SOME DIRECT ESTIMATES OF THE MODE

By Ulf Grenander

University of Stockholm

0. Summary. Consider absolutely continuous unimodel distributions on the real line. A class of estimates is proposed for estimating the mode of the probability distribution. The large sample behavior is studied, and it is found that in the case of main interest the estimates are not even consistent. To remedy this a modification is suggested and it is shown that the new estimates are consistent.

1. Introduction. In recent years there has appeared a number of papers devoted to the study of estimating the frequency function of an absolutely continuous probability distribution on the real line. We refer to the list of references given at the end of this paper. In general one starts from some smoothed form of the empirical distribution function and the estimate is obtained by differentiation. A typical estimate of the frequency function may take the form

$$f^*(y) = n^{-1} \sum_{\nu=1}^{n} K(y - y_\nu),$$

where $y_1, y_2, \cdots, y_n$ represents the sample and $K(y)$ is a function, usually a frequency function concentrated around $y = 0$. A reader familiar with statistical spectral analysis would find a resemblance to the problem of estimating the spectral density of a stationary stochastic process. The analogy does not go very far; mathematically the two estimation problems are distinct.

If one wants to estimate the mode M of $f(x)$, assuming that it is well defined, it would be natural to look for the value or values of y that make $f^*(y)$ as large as possible. This would give a class of reasonable estimates of the mode; see [1].

The present paper deals with the following problem. Is it possible to find direct estimates of M, i.e. analytical expressions which behave reasonably well as estimates of M? Such an expression could be taken as a (quasi-linear) weighted average of the order statistics x_i, $x_1 \leqq x_2 \leqq \cdots \leqq x_n$, such that the weights tend to be large at values where the frequency function is large. The author has been interested in expressions like $M_p^* = B/A$, where

$$B = \tfrac{1}{2} \sum_{\nu=1}^{n-1} (x_{\nu+1} + x_\nu)/(x_{\nu+1} - x_\nu)^p, \qquad A = \sum_{\nu=1}^{n-1} 1/(x_{\nu+1} - x_\nu)^p.$$

The idea is of course that this should be close to

$$M_{p+1} = \int_{-\infty}^{\infty} x f^{p+1}(x)\, dx \Big/ \int_{-\infty}^{\infty} f^p(x)\, dx.$$

The latter quantity will be close to M if we choose the number p sufficiently large, at least under appropriate regularity conditions. We shall investigate the

Received March 9, 1964.

large sample behavior of M_p^* in the next section. Unfortunately it will turn out that M_p^* has a non-degenerated limit distribution if $p > 1$. Since the estimate is not even consistent it is necessary to modify it. This is done in Section 3 where it is also proved that the modified estimates are consistent.

Optimality questions will not be discussed in this paper. T. Dalenius and some of his colleagues at the University of Stockholm have studied various estimates of the mode numerically. Their results will appear in a separate paper that will also contain a discussion of the inference aspects of the problem.

2. A direct estimate. Let us consider a stochastic variable with an absolutely continuous c.d.f. $F(x)$ and the fr.f. $f(x)$, so that $F(x) = \int_{-\infty}^{x} f(\xi)\, d\xi$. We shall assume throughout this paper that

(1) $f(x)$ is positive and has a continuous derivative,

(2) $f(x)$ has a unique maximum, the mode M, so that $f(x) \leqq f(M)$ with equality only for $x = M$, and

(3) $f(x)$ is monotone for large values of $|x|$. We do not assume any parametric representation for $f(x)$, nor do we restrict the choice of $f(x)$ by any symmetry properties.

The task of estimating M is then of non-parametric character. We shall start by investigating the properties of the estimates,

$$M_p^* = B/A,$$

where

$$B = \tfrac{1}{2} \sum_{\nu=1}^{n-1} (x_{\nu+1} + x_\nu)/(x_{\nu+1} - x_\nu)^p, \qquad A = \sum_{\nu=1}^{n-1} 1/(x_{\nu+1} - x_\nu)^p$$

and p is a positive number. The estimate is computed from the ordered sample, $x_1 < x_2 < \cdots < x_n$, the order statistics.

