The Annals of Mathematical Statistics
1971, Vol. 42, No. 6, 1909–1915

A NOTE ON THE ESTIMATION OF THE MODE[1]

BY EDWARD J. WEGMAN
University of North Carolina

Let $X_1, \cdots, X_n$ be a sample from a unimodal distribution, F, and let $\{a_n\}$ be a sequence converging to zero. A nonparametric estimate of the mode is the center of the interval of length $2a_n$ containing the most observations. This estimate is shown to be strongly consistent and conditions on the speed at which a_n may converge to zero are given. This estimator of the mode is related to the naive density estimator, $(F_n(x+a_n)-F_n(x-a_n))/2a_n$, where F_n is the empirical distribution function. A simple strong consistency result for this naive density estimator is given. Also other estimators of the mode are discussed briefly and an application of estimators of the mode is mentioned.

1. Introduction and review. The estimation of the mode of a probability distribution has received the attention of several authors, Chernoff (1964), Dalenius (1965), Grenander (1965), Nadaraya (1965), Parzen (1962), Robertson, Cryer and Hogg (1968), Venter (1967) and Wegman (1970), recently. Without exception, to demonstrate consistency, these authors at least assume continuity of the density function, f, and of course, define the mode, M, as that number which maximizes f. That is, $f(M) > f(x)$ for every $x \neq M$. By a distribution of Type I, we mean that there is an M such that $f(M) > f(x)$ for every $x \neq M$. We shall not require continuity.

One could define the mode in several other cases. If there is exactly one infinite discontinuity in the density function, then the location of this discontinuity is the mode. That is, if there is exactly one M such that one or both of $\lim_{x\uparrow M} f(x)$ and $\lim_{x\downarrow M} f(x)$ is infinite, then M is the mode of f. We shall call this a Type II unimodal distribution function. Even more generally, if the distribution function may be written as the sum of an absolutely continuous distribution and a discrete distribution with isolated mass points, then the mass point with largest probability is the mode. This distribution we shall call a Type III unimodal distribution.

Estimates of the mode may be obtained either directly or indirectly. The indirect estimates are found as by products of density estimation procedures and include estimates found in Parzen (1963), Nadaraya (1965), and Wegman (1970).

A number of direct estimates of the mode have also been given. Chernoff (1964) chooses a sequence a_n (which converges to 0 sufficiently slowly) and picks at his estimate the center of the interval of length $2a_n$ which contains the most observations. This estimate is identical to the estimates of Parzen and Nadaraya when the kernel is chosen to be the uniform kernel. Chernoff and Parzen each present an

Received November 4, 1970.

[1] Research supported by National Science Foundation Grant GP–23520.

AMS 1970 *subject classification numbers:* 60.30, 62.10, 62.15, 62.70, 62.97.
Keywords and phrases: Mode, estimation of the mode, consistency, strong consistency.

argument for weak consistency and Nadaraya gives one to demonstrate the strong consistency. The results of Parzen and Nadaraya are by products of similar results for density estimates.

Venter (1967) and Dalenius (1965) apparently independently propose another estimate. Whereas Chernoff fixes the length of the interval and chooses the interval containing the most observations. Venter and Dalenius fix the number of observations and choose the shortest interval containing this fixed number of observations. Venter also provides a proof of strong consistency. This estimate is related to the "nearest neighbor" density estimate of Loftsgaarden and Quensenberry (1965). Moore and Henrichon (1969) discuss the relation of the Venter-Dalenius estimate of the mode to this nearest neighbor density estimate.

A refinement of the technique proposed by Venter and Dalenius is found in Robertson, Cryer and Hogg (1968), who define a function $k(n)$ and choose the shortest interval containing $k(n)$ observations. Within this interval they choose the smallest interval containing $k[k(n)]$ observations, repeating until a smallest interval contains something close to a fixed number of observations. These authors also provide strong consistency results.

