SEQUENTIAL ESTIMATION OF A CONTINUOUS PROBABILITY DENSITY FUNCTION AND MODE

Yamato, Hajime
Department of Mathematics, Faculty of Science, Kagoshima University

https://doi.org/10.5109/13049

出版情報：統計数理研究. 14 (3/4), pp.1-12, 1971-03. Research Association of Statistical Sciences

SEQUENTIAL ESTIMATION OF A CONTINUOUS PROBABILITY DENSITY FUNCTION AND MODE

By

Hajime YAMATO*

(Received December 17, 1970)

1. Introduction.

The purpose of this paper is to discuss statistical properties of an estimate of a probability density function based on the first n observations under the assumption of continuity or uniform continuity of the probability density function in case where we observe a sequence of random vectors which come from a population with the probability density function.

Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3, \cdots$ be a sequence of independent identically distributed m-dimensional random vectors having a probability density function $f(\boldsymbol{x})$. Suppose the first n observations be denoted by $\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_n$. Then a natural estimate of the probability density function $f(\boldsymbol{x})$ may be denoted as follows for a suitable positive constant h

(1.1) $$\tilde{f}_n(\boldsymbol{x}) = \frac{1}{n(2h)^m} \{\text{no. of observations such that } x_j - h \leqq X_{ij} \leqq x_j + h$$

$$\text{for } i=1, 2, \cdots, n \text{ and } j=1, 2, \cdots, m \text{ among } \boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_n$$

$$\text{where } \boldsymbol{X}_i = (X_{i1}, X_{i2}, \cdots, X_{im}) \text{ for } i=1, 2, \cdots, n \text{ and}$$

$$\boldsymbol{x} = (x_1, x_2, \cdots, x_m)\}.$$

This may be denoted more adequately, by defining $K(\boldsymbol{y})$ as

(1.2) $$K(\boldsymbol{y}) = \begin{cases} 1/2^m & \text{if } -1 \leqq y_i \leqq 1 \text{ for } i=1, 2, \cdots, m \text{ where } \boldsymbol{y} = (y_1, y_2, \cdots, y_m) \\ 0 & \text{otherwise} \end{cases}$$

and

(1.3) $$\tilde{f}_n(\boldsymbol{x}) = \frac{1}{nh^m} \sum_{j=1}^{n} K\left(\frac{\boldsymbol{x} - \boldsymbol{X}_j}{h}\right)$$

In case of one-dimensional random variable, Parzen [4] considered a generalized estimate $\tilde{f}_n(x) = \frac{1}{nh_n} \sum_{j=1}^{n} K\left(\frac{x - X_j}{h_n}\right)$, in the sense $K(y)$ need not be equal to 1/2 for $|y| \leqq 1$, but satisfies all or some of

(1.4) $$\sup_{-\infty<y<\infty} |K(y)| < \infty$$

(1.5) $$\int_{-\infty}^{\infty} |K(y)|\, dy < \infty$$

* Department of Mathematics, Faculty of Science, Kagoshima University, Kagoshima.

(1.6) $$\lim_{y\to\pm\infty} |yK(y)|=0$$

(1.7) $$\int_{-\infty}^{\infty} K(y)dy=1$$

and for a certain class of probability density function he proved the following statistical properties under a certain condition about a sequence $\{h_n\}$ and an additive condition about the measurable function $K(y)$: (1) the estimate $\tilde{f}_n(x)$ is asymptotically unbiased, consistent, uniformly consistent and asymptotically normal and (2) a sample mode obtained from $\tilde{f}_n(x)$ is a consistent estimate of the population mode and asymptotically normal.

Now we shall consider as an estimate of the probability density function $f(\boldsymbol{x})$

(1.8) $$\hat{f}_n(\boldsymbol{x})=\frac{1}{n}\sum_{j=1}^{n}\frac{1}{h_j^m}K\Big(\frac{\boldsymbol{x}-\boldsymbol{X}_j}{h_j}\Big)$$
$$=\frac{n-1}{n}\hat{f}_{n-1}(\boldsymbol{x})+\frac{1}{nh_n^m}K\Big(\frac{\boldsymbol{x}-\boldsymbol{X}_n}{h_n}\Big).$$

The estimate $\hat{f}_n(\boldsymbol{x})$ is considered more suitable than $\tilde{f}_n(\boldsymbol{x})$ in order to correct the estimate successively in case where a sequence of radom vectors or variables are observed. In the following sections, we shall treat the statistical properties of $\hat{f}_n(\boldsymbol{x})$ under the assumption that the measurable function $K(\boldsymbol{y})$ satisfies all or some of

(1.9) $$\sup_{y\in R^m} |K(\boldsymbol{y})|<\infty$$

(1.10) $$\int_{R^m} |K(\boldsymbol{y})|d\boldsymbol{y}<\infty$$

(1.11) $$\int_{R^m} K(\boldsymbol{y})d\boldsymbol{y}=1$$

(1.12) $$\lim_{\|y\|\to\infty} \|\boldsymbol{y}\|^m|K(\boldsymbol{y})|=0$$

where R^m denotes the m-dimensional Euclidian space and

$$\|\boldsymbol{y}\|=(y_1^2+y_2^2+\cdots+y_m^2)^{1/2} \qquad \text{for } \boldsymbol{y}=(y_1, y_2, \cdots, y_m)\in R^m.$$

Wolverton and Wagner [7] pointed out that an estimate of a probability density functioh could be obtained by modifying their asymptotically optimal discriminant function and their estimate turns out to be identical with ours. Their consideration over the statistical properties from point of view of the estimation theory, however, were only on its asymptotically uniform unbiasedness assuming the uniform continuity of the probability density function.

In section 2 we shall prove the asymptotic unbiasedness of the estimate $\hat{f}_n(\boldsymbol{x})$ for the continuous probability density function and moreover at the continuous points $\boldsymbol{x}$ of the probability density function.

In section 3 we shall treat the limits of the variance and the mean square error of the estimate $\hat{f}_n(\boldsymbol{x})$ and the limit of $nh_n^m \operatorname{Var}[\hat{f}_n(\boldsymbol{x})]$. We shall also treat the limit of the ratio of the variance of the estimate $\hat{f}_n(\boldsymbol{x})$ to the one of the estimate $\tilde{f}_n(\boldsymbol{x})$.

