On strong consistency of a sequential estimator of probability density

By
Eiichi Isogai*

(Received October 30, 1977)

1. Introduction and Summary

Let $F(x)$ be a probability distribution function on the real line. It is well known that assuming that the singular part is identically zero, $F(x)$ can be uniquely decomposed into

(1. 1) $$F(x)=F_1(x)+F_2(x),$$

where $F_1(x)$ is an absolutely continuous function and $F_2(x)$ is a pure step function with steps of magnitude, say, S_i at the points $x=x_i$, $i=0, \pm 1, \pm 2, \cdots\cdots$ and finally both $F_1(x)$ and $F_2(x)$ are non-decreasing. If the singular part is identically zero as has been assumed here, $F_1(x)$ has a density function $f(x)$ almost everywhere, namely,

(1. 2) $$dF_1(x)/dx=f(x) \quad \text{a.e.x.}$$

At a point of continuity x_0' of $F(x)$ its density is clearly $f(x_0')$.

Let X_1, X_2, X_3, $\cdots\cdots$ be a sequence of independent identically distributed random variables with the common distribution function $F(x)$. We shall consider the problem of estimating the density $f(x)$ at all points of continuity of $F(x)$ and also of $f(x)$ as has been seen in (1. 2) from X_1, X_2, X_3, $\cdots\cdots$. The kernel estimate of f from X_1, X_2, X_3 $\cdots\cdots$, X_n is given by

(1. 3) $$f_n(x)=(B_n/n)\sum_{j=1}^{n} K(B_n(X_j-x)),$$

where K, the kernel, is an arbitrary bounded probability density on the real line and $\{B_n\}$ is a sequence of positive numbers. For some conditions on K and $\{B_n\}$, Murthy [1] proved that $f_n(x_0')$ 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 $f(x)$ under the condition

(1. 4) $$\sum_i S_i/|x_0'-x_i|<\infty.$$

(That is $f_n(x_0')\longrightarrow f(x_0')$ in prob. as $n\rightarrow\infty$).

---

* Niigata University

In this paper we shall give a class of sequential estimators $\{f_n\}$ such that $f_n(x_0')$ is a strong consistent estimate of $f(x_0')$ at a point of continuity x_0' of the distribution $F(x)$ and also of the density $f(x)$ under the condition (1. 4) in the sence that

(1. 5) $$f_n(x_0') \longrightarrow f(x_0') \quad \text{with probability one as } n \to \infty.$$

In section 2, we shall give some lemmas to be used throughout the paper. In section 3, we shall construct sequential estimators $\{f_n\}_{n=1}^{\infty}$ and prove the strong consistency of f_n and also give the rate of variance of f_n.

2. Auxiliary Results

The following two lemmas are necessary for proving Theorem 1 and 2. Lemma 2. 1 and Lemma 2. 2 can be found in WATANABE [3] and [4], respectively.

LEMMA 2. 1. Let $\{A_n\}_{n=1}^{\infty}$ be a sequence of non-negative numbers. Suppose that there exist three sequences of non-negative numbers $\{a_n\}_{n=1}^{\infty}$, $\{b_n\}_{n=1}^{\infty}$, $\{L_n\}_{n=1}^{\infty}$ and a positive constant L such that

(2. 1) $$A_{n+1} \leqslant (1-a_{n+1})A_n + L \cdot a_{n+1} \cdot b_{n+1} + L_{n+1} \quad \text{for all } n \geqslant 1,$$

(2. 2) $$1 \geqslant a_n \geqslant 0 (n=1, 2, 3, \cdots\cdots), \sum_{n=1}^{\infty} a_n = \infty \text{ and } \lim_{n\to\infty} a_n = 0,$$

(2. 3) $$\lim_{n\to\infty} b_n = 0,$$

(2. 4) $$\sum_{n=1}^{\infty} L_n < \infty.$$

Then, it holds that $\lim_{n\to\infty} A_n = 0$.