Finally, Grenander (1965) provides several estimates, one of which he shows is consistent. Grenander's consistent estimate is based on the fact that raising a density to a power makes the mode more and more pronounced. Grenander's estimate is appealing because it uses all of the data, whereas the other estimates use only a portion of the data. Dalenius (1965) concludes from a Monte Carlo study that the Chernoff-type and Venter-type estimates are on the average closer to the mode than the Grenander-type but the former have larger variances than the latter. Finally, we remark that Chernoff, Venter and Grenander each develop the asymptotic distribution theory for their respective estimates.

In this paper we shall focus our attention on the Chernoff-type estimate and give strong consistency results when the sample is selected from distributions either of Type I, Type II or Type III.

2. Strong consistency. We limit our attention to proving strong consistency results from the Chernoff-type estimate. Let $X_1, X_2, \cdots, X_n$ be a sample from either a Type I, II, or III unimodal distribution. Let a_n be a sequence of numbers to be described later and let (l_n, r_n) be the interval of length $2a_n$ containing the most observations.

We prove two theorems: one applying to Type I and Type II distributions and the second to Type III distributions. The first theorem shall require a somewhat unusual condition on the distribution which we call Condition 1.

CONDITION 1. If (b_n, c_n) is a sequence of intervals with b_n and c_n both converging to $-\infty$ or to $+\infty$ and with $c_n - b_n$ converging to zero, it is clear that the probability measure of (b_n, c_n) converges to zero. In addition, we shall require that the probability measure of (b_n, c_n) eventually be less than the probability measure of I_n where I_n is either of the form $(\alpha, \alpha + c_n - b_n)$ or of the form $(\alpha - c_n + b_n, \alpha)$. Here α is any real number.

This condition is met if the density, $f(x)$, eventually decreases monotonically to zero as $x \to +\infty$ and $-\infty$. Also we assume that f is either left- or right- continuous at each of its finite discontinuities if any.

THEOREM 2.1. *If the distribution F is of Type I or II and satisfies Condition 1 and if a_n converges to 0 more slowly than* $[\log(\log(n))/n]^{\frac{1}{2}}$, *then l_n and r_n converge to the mode, M, with probability one.*

PROOF. Let $\Omega' = [\lim \sup_{n\to\infty} W_n[n/(\log\log n)]^{\frac{1}{2}} = 2^{-\frac{1}{2}}]$ where $W_n = \sup_x |F_n(x) - F(x)|$ and F_n is the empirical distribution function. Smirnov (1944) shows that Ω' has probability 1. (A proof in English may be found in Chung (1949) or in Csáki (1968).) Since $\Omega' \subset [\lim_{n\to\infty} W_n/a_n = 0]$, we shall restrict our attention to points in Ω'.

If l_n fails to converge to M with probability one, by the extended Bolzano-Weierstrauss Theorem, there are points in Ω' for which one of three consequences may occur:

(i) $\lim\sup l_n = +\infty$,

(ii) $\lim\inf r_n = -\infty$,

(iii) there is a subsequence l_{n_j} such that $l_{n_j} \to l \neq M$.

Cases (i) and (ii) are similar so we investigate only $\lim\inf r_n = -\infty$. There is a subsequence r_{n_j} diverging to $-\infty$. Choose $\alpha < M$ so that $0 < f(\alpha-) < f(M)$ (or $\lim_{x\to M} f(x)$). Now

$$F_{n_j}(r_{n_j}-) - F_{n_j}(r_{n_j}) \leqq |F_{n_j}(r_{n_j}-) - F(r_{n_j})| + |F_{n_j}(\ _{nj}) - F(l_{n_j})| + |F(r_{n_j}) - F(l_{n_j})|.$$

Dividing throughout by $2a_{nj}$ and letting $j \to \infty$, we obtain

$$\lim\sup \frac{F_{n_j}(r_{n_j}-) - F_{n_j}(l_{n_j})}{2a_{n_j}} \leqq \lim\sup \frac{F(r_{n_j}) - F(l_{n_j})}{2a_{n_j}}. \tag{2.1}$$

By Condition 1,

$$\frac{F(r_n) - F(l_n)}{2a_n} \leqq \frac{F(\alpha) - F(\alpha - 2a_n)}{2a_n} \qquad \text{eventually.}$$