In section 4 we shall treat the uniform consistency of the estimate $\hat{f}_n(\boldsymbol{x})$, which enable us to obtain the consistency of a sample mode obtained by the estimate $\hat{f}_n(\boldsymbol{x})$.

In section 5 we shall treat the limit distribution of the estimate $\hat{f}_n(\boldsymbol{x})$.

In section 6, for the probability density function known in form except for some parameters, we shall compare the estimate $\hat{f}_n(\boldsymbol{x})$ of $f(\boldsymbol{x})$ with the one based on the estimates of parameters.

Unfortunately the author has not been able to treat the limit distribution of a sample mode obtained by the estimate $\hat{f}_n(\boldsymbol{x})$, which the author wishes to develop on another occasion. The author shall treat the statistical properties of the asymptotically optimal discriminant function constructed by Wolverton and Wagner [7] from point of view of the estimation theory in Yamato [8].

The author expresses his hearty thanks to Professor A. Kudo of Kyushu University for his kind encourangement and advice during this work and also to Professor T. Yanagawa of Kumamoto University for his suggestion.

2. Asymptotic unbiasedness.

THEOREM 1. *We suppose that the probability density function $f(\boldsymbol{x})$ is continuous in R^m and $\{h_n\}$ is a sequence of monotone decreasing positive numbers and satisfies a condition*

$$\lim_{n\to\infty} h_n = 0 . \tag{2.1}$$

Let $K(\boldsymbol{y})$ be a measurable function satisfying (1.10) *and* (1.11). *Then the estimate $\hat{f}_n(\boldsymbol{x})$ defined by* (1.8) *is asymptotically unbiased at all points $\boldsymbol{x}$.*

The following corollary can be essentially found in Wolverton and Wagner [7], which is needed in the proof of Theorem 5.

COROLLARY 1. *If we assume the uniform continuity of the probability density function $f(\boldsymbol{x})$ in Theorem 1, then we have*

$$\lim_{n\to\infty} \sup_{\boldsymbol{x}} |E\hat{f}_n(\boldsymbol{x}) - f(\boldsymbol{x})| = 0 . \tag{2.2}$$

COROLLARY 2. *We suppose that $\{h_n\}$ is a sequence of monotone decreasing positive numbers and satisfies* (2.1) *and that the measurable function $K(\boldsymbol{y})$ satisfies* (1.10), (1.11) *and* (1.12). *Then the estimate $\hat{f}_n(\boldsymbol{x})$ is asymptotically unbiased at all points $\boldsymbol{x}$ at which the probability density function $f(\boldsymbol{x})$ is continuous.*

The proofs of Theorem 1, Corollary 1 and Corollary 2 can be easily obtained by noting, at first,

$$E\hat{f}_n(\boldsymbol{x}) = \frac{1}{n} \sum_{j=1}^{n} \int_{R^m} \frac{1}{h_j^m} K\left(\frac{\boldsymbol{y}}{h_j}\right) f(\boldsymbol{x}-\boldsymbol{y}) d\boldsymbol{y} \tag{2.3}$$

and then by applying following Lemma 1 and Lemma 2 on (2.3). Lemma 1 is essentially identical with Theorem 1A in Parzen [4].

LEMMA 1. *We assume that $\{h_n\}$ is a sequence of monotone decreasing positive numbers satisfying* (2.1) *and the measurable function $K(\boldsymbol{y})$ satisfies* (1.10). *Then we have:* (1) *if $f(\boldsymbol{x})$ continuous, then $\int_{R^m} \frac{1}{h_n^m} K\left(\frac{\boldsymbol{y}}{h_n}\right) f(\boldsymbol{x}-\boldsymbol{y}) d\boldsymbol{y}$ converges to $f(\boldsymbol{x}) \int_{R^m} K(\boldsymbol{y}) d\boldsymbol{y}$ at all points $\boldsymbol{x}$ as n tends to ∞.* (2) *if $f(\boldsymbol{x})$ is uniformly continuous, then*

$\int_{R^m} \frac{1}{h_n^m} K\left(\frac{\boldsymbol{y}}{h_n}\right) f(\boldsymbol{x}-\boldsymbol{y}) d\boldsymbol{y}$ *uniformly converges to* $f(\boldsymbol{x}) \int_{R^m} K(\boldsymbol{y}) d\boldsymbol{y}$ *as n tends to* ∞. (3) *if we assume that* $K(\boldsymbol{y})$ *satisfies* (1.12), *then* $\int_{R^m} \frac{1}{h_n^m} K\left(\frac{\boldsymbol{y}}{h_n}\right) f(\boldsymbol{x}-\boldsymbol{y}) d\boldsymbol{y}$ *converges to* $f(\boldsymbol{x}) \int_{R^m} K(\boldsymbol{y}) d\boldsymbol{y}$ *as n tends to* ∞ *at all points* $\boldsymbol{x}$ *at which the probability density function* $f(\boldsymbol{x})$ *is continus.*

LEMMA 2. *If a sequence of functions* $\{g_n(\boldsymbol{x})\}$ *converges to a function* $g(\boldsymbol{x})$ *at a point* $\boldsymbol{x}$ *as n tends to* ∞, *then* $\frac{1}{n} \sum_{j=1}^{n} g_j(\boldsymbol{x})$ *converges to* $g(\boldsymbol{x})$ *at the same point* $\boldsymbol{x}$ *as n tends to* ∞. *If a sequence of functions* $\{g_n(\boldsymbol{x})\}$ *is uniformly bounded on* R^m *and uniformly converges to a bounded function* $g(\boldsymbol{x})$ *on* R^m *as n tends to* ∞, *then* $\frac{1}{n} \sum_{j=1}^{n} g_j(\boldsymbol{x})$ *uniformly converges to* $g(\boldsymbol{x})$ *in* R^m *as n tends to* ∞ (See, for instance, pp. 122 in Tucker [5]).