Furthermore, if $L_n=0$ for all $n \geqslant 1$ in (2. 1) and there exists a constant $\alpha_0 > 0$ such that

(2. 5) $$(1-a_{n+1})b_n/b_{n+1} \leqslant 1 - \alpha_0 a_{n+1} \text{ for all } n \geqslant \text{ some } n_0,$$

where $\{b_n\}_{n=1}^{\infty}$ need not satisfy the condition (2. 3),
then there exists a constant $C > 0$ such that

(2. 6) $$A_n \leqslant C \cdot b_n \quad \text{for all } n \geqslant 1.$$

LEMMA 2. 2. Let $\{U_n\}_{n=1}^{\infty}$ and $\{V_n\}_{n=1}^{\infty}$ be two sequences of random variables on a probability space (Ω, F, P). Let $\{F_n\}_{n=1}^{\infty}$ be a sequence of σ-fields, $F_n \subset F_{n+1} \subset F$, where U_n and V_n are measurable with respect to F_n for each n. Furthermore, let $\{a_n\}_{n=1}^{\infty}$ be a sequence of real numbers satisfying (2. 2). Suppose that the following conditions are satisfied:

(2. 7) $$U_n \geqslant 0 \quad \text{a.s.} \quad \text{for all } n \geqslant 1,$$

(2. 8) $$E[U_1] < \infty,$$

(2. 9) $$E[U_{n+1}/F_n] \leqslant (1-a_{n+1})U_n + V_n \quad \text{a.s.} \quad \text{for all } n \geqslant 1,$$

(2. 10) $$\sum_{n=1}^{\infty} E[|V_n|] < \infty.$$

Then, it holds that $\lim_{n\to\infty} U_n=0$ a.s. and $\lim_{n\to\infty} E[U_n]=0$.

3. Strong Consistency

In this section, we shall prove two theorems.

Let $K(y)$ be a real-valued Borel measurable function on the real line satisfying

(K1) $\quad K(y)\geqslant 0 \quad$ for all $y\in(-\infty, \infty)$,

(K2) $\quad \int_{-\infty}^{\infty} K(y)\,dy=1$,

(K3) $\quad \sup_{-\infty<y<\infty} K(y)=K_0<\infty$,

(K4) $\quad \lim_{|y|\to\infty} |y|\,K(y)=0$.

Also, let $\{h_n\}_{n=1}^{\infty}$ be a sequence of real numbers satisfying

(H1) $\quad h_n>0 \quad$ for all $n\geqslant 1$,

(H2) $\quad \lim_{n\to\infty} h_n=0$.

Then, we can define the sequence $\{K_n(x, y)\}_{n=1}^{\infty}$ for $x, y\in(-\infty, \infty)$,

(3. 1) $\quad K_n(x, y)=h_n^{-1}\,K(h_n^{-1}(x-y)) \quad$ for $n=1, 2, \cdots\cdots$.

The following lemma can be found in PARZEN [2].

LEMMA 3. 1. Suppose that $K(y)$ is a real-valued Borel function on the real line satisfying (K1), (K3), (K4) and

(K5) $\quad \int_{-\infty}^{\infty} K(y)\,dy<\infty$.

Let $g(y)$ satisfy

(3. 2) $\quad \int_{-\infty}^{\infty} |g(y)|\,dy<\infty$.

Let $\{K_n(x, y)\}_{n=1}^{\infty}$ be defined by (3. 1) where $\{h_n\}_{n=1}^{\infty}$ is a sequence of real numbers satisfying (H1) and (H2). Define

(3. 3) $\quad g_n(x)=\int_{-\infty}^{\infty} K_n(x, y)\,g(y)\,dy$.

Then, at every point x of continuity of $g(\cdot)$,

(3. 4) $\quad \lim_{n\to\infty} \left| g_n(x)-g(x)\int_{-\infty}^{\infty} K(y)\,dy \right| =0$.

