

¹ Department of Statistics, University of North Carolina, Chapel Hill, North Carolina. The research of this author and the preparation of the manuscript were supported by National Science Foundation Grant GP-23420.

² Department of Biomathematics and Medical Engineering, University of North Carolina, Chapel Hill, North Carolina, and Division of Cardiology, Veterans Administration Hospital, Durham, North Carolina. The research of this author was supported by National Institute of Health Training Grant # 5T-01-GM-01-504-03.

Estimation of the mode with an application
to cardiovascular physiology

by

Edward J. Wegman¹

D. Woodrow Benson, Jr.²

Department of Statistics
University of North Carolina at Chapel Hill
Institute of Statistics Mimeo Series No. 702

August, 1970

Estimation of the mode with an application to cardiovascular physiology.

by

Edward J. Wegman¹

D. Woodrow Benson, Jr.²

1. Introduction and Review. The estimation of the mode of a probability distribution has received the attention of several authors, [2], [5], [6], [8], [10], [11], [13] and [14] 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 $x \neq M$. We shall call this type of distribution a unimodal distribution of Type I.

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 $\lim_{x \rightarrow M} f(x) = \infty$, 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.

In the remainder of this section, we shall discuss various proposed estimates of the mode. In Section 2, we shall provide some strong consistency

¹ Department of Statistics, University of North Carolina, Chapel Hill, North Carolina. The research of this author and the preparation of the manuscript were supported by National Science Foundation Grant GP-23520.

² Department of Biomathematics and Medical Engineering, University of North Carolina, Chapel Hill, North Carolina, and Division of Cardiology, Veterans Administration Hospital, Durham, North Carolina. The research of this author was supported by National Institute of Health Training Grant # 5T-01-GM-01-504-03.

results when the underlying distributions are of Type II and Type III. In Section 3, we give an interesting physiological application.

Estimates of the mode are obtained directly or indirectly. The indirect estimates are found as by products of density estimation procedures and include estimates found in Parzen [10], Nadaraya [8], and Wegman [14].

Parzen and Nadaraya use the kernel-type density estimation and Wegman uses a maximum likelihood technique. The indirect estimate of the mode is chosen as the mode of the density estimate.

There are several direct estimates of the mode. Chernoff [2] chooses a sequence a_n converging to 0 sufficiently slowly and picks as 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. Several consistency results are available for this estimate. Chernoff shows consistency. Parzen obtains consistency and Nadaraya strong consistency as by products of similar results for the density estimates.

Venter [13] and Dalenius [5] apparently independently propose the same alternative 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 [7].

A refinement of the technique proposed by Venter and Dalenius is found in Robertson, Cryer and Hogg [11]. They 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 [6] provides several estimates, one of which he shows is consistent. Grenander's 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 throw more or less of the data away. Dalenius [5] 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.

2. Strong Consistency. We shall limit our attention to strong consistency results for the Chernoff-type estimate. Let X_1, X_2, \dots, X_n be a sample from any type 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 shall 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 shall call Condition I. 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 0, it is clear that the probability measure of (b_n, c_n) converges to 0. In addition to this 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 $(\alpha - c_n + b_n, \alpha)$. This condition is met if the density, $f(x)$, eventually decreases monotonically to 0 as $x \rightarrow +\infty$ or $-\infty$. Also we shall 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 I and if a_n converges to 0 more slowly than $\sqrt{\log(\log(n))}/n$, then ℓ_n and r_n converge to the mode, M , with probability one.

PROOF: Let $\Omega' = \left[\limsup_{n \rightarrow \infty} W_n \sqrt{\frac{n}{\log \log n}} = \frac{1}{\sqrt{2}} \right]$ where $W_n = \sup_x |F_n(x) - F(x)|$ and F_n is the empirical distribution function. Smirnov [12] shows that Ω' has probability 1. A proof in English may be found in Chung [3] or in Csáki [4]. It is evident that $\Omega' \subset \left[\lim_{n \rightarrow \infty} \frac{W_n}{a_n} = 0 \right]$. 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

things may happen:

1. $\limsup l_n = +\infty$
2. $\liminf r_n = -\infty$
3. There is a subsequence l_{n_j} such that $l_{n_j} \rightarrow l \neq M$.