In the above theorem and corollaries we do not assume that the function $K(\boldsymbol{y})$ satisfies (1.9), but it is considered to be natural that we restrict the estimate $\hat{f}_n(\boldsymbol{x})$ to the one with $K(\boldsymbol{y})$ satisfying (1.9).

3. Consistency.

We can obtain the consistency of the estimate $\hat{f}_n(\boldsymbol{x})$ by the following theorem.

THEOREM 2. *We suppose that the probability density function* $f(\boldsymbol{x})$ *is continuous in* R^m *and a sequence of monotone decreasing positive numbers* $\{h_n\}$ *satisfies* (2.1) *and*

$$\lim_{n\to\infty} n h_n^m = \infty . \tag{3.1}$$

Let the measurable function $K(\boldsymbol{y})$ *satisfy* (1.9) *and* (1.10). *Then we have*

$$\lim_{n\to\infty} Var[\hat{f}_n(\boldsymbol{x})] = 0 \quad \text{at all points } \boldsymbol{x} . \tag{3.2}$$

Furthermore if $K(\boldsymbol{y})$ *satisfies* (1.11), *then we have*

$$\lim_{n\to\infty} E|\hat{f}_n(\boldsymbol{x}) - f(\boldsymbol{x})|^2 = 0 \quad \text{at all points } \boldsymbol{x} . \tag{3.3}$$

PROOF. We shall note, at first, that

$$\begin{aligned} Var[\hat{f}_n(\boldsymbol{x})] &= \frac{1}{n^2} \sum_{j=1}^{n} \frac{1}{h_j^{2m}} E\left\{ K\left(\frac{\boldsymbol{x}-\boldsymbol{X}_j}{h_j}\right) - E\left[K\left(\frac{\boldsymbol{x}-\boldsymbol{X}_j}{h_j}\right)\right]\right\}^2 \\ &= \frac{1}{n^2} \sum_{j=1}^{n} \frac{1}{h_j^{2m}} \int_{R^m} K^2\left(\frac{\boldsymbol{y}}{h_j}\right) f(\boldsymbol{x}-\boldsymbol{y}) d\boldsymbol{y} \\ &\quad - \frac{1}{n^2} \sum_{j=1}^{n} \left\{ \frac{1}{h_j^m} \int_{R^m} K\left(\frac{\boldsymbol{y}}{h_j}\right) f(\boldsymbol{x}-\boldsymbol{y}) d\boldsymbol{y} \right\}^2 . \end{aligned} \tag{3.4}$$

From (1.9) and (1.10), we have

$$\int_{R^m} |K(\boldsymbol{y})|^s d\boldsymbol{y} < \infty \qquad \text{for } s = 2, 3 . \tag{3.5}$$

Consequently by Lemma 1 $\frac{1}{h_n^m} \int_{R^m} K^2\left(\frac{\boldsymbol{y}}{h_n}\right) f(\boldsymbol{x}-\boldsymbol{y}) d\boldsymbol{y}$ converges to $f(\boldsymbol{x}) \int_{R^m} K^2(\boldsymbol{y}) d\boldsymbol{y}$

at all points $\boldsymbol{x}$ as n tends to ∞. Therefore from (3.1) $\frac{1}{nh_n^{2m}}\int_{R^m}K^2\left(\frac{\boldsymbol{y}}{h_n}\right)f(\boldsymbol{x}-\boldsymbol{y})d\boldsymbol{y}$ converges to zero at all point points $\boldsymbol{x}$ as n tends to ∞ and hence by Lemma 2 $\frac{1}{n}\sum_{j=1}^{n}\frac{1}{jh_j^{2m}}\int_{R^m}K^2\left(\frac{\boldsymbol{y}}{h_j}\right)f(\boldsymbol{x}-\boldsymbol{y})d\boldsymbol{y}$ converges to zero at all points $\boldsymbol{x}$ as n tends to ∞. By applying this fact on an inequality:

$$(3.6)\qquad 0\leqq\frac{1}{n^2}\sum_{j=1}^{n}\frac{1}{h_j^{2m}}\int_{R^m}K^2\left(\frac{\boldsymbol{y}}{h_j}\right)f(\boldsymbol{x}-\boldsymbol{y})d\boldsymbol{y}$$
$$\leqq\frac{1}{n}\sum_{j=1}^{n}\frac{1}{jh_j^{2m}}\int_{R^m}K^2\left(\frac{\boldsymbol{y}}{h_j}\right)f(\boldsymbol{x}-\boldsymbol{y})d\boldsymbol{y}$$

we have

$$(3.7)\qquad \frac{1}{n^2}\sum_{j=1}^{n}\frac{1}{h_j^{2m}}\int_{R^m}K^2\left(\frac{\boldsymbol{y}}{h_j}\right)f(\boldsymbol{x}-\boldsymbol{y})d\boldsymbol{y}\to 0\qquad (n\to\infty).$$

Next, by Lemma 1 we have easily

$$(3.8)\qquad \frac{1}{n^2}\sum_{j=1}^{n}\left\{\frac{1}{h_j^{m}}\int_{R^m}K\left(\frac{\boldsymbol{y}}{h_j}\right)f(\boldsymbol{x}-\boldsymbol{y})d\boldsymbol{y}\right\}^2\to 0\qquad (n\to\infty).$$

By applying (3.7) and (3.8) on (3.4), we have (3.2).

At last, we suppose that $K(\boldsymbol{y})$ satisfies (1.11). By applying Theorem 1 and (3.2) on

$$(3.9)\qquad E|\hat{f}_n(\boldsymbol{x})-f(\boldsymbol{x})|^2=Var[\hat{f}_n(\boldsymbol{x})]+|E\hat{f}_n(\boldsymbol{x})-f(\boldsymbol{x})|^2$$

we have (3.3). Thus the theorem is proved.

Now by (3.3) we have

$$(3.10)\qquad \hat{f}_n(\boldsymbol{x})\xrightarrow{P}f(\boldsymbol{x})\quad \text{at all points } \boldsymbol{x}\ (n\to\infty),$$

so the estimate $\hat{f}_n(\boldsymbol{x})$ is consistent at all points $\boldsymbol{x}$, where (3.10) denotes that $\hat{f}_n(\boldsymbol{x})$ converges to $f(\boldsymbol{x})$ in probability at all points $\boldsymbol{x}$ as n tends to ∞.