We need the following lemma in proving Theorem 1, which is essentially the same as Lemma in MURTHY [1].

LEMMA 3. 2. Let $\{K_n(x, y)\}_{n=1}^{\infty}$ be defined by (3. 1). Let x_i $(i=0, \pm 1, \pm 2, \cdots\cdots)$ be the points of discontinuity of the distribution $F(x)$ and S_i the saltus of $F(x)$ at x_i and x a point of continuity of $F(x)$ and also of $f(x)$ the derivative of the absolutely continuous part of $F(x)$. Then, under the condition

$$(3.5) \qquad \sum_i S_i / |x_i - x| < \infty,$$

we have $\lim_{n\to\infty} |E[K_n(x, X_n)] - f(x)| = 0$.

Now, we shall construct sequential estimators $\{f_n\}_{n=1}^{\infty}$ of f. The following algorithm is found in WATANABE [4].

Algorithm. Let $\{a_n\}_{n=1}^{\infty}$ be a sequence of real numbers satisfying

$$\text{(A1)} \qquad 1 \geqslant a_n > 0 \quad (n=1, 2, \cdots\cdots) \text{ and } \sum_{n=1}^{\infty} a_n = \infty,$$

$$\text{(A2)} \qquad \lim_{n\to\infty} a_n = 0.$$

Then, $f_n(x)$ is given by the recurrence relation as follows:

$$\text{(A)} \qquad \begin{aligned} & f_0(x) = K(x) \qquad \text{for all } x \in (-\infty, \infty) \\ & f_{n+1}(x) = f_n(x) + a_{n+1}\{K_{n+1}(x, X_{n+1}) - f_n(x)\} \end{aligned}$$

for all $n \geqslant 1$ and all $x \in (-\infty, \infty)$.

THEOREM 1. Let x be an arbitrary point of continuity of $F(x)$ and also of $f(x)$ and satisfy the condition (3. 5).

(i) If $\lim_{n\to\infty} a_n h_n^{-1} = 0$, then

$$(3.6) \qquad \lim_{n\to\infty} E[(f_n(x) - f(x))^2] = 0.$$

(ii) If $\sum_{n=1}^{\infty} a_n^2 h_n^{-1} < \infty$, then

$$(3.7) \qquad \lim_{n\to\infty} f_n(x) = f(x) \qquad \text{with probability one}$$

and (3. 6) holds.

PROOF. From the algorithm (A), we have

$$\begin{aligned} E[f_{n+1}(x)] - f(x) & \\ &= (1 - a_{n+1})\{E[f_n(x)] - f(x)\} \\ &+ a_{n+1}\{E[K_{n+1}(x, X_{n+1})] - f(x)\}. \end{aligned}$$

Thus, we obtain

$$(3.8) \qquad \begin{aligned} |E[f_{n+1}(x)] - f(x)| & \\ &\leqslant (1 - a_{n+1})|E[f_n(x)] - f(x)| \\ &+ a_{n+1}|E[K_{n+1}(x, X_{n+1})] - f(x)|. \end{aligned}$$

Let us write that

$$A_{n+1}=|E[f_{n+1}(x)]-f(x)| \tag{3.9}$$

and

$$b_{n+1}=|E[K_{n+1}(x, X_{n+1})]-f(x)|.$$

From (3. 8), we have

$$A_{n+1}\leqslant(1-a_{n+1})A_n+a_{n+1}b_{n+1}. \tag{3.10}$$

By Lemma 3. 2, we get

$$\lim_{n\to\infty} b_n=0. \tag{3.11}$$

In view of (3. 9), (3. 10), (3. 11), (A1) and (A2), the conditions of Lemma 2. 1 can be easily checked. Therefore, we obtain $\lim_{n\to\infty} A_n=0$, that is,

$$\lim_{n\to\infty} |E[f_n(x)]-f(x)|=0. \tag{3.12}$$

Now, from the algorithm (A), we have

$$\begin{aligned} f_{n+1}(x)-E[f_{n+1}(x)] &=(1-a_{n+1})\{f_n(x)-E[f_n(x)]\} \\ &\quad +a_{n+1}\{K_{n+1}(x, X_{n+1})-E[K_{n+1}(x, X_{n+1})]\}. \end{aligned}$$

