The Annals of Mathematical Statistics
1969, Vol. 40, No. 4, 1499–1502

UNIFORM CONSISTENCY OF SOME ESTIMATES OF A DENSITY FUNCTION

By D. S. Moore and E. G. Henrichon

Purdue University

1. Introduction and summary. Let $X_1, \cdots, X_n$ be independent random variables identically distributed with absolutely continuous distribution function F and density function f. Loftsgaarden and Quesenberry [3] propose a consistent nonparametric point estimator $\hat{f}_n(z)$ of $f(z)$ which is quite easy to compute in practice. In this note we introduce a step-function approximation f_n^* to $\hat{f}_n$, and show that both $\hat{f}_n$ and f_n^* converge *uniformly* (in probability) to f, assuming that f is positive and uniformly continuous in $(-\infty, \infty)$. For more general f, uniform convergence over any compact interval where f is positive and continuous follows.

Uniform convergence is useful for estimation of the mode of f, for it follows from our theorem (see [4], section 3) that a mode of either $\hat{f}_n$ or f_n^* is a consistent estimator of the mode of f. The mode of f_n^* is particularly tractable; it is applied in [2] to some problems in pattern recognition. From the point of view of mode estimation, we thus obtain two new estimates which are similar in conception to those proposed by some previous authors. Let $k(n)$ be an appropriate sequence of numbers in each case. Chernoff [1] estimates the mode as the center of the interval of length $2k(n)$ containing the most observations. Venter [5] estimates the mode as the center (or endpoint) of the shortest interval containing $k(n)$ observations. The estimate based on $\hat{f}_n$ is that z such that the distance from z to the $k(n)$th closest observation is least. Finally, the estimate from f_n^* is that *observation* such that the distance from it to the $k(n)$th closest observation is least.

2. The result. Choose a non-decreasing sequence of positive integers, $\{k(n)\}$, such that $k(n) \to \infty$ but $k(n) = o(n)$. For any real number z, let $r_{k(n)}(z)$ be the distance from z to the $k(n)$th closest of the observations $X_1, \cdots, X_n$. Then the univariate form of the Loftsgaarden-Quesenberry estimator is

$$f_n(z) = \{(k(n) - 1)/n\}\{1/2r_{k(n)}(z)\}.$$

We define also the random step-function f_n^* as follows: let $X_{1n} \leqq X_{2n} \leqq \cdots \leqq X_{nn}$ be the order statistics from $X_1, \cdots, X_n$. Then

$$f_n^*(z) = 0, \quad \text{if } z < X_{1n} \quad \text{or } z \geqq X_{nn};$$
$$= \hat{f}_n(X_{in}), \quad \text{if } X_{in} \leqq z < X_{i+1,n}; \qquad i = 1, \cdots, n-1.$$

Theorem. *If $f(z)$ is uniformly continuous and positive on $(-\infty, \infty)$ and $(\log n)/k(n) \to 0$, then for every $\epsilon > 0$*

Received 8 August 1968; revised 7 March 1969.

$$P[\sup_{-\infty<z<\infty} |\hat{f}_n(z) - f(z)| > \epsilon] \to 0 \tag{2.1}$$

and

$$P[\sup_{-\infty<z<\infty} |f_n^*(z) - f(z)| > \epsilon] \to 0. \tag{2.2}$$

PROOF. We will abbreviate (2.1) by $\hat{f}_n \to f$ (UP) and denote convergence in probability by $a_n \to a$ (P). Define

$$U_{k(n)}(z) = F(z + r_{k(n)}(z)) - F(z - r_{k(n)}(z)).$$

We show first that

$$\{n/(k(n) - 1)\} U_{k(n)}(z) \to 1 \text{ (UP)}. \tag{2.3}$$

By definition of $r_{k(n)}(z)$, the interval $[z - r_{k(n)}(z), z + r_{k(n)}(z)]$ contains exactly $k(n)$ observations, one of which falls at an endpoint of the interval. Suppose the order statistic X_{qn} is the lower endpoint. Then