COROLLARY 3. *If we assume the uniform continuity of the probability density function* $f(\boldsymbol{x})$ *in Theorem* 2, *then we have* $\lim_{n\to\infty}\sup_{x} Var[f_n(\boldsymbol{x})]=0$ *and* $\lim_{n\to\infty}\sup_{x} E|f_n(\boldsymbol{x})-f(\boldsymbol{x})|^2=0.$

COROLLARY 4. *We suppose that a sequence of monotone decreasing positive numbers* $\{h_n\}$ *satisfies* (2.1) *and* (3.1) *and that the measurable function* $K(\boldsymbol{y})$ *satisfies* (1.9), (1.10) *and* (1.12). *Then we have*

$$(3.11)\qquad \lim_{n\to\infty}Var[\hat{f}_n(\boldsymbol{x})]=0$$

at all points of continuity of $f(\boldsymbol{x})$. *Furthermore if* $K(\boldsymbol{y})$ *satisfies* (1.11), *then we have*

$$(3.12)\qquad \lim_{n\to\infty}E|\hat{f}_n(\boldsymbol{x})-f(\boldsymbol{x})|^2=0$$

at all points of continuity of $f(\boldsymbol{x})$.

We can prove the above corollary in a same manner as the proof of Theorem 2.

THEOREM 3. *We suppose that the probability density function* $f(\boldsymbol{x})$ *is continuous in* R^m *and for the sequence of monotone decreasing positive numbers* $\{h_n\}$ *satisfying* (2.1) *there exists a limit with*

$$\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n}\frac{h_n^m}{h_j^m}=\alpha \qquad (0\leqq\alpha\leqq 1)\,. \tag{3.13}$$

Let the measurable function $K(\boldsymbol{y})$ *satisfy* (1.9) *and* (1.10). *Then we have*

$$\lim_{n\to\infty} nh_n^m Var[\hat{f}_n(\boldsymbol{x})]=\alpha f(\boldsymbol{x})\int_{R^m}K^2(y)d\boldsymbol{y} \tag{3.14}$$

PROOF. From (3.4) we have

$$nh_n^m Var[\hat{f}_n(\boldsymbol{x})]=\frac{h_n^m}{n}\sum_{j=1}^{n}\int_{R^m}K^2\Big(\frac{\boldsymbol{y}}{h_j}\Big)f(\boldsymbol{x}-\boldsymbol{y})d\boldsymbol{y} -\frac{h_n^m}{n}\sum_{j=1}^{n}\Big\{\frac{1}{h_j^m}\int_{R^m}K\Big(\frac{\boldsymbol{y}}{h_j}\Big)f(\boldsymbol{x}-\boldsymbol{y})d\boldsymbol{y}\Big\}^2. \tag{3.15}$$

At first, we have

$$\frac{h_n^m}{n}\sum_{j=1}^{n}\Big\{\frac{1}{h_j^m}\int_{R^m}K\Big(\frac{\boldsymbol{y}}{h_j}\Big)f(\boldsymbol{x}-\boldsymbol{y})d\boldsymbol{y}\Big\}^2 \leqq h_n^m\|f\|^2\Big\{\int_{R^m}|K(\boldsymbol{y})|\,d\boldsymbol{y}\Big\}^2, \tag{3.16}$$

which tends to zero as n tends to ∞, where $\|f\|=\max_{\boldsymbol{x}} f(\boldsymbol{x})$, whose existence is secured by the continuity of $f(\boldsymbol{x})$.

Next, we have

$$\frac{h_n^m}{n}\sum_{j=1}^{n}\frac{1}{h_j^{2m}}\int_{R^m}K^2\Big(\frac{\boldsymbol{y}}{h_j}\Big)f(\boldsymbol{x}-\boldsymbol{y})d\boldsymbol{y} =\int_{R^m}K^2(y)\Big\{\frac{1}{n}\sum_{j=1}^{n}\frac{h_n^m}{h_j^m}f(\boldsymbol{x}-h_j\boldsymbol{y})\Big\}d\boldsymbol{y}\,. \tag{3.17}$$

We shall show that $\frac{1}{n}\sum_{j=1}^{n}\frac{h_n^m}{h_j^m}f(\boldsymbol{x}-h_j\boldsymbol{y})$ uniformly converges to $\alpha f(\boldsymbol{x})$ in $\boldsymbol{y}\in\{\boldsymbol{y}\mid\|\boldsymbol{y}\|\leqq M\}$ for any $M>0$. It follows from the continuity of $f(\boldsymbol{x})$ that for arbitrary $\varepsilon>0$ there exists a positive integer N_0 not depending on $\boldsymbol{y}\in\{\boldsymbol{y}\mid\|\boldsymbol{y}\|\leqq M\}$ such that

$$|f(\boldsymbol{x}-h_j\boldsymbol{y})-f(\boldsymbol{x})|<\varepsilon \qquad \text{for } j>N_0 \text{ and } \boldsymbol{y}\in\{\boldsymbol{y}\mid\|\boldsymbol{y}\|\leqq M\}\,. \tag{3.18}$$