Thus

$$\lim\sup \frac{F_{n_j}(r_{n_j}-) - F_{n_j}(l_{n_j})}{2a_{n_j}} \leqq f(\alpha-).$$

Thus we may obtain a subsequence of n_j (for convenience let us relabel it n_j) such that

$$\lim_{j\to\infty} \frac{F_{n_j}(r_{nj}-) - F_{n_j}(l_{nj})}{2a_{n_j}} \leqq f(\alpha-).$$

On the other hand, let $(l_n{}^*, r_n{}^*)$ be the interval of length $2a_n$ with center at the mode, M. If the Type I distribution has a jump at the mode or if only one of $\lim_{x\uparrow M} f(x)$ and $\lim_{x\downarrow M} f(x)$ is infinite, a slight change in the choice of $(l_n{}^*, r_n{}^*)$ is necessary. Consider

$$F(r_{nj}^*)-F(l_{nj}^*) \leqq |F(r_{nj}^*)-F_{n_j}(r_{n_j}^*-)|+|F(l_{nj}^*)-F_{n_j}(l_{nj}^*)|+|F_{n_j}(r_{n_j}^*-)-F_{n_j}(l_{nj}^*)|.$$

As before

$$\text{(2.2)} \qquad \liminf \frac{F(r_{nj}^*)-F(l_{nj}^*)}{2a_{n_j}} \leqq \liminf \frac{F_{n_j}(r_{nj}^*-)-F_{n_j}(l_{nj}^*)}{2a_{n_j}}$$

But the left-hand side is either $f(M)$ in the Type I distribution or ∞ in the Type II distribution. In either case we have

$$F_n(r_n-)-F_n(l_n) < F_n(r_n{}^*-)-F_n(l_n{}^*) \text{ i.o.}$$

That is to say, the number of observations in (l_n, r_n) is less than the number of observations $(l_n{}^*, r_n{}^*)$ infinitely often. But (l_n, r_n) was chosen to be the interval with the most observations so that we have a contradiction. Thus $\liminf r_n \neq -\infty$. By a similar argument, we obtain $\limsup l_n \neq +\infty$.

Suppose then there is a subsequence l_{n_j} such that $l_{n_j} \to l \neq M$. By an analysis similar to that we used to obtain (2.1), we obtain

$$\text{(2.3)} \qquad \limsup \frac{F_{n_j}(r_{n_j}-)-F_{n_j}(l_{n_j})}{2a_{n_j}} \leqq \limsup \frac{F(r_{n_j})-F(l_{n_j})}{2a_{n_j}}.$$

The right-hand side is less than or equal to the maximum of $f(l-)$ or $f(l+)$, both of which are less than $f(M)$. Thus we may obtain a subsequence of n_j (which as before we relabel n_j) such that

$$\lim_{j\to\infty} \frac{F_{n_j}(r_{n_j}-)-F_{n_j}(l_{n_j})}{2a_{n_j}} < f(M).$$

By arguments similar to those leading to (2.2), we obtain

$$\text{(2.4)} \qquad \liminf \frac{F_{n_j}(r_{nj}^*-)-F_{n_j}(l_{nj}^*)}{2a_{n_j}} \geqq f(M).$$

As before, this leads us to a contradiction, so that l_n can converge only to M. Since $r_n = l_n+2a_n$, and a_n converges to zero, r_n converges to M, which completes the proof.

Note that the proof just given does not require continuity of the density.

THEOREM 2.2. *If F is a Type* III *distribution and a_n is* any *sequence converging to* 0, *then l_n and $r_n \to M$ with probability one.*

PROOF. Let $\Omega' = [\lim_{n\to\infty} \sup_x |F_n(x) - F(x)| = 0]$. Ω' has probability one by the well-known Glivenko-Cantelli Theorem. We restrict our attention to $\omega \in \Omega'$. As in the proof of Theorem 2.1, one of three things may happen if $l_n \neq M$:

(i) $\limsup l_n = +\infty$

(ii) $\liminf r_n = -\infty$

(iii) there is a subsequence l_{n_j} such that $l_{n_j} \to l \neq M$

Again cases (i) and (ii) are similar and we will only investigate case (ii).