Hence, we get

$$\begin{aligned} |f_{n+1}(x)-E[f_{n+1}(x)]|^2 &=(1-a_{n+1})^2|f_n(x)-E[f_n(x)]|^2 \\ &\quad +a_{n+1}^2|K_{n+1}(x, X_{n+1})-E[K_{n+1}(x, X_{n+1})]|^2 \\ &\quad +2(1-a_{n+1})a_{n+1}\cdot\{f_n(x)-E[f_n(x)]\} \\ &\quad \times\{K_{n+1}(x, X_{n+1})-E[K_{n+1}(x, X_{n+1})]\}. \end{aligned} \tag{3.13}$$

By using the independence of $\{X_n\}_{n=1}^{\infty}$, we have

$$\begin{aligned} &E[\{f_n(x)-E[f_n(x)]\}\{K_{n+1}(x, X_{n+1})-E[K_{n+1}(x, X_{n+1})]\} \\ &\quad /X_1, X_2, \cdots\cdots, X_n] \\ &=\{f_n(x)-E[f_n(x)]\}E[K_{n+1}(x, X_{n+1})-E[K_{n+1}(x, X_{n+1})]] \\ &=0. \end{aligned} \tag{3.14}$$

From (A1), we have

$$(1-a_{n+1})^2\leqslant 1-a_{n+1}. \tag{3.15}$$

Combining (3. 13), (3. 14) and (3. 15), we get

$$(3.16)\quad E[(f_{n+1}(x)-E[f_{n+1}(x)])^2/X_1, X_2, \cdots\cdots, X_n]$$
$$\leqslant(1-a_{n+1})\cdot(f_n(x)-E[f_n(x)])^2$$
$$+a_{n+1}^2 E[(K_{n+1}(x, X_{n+1})-E[K_{n+1}(x, X_{n+1})])^2]$$
$$=(1-a_{n+1})\cdot(f_n(x)-E[f_n(x)])^2$$
$$+a_{n+1}^2 \mathrm{Var}[K_{n+1}(x, X_{n+1})].$$

We shall evaluate $\mathrm{Var}[K_{n+1}(x, X_{n+1})]$.

$$(3.17)\quad \mathrm{Var}[K_{n+1}(x, X_{n+1})]$$
$$\leqslant E[K_{n+1}^2(x, X_{n+1})]$$
$$=\int_{-\infty}^{\infty} K_{n+1}^2(x, y)\,d(F_1(y)+F_2(y))$$
$$=\int_{-\infty}^{\infty} K_{n+1}^2(x, y)f(y)dy+\sum_i K_{n+1}^2(x, x_i)S_i.$$

In view of (K1), (K2), (K3), (K4) and Lemma 3. 1, we have

$$(3.18)\quad \lim_{n\to\infty}\int_{-\infty}^{\infty} h_{n+1}^{-1} K^2(h_{n+1}^{-1}(x-y))f(y)dy$$
$$=f(x)\int_{-\infty}^{\infty} K^2(y)dy.$$

Therefore, there exists a constant $C_1(x)>0$ such that

$$\int_{-\infty}^{\infty} h_{n+1}^{-1} K^2(h_{n+1}^{-1}(x-y))f(y)dy\leqslant C_1(x) \qquad \text{for all } n\geqslant 1.$$

Thus, we obtain

$$(3.19)\quad \int_{-\infty}^{\infty} K_{n+1}^2(x, y)f(y)dy\leqslant C_1(x)\cdot h_{n+1}^{-1} \qquad \text{for all } n\geqslant 1.$$