$$\sum_{j=1}^{k(n)-1} \{F(X_{q+j,n}) - F(X_{q+j-1,n})\} \leqq U_{k(n)}(z) \leqq \sum_{j=1}^{k(n)} \{F(X_{q+j,n}) - F(X_{q+j-1,n})\} \tag{2.4}$$

with the conventions $F(X_{0,n}) = 0$ and $F(X_{n+1,n}) = 1$. Upper and lower bounds having the same distribution as those in (2.4) exist when X_{qn} is an upper endpoint. (It is stated in [3] that $U_{k(n)}$ has the beta distribution of one of the sums of elementary coverages in (2.4). This is false, since with probability one only one endpoint of the interval coincides with an observation; the modifications required to correct the proof of [3] are trivial.)

It is well known that

$$F(X_{1n}),\ F(X_{2n}) - F(X_{1n}),\ \cdots,\ 1 - F(X_{nn})$$

have the same joint distribution as

$$Y_1/S_{n+1},\ \cdots,\ Y_{n+1}/S_{n+1},$$

where $Y_1, \cdots, Y_{n+1}$ are independent exponential random variables with mean 1 and $S_{n+1} = Y_1 + \cdots + Y_{n+1}$. So the upper and lower bounds for $\{n/(k(n) - 1)\} U_{k(n)}$ obtained from (2.4) will converge to 1 (UP) if we can prove that

$$\max_{0 \leqq i \leqq n-k(n)+1} |\{k(n)\}^{-1} \sum_{j=i+1}^{i+k(n)} Y_j/(n^{-1}S_{n+1})\} - 1| \to 0 \text{ (P)}. \tag{2.5}$$

Since $n^{-1}S_{n+1} \to 1$ with probability one by the law of large numbers, (2.5) will follow if we can show that the sums $\{k(n)\}^{-1} \sum_{i+1}^{i+k(n)} Y_j$ are uniformly near 1 in probability. For any $\epsilon > 0$,

$$\begin{aligned} P_n &= P[\text{for some } i, |\textstyle\sum_{j=i+1}^{i+k(n)} (Y_j - 1)| > k(n)\epsilon] \\ &\leqq \textstyle\sum_{i=1}^{n} P[\sum_{j=i+1}^{i+k(n)} (Y_j - 1) > k(n)\epsilon] \\ &\quad + \textstyle\sum_{i=1}^{n} P[\sum_{j=i+1}^{i+k(n)} (Y_j - 1) < -k(n)\epsilon]. \end{aligned} \tag{2.6}$$

Using the fact that $P[X > 0] \leqq E[e^{tX}]$ for any random variable X and $t > 0$

such that the right side is finite, we obtain

$$P[\textstyle\sum_{j=i+1}^{i+k(n)} (Y_j - 1) > k(n)\epsilon] \leqq E[\exp\{t(\sum Y_j - k(n) - k(n)\epsilon)\}]$$
$$= \{e^{-t(1+\epsilon)}/(1-t)\}^{k(n)}, \qquad 0 < t < 1.$$

(Recall that a sum of $k(n)Y_j$'s has the gamma distribution with parameter $k(n)$.) Choosing the minimizing value $t = 1 - (1+\epsilon)^{-1}$ gives the bound $\{(1+\epsilon)e^{-\epsilon}\}^{k(n)}$. A similar bound holds for each term of the second sum on the right side of (2.6). Therefore $P_n \leqq (n+1)a(\epsilon)^{-k(n)}$, where $a(\epsilon) > 1$ for $\epsilon > 0$. Since $(\log n)/k(n) \to 0$, $P_n \to 0$ and (2.5) is proved.

It follows from (2.3) that $U_{k(n)} \to 0$ (UP) and hence, since f is everywhere positive, that $r_{k(n)} \to 0$ (UP).