Hence for sufficiently large n, we have

$$\begin{aligned}&\Big|\frac{1}{n}\sum_{j=1}^{n}\frac{h_n^m}{h_j^m}f(\boldsymbol{x}-h_j\boldsymbol{y})-\alpha f(\boldsymbol{x})\Big|\\ &\leqq\frac{1}{n}\sum_{j=1}^{n}\frac{h_n^m}{h_j^m}|f(\boldsymbol{x}-h_j\boldsymbol{y})-f(\boldsymbol{x})|+\Big|\frac{1}{n}\sum_{j=1}^{n}\frac{h_n^m}{h_j^m}-\alpha\Big|f(\boldsymbol{x})\\ &\leqq\frac{2}{n}\sum_{j=1}^{N_0}\frac{h_n^m}{h_j^m}\|f\|+\varepsilon\frac{1}{n}\sum_{j=N_0+1}^{n}\frac{h_n^m}{h_j^m}+\Big|\frac{1}{n}\sum_{j=1}^{n}\frac{h_n^m}{h_j^m}-\alpha\Big|f(\boldsymbol{x}),\end{aligned} \tag{3.19}$$

which is not depend on $\boldsymbol{y}\in\{\boldsymbol{y}\mid\|\boldsymbol{y}\|\leqq M\}$. As n tends to ∞, the second term of the last expression of (3.19) tends to $\alpha\varepsilon$ ($\leqq\varepsilon$) and the rest of the expression tend to zero. Since ε is arbitrary, we have

$$\frac{1}{n}\sum_{j=1}^{n}\frac{h_n^m}{h_j^m}f(\boldsymbol{x}-h_j\boldsymbol{y})\to\alpha f(\boldsymbol{x}) \qquad (n\to\infty) \tag{3.20}$$

uniformly in $\boldsymbol{y}\in\{\boldsymbol{y}\mid\|\boldsymbol{y}\|\leqq M\}$ and also by (1.9)

(3.21) $$K^2(\boldsymbol{y})\frac{1}{n}\sum_{j=1}^{n}\frac{h_n^m}{h_j^m}f(\boldsymbol{x}-h_j\boldsymbol{y})\to\alpha f(\boldsymbol{x})K^2(\boldsymbol{y}) \qquad (n\to\infty)$$

uniformly in $\boldsymbol{y}\in\{\boldsymbol{y}\mid\|\boldsymbol{y}\|\leqq M\}$. Therefore for arbitrary $M>0$ we have

(3.22) $$\left|\int_{R^m}K^2(\boldsymbol{y})\left\{\frac{1}{n}\sum_{j=1}^{n}\frac{h_n^m}{h_j^m}f(\boldsymbol{x}-h_j\boldsymbol{y})\right\}d\boldsymbol{y}-\alpha f(\boldsymbol{x})\int_{R^m}K^2(\boldsymbol{y})d\boldsymbol{y}\right|$$

$$\leqq\left|\int_{\|y\|\leqq M}K^2(\boldsymbol{y})\left\{\frac{1}{n}\sum_{j=1}^{n}\frac{h_n^m}{h_j^m}f(\boldsymbol{x}-h_j\boldsymbol{y})\right\}d\boldsymbol{y}-\alpha f(\boldsymbol{x})\int_{R^m}K^2(\boldsymbol{y})d\boldsymbol{y}\right|$$

$$+\|f\|\int_{\|y\|\geqq M}K^2(\boldsymbol{y})d\boldsymbol{y}+\alpha f(\boldsymbol{x})\int_{\|y\|\geqq M}K^2(\boldsymbol{y})d\boldsymbol{y}\,,$$

whose first term tends to zero by (3.21) as one lets n tend to ∞ at first and secondly the remaining terms tends to zero by (3.5) as one lets M tend to ∞. Therefore by applying (3.16), (3.17) and (3.22) on (3.15), we have (3.14). Thus the theorem is proved.

This theorem enables us to compare asymptotically the variance of our estimate $\hat{f}_n(\boldsymbol{x})$ with that of the estimate $\tilde{f}_n(\boldsymbol{x})$ proposed by Parzen [4] in one-dimensional case and quated as (1.3) in this paper. Our estimate is asymptotically at least as good as $\tilde{f}_n(\boldsymbol{x})$, as stated in the following corollary.

COROLLARY 5. *Under the same assumption as in Theorem 4, the variance of* $\hat{f}(\boldsymbol{x})$ *is asymptotically at most as large as the one of* $\tilde{f}(\boldsymbol{x})$.

PROOF. By comparing (3.11) and Theorem 2A in Parzen [4], we have

(3.23) $$\lim_{n\to\infty}\frac{Var[\hat{f}_n(\boldsymbol{x})]}{Var[\tilde{f}_n(\boldsymbol{x})]}=\lim_{n\to\infty}\frac{nh_n^m Var[\hat{f}_n(\boldsymbol{x})]}{nh_n^m Var[\tilde{f}_n(\boldsymbol{x})]}=\alpha\leqq 1\,.$$

Thus the corollary is proved.

4. Uniform consistency of $\hat{f}_n(x)$ and consistency of a sample mode.

In this section we shall show that the estimate $\hat{f}_n(\boldsymbol{x})$ is uniformly consistent and by using this fact it turns out that a sample mode obtained from the estimate $\hat{f}_n(\boldsymbol{x})$ is a consistent estimate of the population mode.

THEOREM 4. *We suppose that the probability density function* $f(\boldsymbol{x})$ *is uniformly continuous in* R^m *and that the sequence of monotone decreasing positive numbers* $\{h_n\}$ *satisfies* (2.1) *and*

(4.1) $$\lim_{n\to\infty}n^{1/2}h_n^m=0\,.$$

Let the measurable function $K(\boldsymbol{y})$ *satisfy* (1.10) *and* (1.11), *its Fourier transform*

(4.2) $$k(\boldsymbol{u})=\int_{R^m}e^{iu'y}K(\boldsymbol{y})d\boldsymbol{y}$$

be absolutely integrable and $k(\boldsymbol{u})$ *be nondecreasing in negative part and nonincreasing in positive part for each argument. Then we have*

(4.3) $$\sup_x|\hat{f}_n(\boldsymbol{x})-f(\boldsymbol{x})|\overset{P}{\longrightarrow}0 \qquad (n\to\infty)\,.$$

Furthermore if $f(\boldsymbol{x})$ has a unique made θ defined by

$$f(\theta)=\max_{x} f(\boldsymbol{x}) \tag{4.4}$$

then a sample mode θ_n defined by

$$\hat{f}_n(\theta_n)=\max_{x} \hat{f}_n(\boldsymbol{x}) \tag{4.5}$$

converges to θ in probability as n tends to ∞.

