_0$ $\log(\hat{f}_n(x))$ diverges uniformly to $-\infty$ for $x \le x_0$. Thus

$$\lim\sup n^{-1} \cdot \sum_{y_i \le x_0} \log(\hat{f}_n(y_i)) = -\infty.$$

Finally, we may write

$$\lim\sup L_n(\hat{f}_n) = -\infty.$$

Let us consider $f_m = E(f \mid \mathcal{L}([L, R]))$ where $L = m - \frac{1}{2}\epsilon$ and $R = m + \frac{1}{2}\epsilon$. Clearly by the Kolmogorov Strong Law with probability one

$$\lim_{n\to\infty} L_n(f_{m'}) = \int \log f_{m'} \cdot f \, d\lambda.$$

This quantity is finite since the entropy is finite. Hence

$$\lim\inf \left[L_n(f_{m'}) - L_n(\hat{f}_n)\right] = +\infty.$$

This is equivalent to saying, eventually $f_{m'}$ is more likely than $\hat{f}_n$ which is a contradiction to the choice of $\hat{f}_n$. Hence $\limsup m_n < \infty$. A similar argument holds for the other inequality.

It may be that the support is bounded below, bounded above or both. Let $\alpha = \inf\{x: f(x) > 0\}$ and $\beta = \sup\{x: f(x) > 0\}$. Using methods similar to those found in Lemma 5.2, we may obtain

Lemma 5.3. If $\frac{1}{2}\epsilon < \min\{M - \alpha, \beta - M\}$ and the entropy of the continuous unimodal density f is finite, then with probability one

$$\alpha + \frac{1}{2}\epsilon < \liminf m_n \le \limsup m_n < \beta - \frac{1}{2}\epsilon.$$

Let m' be any cluster point of $\{m_n\}$. Then m' has the property $\alpha + \frac{1}{2}\epsilon < m' < \beta - \frac{1}{2}\epsilon$. (Here let us permit $\alpha = -\infty$ and $\beta = +\infty$). Now let $f_m = E(f|L([L, R]))$ with $L = m - \frac{1}{2}\epsilon$ and $R = m + \frac{1}{2}\epsilon$ and let $f_{m'} = E(f|L([L', F']))$ with $L' = m' - \frac{1}{2}\epsilon$ and $R' = m' + \frac{1}{2}\epsilon$. By Theorem 3.1, f_m and $f_{m'}$ each agree with f on the complement of some interval. Let c be the smallest endpoint of those two intervals and d be the largest. Hence on $[c, d]^c$, $f_m = f_{m'} = f$. Moreover, since both m and m' are in $(\alpha + \frac{1}{2}\epsilon, \beta - \frac{1}{2}\epsilon)$, (recall the bounds on M are $M - \frac{1}{2}\epsilon$ and $M + \frac{1}{2}\epsilon)$, both c and d are in the support of f. We may pick $\eta > 0$ sufficiently small so that $c - \eta$ and $d + \eta$ are in the support. In the next lemma, when we write $\hat{f}_n$ we shall mean that subsequence of the sequence of maximum likelihood estimates for which the center modes converge to m'. By f_n^* we shall mean $\hat{f}_{nm}$, the maximum likelihood estimate whose mode is m.

Lemma 5.4. With probability one, for sufficiently large n, $f_n^* = \hat{f}_n$ on $(c - \eta, d + \eta)^c$.

Proof: The set of probability one is Ω' as described in 4. Pick t_0 in $(c - \eta, c)$. Let $\delta \in (0, c - t_0)$. Since f has a point of increase in $(t_0, c - \delta)$, $\hat{f}_n$ and f_n^* must both eventually have a jump in $(t_0, c - \delta)$. Let y_i be the smallest observation greater than t_0 for which $\hat{f}_n$ has a jump and let y_{i*} be the smallest observation greater than t_0 for which f_n^* has a jump. Without loss of generality, we may assume $y_{i*} \geq y_i$. Let y_{j*} be the largest observation smaller than or equal to t_0 for which there is a jump in f_n^*. Thus we have $y_{j*} \leq t_0 < y_i \leq y_{i*}$. We wish to show equality holds in the last inequality. Assume $y_i < y_{i*}$. Let $f = f_n^*(t_0)$ and $T_t = [f_n^* \geq t]$, so $T_t = [y_{j*}, y_{k*}]$ for some observation $y_{k*} \geq m + \frac{1}{2}\epsilon$. Then $[y_i, y_{k*}] \in L([m - \frac{1}{2}\epsilon, m + \frac{1}{2}\epsilon])$. Hence by (4.1)

