TECHNICAL REPORT NO. 574

July 1979

TWO-STAGE ESTIMATION OF THE MODE

K.N. Andriano

University of Wisconsin-Madison

Two-Stage Estimation of the Mode

K.N. Andriano

Statistics Department, University of Wisconsin-Madison
Madison, Wisconsin, 53706 U.S.A.

Summary

Mode estimation for univariate continuous distributions has received considerable attention in recent years. Most techniques estimate the mode from a single random sample of size n. It is the purpose of this paper to develop a two-stage estimator of the mode and show that the resultant estimator is strongly consistent. Some discussion is also given as to when the estimator might be an improvement over estimators based on single random samples.

Key words and phrases: mode, two-stage estimation, strong consistency

1. Introduction

Beginning with the classical work of Parzen (1962) the interest in univariate mode estimation for continuous distributions has continually grown and has attracted the attention of such people as Chernoff (1964), Dalenius (1965), Grenander (1965), Venter (1967), Wegman (1971), Robertson and Cryer (1974), Sager (1975), Eddy (1976) and Andriano (1978) to name a few. Some work on mode estimation for multivariate distributions has also been done; however, this paper is concerned exclusively with mode estimation for univariate distributions. The current interest in mode estimation has resulted not only from its intellectual appeal but also from its practical appeal. Dalenius (1965), Wegman (1971) and others have given practical examples where the mode would be a useful statistical parameter.

Many of the estimators developed so far are based on a single random sample of size n from an unknown distribution. In the following a two-stage estimator of the mode is developed which is shown to be strongly consistent. It is assumed that the distribution function, $F(y)$, is absolutely continuous and that its density, $f(y)$, satisfies the following regularity conditions:

(1.1)
- (a) $f(y)$ is continuous on $(-\infty,\infty)$ and positive on some interval (a,b) where $[a,b]$ is the closure of the support of $f(y)$.
- (b) $f(y)$ has a unique maximum, the mode M, such that $a < M < b$.
- (c) $f(y)$ is strictly monotone (on its support) for $y < M$ and $y > M$.

Many distributions encountered in practice satisfy (1.1).

2. A two-stage estimator.

2.1 Preliminaries. For $y_o \in [a,M)$ and $f(y_o) > 0$ define $V(y_o)$ by

$$V(y_o) = \inf\{y \geq M\colon f(y) < f(y_o)\};$$

if $f(y_o) = 0$, let $V(y_o) = b$. For each number q, $0 < q \leq 1$, it can easily be shown that there exists a $y \in [a,M)$ such that $\int_y^{V(y)} f(t)dt = q$. The interval $[y,V(y)]$ will be referred to as the interval $[L(q),R(q)]$ in the following.

It can also be shown that $[L(q),R(q)]$ is the unique shortest set containing probability q in the sense that if B is any other Borel set such that $\int_B f(t)dt \geq q$ and if $\int_{B\Delta[L(q),R(q)]} f(t)dt > 0$, then the Lebesque measure of B is strictly greater than the Lebesque of $[L(q),R(q)]$. It is also easy to see that $M \in (L(q),R(q))$.

2.2. The estimator. Let $y_1,y_2,\ldots,y_m$ be a random sample of size m from F(y) and let k(m) be an integer-valued function such that $k(m) < m$ and $k(m)/m \to q$ as m tends to infinity. If $[L(m),R(m)]$ is defined to be the smallest interval containing k(m) observations, then Robertson and Cryer (1974) have shown that $L(m) \to L(q)$ and $R(m) \to R(q)$ with probability one (wp 1). Let the y's contained in $[L(m),R(m)]$ be denoted by $z_1,z_2,\ldots,z_{k(m)}$.