The second term of the last equation in (3. 17) is evaluated as follows.

$$\sum_i K_{n+1}^2(x, x_i)S_i$$
$$=\sum_i h_{n+1}^{-2}\cdot K^2(h_{n+1}^{-1}(x-x_i))S_i$$
$$\leqslant \sup_{-\infty<y<\infty} K(y)\cdot h_{n+1}^{-1}$$
$$\times\sum_i h_{n+1}^{-1}|x-x_i|K(h_{n+1}^{-1}(x-x_i))S_i/|x-x_i|,$$

where since x and $x_i(i=0, \pm 1, \pm 2, \cdots\cdots)$ are points of continuity and discontinuity of the distribution $F(x)$, respectively, $|x-x_i|\neq 0$ for all i. Since $|y|K(y)\longrightarrow 0$ as $|y|\longrightarrow\infty$, it follows that $|y|K(y)$ is bounded. Hence, $|y|K(y)\leqslant M(<\infty)$ for all y. Therefore,

$h_{n+1}^{-1}|x-x_i|K(h_{n+1}^{-1}(x-x_i))\leqslant M$ for all $n\geqslant 1$ and all i. Thus, we obtain

(3. 20) $$\sum_i K_{n+1}^2(x, x_i)S_i$$

$$\leqslant K_0 \cdot M \cdot (\sum_i S_i/|x-x_i|) \cdot h_{n+1}^{-1}.$$

Let us write $C_2(x)=C_1(x)+K_0 \cdot M \sum_i S_i/|x-x_i|$.

From (3. 5), it is easy to see that $0<C_2(x)<\infty$. Combining (3. 17), (3. 19) and (3. 20), we get

(3. 21) $$\mathrm{Var}\,[K_{n+1}(x, X_{n+1})] \leqslant C_2(x) \cdot h_{n+1}^{-1}.$$

In view of (3. 16) and (3. 21), we obtain

$$E[(f_{n+1}(x)-E[f_{n+1}(x)])^2/X_1, \cdots\cdots, X_n]$$
$$\leqslant (1-a_{n+1})(f_n(x)-E[f_n(x)])^2+C_2(x) \cdot a_{n+1}^2 \cdot h_{n+1}^{-1}$$

for all $n \geqslant 1$. Putting $U_n(x)=(f_n(x)-E[f_n(x)])^2$ and $V_n(x)=C_2(x) \cdot a_{n+1}^2 \cdot h_{n+1}^{-1}$, we have

(3. 22) $$E[U_{n+1}(x)/X_1, \cdots\cdots, X_n]$$
$$\leqslant (1-a_{n+1})U_n(x)+V_n(x) \qquad \text{a.s.} \qquad \text{for all } n \geqslant 1.$$

If $\sum_{n=1}^{\infty} a_n^2 \cdot h_n^{-1}<\infty$, then it holds that

(3. 23) $$\sum_{n=1}^{\infty} E[|V_n(x)|]<\infty.$$

In Lemma 2. 2, let F_n be a σ-field generated by $X_1, \cdots\cdots, X_n$ for each n. Combining (3. 22) and (3. 23) and using Lemma 2. 2, we have $\lim_{n\to\infty} U_n(x)=0$ with probability one and $\lim_{n\to\infty} E[U_n(x)]=0$, that is,

(3. 24) $\lim_{n\to\infty} |f_n(x)-E[f_n(x)]|=0$ with probability one and

(3. 25) $$\lim_{n\to\infty} E[(f_n(x)-E[f_n(x)])^2]=0,$$

provided $\sum_{n=1}^{\infty} a_n^2 \cdot h_n^{-1}<\infty$.

Taking expectations on both sides of (3. 22), we obtain

$$E[U_{n+1}(x)]$$
$$\leqslant (1-a_{n+1}) \cdot E[U_n(x)]+C_2(x)a_{n+1} \cdot a_{n+1}h_{n+1}^{-1}$$

for all $n \geqslant 1$. By using Lemma 2. 1, we get

(3. 26) $$\lim_{n\to\infty} E[(f_n(x)-E[f_n(x)])^2]=0,$$

provided $\lim_{n\to\infty} a_n h_n^{-1}=0$.