The cases 1 and 2 are similar so we shall investigate $\liminf r_n = -\infty$. There is a subsequence r_{n_j} diverging to $-\infty$. Choose $\alpha < M$ so that $0 < f(\alpha-) < f(M)$ (or $\lim_{x \rightarrow M} f(x)$). Now

$$F_{n_j}(r_{n_j}-) - F_{n_j}(l_{n_j}) \leq \\ |F_{n_j}(r_{n_j}-) - F(r_{n_j})| + |F_{n_j}(l_{n_j}) - F(l_{n_j})| + |F(r_{n_j}) - F(l_{n_j})|.$$

Dividing by $2a_{n_j}$ and letting $j \rightarrow \infty$,

$$(2.1) \quad \limsup \frac{F_{n_j}(r_{n_j}-) - F_{n_j}(l_{n_j})}{2a_{n_j}} \leq \limsup \frac{F(r_{n_j}) - F(l_{n_j})}{2a_{n_j}}.$$

By Condition I,

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

Thus

$$\limsup \frac{F_{n_j}(r_{n_j}-) - F_{n_j}(l_{n_j})}{2a_{n_j}} \leq f(\alpha-).$$

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

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

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, a slight change in the choice of (l_n^*, r_n^*) is necessary. Consider

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

As before

$$(2.2) \quad \liminf \frac{F(r_{n_j}^*) - F(l_{n_j}^*)}{2a_{n_j}} \leq \liminf \frac{F_{n_j}(r_{n_j}^*-) - F_{n_j}(l_{n_j}^*)}{2a_{n_j}}.$$

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

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

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

$\liminf r_n \neq -\infty$. Similarly $\limsup l_n \neq +\infty$.

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

$$(2.3) \quad \limsup \frac{F_{n_j}(r_{n_j}^-) - F_{n_j}(l_{n_j})}{2a_{n_j}} \leq \limsup \frac{F(r_n) - F(l_n)}{2a_{n_j}}.$$

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

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

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

$$(2.4) \quad \liminf \frac{F_{n_j}(r_{n_j}^{*-}) - F_{n_j}(l_{n_j}^*)}{2a_{n_j}} \geq f(M).$$

As before, this leads us to a contradiction, so that we must conclude that l_n can converge only to M . But $r_n = l_n + 2a_n$, so that r_n also converges to M . This completes the proof of Theorem 2.1.

We remark here 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 $r_n \rightarrow M$ with probability one.

PROOF: Let $\Omega' = [\lim_{n \rightarrow \infty} \sup_x |F_n(x) - F(x)| = 0]$. 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:

1. $\limsup l_n = +\infty$,
2. $\liminf r_n = -\infty$,
3. there is a subsequence l_{n_j} such that $l_{n_j} \rightarrow l \neq M$.

Again cases 1 and 2 are similar. We will investigate case 2.

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

$$(2.5) \quad F_{n_j}(r_{n_j}^-) - F_{n_j}(l_{n_j}) \leq F_{n_j}(\alpha-).$$

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

$$(2.6) \quad F_{n_j}(\alpha) < F_{n_j}(M+) - F_{n_j}(M-) \leq F_{n_j}(r_{n_j}^{*-}) - F_{n_j}(\ell_{n_j}^*).$$

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 $\limsup r_n \neq +\infty$.