Suppose now m_1 additional random observations $y_{m+1},y_{m+2},\ldots,y_{m+m_1}$ are sampled from F(y). Let $z_{k(m)+1},\ldots,z_{k(m)+s}$ $(s \leq m_1)$ represent those y_i's $(m+1 \leq i \leq m+m_1)$ that are contained in $[L(m),R(m)]$ and let $x_1,x_2,\ldots,x_t$ represent the order statistics of $z_1,z_2,\ldots,z_t$ where $t = k(m) + s$.

Now, for $n = m + m_1$, define the discrete random variable K(n) as follows: If $t \geq r(n) + 1$ define $x_{K(n)+r(n)} - x_{K(n)} = \min\{x_{j+r(n)} - x_j;\ j = 1,2,\ldots,t-r(n)\}$, where r(n) is a nonrandom sequence of integers to be specified further; if $t < r(n) + 1$, let $K(n) = 1$.

For any value of n, the two-stage estimator, $M(n,\alpha)$, is defined by

$$M(n,\alpha) = \alpha x_{K(n)+r(n)} + (1-\alpha)x_{K(n)}$$

where α is a "free parameter" subject to the restriction that $0 \leq \alpha \leq 1$. The appropriate value(s) of α to use will obviously depend on the characteristics of F(y), the sample size, etc. The actual mechanics of choosing an α will not be pursued in this paper.

4. Consistency.

Theorem 4.1. If the sequence of integers r(n) is chosen such that

(4.1) $$\lim_{n\to\infty} r(n)/n = 0$$

and

(4.2) $$\sum_{n=1}^{\infty} n\lambda^{r(n)} < \infty, \text{ for all } \lambda,\ 0 < \lambda < 1,$$

then $M(n,\alpha) \to M$ wp 1.

Proof. Choose q,m_1 and α and then keep them fixed. Also, in the following, all limits are taken as n tends to infinity.

For each n let J(n) be a discrete random variable defined by the following: If $R(m) \leq M$ let $J(n) = 1$; if $R(m) > M$ and $[M,R(m)]$

contains at least $r(n)+1$ observations, let

$x_{J(n)+r(n)} = \min\{x_{j+r(n)}: j = 1,\ldots,t-r(n);\ M \le x_j \le x_{j+r(n)} \le R(m)\}$.

If $[M,R(m)]$ contains fewer than $r(n)+1$ observations, let $J(n) = 1$. It should be noted that since $R(q) > M$ and (4.1) holds, the strong law of large numbers and the fact that $R(m) \to R(q)$ wp 1 guarantees that $[M,R(m)]$ will eventually contain $r(n)+1$ observations wp 1.

Consider the following events:

$$S_0 = [\lim x_{J(n)+r(n)}=M,\ \lim x_{J(n)}=M],$$

$$S_1 = [\lim\{F(x_{J(n)+r(n)})-F(x_{J(n)})\}nr(n)^{-1}=1],$$

$$S_2 = [\lim\{F(x_{K(n)+r(n)})-F(x_{K(n)})\}nr(n)^{-1}=1], \tag{4.3}$$

$$S_3 = [\lim\{F(x_{J(n)+r(n)})-F(x_{J(n)})\}/(x_{J(n)+r(n)}-x_{J(n)})=f(M)],$$

$$S_4 = [\underline{\lim}\{F(x_{K(n)+r(n)})-F(x_{K(n)})\}/(x_{K(n)+r(n)}-x_{K(n)})=f(M)].$$

The method of proof will be to show $P(S_0) = P(S_1) = P(S_2) = P(S_3) = 1$ and that $S_1 \cap S_2 \cap S_3 \subset S_4$ which will imply $x_{K(n)}, x_{K(n)+r(n)} \to 1$ wp 1 and hence the result.

Since (1.1) holds, every interval of the form $[M,M+\delta]$, $\delta > 0$, has positive probability. Hence, since (4.1) holds and $R(m) \to R(q)$ wp 1, the strong law of large numbers guarantees that $x_{J(n)}$ and $x_{J(n)+r(n)}$ converge to M wp 1. But this means $P(S_0) = 1$.