In the usual way one can consider the ordered sample $(x_1, x_2, \cdots, x_n)$ as obtained from an ordered sample $(z_1, z_2, \cdots z_n)$ from a rectangular distribution $R(0, 1)$ via the transformation $G(z) : x_\nu = G(z_\nu)$, where G denotes the inverse function corresponding to $F(x)$, $G = F^{-1}$. It is now necessary to get around the difficulty inherent in the dependence between the various z-variables. To do this one can appeal to the following time-honoured idea. It is known, or can be directly verified, that the z's can be represented as

$$z_1 = \xi_1/\xi_1 + \xi_2 + \cdots + \xi_{n+1} = \eta_1/\eta_{n+1}$$

$$z_2 = \xi_1 + \xi_2/\xi_1 + \xi_2 + \cdots + \xi_{n+1} = \eta_2/\eta_{n+1}$$

$$\cdots$$

$$z_n = \xi_1 + \xi_2 + \cdots + \xi_n/\xi_1 + \xi_2 + \cdots + \xi_{n+1} = \eta_n/\eta_{n+1}$$

$$\eta_\nu = \xi_1 + \xi_2 + \cdots + \xi_\nu$$

where $\xi_1, \xi_2, \cdots \xi_{n+1}$ are independent observations from a negative exponential population with the frequency function $e^{-\xi}$, $\xi \geqq 0$ and 0 for $\xi < 0$.

Let $\psi_n(t)$ be the c.d.f. of the estimate M_p^*. Then

$$\psi_n(t) = P(B/A \leqq t) = P(B - At \leqq 0)$$

so that we would investigate the probability distribution of the stochastic variable

$$S = \sum_{\nu=1}^{n-1} n^{-2p} \frac{\frac{1}{2}\left[G\left(\frac{\eta_{\nu+1}}{\eta_{n+1}}\right) + G\left(\frac{\eta_\nu}{\eta_{n+1}}\right)\right] - t}{\left[G\left(\frac{\eta_{\nu+1}}{\eta_{n+1}}\right) - G\left(\frac{\eta_\nu}{\eta_{n+1}}\right)\right]^p} = \sum_{\nu=1}^{n-1} C_\nu$$

The factor n^{-2p} has been inserted for convenience in the following derivation. Now we have arrived at something very close to a sum of independent and identically distributed stochastic variables to which we can apply the classical results on probabilistic limit laws. To reduce the non-linear behavior of G split the sum into a number of partial sums,

$$S = S_0 + S_1 + \cdots + S_{s-1}$$

where

$$S_0 = \sum_{1}^{[\alpha_1 n]-1} C_\nu$$

$$S_r = \sum_{[\alpha_r n]}^{[\alpha_{r+1} n]-1} C_\nu, \qquad r = 1, 2, \cdots, s-2$$

$$S_{s-1} = \sum_{[\alpha_{s-1} n]}^{n-1} C_\nu$$

and $0 < \alpha_1 < \alpha_2 < \cdots < \alpha_{s-1} < 1$ is a division D of the unit interval. In the following D is chosen sufficiently fine and then kept fixed.

It is now easy to find suitable bounds for S_r. For $0 < r < s - 1$ with the obvious modifications for $r = 0, s - 1$,

$$n^{-2p} \frac{G[\eta_{[\alpha_r n]}/\eta_{n+1}] - t}{\max g^p(z)} \sum_{[\alpha_r n]}^{[\alpha_{r+1} n]-1} \frac{\eta_{n+1}^p}{\xi_{\nu+1}^p} \leqq S_r$$

$$\leqq n^{-2p} \frac{G[\eta_{[\alpha_{r+1} n]}/\eta_{n+1}] - t}{\min g^p(z)} \sum_{[\alpha_r n]}^{[\alpha_{r+1} n]-1} \frac{\eta_{n+1}^p}{\xi_{\nu+1}^p},$$

where $\max g^p(z)$, $\min g^p(z)$ denote the maximum and minimum respectively of the function $g^p(z)$ when z takes the values from $z_{[\alpha_r n]} = \eta_{[\alpha_r n]}/\eta_{n+1}$ to $z_{[\alpha_{r+1} n]} = \eta_{[\alpha_{r+1} n]}/\eta_{n+1}$, and $g(z) = G'(z)$. Note that $g(z) = 1/f(x)$ with $z = G(x)$. But as n tends to infinity we know that $z_{[\alpha_r n]}$ tends to α_r in probability. We also know that

