ESTIMATION OF PROBABILITY DENSITY

By V. K. Murthy

Douglas Aircraft Company

0. Summary. Assuming that the distribution being sampled is absolutely continous, Parzen [3] has established the consistency and asymptotic normality of a class of estimators $\{f_n(x)\}$ based on a random sample of size n, for estimating the probability density. In this paper, we relax the assumption of absolute continuity of the distribution $F(x)$ and show that the class of estimators $\{f_n(x)\}$ still consistently estimate the density at all points of continuity of the distribution $F(x)$ where the density $f(x)$ is also continuous. It is further shown that the sequence of estimators $\{f_n(x)\}$ are asymptotically normally distributed. The extension of these results to the bi-variate and essentially the multi-variate case with applications and a discussion on the construction of higher dimensional windows will be presented at the International Symposium in Multivariate Analysis to be held in Dayton, Ohio during June 1965.

1. A class of estimators for the density at a point of continuity of $F(x)$ and the density $f(x)$. Let $F(x)$ be a probability distribution function. Assuming that the singular part is identically zero, $F(x)$ can be decomposed into (see e.g. Cramér [1] pp. 52, 53)

$$F(x) = F_1(x) + F_2(x) \tag{1.1}$$

where $F_1(x)$ is an everywhere continuous function and $F_2(x)$ is a pure step function with steps of magnitude, say, S_ν at the points $x = x_\nu$, $\nu = 1, 2, \cdots$ and finally both $F_1(x)$ and $F_2(x)$ are non-decreasing and are uniquely determined. If the singular part is not identically zero as has been assumed here, the results are only valid almost everywhere.

Let

$$dF_1(x) = f(x)\,dx. \tag{1.2}$$

At a point of continuity x_0 of $F(x)$ its density is clearly $f(x_0)$. Let $X_1, X_2, \cdots X_n$ be a random sample of size n from the distribution $F(x)$, i.e., $X_1, X_2, \cdots X_n$ are independently identically distributed random variables with the distribution $F(x)$.

Let

$$F_n(x) = 1/n \quad [\text{number of observations} \leqq x \text{ among } X_1, X_2, \cdots X_n]. \tag{1.3}$$

Clearly $F_n(x)$ is a binomially distributed random variable with

$$E[F_n(x)] = F(x), \quad \text{and} \quad \operatorname{Var}[F_n(x)] = (1/n)F(x)[1 - F(x)]. \tag{1.4}$$

Received 14 November 1963; revised 20 December 1964.

A function $K(\omega)$ is called a window if it satisfies the following condition

$$(1.5) \qquad K(\omega) \geqq 0, \qquad K(\omega) = K(-\omega),$$
$$\lim_{|\omega| \to \infty} \omega K(\omega) = 0, \quad \text{and} \quad \int_{-\infty}^{\infty} K(\omega)\, d\omega = 1.$$

Following Parzen [3] and Murthy [2] let us propose

$$(1.6) \qquad f_n(x_0) = \int_{-\infty}^{\infty} B_n K(B_n(x - x_0))\, dF_n(x),$$

as an estimate of the density $f(x_0)$ at a point of continuity x_0 of the distribution $F(x)$ and $\{B_n\}$ is a sequence of non-negative constants depending on the sample size n such that

$$(1.7) \qquad \lim_{n \to \infty} B_n = \infty.$$

2. Asymptotic unbiassedness of $f_n(x)$ at a continuity point of $F(x)$ and $f(x)$. We have from (1.6)

$$(2.1) \qquad f_n(x_0) = \int_{-\infty}^{\infty} B_n K(B_n(x - x_0))\, dF_n(x)$$
$$= (B_n/n) \sum_{j=1}^{n} K(B_n(X_j - x_0)),$$

where x_0 is a point of continuity of the distribution $F(x)$ at which the density $f(x)$ is also continuous. Taking expectation on both sides of (2.1) we obtain

$$(2.2) \qquad E(f_n(x_0)) = B_n \int_{-\infty}^{\infty} K(B_n(x - x_0))\, dF(x).$$

We will now prove that

$$(2.3) \qquad \lim_{n \to \infty} E(f_n(x_0)) = f(x_0).$$

To prove (2.3) we need the following lemma.