PROOF. Since $K(\boldsymbol{y})$ and $k(\boldsymbol{u})$ are absolutely integrable, we have

$$K(\boldsymbol{y})=\frac{1}{(2\pi)^m}\int_{R^m} e^{-i\boldsymbol{u}'\boldsymbol{y}}k(\boldsymbol{u})d\boldsymbol{u}\,. \tag{4.6}$$

In term of $k(\boldsymbol{u})$, the Fourier transform of $K(\boldsymbol{y})$, we have

$$\hat{f}_n(\boldsymbol{x})=\frac{1}{(2\pi)^m}\int_{R^m}\left\{\frac{1}{n}\sum_{j=1}^{n} e^{i\boldsymbol{u}'X_j}k(h_j\boldsymbol{u})\right\}e^{-i\boldsymbol{u}'\boldsymbol{x}}d\boldsymbol{u}\,. \tag{4.7}$$

Hence if we let $\varphi(\boldsymbol{u})$ be the population characteristic function then

$$|\hat{f}_n(\boldsymbol{x})-E\hat{f}_n(\boldsymbol{x})|=\left|\frac{1}{(2\pi)^m}\int_{R^m}\frac{1}{n}\sum_{j=1}^{n}\left\{e^{i\boldsymbol{u}'X_j}-\varphi(\boldsymbol{u})\right\}k(h_j\boldsymbol{u})e^{-i\boldsymbol{u}'\boldsymbol{x}}d\boldsymbol{u}\right|. \tag{4.8}$$

Therefore we have

$$\sup_{x}|\hat{f}_n(\boldsymbol{x})-E\hat{f}_n(\boldsymbol{x})|\leqq\frac{1}{(2\pi)^m}\int_{R^m}\left|\frac{1}{n}\sum_{j=1}^{n}\left\{e^{i\boldsymbol{u}'X_j}-\varphi(\boldsymbol{u})\right\}k(h_j\boldsymbol{u})\right|d\boldsymbol{u}\,. \tag{4.9}$$

It follows from (4.9) and Schwarz's inequality that

$$E[\sup_{x}|\hat{f}_n(\boldsymbol{x})-E\hat{f}_n(\boldsymbol{x})|]\leqq\frac{1}{(2\pi)^m}\int_{R^m}\left\{\frac{1}{n^2}\sum_{j=1}^{n}E|e^{i\boldsymbol{u}'X_j}-\varphi(\boldsymbol{u})|^2|k(h_j\boldsymbol{u})|^2\right\}^{1/2}d\boldsymbol{u}\,. \tag{4.10}$$

Since $E|e^{i\boldsymbol{u}'X_j}-\varphi(\boldsymbol{u})|^2\leqq 1$, $\{h_n\}$ is the sequence of monotone decreasing positive numbers and $k(\boldsymbol{u})$ is nondecreasing in negative part and nonincreasing in positive part for each argument, by (4.10) we have

$$\begin{aligned} E[\sup|\hat{f}_n(\boldsymbol{x})-E\hat{f}_n(\boldsymbol{x})|]&\leqq\frac{1}{(2\pi)^m n}\int_{R^m}\{n|k(h_n\boldsymbol{u})|^2\}^{1/2}d\boldsymbol{u} \\ &=\frac{1}{n^{1/2}h_n^m(2\pi)^m}\int_{R^m}|k(\boldsymbol{u})|d\boldsymbol{u}\,. \end{aligned} \tag{4.11}$$

By applying (4.1) on (4.11), we have

$$\lim_{n\to\infty} E[\sup_{x}|\hat{f}_n(\boldsymbol{x})-E\hat{f}_n(\boldsymbol{x})|]=0\,. \tag{4.12}$$

It follows from (4.12) and Markov's inequality (See, for instance, p. 158 in Loeve [2]) that

$$\sup_{x}|\hat{f}_n(\boldsymbol{x})-E\hat{f}_n(\boldsymbol{x})|\overset{P}{\longrightarrow}0 \qquad (n\to\infty)\,. \tag{4.13}$$

In the inequality:

$$\sup_{x}|\hat{f}_n(\boldsymbol{x})-\hat{f}(\boldsymbol{x})|\leqq\sup_{x}|\hat{f}_n(\boldsymbol{x})-E\hat{f}_n(\boldsymbol{x})|+\sup_{x}|E\hat{f}_n(\boldsymbol{x})-f(\boldsymbol{x})| \tag{4.14}$$

the righthand side converges to zero in probability as n tends to ∞ by (4.13) and (2.2). Thus we obtain (4.3).

Next, from the absolute integrability of $k(\boldsymbol{u})$, it turns out by the similar method to the proof of Theorem 31 in Bochner and Chandrasekharan [1] that $\hat{f}_n(\boldsymbol{x})$ tends to zero as $\|\boldsymbol{x}\|$ tends to ∞ and by the similar method to the proof of (1.3) in Bochner and Chandrasekharan [1] we can obtain the uniform continuity of $\hat{f}_n(\boldsymbol{x})$ by using an inequality $|u_1+\cdots+u_m|\leqq\sqrt{m}(u_1^2+\cdots+u_m^2)^{1/2}$. Therefore there exists a random vector θ_n defined by (4.4). By the same reason as the proof of Theorem 3A in Parzen [4] we are able to obtain

$$\theta_n \xrightarrow{P} \theta \qquad (n\to\infty), \tag{4.15}$$

which completed the proof of the theorem.

Thus the uniform consistency of $\hat{f}_n(\boldsymbol{x})$ and the consistency of θ_n are established. Concerning the measurable function $K(\boldsymbol{y})$ satisfying the assumption in Theorem 4, for example, $K(\boldsymbol{y})=\frac{1}{(2\pi)^{m/2}}e^{-\boldsymbol{y}'\boldsymbol{y}/2}$ satisfies the assumption and in the case $k(\boldsymbol{u})=e^{-\boldsymbol{u}'\boldsymbol{u}/2}$.

5. Asymptotic normality of $\hat{f}_n(\boldsymbol{x})$.

We shall consider the asymptotic normality of $\hat{f}_n(\boldsymbol{x})$.