$$(n^p)^{-1} \zeta_{n+1}^p = \left(\frac{\xi_1 + \xi_2 + \cdots + \xi_{n+1}}{n}\right)^p \to (E\xi)^p = 1$$

in probability. Hence we have

LEMMA 1. *For a given positive* ϵ *there exists* n_0 *such that for* $n > n_0$ *with probability arbitrarily close to* 1, *the simultaneoss inequalities hold*:

$$n^{-p} \frac{G(\alpha_r) - \epsilon - t}{\max\limits_{\alpha_r \leqq z \leqq \alpha_{r+1}} g^p(z) + \epsilon} \sum_{[\alpha_r n]}^{[\alpha_{r+1} n]-1} \frac{1}{\xi_{\nu+1}^p} \leqq S_r$$

$$\leqq n^{-p} \frac{G(\alpha_{r+1}) + \epsilon - t}{\min\limits_{\alpha_r \leqq z \leqq \alpha_{r+1}} g^p(z) - \epsilon} \sum_{[\alpha_r n]}^{[\alpha_{r+1} n]-1} \frac{1}{\xi_{\nu+1}^p}.$$

To obtain the asymptotic distribution of the left and right members of the inequality in the lemma one should observe that the terms $\zeta_{\nu+1} = 1/\xi_{\nu+1}^p$ are all independent and have the same frequency function

$$h(\zeta) = (1/p\zeta^{(1/p)+1}) \exp(- \zeta^{-1/p}), \; \zeta > 0.$$

This function vanishes at $\zeta = 0$ together with all its derivatives. For large values of ζ the function behaves as $1/p\zeta^{(1/p)+1}$. This means that if p is large enough the appropriate limit theorems will lead us to the stable laws.

In order to get the limit law with all its parameters for the variable $Z = 1/m^p \sum_{\nu=1}^m \zeta_\nu$, it is just as simple to do this directly as to get it by appealing to the classical results. The characteristic function $\varphi(z)$ of ζ is

$$\varphi(z) = p^{-1} \int_0^\infty \exp\,(iz\zeta - \zeta^{-1/p}) \frac{d\zeta}{\zeta^{(1/p)+1}}$$

and

$$\varphi(z) - 1 = p^{-1} \int_0^\infty \exp(-\zeta^{-1/p})[\exp(iz\zeta) - 1] \frac{d\zeta}{\zeta^{(1/p)+1}}.$$

For $z > 0$ this turns out to be

$$\varphi(z) - 1 = \frac{z^{1/p}}{p} \int_0^\infty \exp[-(u/z)^{-1/p}][e^{iu} - 1] \frac{du}{u^{(1/p)+1}}.$$

If $p > 1$ we have for small values of z because of dominated convergence

$$\varphi(z) - 1 = (z^{1/p}/p)[\gamma + o(1)]$$

with

$$\gamma = \int_0^\infty [e^{iu} - 1] \frac{du}{u^{(1/p)+1}} = e^{-\pi i/2p}\Gamma(- p^{-1}).$$

This gives for $z > 0$

$$\log E \exp izZ = m/p(m^{-p}z)^{1/p}[\gamma + o(1)] \rightarrow \gamma z^{1/p}/p,$$

which completes the proof of

LEMMA 2. *The normed sums* Z *have, for* $p > 1$, *as limiting distributions the stable* *laws with the characteristic function*

$$\exp \gamma z^{1/p}/p,\ z \geqq 0$$
$$\exp \bar{\gamma}|z|^{1/p}/p,\ z \leqq 0.$$

From Lemma 2 it follows that the limiting distribution of the sum

$$1/n^p \sum_{[\alpha_r n]}^{[\alpha_{r+1}n]-1} 1/\zeta_{\nu+1}^p$$

has a characteristic function whose logarithm is given by

$$(\alpha_{r+1} - \alpha_r)(\gamma/p)z^{1/p},\ z > 0.$$

It is necessary to deal separately with S_0 and S_r ; let us look at S_0. But

$$S_0 = n^{-2p} \sum_1^{[\alpha_1 n]-1} \frac{x_{\nu+1} - x_\nu - 2t}{2[x_{\nu+1} - x_\nu]^p} = \frac{n^{-2p}}{2} \sum_1^{[\alpha_1 n]-1} \frac{1}{[x_{\nu+1} - x_\nu]^{p-1}}$$
$$+ n^{-2p} \sum_1^{[\alpha_1 n-1]} \frac{(x_\nu - t)}{[x_{\nu+1} - x_\nu]^p} = \Sigma_1 + \Sigma_2 .$$