To conclude (2.1) we need only (2.3) and the fact that $U_{k(n)}/2r_{k(n)} \to f$ (UP). Since f is uniformly continuous and $r_{k(n)} \to 0$ (UP), this is immediate from the estimate

$$|[U_{k(n)}(z)]/[2r_{k(n)}(z)] - f(z)| = |\,[2r_{k(n)}(z)]^{-1}\textstyle\int_{z-r}^{z+r}[f(t) - f(z)]\,dt|$$
$$\leqq \max\{|f(t) - f(z)| : z - r_{k(n)}(z) \leqq t \leqq z + r_{k(n)}(z)\}. \tag{2.7}$$

The argument for (2.2) is slightly longer. Let $i(z)$ be the index such that

$$X_{i(z),n} \leqq z < X_{i(z)+1,n}$$

For any compact interval I, the probability that X_{1n} and X_{nn} fall outside I approaches 1 as $n \to \infty$, by positivity of f. Thus $i(z)$ is defined for all $z \epsilon I$ with probability approaching 1 for large n. The Glivenko-Cantelli theorem and uniform continuity of F^{-1} on $[\alpha, 1-\alpha]$ for any $\alpha > 0$ give that

$$\sup_{z\epsilon I} |X_{i(z),n} - z| \to 0 \text{ (P)}. \tag{2.8}$$

From (2.8) and the fact that $r_{k(n)} \to 0$ (UP), we can conclude by an estimate analogous to (2.7) that

$$\sup_{z\epsilon I} |[U_{k(n)}(X_{i(z),n})]/[2r_{k(n)}(X_{i(z),n})] - f(z)| \to 0 \text{ (P)}$$

and hence, using (2.3), that for any compact interval I and any $\epsilon > 0$,

$$\lim_{n\to\infty} P[\sup_{z\epsilon I} |f_n^*(z) - f(z)| > \epsilon] = 0. \tag{2.9}$$

If we can establish that for any $\epsilon > 0$ there is a compact interval I_ϵ such that

$$\lim_{n\to\infty} P[\sup_{z \notin I_\epsilon} |f_n^*(z) - f(z)| > \epsilon] = 0, \tag{2.10}$$

this with (2.9) will imply (2.2).

Since $f(z) \to 0$ as $z \to \pm\infty$, we can choose a compact interval $I^* = [a, b]$ such that $f(z) < \epsilon/2$ outside I^*. Then by (2.1), $f_n(z) < \epsilon$ for all $z \notin I^*$ with probability approaching 1 as $n \to \infty$. Let $I_\epsilon = [a, b + c]$ for some $c > 0$. Then

by (2.8) and the fact that $P[X_{1n} < a, X_{nn} > b + c] \to 1$, we have that

$$P[X_{i(z),n} \notin I^* \text{ for all } z \notin I_\epsilon \text{ with } X_{1n} \leqq z < X_{nn}] \to 1.$$

Thus with probability approaching 1, $f_n^*(z)$ is either 0 or $\hat{f}_n(X_{in})$ for some $X_{in} \notin I^*$, for all $z \notin I_\epsilon$. This establishes (2.10).

3. Acknowledgment. We thank the referee for the remarks in the Introduction concerning the relationship of several estimates of the mode.

REFERENCES

[1] CHERNOFF, H. (1964). Estimation of the mode. *Ann. Inst. Statist. Math.* **16** 31–41.
[2] HENRICHON, E. G. and FU, K. S. (1968). On mode estimation for pattern recognition. Submitted to *IEEE Trans. Information Theory*.
[3] LOFTSGAARDEN, D. O. and QUESENBERRY, C. P. (1965). A nonparametric estimate of a multivariate density function. *Ann. Math. Statist.* **36** 1049–1051.
[4] PARZEN, EMANUEL (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.* **33** 1065–1076.
[5] VENTER, J. H. (1967). On estimation of the mode. *Ann. Math. Statist.* **38** 1446–1455.