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

$$\liminf F_{n_j}(r_{n_j}^-) - F_{n_j}(\ell_{n_j}) \leq P(X=\ell)$$

otherwise

$$\liminf F_{n_j}(r_{n_j}^-) - F_{n_j}(\ell_{n_j}) \leq 0.$$

On the other hand

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

Since $P(X=M) > P(X=\ell)$ 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 estimate 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 $f_n(x) = \frac{F_n(x+h_n) - F_n(x-h_n)}{2h_n}$, where h_n converges to 0 more slowly than $\sqrt{\log(\log(n))/n}$, then at every continuity point x of f ,

$f_n(x) \rightarrow f(x)$ with probability 1.

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

$$\frac{F_n(x+h_n) - F_n(x-h_n)}{2h_n} \leq$$

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

and letting n diverge to infinity, we have

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

Similarly

$$\frac{F(x+h_n) - F(x-h_n)}{2h_n} \leq$$

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

Letting n diverge to infinity, we have

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

which is sufficient to complete the proof.

3. An Application. An interesting application of estimates of the mode occurs in determining the baseline calibration for ascending aortic flow as obtained with an electromagnetic flowmeter. The physiological details may be found in [1].

Records of ascending aortic flow are illustrated in Figure 1. For purposes of this discussion, each cycle of the flow record can be discussed in terms of two intervals: systole and diastole. The prominent, positive deflection in the flow record corresponds to the ejection of blood by the heart into the ascending aorta during systole. During diastole, there is a small retrograde (negative) flow associated with closure of the aortic valves. During the remainder of diastole, flow is virtually zero.

The electromagnetic flowmeters are stable for short periods of time, say 1 minute, but over longer periods, say 10 - 20 minutes, the baseline corresponding to zero flow may be expected to drift. In order to process the flow data with a digital computer, it is essential to identify zero flow. It is known physiologically that the interval of virtually zero flow, diastole, is longer than systole. Therefore, when discrete equally spaced samples of the flow record are available, the most frequently occurring value should correspond to zero flow. This suggests that the problem of determining the zero flow baseline somehow corresponds to estimating a mode. Since a parametric form of the true flow curve is not known, exact arguments cannot be given. We present, however, a heuristic discussion.

Denote the flow curve without noise by $f\ell(t)$; we will assume $f\ell(t)$ has a derivative, so that it is measurable. The time interval required to complete one cycle will be denoted by $[t_0, t_0 + \tau]$; we will choose a point, T , at random in this interval. Thus, T is a random variable distributed uniformly on $[t_0, t_0 + \tau]$. We wish to determine the density function of the random variable $f\ell(T)$. The exact form of the density depends on the nature of $f\ell(t)$, but we

observe that $f\ell(t)$ has zero derivative at the peak positive flow, at the peak negative flow, and throughout the zero flow portion of the cycle. Thus, the Jacobian of the transformation, $f\ell$, will be infinite in these places, and the density will have a shape as in Figure 2. In addition, the point corresponding to zero flow will have probability mass proportional to the length of time there is zero flow. Thus, $f\ell(T)$ has a distribution of Type III. According to Theorem 2.2, we could consistently estimate the zero flow baseline by choosing points, T_1, T_2, \dots, T_n , at random from $[t_0, t_0 + \tau]$ and applying the Chernoff estimate of the mode to $f\ell(T_1), f\ell(T_2), \dots, f\ell(T_n)$.

In practice, samples $f\ell(T_1), f\ell(T_2), \dots, f\ell(T_n)$ are obtained with an analog to digital converter and $f\ell(T)$ is contaminated by electrical noise, errors in conversion, electrical artifacts arising from artificial or spontaneous activation of the heart, and other unidentified noise. None the less, this technique has proved successful in accurately locating the zero flow baseline.

A frequently used alternative is the average diastolic flow; O'Rourke and Taylor [9] have used this approach to obtain a baseline calibration. A disadvantage with this method is that the beginning and end of diastole must be accurately known. Furthermore, the presence of large artifacts during diastole (see Figure 1-B) would reduce the effectiveness of a calibration based on an average diastolic value. These problems are not encountered when an estimate of the mode is used.

AORTIC FLOW (cm³/sec.)