THEOREM 5. *We suppose that the probability density function $f(\boldsymbol{x})$ is continuous in R^m and that for the sequence of monotone decreasing positive numbers $\{h_n\}$ satisfying* (2.1) *and* (3.1), *there exists a non-zero limit with* (3.13). *Let the measurable function $K(\boldsymbol{y})$ satisfy* (1.9) *and* (1.10). *Then the distribution function of*

$$\frac{\hat{f}_n(\boldsymbol{x})-E\hat{f}_n(\boldsymbol{x})}{\sqrt{Var[\hat{f}_n(\boldsymbol{x})]}} \tag{5.1}$$

converges to the standardized normal distribution function at all points $\boldsymbol{x}$.

$h_n=1/(\log n)^{1/m}$ and $h_n=1/n^{r/m}$ $(0<r<1/2)$ are examples of sequences of monotone decreasing positive numbers $\{h_n\}$ satisfying (2.1), (3.1), (4.1) and (3.13) with $\alpha=1$ and $\alpha=1/(r+1)$ respectively. In case of $h_n=1/n^{r/m}$ $(0<r<1/2)$,

$$\frac{2}{3}<\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n}\frac{h_n^m}{h_j^m}<1\,.$$

Thus the assumption of the existence of non-zero limit with (3.13) does not appear to be very restrictive.

PROOF. If we put for any fixed $\boldsymbol{x}$

$$V_j=\frac{1}{h_j^m}K\Big(\frac{\boldsymbol{x}-\boldsymbol{X}_j}{h_j}\Big) \qquad \text{for } j=1, 2, 3, \cdots \tag{5.1}$$

then $V_1, V_2, V_3, \cdots$ are a sequence of independent random variables and we have

$$\frac{\hat{f}_n(\boldsymbol{x})-E\hat{f}_n(\boldsymbol{x})}{\sqrt{Var[\hat{f}_n(\boldsymbol{x})]}}=\frac{\sum_{j=1}^{n}(V_j-EV_j)}{\sqrt{\sum_{j=1}^{n}Var[V_j]}} \tag{5.3}$$

Therefore, by virtue of Lyapunov's condition (See, for instance, p. 432 in Parzen [3]), it is enough to show that

(5.4)
$$\lim_{n\to\infty}\frac{\sum_{j=1}^{n}E|V_j-EV_j|^3}{(\sum_{j=1}^{n}Var[V_j])^{3/2}}=0.$$

From Theorem 3, we have

(5.5)
$$\frac{h_n^m}{n}\sum_{j=1}^{n}Var[V_j]=nh_n^m Var[\hat{f}_n(\boldsymbol{x})]\longrightarrow \alpha f(\boldsymbol{x})\int_{R^m}K^2(\boldsymbol{y})d\boldsymbol{y}\qquad (n\to\infty).$$

On the other hand, by using an inequality:

(5.6)
$$(a+b)^3\leqq 4(a^3+b^3)\qquad \text{for } a,b\geqq 0,$$

we have

(5.7)
$$\sum_{j=1}^{n}E|V_j-EV_j|^3\leqq 4\sum_{j=1}^{n}\{E|V_j|^3+|EV_j|^3\}$$

and further

(5.3)
$$E|V_j|^3=\int_{R^m}\left|\frac{1}{h_j^m}K\left(\frac{\boldsymbol{y}}{h_j}\right)\right|^3 f(\boldsymbol{x}-\boldsymbol{y})d\boldsymbol{y}$$
$$\leqq \|f\|\frac{1}{h_j^{2m}}\int_{R^m}|K(\boldsymbol{z})|^3 d\boldsymbol{z}$$

(5.9)
$$|EV_j|\leqq E|V_j|\leqq \|f\|\int_{R^m}|K(\boldsymbol{z})|d\boldsymbol{z}.$$

As $\{h_n\}$ is a sequence of monotone decreasing positive numbers, we can evaluate the lefthand side of (5.7) by using (5.8) and (5.9),

(5.10)
$$\sum_{j=1}^{n}E|V_j-EV_j|^3\leqq 4\|f\|\frac{n}{h_n^{2m}}\int_{R^m}|K(\boldsymbol{z})|^3 d\boldsymbol{z}+4n\|f\|^3\left\{\int_{R^m}|K(\boldsymbol{z})|d\boldsymbol{z}\right\}^3.$$

Hence by (5.10) we have

(5.11)
$$\lim_{n\to\infty}\frac{\sum_{j=1}^{n}E|V_j-EV_j|^3}{(\sum_{j=1}^{n}Var[V_j])^{3/2}}$$
$$=\lim_{n\to\infty}\frac{\left(\frac{h_n^m}{n}\right)^{3/2}\sum_{j=1}^{n}E|V_j-EV_j|^3}{\left(\frac{h_n^m}{n}\sum_{j=1}^{n}Var[V_j]\right)^{3/2}}$$
$$\leqq \lim_{n\to\infty}\frac{1}{\left(\frac{h_n^m}{n}\sum_{j=1}^{n}Var[V_j]\right)^{3/2}}\left[\frac{1}{(nh_n^m)^{1/2}}\int_{R^m}|K(\boldsymbol{z})|^3 d\boldsymbol{z}\right.$$
$$\left.+4\frac{h_n^{3m/2}}{n^{1/2}}\|f\|^3\left\{\int_{R^m}|K(\boldsymbol{z})|d\boldsymbol{z}\right\}^3\right].$$

By applying (2.1), (3.1), (1.10), (3.5) and (5.4) on (5.10) we have (5.3), which leads us to the completion of the theorem.

6. Comparison with parametric estimation.

In the above sections we discussed the statistical properties of the estimate $\hat{f}_n(\boldsymbol{x})$ under the assumption that the probability density function is continuous at a point $\boldsymbol{x}$, continuous in R^m or uniformly continuous in R^m. Now, we shall consider the situation where the underlying distribution is assumed to be known in form. Let the probability density function be known in form except for the values of some real parameters, $\theta=(\theta_1, \theta_2, \cdots, \theta_k)$, belonging to a space ω. Corresponding to a given $\theta \in \omega$, we shall denote the probability density function by $f_\theta(\boldsymbol{x})$.