The first of these two terms can be written as

$$\Sigma_1 = 1/2n^2 \sum_1^{[\alpha_1 n-1]} n^{-2(p-1)}\{1/g^{p-1}(\theta_\nu)[z_{\nu+1} - z_\nu]^{p-1}\}$$

where θ_ν is some number in the interval $(z_\nu , z_{\nu+1})$. But expressions of the form

$$n^{-1} \sum_1^n n^{-2q}\, 1/[z_{\nu+1} - z_\nu]q;\ q > 0;$$

have limit distributions as $n \to \infty$; see the proof of Lemma 2. Since the values of $g(\theta_\nu)$ are bounded from below it follows that Σ_1 tends to zero in probability. The second term Σ_2 can be dominated, if n is large enough,

$$|\Sigma_2| \leqq \sum_1^{[\alpha_1 n]-1} n^{-2p} \frac{|G(z_\nu)| + |t|}{g^p(z_\nu)[z_{\nu+1} - z_\nu]^p}.$$

Hence

$$|\Sigma_2| \leqq \frac{|G(\varphi)| + |t|}{g^p(\varphi)} \sum_1^{[\alpha_1 n]-1} n^{-2p} 1/[z_{\nu+1} - z_\nu]^p,$$

where φ is in the interval $(z_1 , z_{[\alpha_1 n]})$. We now assume that $f(x) = O(|x|^{-\beta})$, $\beta > 1$, as $x \to -\infty$. Then the function

$$\frac{|G(z)| + |t|}{g^p(z)} = (|x| + t)f^p(x) = O(x^{1-p\beta}) = O(1).$$

But if α_1 is small, the limiting distribution of

$$\sum_1^{[\alpha_1 n]-1} n^{-2p}\ 1/[z_{\nu+1} - z_\nu]^p$$

is concentrated around the value zero. This shows that $\Sigma_1 + \Sigma_2$ can be neglected as $n \to \infty$ if α_1 is small enough. The term S_r is treated analogously.

Combining this with Lemma 1, it follows that S is with probability arbitrarily close to 1 contained between two stochastic variables whose characteristic functions have logarithms close to

$$\frac{\bar{\gamma} z^{1/p}}{p} \sum_{\alpha_{r+1} \leqq F(t \mp \epsilon)} \left[\frac{t \mp \epsilon - G(\alpha_{r+1})}{\min\limits_{\alpha_r \leqq z \leqq \alpha_{r+1}} \text{or} \max g^p(z) \mp \epsilon} \right]^{1/p} (\alpha_{r+1} - \alpha_r)$$
$$+ \frac{\gamma z^{1/p}}{p} \sum_{\alpha_{r+1} > F(t \mp \epsilon)} \left[\frac{G(\alpha_{r+1}) \pm \epsilon - t}{\min\limits_{\alpha_r \leqq z \leqq \alpha_{r+1}} \text{or} \max g^p(z) \mp \epsilon} \right]^{1/p} (\alpha_{r+1} - \alpha_r);\ z > 0.$$

This implies that the logarithm of the limiting characteristic function of S is

$$\frac{\bar{\gamma} z^{1/p}}{p} \int_{\alpha \leqq F(t)} \frac{[t - G(\alpha)]^{1/p}}{g(\alpha)} d\alpha + \frac{\gamma z^{1/p}}{p} \int_{\alpha > F(t)} \frac{[G(\alpha) - t]^{1/p}}{g(\alpha)} d\alpha$$
$$= \frac{\bar{\gamma} z^{1/p}}{p} \int_{-\infty}^{t} (t - x)^{1/p} f^2(x)\, dx + \frac{\gamma z^{1/p}}{p} \int_{t}^{\infty} (x - t)^{1/p} f^2(x)\, dx.$$

Indeed for the finite part of the integrals the proof is complete. It remains to verify that the contributions of the tails of the integrals can be made small. Take e.g.

$$\int_{-\infty}^{-A} (t - x)^{1/p} f^2(x)\, dx = O\left(\int_{-\infty}^{-A} (t - x)^{1/p} x^{-2\beta}\, dx\right)$$
$$= O\left(\int_{-\infty}^{-A} x^{-2\beta+1}\, dx\right) = O(A^{-2(\beta-1)})$$

which tends to zero as $A \to -\infty$. We have proved the following.

