The research for this work was partially sponsored by the National Science Foundation under Grant No. GU-2059 and the United States Air Force under Contract No. AFOSR-68-1415.

MAXIMUM LIKELIHOOD ESTIMATION OF A UNIMODAL DENSITY, II

by

Edward J. Wegman

Department of Statistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 647

NOVEMBER 1969

Maximum Likelihood Estimation of a Unimodal Density, II

by

Edward J. Wegman

1. Introduction. Several authors, Grenander [3], Robertson [6], and Rao [4], have described the MLE for a unimodal density when the mode was known as well as some of the estimate's properties. The MLE for a unimodal density when the mode is unknown was described in [7]. Strong consistency was also established in [7]. We wish to describe some additional properties in this paper.

2. Asymptotic Distribution. Let $\hat{f}_n$ be the maximum likelihood estimate with unknown mode and f_n^* the maximum likelihood estimate with known mode. In defining $\hat{f}_n$, $\epsilon > 0$ was a predetermined number. Let $y_1 < y_2 < \dots < y_n$ be the ordered observations sampled according to the density f and let $A_1 = [y_1, y_2)$, $A_2 = [y_2, y_3)$, ..., $A_{\ell(n)} = [y_{\ell(n)}, L_n)$, $A_{\ell(n)+1} = [L_n, R_n]$, $A_{\ell(n)+2} = (R_n, y_{r(n)}]$, ..., $A_k = (y_{n-1}, y_n]$. Here $R_n - L_n = \epsilon$; the sequences $\{L_n\}$ and $\{R_n\}$ converge to L and R respectively; and $y_{\ell(n)}$ and $y_{r(n)}$ are respectively the largest observation smaller than L_n and the smallest observation larger than R_n. L_n and R_n are determined by the maximum likelihood procedure and at least one of L_n or R_n is an observation for each n. If $L([L_n, R_n])$ is the σ-lattice

---

The research for this work was partially sponsored by the National Science Foundation under Grant No. GU-2059 and the United States Air Force under Contract No. AFOSR-68-1415.

of intervals containing $[L_n, R_n]$, the maximum likelihood estimate, $\hat{f}_n$, is given by the conditional expectation, $E(\hat{g}_n | L([L_n, R_n]))$, where

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

Here n_i is the number of observations in A_i, λ is Lebesgue measure and I_{A_i} is the indicator of A_i.

In a similar manner, let $A_i^* = [y_1, y_2)$, ..., $A_{q(n)} = [y_{q(n)}, M)$, $A_{q(n)+1} = [M, y_{q(n)+L}]$, $A_{q(n)+2} = (y_{q(n)+1}, y_{q(n)+2}]$, ..., $A_n = (y_{n-1}, y_n]$. Here M is the known mode and $y_{q(n)}$ is the largest observation smaller than M. Notice with probability one, $M \neq y_j$ for each j. If $L(M)$ is the σ-lattice of intervals containing M, the maximum likelihood estimate, f_n^*, is given by $E(g_n^* | L(M))$ where

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

Of course, n_i^* is the number of observations in A_i^*. In [7], it is shown that $M \in (L,R)$, hence $\hat{g}_n$ and g_n^* agree except possibly on $[y_{\ell(n)}, y_{r(n)}]$. A similar situation was the case in Lemma 5.4 in [7]. If we require only that some neighborhood of L, say N_L, is a set of points of increase of f and similarly some neighborhood of R, say N_R, is a set of points of decrease of f, we may use the arguments of Lemma 5.4 in [7] to obtain

Lemma 2.1. Let $\eta > 0$ be an arbitrary number such that $L-\eta$ and $R+\eta$ are elements of N_L and N_R respectively. Then with probability one, for sufficiently large n, $\hat{f}_n$ and f_n^* agree on $(L-\eta, R+\eta)^c$.

Hence, for any $x \notin [L,R]$, for sufficiently large n, $\hat{f}_n(x) = f_n^*(x)$. An immediate theorem follows

Theorem 2.1: For $x \notin [L,R]$, $\hat{f}_n(x)$ has the same asymptotic distribution as $f_n^*(x)$.

Rao [4] through some very clever, but rather tedious arguments develops the asymptotic distribution of $f_n^*(x)$. Arguments similar to these could be applied to $\hat{f}_n(x)$, but are avoided by use of Lemma 2.1. Rao assumes a non-zero derivative of the density, f, at each point x where the asymptotic distribution is to be found.