$$(5.1) \qquad t \leq (y_i - y_{j*})^{-1} \int_{[y_{j*}, y_i]} g_n^* \, d\lambda$$

where $g_n^* = \hat{g}_{nm}$ as defined in Section 2. In a similar manner using (4.2), we can show

$$(5.2) \qquad t \geq (y_{i*} - y_i)^{-1} \cdot \int_{[y_i, y_{i*}]} g_n^* \, d\lambda.$$

Let $\hat{g}_n = \hat{g}_{nm_n}$ as defined in Section 2. For sufficiently large n, $\hat{g}_n$ and g_n^* will agree. This is easily seen, since $m_n - \frac{1}{2}\epsilon$ converges to $m' - \frac{1}{2}\epsilon > c - \delta$. As soon as $m_n - \frac{1}{2}\epsilon > c - \delta$, $\hat{g}_n$ and g_n^* agree. This together with (5.1) and (5.2) shows

$$(5.3) \qquad (y_{i*} - y_i)^{-1} \cdot \int_{[y_i, y_{i*}]} \hat{g}_n \, d\lambda \leq (y_i - y_{j*})^{-1} \cdot \int_{[y_{j*}, y_i]} \hat{g}_n \, d\lambda.$$

Now let $t = \hat{f}_n(t_0)$ and $P_t = [\hat{f}_n > t]$. Again by use of (4.2), we have

$$(5.4) \qquad \hat{f}_n(t_0) \geq (y_i - y_{j*})^{-1} \cdot \int_{[y_{j*}, y_i]} \hat{g}_n \, d\lambda.$$

Finally letting $t = \hat{f}_n(y_i)$ and using (4.1), we have

$$\hat{f}_n(y_i) \le (y_{i*} - y_i)^{-1} \cdot \int_{[y_i, y_{i*}]} \hat{g}_n \, d\lambda. \tag{5.5}$$

Using (5.3), (5.4) and (5.5),

$$\hat{f}_n(t_0) \ge \hat{f}_n(y_i).$$

But y_i is a jump point in $\hat{f}_n$, so

$$\hat{f}_n(t_0) < \hat{f}_n(y_i).$$

Thus our assumption $y_i \neq y_{i*}$ is false. Let us now consider $t = f_n^*(t_0)$ and $P_t = [f_n^* > t]$. Recall now that the definition of f_n^* involves the conditional expectation of g_n^* with respect to $L_n([m - \frac{1}{2}\epsilon, m + \frac{1}{2}\epsilon])$. Since there are only a finite number of sets in this lattice, the supremum in (4.2) is really a maximum. Let $L = [u, v]$ be the member of the lattice such that

$$f_n^*(t_0) = (L - P_t)^{-1} \cdot \int_{L-P_t} g_n^* \, d\lambda.$$

Rewriting this

$$f_n^*(t_0) = (y_i - u + v - y_{k*})^{-1} \cdot \int_{[u,y_i] \cup [y_{k*},v]} g_n^* \, d\lambda.$$

But by (4.2),

$$f_n^*(t_0) \ge (v - y_{k*})^{-1} \cdot \int_{[y_{k*},v]} g_n^* \, d\lambda .$$

Combining these two displays, we obtain

$$f_n^*(t_0) \le (y_i - u)^{-1} \cdot \int_{[u,y_i]} g_n^* \, d\lambda.$$

But also by (4.2),

$$f_n^*(t_0) \ge (y_i - u)^{-1} \cdot \int_{[u,y_i]} g_n^* \, d\lambda.$$

Hence, for sufficiently large n,

$$f_n^*(t_0) = \sup_{u < y_i} \{(y_i - u)^{-1} \cdot \int_{[u,y_i]} g_n^* \, d\lambda\}.$$