It is easy to see that

$$(3.27)\qquad |f_n(x)-f(x)|\leqslant |f_n(x)-E[f_n(x)]|+|E[f_n(x)]-f(x)|.$$

From (3. 12), we get

$$(3.28)\qquad \lim_{n\to\infty}(E[f_n(x)]-f(x))^2=0.$$

By (3. 27) and the inequality $(a+b)^2\leqslant 2(a^2+b^2)$, we obtain

$$(3.29)\qquad E[(f_n(x)-f(x))^2]$$
$$\leqslant 2\{E[(f_n(x)-E[f_n(x)])^2]+(E[f_n(x)]-f(x))^2\}.$$

Combining (3. 26), (3. 28) and (3. 29), we have

$$\lim_{n\to\infty}E[(f_n(x)-f(x))^2]=0,$$

provided $\lim_{n\to\infty}a_nh_n^{-1}=0$.

Thus, the first statement of the theorem is proved.

Now, we suppose that $\sum_{n=1}^{\infty}a_n^2\cdot h_n^{-1}<\infty$. From (3. 12), (3. 24) and (3. 27), we have

$$\lim_{n\to\infty}|f_n(x)-f(x)|=0 \qquad \text{with probability one.}$$

Combining (3. 25), (3. 28) and (3. 29), we get

$$\lim_{n\to\infty}E[(f_n(x)-f(x))^2]=0.$$

Thus, the second statement of the theorem is proved. Therefore, the proof of the theorem is completed.

The following theorem presents the rate of variance of $f_n(x)$.

THEOREM 2. Let x be an arbitrary point of continuity of $F(x)$ and also of $f(x)$. Suppose that the condition (3. 5) holds. If there exists a constant $\alpha_0>0$ such that

$$(3.30)\qquad (1-a_{n+1})\cdot a_nh_n^{-1}/a_{n+1}h_{n+1}^{-1}\leqslant 1-\alpha_0a_{n+1} \text{ for all } n\geqslant \text{ some } n_0,$$

then there exists a constant $C(x)>0$ such that

$$\mathrm{Var}\,[f_n(x)]\leqslant C(x)\cdot a_nh_n^{-1} \qquad \text{for all } n\geqslant 1.$$

PROOF. Proceeding in the same way as in the proof of Theorem 1, we have

$$(3.31)\qquad E[(f_{n+1}(x)-E[f_{n+1}(x)])^2]$$
$$\leqslant(1-a_{n+1})E[(f_n(x)-E[f_n(x)])^2]$$
$$+C_1(x)\cdot a_{n+1}\cdot a_{n+1}h_{n+1}^{-1} \qquad \text{for all } n\geqslant 1,$$

where $C_1(x)$ is some positive constant depending on x. In view of (3. 30), (3. 31) and Lemma 2. 1, we obtain that there exists a constant $C(x)>0$ such that

$$E[(f_n(x)-E[f_n(x)])^2] \leqslant C(x) \cdot a_n h_n^{-1} \qquad \text{for all } n \geqslant 1.$$

Thus, the proof of the theorem is completed.

4. Acknowledgement

The author expresses his heartly thanks to Professor K. TANAKA of Niigata University for his kind encouragements and advices.

References

[1] MURTHY, V. K. (1965). *Estimation of probability.* Ann. Math. Statist. 34 1027–1031.

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

[3] WATANABE, M. (1973). *On asymptotically optimal algorithms for pattern classification problems.* Bull. Math. Stat. 15, 3–4, 31–48.

[4] WATANABE, M. (1974). *On convergence of asymptotically optimal descriminant functions for pattern classification problems.* Bull. Math. Stat. 16, 1–2, 23–34.