3. A Characterization of $\hat{f}_n$. Reid (see [1] and [2]) gave a geometrical interpretation of a conditional expectation with respect to a σ-lattice, L, when L consists of intervals with the right (or left) endpoint fixed. If the σ-lattice is $L(M)$, the conditional expectation may be characterized by applying Reid's method individually to the right and to the left of M. To find $E(h|L(M))$, the conditional expectation of some function h with respect to $L(M)$, determine $H(x) = \int_{(-\infty,x]} h d\lambda$. To the left of M, $E(h|L(M))$ is given by the slope of the greatest convex minorant of H and to the right of M, by the slope of the least concave majorant of H.

Let us assume that h has bounded support, $\{x: h(x) \neq 0\}$. Let L and R be fixed with $R-L = \epsilon$. We want a geometrical interpretation of the conditional expectation of h with respect to $L([L,R])$. Robertson

[5] gives a representation of conditional expectations. This representation holds on finite measure spaces, hence the requirement that h has bounded support. Let $E(h|\mathcal{L})(x) = y$ and $P_y = \{x: E(h|\mathcal{L})(x) > y\}$. Let $\mathcal{H} = \{L^* \in \mathcal{L}: \lambda(L^*-P_y) > 0\}$. Then

$$y = \sup_{L^*\in\mathcal{H}} [\lambda(L^*-P_y)]^{-1} \cdot \int_{L^*-P_y} h d\lambda.$$

Letting $\mathcal{L} = \mathcal{L}([L,R])$ and $x \in [L,R]$, it is clear that P_y is empty, so that

$$y = \sup_{L^*\in\mathcal{H}} (\lambda(L^*))^{-1} \cdot \int_{L^*} h d\lambda.$$

In fact, this supremum is a maximum and $\mathcal{H} \equiv \mathcal{L}([L,R])$. If L^* is the maximizing interval, for all $x\in L^*$, $E(h|\mathcal{L}([L,R]))(x) = \lambda(L^*)^{-1} \cdot \int_{L^*} h d\lambda$. Let $a = \inf L^*$ and $b = \sup L^*$. As in the case of the conditional expectation with respect to $\mathcal{L}(M)$, it is not difficult to see we may apply Reid's method individually to the left of a and to the right of b. Thus we have,

Theorem 3.1: The conditional expectation of a function, h, with bounded support, with respect to a σ-lattice, $\mathcal{L}([L,R])$, is given by the following procedure.

Find the interval $[a,b]$ containing $[L,R]$ such that $(H(b) - H(a))/(b-a)$ is maximized. On $[a,b]$, the conditional expectation is given by $(H(b) - H(a))/(b-a)$. To the left of a, it is the slope of the greatest convex minorant of H and to the right of b, it is the slope of the least concave majorant of H.

If $h = \hat{g}_n$, the theorem applies since the support of $\hat{g}_n$ is $[y_1,y_n]$. Let $\hat{G}_n(x) = \int_{(-\infty,x]} \hat{g}_n d\lambda$.

Corollary 3.1. If $h = \hat{g}_n$ in Theorem 3.1, $\hat{G}_n$ may be replaced by F_n, the empirical distribution function.

The proof is straightforward and is left to the reader. It is interesting to note that Theorem 3.1 implies Theorem 3.1 of [7] if the condition of f being continuous is exchanged for f having bounded support. The author is indebted to the referee of [7] for pointing out this fact.

References.

[1] Brunk, H.D. (1956), "On an inequality for convex functions," Proc. Amer. Math. Soc., 7, pp. 817-824.

[2] Brunk, H.D., Ewing, G.M., Reid, W.T., "The mimimum of a certain definite integral suggested by the maximum likelihood estimate of a distribution function," Bull. Amer. Math. Society, Abstract 60-6-684.

[3] Grenander, Ulf. (1956), "On the theory of mortality measurement," Skan. Aktuarieltidskr., 39, pp. 125-153.

[4] Rao, B.L.S. Prakasa (1966), "Asymptotic distributions in some non-regular statistical problems," Technical Report No. 9, Department of Statistics and Probability, Michigan State University.

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

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

[7] Wegman, E.J. (1970), "Maximum likelihood estimation of a unimodal density function," to appear Ann. Math. Statist., April, 1970.

Maximum Likelihood Estimation of a Unimodal Density, II

by

Edward J. Wegman

Abstract. This paper is a sequel to the earlier paper, "Maximum Likelihood Estimation of a Unimodal Density Function." The MLE of a unimodal density with unknown mode is shown to agree, for sufficiently large n and on certain regions, with the MLE of a unimodal density with known mode. The asymptotic distributions of the MLE's then agree. Also a geometrical interpretation of the MLE of a unimodal density with unknown mode is given.

Keywords: Maximum Likelihood Estimation
Unimodal Density
Asymptotic Distribution
Geometric Characterization
Mode

Subject Classification Numbers:
62.15
62.70
60.05