Similarly, for sufficiently large n,

$$\hat{f}_n(t_0) = \sup_{u < y_i} \{(y_i - u)^{-1} \cdot \int_{[u,y_i]} \hat{g}_n \, d\lambda\}.$$

Since $\hat{g}_n$ and g_n^* eventually agree in this region, $\hat{f}_n$ and f_n^* must be equal at t_0 eventually. For any $t < t_0$, by virtue of the fact, $\hat{f}_n(t_0) = f_n^*(t_0)$, and by use of (4.1) and (4.2), we obtain the desired conclusion.

We may now state and prove the theorem.

THEOREM 5.1. If $\frac{1}{2}\epsilon < \min\{M - \alpha, \ \beta - M\}$ and the entropy of the continuous unimodal density f is finite, then with probability one

$$\lim_{n \to \infty} m_n = m, \quad \text{where } m \text{ is the mode of } H.$$

Proof: If not, a subsequence of $\{m_n\}$, (which we again label $\{m_n\}$), converges to $m' \neq m$, with $\alpha + \frac{1}{2}\epsilon < m' < \beta - \frac{1}{2}\epsilon$. Let $f_m = E(f \mid L([m - \frac{1}{2}\epsilon, \ m + \frac{1}{2}\epsilon]))$ and $f_{m'} = E(f \mid L([m' - \frac{1}{2}\epsilon, \ m' + \frac{1}{2}\epsilon]))$. Let f_n^* be the maximum likelihood estimate when the mode is known to occur at m and let $\hat{f}_n$ be the maximum likelihood estimate when the mode in unknown. We intend to show that f_n^* has a larger likelihood product than $\hat{f}_n$. This will be a contradiction, so that with probability one, m_n converges to m. (It is understood that $\{f_n^*\}$ and $\{\hat{f}_n\}$ refer to the subsequences corresponding to the subsequence $\{m_n\}$ converging to m'). Recall that on $[c, d]^c$, $f = f_{m'} = f_m$. Let $\eta > 0$ be sufficiently small so that $c - \eta$ and $d + \eta$ are in the support of f.

For sufficiently large n, f_n^* and $\hat{f}_n$ agree on $(c - \eta, \ d + \eta)^c$ with probability one. Let $k = \min\{f(c - \eta), f(d + \eta)\}$. Because $f_m > \frac{1}{2}k$ and $f_{m'} > \frac{1}{2}k$ on $[c - \eta, \ d + \eta]$ and because $\hat{f}_n$ converges to $f_{m'}$ almost uniformly with probability one and f_n^* converges to f_m almost uniformly with probability one,

$\log(\hat{f}_n)$ converges to $\log(f_{m'})$ almost uniformly with probability one and $\log(f_n^*)$ converges to $\log(f_m)$ almost uniformly with probability one. Let $\delta > 0$. Pick the set A of arbitrarily small measure so that $P(A) \cdot \log(k\epsilon/2) > -\delta$. Now if $y_1 \le \ldots \le y_n$ are the ordered observations, let

$$L_n(f_n^*) - L_n(\hat{f}_n) = \frac{1}{n} \sum_{i=1}^{n} \{\log(f_n^*(y_i)) - \log(\hat{f}_n(y_i))\}.$$

For sufficiently large n with probability one, f_n^* and $\hat{f}_n$ agree on $(c - \eta, \ d + \eta)^c$. Thus if we let Σ_1 be the summation over y_i in $[c - \eta, \ d + \eta]$, for sufficiently large n with probability one,

$$L_n(f_n^*) - L_n(\hat{f}_n) = \frac{1}{n} \Sigma_1 \{\log(f_n^*(y_i)) - \log(\hat{f}_n(y_i))\} .$$