Lemma. *Let $K(x)$ be a window satisfying* (1.5). *Let $x_i (i = 0, \pm 1, \pm 2, \cdots)$ be the points of discontinuity of the distribution $F(x)$ and S_i the saltus of $F(x)$ at x_i. Further, let $A_n(x) = B_n K(B_n(x - x_0'))$ where B_n is a sequence of non-negative constants tending to infinity as $n \to \infty$, and x_0' a point of continuity of $F(x)$ and also of $f(x)$ the derivative of the absolutely continuous part of $F(x)$. Then*

$$(2.4) \qquad \lim_{n \to \infty} J(A_n) = \lim_{n \to \infty} \int_{-\infty}^{\infty} A_n(x)\, dF(x) = f(x_0')$$

provided the series $\sum_i S_i/|x_i - x_0'|$ converges.

Proof. We have

$$(2.5) \quad J(A_n) = \int_{-\infty}^{\infty} B_n K(B_n(x - x_0'))\, dF(x)$$
$$= \int_{-\infty}^{\infty} B_n K(B_n(x - x_0')) f(x)\, dx + \sum_i B_n K(B_n(x_i - x_0')) S_i.$$

Now

$$\sum_i B_n K(B_n(x_i - x_0')) S_i = \sum_{|i| \leqq m} B_n K(B_n(x_i - x_0')) S_i$$
$$+ \sum_{|i| > m} B_n K(B_n(x_i - x_0')) S_i = \Sigma_1 + \Sigma_2, \qquad \text{say}.$$

Since x_0' is a point of continuity, $x_i \neq x_0'$ for all i. Since $|xK(x)| \to 0$ as $x \to \pm \infty$, we can choose an $N_0 > 0$ such that

$$|B_n(x_i - x_0')K(B_n(x_i - x_0'))| < \epsilon \qquad \text{for } n > N_0 .$$

Hence

$$|\Sigma_1| < \epsilon \sum\nolimits_{|i| \leqq m} S_i/|x_i - x_0'| \leqq \quad ,$$

where $A = \sum_i S_i/|x_i - x_0'| < \infty$, by assumption. Since $xK(x) \to 0$ as $|x| \to \infty$ it follows that $|xK(x)|$ is bounded. Hence $|xK(x)| \leqq K_0$ (finite) for all x. Therefore

$$|\Sigma_2| \leqq K_0 \sum\nolimits_{|i|>m} S_i/|x_i - x_0'|.$$

Since $\sum_i S_i/|x_i - x_0'|$ converges, we can choose m such that

$$\sum\nolimits_{|i|>m} S_i/(|x_i - x_0'|) < \epsilon.$$

Therefore $\lim_{n\to\infty} \sum_i B_n K(B_n(x_i - x_0'))S_i = 0$. Hence

$$\text{(2.6)} \quad \lim_{n\to\infty} J(A_n) = \lim_{n\to\infty} \int_{-\infty}^{\infty} B_n K(B_n(x - x_0'))f(x)\, dx = f(x_0').$$

It may be noted that if the points of discontinuity of the distribution function are isolated points then the condition $\sum_i S_i/|x_i - x_0'| < \infty$ is automatically satisfied. For, in that case $\inf_i |x_i - x_0'| > 0$ for every point of continuity x_0' and consequently

$$\sum\nolimits_i S_i/|x_i - x_0'| \leqq (1/x'') \sum\nolimits_i S_i \leqq 1/x''$$

where $x'' = \inf_i |x_i - x_0'|$.

It may also be mentioned that if as assumed in the lemma $\int_{-\infty}^{\infty} K(x)\, dx \neq 1$ but is finite i.e. $\int_{-\infty}^{\infty} K(x)\, dx < \infty$, then the limit in (2.6) will be

$$\text{(2.7)} \qquad \lim_{n\to\infty} J(A_n) = f(x_0) \int_{-\infty}^{\infty} K(x)\, dx.$$

Using the lemma it is at once clear that

$$\text{(2.8)} \qquad \lim_{n\to\infty} E(f_n(x_0)) = f(x_0).$$

at a point of continuity x_0 of the distribution $F(x)$ and also of $f(x)$. The proof of asymptotic unbiassedness is now complete.

3. The consistency of $\{f_n(x)\}$ at a point of continuity of $F(x)$ and also of $f(x)$. We will prove the consistency of $f_n(x)$ at a continuity point of $F(x)$ and $f(x)$ by showing that the variance of $f_n(x)$ tends to zero as $n \to \infty$. This together with the property of asymptotic unbiassedness proved earlier will establish the consistency.