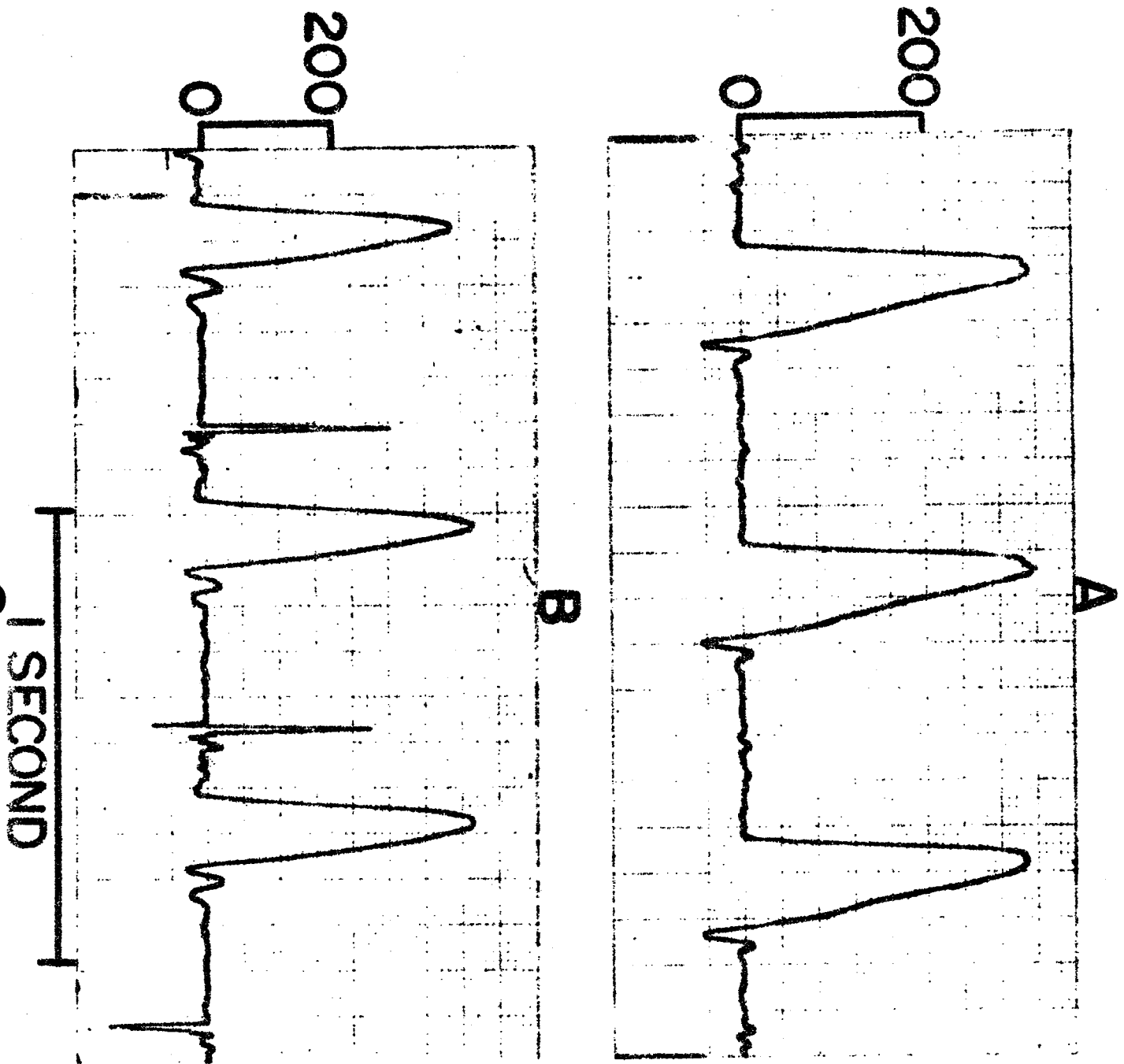


FIGURE 1. This figure illustrates portions of two records of ascending aortic blood flow obtained from dogs with chronically implanted electromagnetic flowmeters. The record in Panel B demonstrates a large artifact during diastole; this corresponds to application of a pacing voltage to the heart.