There is a subsequence r_{n_j} diverging to $-\infty$. Pick α so small that $F(\alpha-) < P(X = M)$. But since eventually $(l_{n_j}, r_{n_j}) \subset (-\infty, \alpha)$,

$$F_{n_j}(r_{n_j}-) - F_{n_j}(l_{n_j}) \leqq F_{n_j}(\alpha-). \tag{2.5}$$

Also eventually by the Glivenko-Cantelli Theorem,

$$F_{n_j}(\alpha) < F_{n_j}(M+) - F_{n_j}(M-).$$

Let (l_n^*, r_n^*) be as in Theorem 2.1 so that

$$F_{n_j}(\alpha) < F_{n_j}(M+) - F_{n_j}(M-) \leqq F_{n_j}(r_{n_j}^*-) - F_{n_j}(l_{n_j}^*). \tag{2.6}$$

Combining (2.5) and (2.6) we have the contradiction that (l_n^*, r_n^*) contains more observations than (l_n, r_n) infinitely often. Thus we have $\liminf r_n \neq -\infty$. Similarly, we may conclude $\limsup l_n \neq +\infty$.

Let us suppose that l_{n_j} converges to $l \neq M$. If l is a mass point of the discrete part of the distribution, then

$$\liminf \{F_{n_j}(r_{n_j}-) - F_{n_j}(l_{n_j})\} \leqq P(X = l),$$

otherwise

$$\liminf \{F_{n_j}(r_{n_j}-) - F_{n_j}(l_{n_j})\} = 0.$$

On the other hand

$$\limsup \{F_{n_j}(r_{n_j}^*-) - F_{n_j}(l_{n_j}^*)\} = P(X = M).$$

Since $P(X = M) > P(X = l)$, we again have the contradiction that the number of observations in (l_n, r_n) is less than the number of observations in (l_n^*, r_n^*) infinitely often. Thus we can only conclude that l_n and r_n converge to M for all points in Ω'.

In Section 1, we pointed out that this estimator of the mode suggested by Chernoff is related to the kernel estimators of the density, where the kernel is chosen to be the uniform kernel. The type of analysis found in Theorem 2.1 can be extended to show a simple proof of strong consistency for the kernel estimate with uniform kernel.

THEOREM 2.3. *If* $f_n(x) = [F_n(x+h_n) - F_n(x-h_n)]/(2h_n)$, *where* h_n *converges to* 0 *more slowly than* $[\log(\log(n))/n]^{\frac{1}{2}}$, *then at every continuity point* x *of* f, $f_n(x) \to f(x)$ *with probability one. In addition, if* $f(x)$ *is uniformly continuous, then* $f_n(x) \to f(x)$ *uniformly in* x *with probability one.*

PROOF. Let Ω' be defined as in Theorem 2.1. Fix $\omega \in \Omega'$. By noting that

$$(2.7)\quad \frac{F_n(x+h_n)-F_n(x-h_n)}{2h_n} \leqq \frac{|F_n(x+h_n)-F(x+h_n)|}{2h_n}+\frac{|F_n(x-h_n)-F(x-h_n)|}{2h_n}+\frac{F(x+h_n)-F(x-h_n)}{2h_n}$$

and letting n diverge to infinity, we have

$$\limsup \frac{F_n(x+h_n)-F_n(x-h_n)}{2h_n} \leqq f(x).$$

Similarly

$$(2.8)\quad \frac{F(x+h_n)-F(x-h_n)}{2h_n} \leqq \frac{|F(x+h_n)-F_n(x+h_n)|}{2h_n}+\frac{|F(x-h_n)-F_n(x-h_n)|}{2h_n}+\frac{F_n(x+h_n)-F_n(x-h_n)}{2h_n}$$

Letting n diverge to infinity, we have

$$f(x) \leqq \liminf \frac{F_n(x+h_n)-F_n(x-h_n)}{2h_n},$$

which is sufficient to complete the first part of the theorem.