To show that $P(S_1) = P(S_2) = P(S_3) = 1$, the following lemma of Sager (1975) is needed.

Lemma 4.1. Let $A_1,A_2,\ldots;\ B_1,B_2,\ldots$ be sequences of random variables such that $A_n < B_n$ for each n and $[A_n,B_n]$ contains $r(n)+1$ observations. Then if (4.2) holds, $\{F(B_n)-F(A_n)\}\{F_n(B_n)-F_n(A_n)\}^{-1}$ converges to one wp 1 where $F_n(\cdot)$ denotes the empirical distribution function.

From Lemma 4.1 it is obvious that $P(S_1)=P(S_2)=1$. To see that $P(S_3) = 1$, let $U_n = (x_{J(n)+r(n)}-M)/[F(x_{J(n)+r(n)})-F(M)]$ and $V_n = (x_{J(n)}-M)/[F(x_{J(n)})-F(M)]$ and write

$$\frac{x_{J(n)+r(n)}-x_{J(n)}}{F(x_{J(n)+r(n)})-F(x_{J(n)})} = \frac{U_n[F(x_{J(n)+r(n)})-F(M)]nr(n)^{-1}}{[F(x_{J(n)+r(n)})-F(x_{J(n)})]nr(n)^{-1}} + V_n - \frac{V_n[F(x_{J(n)+r(n)})-F(M)]nr(n)^{-1}}{[F(x_{J(n)+r(n)})-F(x_{J(n)})]nr(n)^{-1}}. \tag{4.4}$$

Since $P(S_0) = 1$ and in view of (1.1), it is evident that U_n and $V_n \to [f(M)]^{-1}$ wp 1. Thus, set $A_n = M$ and $B_n = x_{J(n)+r(n)}$ in Lemma 4.1 and using $P(S_1) = 1$, it is the case that (4.4) converges to $[f(M)]^{-1}$ wp 1. Thus, $P(S_3) = 1$.

To show the containment relationship let

$$a(n) = F(x_{J(n)+r(n)})-F(x_{J(n)}),$$
$$b(n) = x_{J(n)+r(n)}-x_{J(n)},$$
$$c(n) = F(x_{K(n)+r(n)})-F(x_{K(n)}),$$

and

$$d(n) = x_{K(n)+r(n)}-x_{K(n)}.$$

Let $w \in S_1 \cap S_2 \cap S_3$. Using (4.3) and the relationship $0 < x_{K(n)+r(n)}-x_{K(n)} \le x_{J(n)+r(n)}-x_{J(n)}$ (the first inequality is valid wp 1 since $F(y)$ is absolutely continuous; the second inequaltiy is a direct consequence of the definition of $K(n)$) it is the case that

$$\lim(a(n)/b(n))/\underline{\lim}(c(n)/d(n))$$

$$= \underline{\lim}(a(n)/b(n))\overline{\lim}(d(n)/c(n))$$

$$\leq \overline{\lim}[a(n)d(n)/(b(n)c(n))]$$

$$\leq \overline{\lim}(d(n)/b(n))\overline{\lim}\left(\frac{a(n)r(n)^{-1}n}{c(n)r(n)^{-1}n}\right)$$

$$\leq \frac{\overline{\lim}(d(n)/b(n))\overline{\lim}(a(n)r(n)^{-1}n)}{\underline{\lim}(c(n)r(n)^{-1}n)}$$

$$\leq 1$$

Hence,

$$\lim(a(n)/b(n)) = f(M) \leq \underline{\lim}(c(n)/d(n))$$

$$\leq \overline{\lim}(c(n)/d(n))$$

$$= \overline{\lim}\{(\textstyle\int_{[x_{K(n)},x_{K(n)+r(n)}]} f(t)dt)/(x_{K(n)+r(n)}-x_{K(n)})\}$$