Now on A, $\log(f_n^*) - \log(\hat{f}_n) > \log(k\epsilon/2)$, eventually. This follows since $f_n^* > \frac{k}{2}$ eventually and $\hat{f}_n \le \frac{1}{\epsilon}$. Thus

$$\frac{1}{n} \Sigma_2 \{\log(f_n^*(y_i)) - \log(\hat{f}_n(y_i))\} > \frac{n^*}{n} \log(k\epsilon/2) \quad \text{eventually,}$$

where n^* is the number of observations in A and Σ_2 is the summation over y_i in $[c - \eta, \ d + \eta] \cap A$. Since (n^*/n) converges to $P(A)$ with probability one, $\frac{1}{n} \Sigma_2\{\log(f_n^*(y_i)) - \log(\hat{f}_n(y_i))\} > -\delta$ with probability one for sufficiently large n.

In a similar manner, one may show eventually

$$-\frac{1}{n}\Sigma_2\{\log(f_m(y_i)) - \log(f_{m'}(y_i))\} > -\delta .$$

Then if Σ_3 is the sum over y_i in $[c-\eta,\ d+\eta] - A$,

$$L_n(f_n^*) - L_n(\hat{f}_n) \geq \frac{1}{n}\Sigma_3 \{\log(f_n^*(y_i)) - \log(\hat{f}_n(y_i))\}$$

$$+ \frac{1}{n}\Sigma_2 \{\log(f_m(y_i)) - \log(f_{m'}(y_i))\} - 2\delta$$

with probability one for sufficiently large n. By the uniform convergence properties of $\log(f_n^*)$ and $\log(\hat{f}_n)$ on $[c-\eta,\ d+\eta] - A$, we have eventually with probability one,

$$L_n(f_n^*) - L_n(\hat{f}_n) \geq \frac{1}{n}\Sigma_1 \{\log(f_m(y_i)) - \log(f_{m'}(y_i))\} - 3\delta.$$

On $[c-\eta,\ d-\eta]^c$, $f = f_m = f_{m'}$ so that with probability one and for sufficiently large n,

$$L_n(f_n^*) - L_n(\hat{f}_n) \geq \frac{1}{n}\sum_{i=1}^{n} \{\log(f_m(y_i)) - \log(f_{m'}(y_i))\} - 3\delta .$$

By the Kolmogorov Strong Law, with probability one

$$\lim_{n\to\infty} (L_n(f_n^*) - L_n(\hat{f}_n)) \geq \int \log\left(\frac{f_m}{f_{m'}}\right) f\, d\lambda - 3\delta.$$

Since m is the unique mode of $\int \log(f_m) f\, d\lambda$, and since δ was arbitrary, with probability one

$$\lim_{n\to\infty} (L_n(f_n^*) - L_n(\hat{f}_n)) > 0 .$$

This is equivalent to f_n^* having a larger likelihood product than $\hat{f}_n$. This is a contradiction. Hence with probability one, $\{m_n\}$ converges to m.

6. Summary. For $\epsilon > 0$, we have given a maximum likelihood estimate of a unimodal density. This estimate is a strongly consistent estimate of $f_m = E(f \mid L([m - \frac{1}{2}\epsilon, m + \frac{1}{2}\epsilon]))$. Of course, $f_m = f$ except on an interval of length ϵ. Here m is the mode of $H(m) = \int \log(f_m) f \, d\lambda$. If $\epsilon = 0$, the same method yields a maximum likelihood estimate. The lack of a uniform bound causes the consistency arguments described in this paper to fail. Hence the asymptotic behavior of this sort of estimate is unknown.

Wegman [5] describes a consistency argument for a related continuous estimate. The analogue for $\hat{f}_n$ is not difficult.

7. Acknowledgements. 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] Robertson, Tim (1966). "A representation for conditional expectations given σ-lattices," Ann. Math. Statist. 37, 1279-1283.

[4] Robertson, Tim (1967). "On estimating a density which is measurable with respect to a σ-lattice," Ann. Math. Statist. 38, 482-493.

[5] Wegman, E.J. (1968). "A note on estimating a unimodal density," Institute of Stat. Mimeo Series No. 599, The Univ. of North Carolina at Chapel Hill. (Sub. to Ann. Math. Statist.)