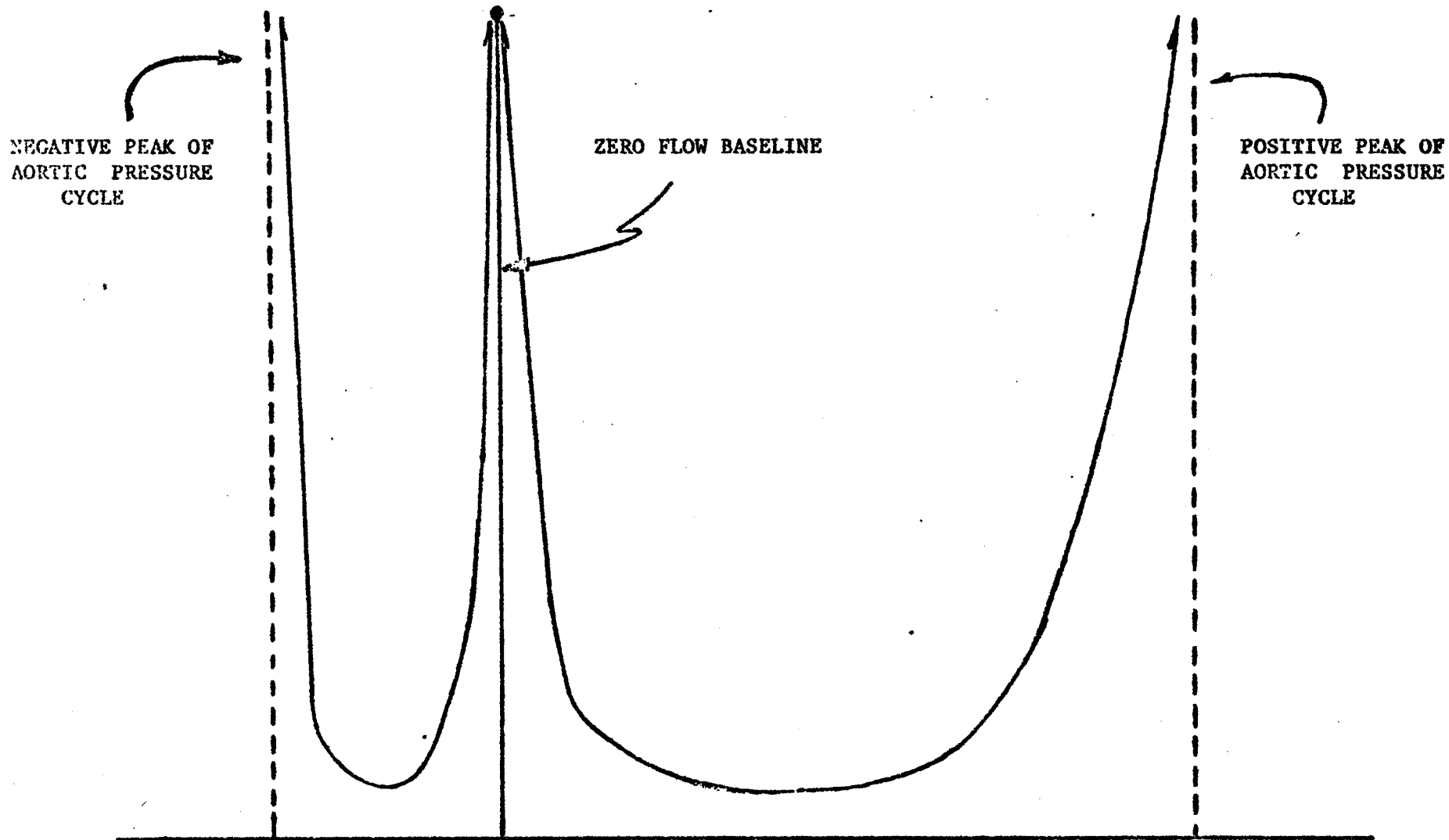


FIGURE 2. This figure illustrates the approximate shape of the density of $fZ(T)$, where T is uniformly distributed.