THEOREM 1. *Under the conditions given in the beginning of this section and if* $p > 1$ *and*

$$f(x) = O(1/|x|^{\beta}),\ x \to \pm\infty$$

with $\beta > 1$, *then the limiting distribution function* $\psi(t)$ *of the estimate* M_p^* *can be computed as* $P(U_t \leqq 0)$ *where the stochastic variable* U_t *has the characteristic function given by*

$$\log E \exp izU_t = \frac{z^{1/p}}{p} [\bar{\gamma} K(t) + \gamma L(t)],\ z \geqq 0$$
$$= \frac{|z|^{1/p}}{p} [\gamma K(t) + \bar{\gamma} L(t)],\ z \leqq 0.$$

The functions $K(t)$ and $L(t)$ are defined as

$$K(t) = \int_0^{\infty} f^2(t - u) u^{1/p}\, du, \qquad L(t) = \int_0^{\infty} f^2(t + u) u^{1/p}\, du.$$

But this means that M^* is *not even consistent*, which is disappointing to put it mildly. To deal with the case $p < 1$, we norm the sums appearing in S by n^{-1-p}

instead of n^{-2p}. Since the stochastic variables $1/\xi^p$ have existing mean values in the present case,

$$E1/\xi^p = \int_0^\infty e^{-\xi}\xi^{-p}\, d\xi = \Gamma(1-p)$$

we get easily, assuming that $f(x) = O(1/|x|^{1/p})$ for $x \to \pm\infty$, S tends in probability to $\Gamma(1-p) \int_0^1 [G(\alpha) - t]/[g^p(\alpha)]\, d\alpha = \Gamma(1-p) \int_{-\infty}^{\infty} (x-t) f^{p+1}(x)\, dx$. Hence

$$\psi(t) = \lim_{n\to\infty} P(M_p^* \leqq t) = 1 \text{ if } \int_{-\infty}^{\infty} (x-t) f^{p+1}(x)\, dx < 0$$
$$= 0 \text{ otherwise}$$

which proves

THEOREM 2. *If $0 < p < 1$ and if $f(x) = O(1/|x|^{1/p})$ for $x \to \pm\infty$, the estimates M_p^* converge in probability to the value*

$$M_{p+1} = \int_{-\infty}^{\infty} x f^{p+1}(x)\, dx \Big/ \int_{-\infty}^{\infty} f^{p+1}(x)\, dx.$$

In the last case the estimate *is* consistent. However, the limiting value M_{p+1} is in general close to M only if p is large. Therefore M_p^* can not be used in its original form, but we shall see in the next section that it can be modified to yield a useful estimate.

3. A modified direct estimate. It is obvious from the proof that the lack of consistency stems from the fact that the stochastic variables $\zeta_\nu = 1/\xi_\nu^p$ do not have finite mathematical expectation. To remedy this we shall construct our estimate not from the first differences $x_{\nu+1} - x_\nu$ but from the kth differences $x_{\nu+k} - x_\nu$, where the integer k will be chosen later on to guarantee consistency. Introduce $M_{p,k}^* = B_k/A_k$ where

$$B_k = \tfrac{1}{2}\sum_{\nu=1}^{n-k} (x_{\nu+k} + x_\nu)/(x_{\nu+k} - x_\nu)^p, \qquad A_k = \sum_{\nu=1}^{n-k} 1/(x_{\nu+k} - x_\nu)^p.$$