If the probability density function $f_\theta(\boldsymbol{x})$ is continuous as a function of θ at a point $\boldsymbol{x}$, then for a consistent estimate $\hat{\theta}$ of θ, $f_{\hat{\theta}}(\boldsymbol{x})$ is a consistent estimate of $f_\theta(\boldsymbol{x})$ at the point $\boldsymbol{x}$. When we compare two consistent estimates $\hat{f}_n(\boldsymbol{x})$ and $f_{\hat{\theta}}(\boldsymbol{x})$, it is expected that the estimate $f_{\hat{\theta}}(\boldsymbol{x})$ is preferable to the estimate $\hat{f}_n(\boldsymbol{x})$. Because the estimate $f_{\hat{\theta}}(\boldsymbol{x})$ utilizes the assumption that the probability density function is known in form whereas the estimate $\hat{f}_n(\boldsymbol{x})$ does not. We shall show this fact by an example in case where the underlying distribution is univariate normal $N(\theta, 1)$ $(-\infty<\theta<\infty)$. By using a consistent estimate of θ, $\bar{X}$, we can obtain a consistent estimate of $f_\theta(x)$,

(6.1) $$f_{\bar{X}}(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-\bar{X})^2}{2}}.$$

It is easy to show that

(6.2) $$Ef_{\bar{X}}(x)=\frac{1}{\sqrt{2\pi}}\sqrt{\frac{n}{n+1}}e^{-\frac{n}{2(n+1)}(x-\theta)^2}$$

(6.3) $$Var[f_{\bar{X}}(x)]=\frac{1}{2\pi}\sqrt{\frac{n}{n+2}}e^{-\frac{n}{n+2}(x-\theta)^2}-\frac{1}{2\pi}\frac{n}{n+1}e^{-\frac{n}{n+1}(x-\theta)^2}$$
$$=\frac{1}{2\pi}\sqrt{\frac{n}{n+2}}e^{-\frac{n}{n+2}(x-\theta)^2}\left\{1-\sqrt{\frac{n+2}{n}}\frac{n}{n+1}e^{-\frac{n(x-\theta)^2}{(n+1)(n+2)}}\right\}$$

(6.4) $$1-\sqrt{\frac{n+2}{n}}\frac{n}{n+1}e^{-\frac{n(x-\theta)^2}{(n+1)(n+2)}}.$$
$$=\sqrt{\frac{n}{n+2}}\frac{n^3}{(n+1)^2(n+2)}(x-\theta)^2+0\left(\frac{1}{n^2}\right) \quad \text{for } x\neq\theta.$$

Hence we have

(6.5) $$\lim_{n\to\infty} Ef_{\bar{X}}(x)=f_\theta(x) \quad \text{for all } x$$

(6.6) $$\lim_{n\to\infty} E|f_{\bar{X}}(x)-f_\theta(x)|^2=0 \quad \text{for all } x$$

(6.7) $$\lim_{n\to\infty} n\,Var[f_{\bar{X}}(x)]=\frac{1}{2\pi}(x-\theta)^2e^{-(x-\theta)^2} \quad \text{for } x\neq\theta$$

(6.8) $$\lim_{n\to\infty} n^2 Var[f_{\bar{X}}(\theta)]=\frac{1}{4\pi}.$$

Therefore, under the assumption that for the sequence of monotone decreasing positive numbers satisfying (2.1) and (3.1) their exists a nonzero limit with (3.13) and that the measurable function $K(y)$ satisfies (1.9) and (1.10), from (3.14), (6.7) and (6.8) we have

$$\lim_{n\to\infty}\frac{Var[f_{\bar{X}}(x)]}{Var[\hat{f}_n(x)]}=0 \qquad \text{for all points } x.$$

As was expected it was shown that the variance of tne estimate $f_{\bar{X}}(x)$ is asymptotically smaller than the one of the estimate $\hat{f}_n(x)$.

At last, we shall note the unique unbiased sufficient estimate of $f_\theta(x)$ (Washio, Morimoto and Ikeda [6]),

$$\tilde{f}_\theta(x)=\frac{1}{\sqrt{2\pi\left(1-\frac{1}{n}\right)}}e^{-\frac{(x-\bar{X})^2}{2\left(1-\frac{1}{n}\right)}}.$$

We can easily verify that

$$\lim_{n\to\infty} n\,Var[\tilde{f}_\theta(x)]=\frac{1}{2\pi}(x-\theta)^2e^{-(x-\theta)^2} \qquad \text{for } x\neq\theta$$

$$\lim_{n\to\infty} n^2 Var[\tilde{f}_\theta(\theta)]=\frac{1}{4\pi},$$

which coincides with the limits in case of $f_{\bar{X}}(x)$. Therefore it can be also shown that the variance of the estimate $\tilde{f}_\theta(x)$ is asymptotically smaller than the one of the estimate $\hat{f}_n(x)$.

As is usually the case, however, in the situation where the underlying probability density function is not known in form except for its continuity, our estimate $\hat{f}_n(x)$ is considered preferable.

References

[1] S. Bochner and K. Chandrasekharen (1949), Fourier Transforms, Princeton University Press.
[2] M. Loéve (1960), Probability Theory, Van Nostrand, Princeton.
[3] E. Parzen (1960), Modern Probability Theory and Its Application, John Wiley and Sons.
[4] E. Parzen (1962), On Estimation of A Probability Density Function and Mode, Ann. Math. Statist., Vol. 33, pp. 1065-1076.
[5] H. G. Tucker (1967), A Graduate Course in Probability, Academic Press.
[6] Y. Washio, H. Morimoto and N. Ikeda (1956), Unbiased Estimation Based on Sufficient Statistics. Bull. Math. Stat., Vol. 6, No. 3-4, pp. 69-93.
[7] C. T. Wolverton and T. J. Wagner (1969), Asymptotically Optimal Discriminant Functions for Pattern Classification. IEEE Trans. Information Theory, Vol. IT-15, No. 2, pp. 258-265.
[8] H. Yamato (1970), On the Wolverton and Wagner's Asymptotically Optimal Discriminant Function (to appear in Rep. Fac. Sci. Kagoshima Univ., (Math. Phys. Chem.) No. 3).