4. Conclusions. The results of Section 2 are really not surprising. All of the estimates of the mode rely on the fact that probability is most concentrated around the mode. This is even more true in distributions of Types II and III than in distributions of Type I. One can speculate, therefore, that results analogous to Theorems 2.1 and 2.2 should hold for the other estimates of the mode.

We remark also on the choice of a_n . It is desirable to choose a_n as small as possible to have the interval estimate of the mode as short as possible. Choice of too short an interval, however, may lead to spurious results. Theorem 2.1 and the result of Nadaraya [8] both indicate $1/a_n$ should be chosen to be $o(n^\alpha/f(n))$ where $f(n)$ is some slowly increasing function of n and $\alpha < \frac{1}{2}$. The consistency result of Chernoff suggests that $1/a_n$ can be $o(n^\alpha/f(n))$ where $f(n)$ is as above and $\alpha < 1$. This allows a_n to be somewhat shorter but the consistency result is correspondingly somewhat weaker. Of course, Theorem 2.1 is less restrictive on the density, f .

Finally, we remark that the proof of Theorem 2.2 holds when the distribution is a discrete distribution with isolated mass points.

References

- [1] Benson, D.W., Jr. (1970), *A Relationship Between Aortic Pressure and Flow*, a Ph.D. dissertation submitted to the faculty of Biomathematics and Medical Engineering, University of North Carolina at Chapel Hill.
- [2] Chernoff, H. (1964), "Estimation of the mode," *Ann. of the Inst. of Stat. Math.*, 16, pp. 31-41.
- [3] Chung, K.L. (1949), "An estimate concerning the Kolmogorov limit distribution," *Trans. Amer. Math. Soc.*, 67, pp. 36-50.
- [4] Csáki, E. (1968), "An iterated logarithm law for semimartingales and its application to empirical distribution function," *Studia Scient. Math. Hung.*, 3, pp. 287-292.
- [5] Dalenius, Tore (1965), "The mode—a neglected statistical parameter," *JRSS(A)*, 128, pp. 110-117.
- [6] Grenander, Ulf (1965), "Some direct estimates of the mode," *Ann. Math. Statist.*, 36, pp. 131-138.
- [7] Loftsgaarden, D.O. and Quensenberry, C.P. (1965), "A non-parametric estimate of a multivariate density function," *Ann. Math. Statist.*, 38, pp. 1261-1265.
- [8] Nadaraya, É.A. (1965), "On nonparametric estimates of density functions and regression curves," *Theory Prob. Appl.*, 10, pp. 186-190.
- [9] O'Rourke, M.F. and Taylor, M.G. (1967), "Input impedance of the systemic circulation," *Circulation Res.*, 20, pp. 365-380.
- [10] Parzen, E. (1962), "On estimation of a probability density function and its mode," *Ann. Math. Statist.*, 33, pp. 1065-1076.
- [11] Robertson, Tim; Cryer, J.D; and Hogg, R.V. (1968), "On non-parametric estimation of distributions and their modes," an unpublished manuscript (see their appendix).
- [12] Smirnov, N.V. (1944), "Approximation of distribution laws of random variables by empirical data," *Yenezu Mam. Hayk.*, 10, pp. 179-206 (in Russian).
- [13] Venter, J.H. (1967), "On estimation of the mode," *Ann. Math. Statist.*, 38, pp. 1446-1455.
- [14] Wegman, E.J. (1970), "Maximum likelihood estimation of a unimodal density function," *Ann. Math. Statist.*, 41, pp. 457-471.