Taking variance on both sides of (2.1), we obtain

$$\text{(3.1)} \quad \operatorname{Var}[f_n(x_0)] = (B_n^2/n)[E(K^2(B_n(x - x_0))) - E^2(K(B_n(x - x_0)))].$$

Taking limit as $n \to \infty$ on both sides of (3.1), we have in view of (2.3) that

$$\text{(3.2)} \qquad \lim_{n\to\infty} \mathrm{Var}\,[f_n(x_0)] = \lim_{n\to\infty} (B_n^2/n)E(K^2(B_n(x - x_0)))$$
$$= \lim_{n\to\infty} (B_n^2/n) \int_{-\infty}^{\infty} K^2(B_n(x - x_0))\, dF(x).$$

We now observe that the function $K^2(x)$ has all but only one of the properties of $K(x)$ namely $\int_{-\infty}^{\infty} K^2(x)\, dx \neq 1$, but is finite. The lemma therefore holds for $K^2(x)$ with the limit as given by (2.7).

We have therefore proved that

$$\text{(3.3)} \qquad \lim_{n\to\infty} B_n \int_{-\infty}^{\infty} K^2(B_n(x - x_0))\, dF(x) = f(x_0) \int_{-\infty}^{\infty} K^2(x)\, dx.$$

at a point of continuity x_0 of the distribution $F(x)$ and also of $f(x)$.

Combining (3.2) and (3.3), we discover that

$$\text{(3.4)} \qquad \lim_{n\to\infty} (n/B_n)\, \mathrm{Var}\,[f_n(x_0)] = f(x_0) \int_{-\infty}^{\infty} K^2(x)\, dx,$$

at a point of continuity x_0 of $F(x)$ and $f(x)$. Assuming now that $B_n \to \infty$ more slowly than n in such a way that $(B_n/n) \to 0$ as $n \to \infty$ we obtain

$$\lim_{n\to\infty} \mathrm{Var}\,[f_n(x_0)] = 0.$$

We have thus proved the following

THEOREM 1. *Let $K(x)$ be a window satisfying* (1.5). *Let B_n be a sequence of non-negative constants depending on the sample size n such that $B_n \to \infty$ as $n \to \infty$ in such a way $(B_n/n) \to 0$ as $n \to \infty$. Then the estimator*

$$f_n(x_0) = \int_{-\infty}^{\infty} B_n K(B_n(x - x_0))\, dF_n(x)$$

is a consistent estimate of $f(x_0)$ at a point of continuity x_0 of the distribution $F(x)$ and also of the density $f(x)$.

4. Asymptotic normality of the sequence $\{f_n(x_0)\}$ at a point of continuity of $F(x)$ and $f(x)$. The estimator $f_n(x_0)$ given by (1.6) can be written as

$$\text{(4.1)} \qquad f_n(x_0) = (1/n) \sum_{j=1}^{n} V_j\, .$$

where $V_j = B_n K(B_n(X_j - x_0))$. The sequence $\{V_j\}$ are independently identically distributed as a random variable

$$\text{(4.2)} \qquad V_n = B_n K(B_n(x - x_0)).$$

A sufficient condition for the sequence $\{f_n(x_0)\}$ to be asymptotically normally distributed is (see Parzen [3] p. 1069) that for some $\delta > 0$

$$\text{(4.3)} \qquad E|V_n - E(V_n)|^{2+\delta}/[n^{\delta/2}[\mathrm{Var}\,(V_n)]^{1+\delta/2}] \to 0 \qquad \text{as } n \to \infty.$$

Applying the lemma we obtain

$$\text{(4.4)} \qquad E|V_n|^{2+\delta} \sim B_n^{1+\delta} f(x_0) \int_{-\infty}^{\infty} [K(x)]^{2+\delta}\, dx,$$
$$\mathrm{Var}\,[V_n] \sim B_n f(x_0) \int_{-\infty}^{\infty} K^2(x)\, dx,$$

at every point of continuity x_0 of $F(x)$ and $f(x)$. In view of $\int_{-\infty}^{\infty} K(x)\, dx = 1$ we have

$$\int_{-\infty}^{\infty} (K(x))^{2+\delta}\, dx < \infty \qquad \text{for all } \delta \geqq 0. \tag{4.5}$$

Taking (4.4), (4.5) and the condition $(B_n/n) \to 0$ as $n \to \infty$ it is easily verified that (4.3) is satisfied. Hence

THEOREM 2. *The sequence of estimators* $\{f_n(x_0)\}$ *are asymptotically normal where* x_0 *is a point of continuity of* $F(x)$ *and also the density* $f(x)$.

Acknowledgment. The author is greatly indebted to the referee for substantially improving the basic lemma in the paper.

REFERENCES

[1] CRAMÉR, H. (1946). *Mathematical Methods of Statistics.* Princeton Univ. Press.

[2] MURTHY, V. K. (1963). Estimation of the cross-spectrum. *Ann. Math. Statist.* **34** 1012–1021.

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