$$\leq \overline{\lim}\{(\textstyle\int_{[x_{K(n)},x_{K(n)+r(n)}]} f(M)dt)/(x_{K(n)+r(n)}-x_{K(n)})\}$$

$$= f(M).$$

Thus, $S_1 \cap S_2 \cap S_3 \subset S_4$ which implies $P(S_4) = 1$ where S_4 can now be redefined as

$$S_4 = \{\lim(c(n)/d(n)) = f(M)\}.$$

It remains to show that $x_{K(n)}, x_{K(n)+r(n)} \to M$ wp 1. Let $w\varepsilon S_0 \cap S_4$ where it is easily seen that $P(S_0 \cap S_4) = 1$. Hence,

$$\frac{F(x_{K(n)+r(n)})-F(x_{K(n)})}{x_{K(n)+r(n)}-x_{K(n)}} \to f(M).$$

By the Mean Value Theorem, there exists a $\theta(n)$, $x_{K(n)} < \theta(n) < x_{K(n)+r(n)}$, such that

$$\frac{F(x_{K(n)+r(n)})-F(x_{K(n)})}{x_{K(n)+r(n)}-x_{K(n)}} = f(\theta(n)) .$$

Thus, $f(\theta(n)) \to f(M)$ which implies $\theta(n) \to M$ since $f(M)$ is unique by (1.1). Since $w\varepsilon S_0$ is is the case that $x_{K(n)+r(n)}-x_{K(n)}$ converges to zero which implies $x_{K(n)}$ and $x_{K(n)+r(n)}$ converge to M. The result of the theorem follows since $M(n,\alpha)$ is a convex combination of $x_{K(n)}$ and $x_{K(n)+r(n)}$.

5. Concluding remarks.

The two-stage estimator $M(n,\alpha)$ can be thought of as a sequential estimator of the mode with a nonrandomized stopping rule (by a non-randomized stopping rule it is meant that the estimation scheme is always terminated after a specified number of independent samples of equal, or unequal, sizes have been taken). The estimators of Grenander (1965), Venter (1967) and others referred to earlier can technically be considered sequential estimators with a nonrandomized stopping rule which says stop after one independent sample of size n is taken. Sequential estimators with nonrandomized stopping rules are frequently called multi-stage estimators (in this definition of multi-stage estimators it is assumed that at least two independent samples of equal, or unequal, sizes are taken).

Multi-stage estimators (as well as sequential estimators with randomized stopping rules) are generally used when the cost of sampling is high and/or the estimates are to have good precision. In addition to these cost and precision considerations, Dalenius (1965) conjectures that multi-stage estimators of the mode would be useful when the density, f(y), does not exhibit a sharp peak at its mode. Distributions such as the chi-square would be of this type.

REFERENCES

[1] Andriano, K.N. (1978). Aspects of univariate mode estimation, Ph.D. Thesis. Iowa State University

[2] Chernoff, H. (1964). Estimation of the mode. Ann. Inst. Statist. Math., 16, 31-41.

[3] Dalenius, T. (1965). The mode-a neglected statistical parameter. J.R. Statist. Soc. A, 128, 110-117.

[4] Eddy, W.F. (1976). Optimal kernel estimators of the mode, Ph.D. Thesis. Yale University.

[5] Grenander, U. (1965). Some direct estimates of the mode. Ann. Math. Statist., 36, 131-138.

[6] Parzen, E. (1962). On estimation of a probability density function and mode. Ann. Math. Statist., 33, 1065-1076.

[7] Robertson, T. and Cryer, J. (1974). An iterative procedure for estimating the mode. J. Amer. Statist. Assoc., 69, 1012-1016.

[8] Sager, T.W. (1975). Consistency in nonparametric estimation of the mode. Ann. Statist., 3, 698-706.

[9] Venter, J.H. (1967). On estimation of the mode. Ann. Math. Statist., 38, 1446-1455.

[10] Wegman, E.J. (1971). A note on the estimation of the mode. Ann. Math. Statist., 43, 1909-1915.