Let us now assume that f is uniformly continuous. By combining (2.7) and (2.8),

$$\left|f_n(x)-\frac{F(x+h_n)-F(x-h_n)}{2h_n}\right| \leqq \frac{|F_n(x-h_n)-F(x-h_n)|}{2h_n}+\frac{|F_n(x+h_n)-F(x+h_n)|}{2h_n}.$$

Taking supremums over x and then limits as $n \to \infty$,

$$(2.9)\quad \lim_{n\to\infty} \sup_x \left|f_n(x)-\frac{F(x+h_n)-F(x-h_n)}{2h_n}\right| = 0.$$

But by the Mean Value Theorem,

$$\frac{F(x+h_n)-F(x-h_n)}{2h_n} = f(\theta)$$

for some θ with $|\theta-x| \leqq 2h_n$. Let $\varepsilon > 0$. Since f is uniformly continuous, there is a $\delta > 0$ such that for every x and θ with $|x-\theta| < \delta$, then $|f(x)-f(\theta)| < \varepsilon/2$. Choose n sufficiently large so that $2h_n < \delta$ and such that

$$\sup_x \left|f_n(x)-\frac{F(x+h_n)-F(x-h_n)}{2h_n}\right| < \varepsilon/2.$$

Then,

$$\sup_x |f_n(x)-f(x)| \leqq \sup_x |f_n(x)-f(\theta)|+\sup_x |f(\theta)-f(x)| < \varepsilon.$$

That is $\lim_{n\to\infty} \sup_x |f_n(x)-f(x)| = 0$.

The results of this section are not surprising since this estimate (as well as the others mentioned in Section 1) rely on the fact probability is concentrated near the mode. This is even more the case with distributions of Types II and III than with those of Type I. We note however that to our knowledge there are no other proofs of strong consistency without assuming continuity of the density.

Finally we should like to point out that one can use estimates of the mode to calibrate blood flow curves. The details of this application can be found in Benson (1970). The intuition behind the application is interesting. Clearly, there is a mean flow of blood through the body (or else we should all perish immediately). However, in some arteries, the most frequently occurring value of flow is zero corresponding to the rest period (diastole) of the heart. Knowledge of the latter is of much more interest to physiologists because the zero flow is much less dependent on the shape of the flow curve than is the mean flow.

Acknowledgments. I thank the referee for his helpful suggestions.

REFERENCES

BENSON, D. W. Jr. (1970). A relationship between aortic pressure and flow. Ph.D. dissertation, Univ. of North Carolina at Chapel Hill.

CHERNOFF, H. (1964). Estimation of the mode. *Ann. Inst. Statist. Math.* **16** 31–41.

CHUNG, K. L. (1949). An estimate concerning the Kolmogorov limit distribution. *Trans. Amer. Math. Soc.* **67** 36–50.

CSÁKI, E. (1968). An iterated logarithm law for semimartingales and its application to empirical distribution function. *Studia Sci. Math. Hungar.* **3** 287–292.

DALENIUS, TORE (1965). The mode—a neglected statistical parameter. *J. Roy. Statist. Soc. Ser. A* **128** 110–117.

GRENANDER, ULF (1965). Some direct estimates of the mode. *Ann. Math. Statist.* **36** 131–138.

LOFTSGAARDEN, D. O. and QUENSENBERRY, C. P. (1965). A non-parametric estimate of a multivariate density function. *Ann. Math. Statist.* **38** 1261–1265.

MOORE, P. S. and HENRICHEN, E.G. (1969). Uniform consistency of some estimates of a density function. *Ann. Math. Statist.* **40** 1499–1502.

NADARAYA, É. A. (1965). On non-parametric estimates of density functions and regression curves. *Theor. Probability Appl.* **10** 186–190.

PARZEN, E. (1962). On estimation of a probability density function and its mode. *Ann. Math. Statist.* **33** 1065–1076.

ROBERTSON, TIM, CRYER, J. D. and HOGG, R. V. (1968). On non-parametric estimation of distributions and their modes. Unpublished manuscript.

SMIRNOV, N. V. (1944). Approximation of distribution laws of random variables by empirical data. *Ycnexu Mam. Hayk.* **10** 179–206. (In Russian.)

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

WEGMAN, E. J. (1970). Maximum likelihood estimation of a unimodal density function. *Ann. Math. Statist.* **41** 457–471.