To compute the limit $\psi(t)$ of

$$\psi_n(t) = P(M_{p,k}^* \leqq t) = P(B_k - A_k t \leqq 0)$$

we consider the new sum

$$S = \sum_{\nu=1}^{n-k} n^{-p-1} \frac{\frac{1}{2}(x_{\nu+k} + x_\nu) - t}{[x_{\nu+k} - x_\nu]^p} = \sum_{\nu=1}^{n-k} C_\nu\, .$$

But the stochastic variables $\xi = \xi_{\nu+k} + \xi_{\nu+k-1} + \cdots + \xi_{\nu+1}$ have the frequency function $1/(k-1)!e^{-\xi}\xi^{k-1}$ so that

$$E\frac{1}{\xi^p} = \frac{1}{(k-1)!}\int_0^\infty e^{-\xi}\xi^{k-1-p}\, d\xi = \frac{\Gamma(k-p)}{(k-1)!}\, .$$

We shall assume that $p < k$ in the following. Now we can proceed in the same way as in the last section with one modification. The terms in a sum of the form

$$\Sigma = n^{-1} \sum_{[\alpha n]}^{[\beta n]} 1/[\xi_{\nu+k} + \cdots + \xi_{\nu+1}]^p$$

are no longer independent. Let us split the summation interval $([\alpha n], [\beta n])$ into consecutive blocks $A_1 \cup B_1 \cup A_2 \cup B_2 \cup \cdots \cup A_s$ where all B_i contain k points and the A_i contain K points where K is a number chosen much larger than k and kept fixed afterwards. The number s should satisfy $sK + (s - 1)k = [\beta n] - [\alpha n] + 1$. We then have

$$\Sigma = n^{-1}\sum_{i=1}^{s} S_{A_i} + n^{-1}\sum_{i=1}^{s-1} S_{B_i},$$

where S_{A_i}, S_{B_i} denote partial sums over the summation intervals A_i and B_i respectively. But the S_{A_i} are independent and identically distributed with mean value $K[\Gamma(k - p)/(k - 1)!]$ so that

$$\frac{1}{n}\sum_{i=1}^{s} S_{A_i} = \frac{s}{n}\cdot\frac{1}{s}\sum_{i=1}^{s} S_{A_i} \to \frac{\Gamma(k-p)}{(k-1)!}\,\frac{K}{K+k}(\beta - \alpha)$$

in probability as $n \to \infty$. On the other hand the non negative terms S_{B_i} have expected value $k\,[\Gamma(k - p)/(k - 1)!]$ so that

$$En^{-1}\sum_{i=1}^{s-1} S_{B_i} = [k(s-1)\Gamma(k-p)]/[n(k-1)!]$$

which is small if K is large enough. Hence Σ tends in probability to $(\beta - \alpha)[\Gamma(k - p)/(k - 1)!]$. One can now proceed as in the last section and get the following result.

THEOREM 3. *Consider the modified estimate*

$$M^*_{p,k} = \frac{\frac{1}{2}\sum_{\nu=1}^{n-k} (x_{\nu+k} + x_\nu)/(x_{\nu+k} - x_\nu)^p}{\sum_{\nu=1}^{n-k} 1/(x_{\nu+k} - x_\nu)^p}$$

with $1 < p < k$. *Under the given conditions, and if* $f(x) = O(1/|x|^\beta)$, $\beta > 1$, *for* $x \to \pm\infty$, *it follows that* $M^*_{p,k}$ *is a consistent estimate of the parameter*

$$M_{p+1} = \int_{-\infty}^{\infty} xf^{p+1}(x)\,dx \Big/ \int_{-\infty}^{\infty} f^{p+1}(x)\,dx.$$

It is conjectured that if $k > 2p$ the estimate is asymptotically normally distributed since the sums involved will be close to sums of k-dependent stochastic variables with finite variance. It would be useful to have asymptotic expressions for the bias and standard deviation of $M^*_{p,k}$.

REFERENCES

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

[2] ROSENBLATT, M. (1956). Remarks on some non-parametric estimates of a density function. *Ann. Math. Statist.* **27** 832–837.

[3] WATSON, G. S. and LEADBETTER, M. R. (1963). On the estimation of the probability density, I. *Ann. Math. Statist.* **34** 480–491.

[4] WHITTLE, P. (1958). On the smoothing of probability density functions. *J. Roy. Statist. Soc., Ser. B.* **20** 334–343.