STRONG CONSISTENCY OF A SEQUENTIAL ESTIMATOR OF A PROBABILITY DENSITY FUNCTION

Davies, H. I.
Department of Statistics, University of North Carolina

https://doi.org/10.5109/13071

出版情報：統計数理研究. 15 (3/4), pp.49-54, 1973-03. Research Association of Statistical Sciences
バージョン：
権利関係：

STRONG CONSISTENCY OF A SEQUENTIAL ESTIMATOR OF A PROBABILITY DENSITY FUNCTION

By

H. I. DAVIES[1*]

(Received April 5, 1972)

1. Introduction.

Consider a sequence $X_1, X_2, X_3 \cdots$ of independent identically distributed m-dimensional random vectors having a probability density function, $f(x)$. Van Ryzin (1969) has shown that under appropriate conditions, estimates of the form

$$f_n(x) = n^{-1} \sum_{j=1}^{n} K_n(x, X_j), \tag{1}$$

where

$$K_n(x, X_j) = h_n^{-m} K(h_n^{-1}(x - X_j)),$$

are strongly consistent. That is, if x is contained in the continuity set $C(f)$, then $f_n(x) \to f(x)$ with probability one. Here $K(u) = K(u_1, u_2, \cdots, u_m)$ is a real valued, Borel measurable function on R^m, where R^m is the m-fold convolution of the real line, such that

$$K(u) \text{ is a density on } R^m \tag{2}$$

$$\sup_{u \in R^m} |K(u)| < \infty \tag{3}$$

$$\|u\| K(u) \to 0 \quad \text{as} \quad \|u\|^2 = \sum_{i=1}^{m} u_i^2 \to \infty \tag{4}$$

and $\{h_n\}$ is a sequence of numbers such that

$$h_n > 0, \quad n = 1, 2, \cdots; \quad \lim_{n\to\infty} h_n = 0, \quad \text{and} \quad \lim_{n\to\infty} n h_n = \infty. \tag{5}$$

Yamato, (1970) considers a modified estimator of the form

$$\hat{f}(x) = n^{-1} \sum_{j=1}^{n} h_j^{-m} K(h_j^{-m}(x - X_j)) \tag{6}$$

and shows the weak consistency of this estimator. (That is $\lim_{n\to\infty} E|\hat{f}_n(x) - f(x)|^2 = 0$ at all points x). It is our purpose to show, using Van Ryzin's method that this modified estimator is also strongly consistent. However, because of the special form that this estimator takes, Van Ryzin's conditions can be somewhat weakened.

---

1 This research was supported by a C. S. I. R. O. postgraduate studentship. This research forms part of the author's doctoral dissertation.

* Department of Statistics, University of North Carolina.

2. Strong Consistency.

Pointwise: In order to show strong consistency we will use the following lemma which is proved in Van Ryzin (1969).

LEMMA 1. *Let* $\{Y_n\}$ *and* $\{Y_n'\}$ *be two sequences of random variables on a probability space* (Ω, F, P). *Let* $\{F_n\}$ *be a sequence of Borel fields,* $F_n \subset F_{n+1} \subset F$, *where* Y_n *and* Y_n' *are measurable with respect to* F_n. *If*

(i) $0 \leqq Y_n$ *a.e.*

(ii) $EY_1 < \infty$

(iii) $E[Y_{n+1} | F_n] \leqq Y_n + Y_n'$ *a.e.*

(iv) $\sum_{n=1}^{\infty} E|Y_n'| < \infty$

then Y_n *converges a.e. to a finite limit.*

THEOREM 1. *If* $K(u)$ *satisfies* (2)—(4) *and* $\{h_n\}$ *is a monotone decreasing sequence of numbers satisfying* (5) *and if in addition*

$$\sum_{n=1}^{\infty} n^{-2} h_n^{-m} < \infty \tag{7}$$

$$\lim_{n\to\infty} n^{-1} \sum_{j=1}^{n} h_j^{-m} h_n^m = \alpha, \qquad 0 \leqq \alpha \leqq 1 \tag{8}$$

then $\hat{f}_n(x) \to f(x)$ *with probability one if* $x \in C(f)$.

PROOF. $E\hat{f}_n(x) \to f(x)$ as $n \to \infty$ (Yamato (1970)). Thus, it is sufficient to show $\hat{f}_n(x) - E\hat{f}_n(x) \to 0$ with probability one as $n \to \infty$.

Let $Y_n = \{\hat{f}_n(x) - E\tilde{f}_n(x)\}^2$ and $F_n =$ Borel field generated by $X_1, X_2, \cdots, X_n$ and note that $\lim_{n\to\infty} E(Y_n) = 0$ (Yamato (1970)). Since,

$$\hat{f}_n(x) = n^{-1}(n-1)\hat{f}_{n-1}(x) + n^{-1} h_n^{-m} K(h_n^{-1}(x - X_n)), \tag{9}$$

it follows that

$$\begin{aligned} \hat{f}_{n+1}(x) - E\hat{f}_{n+1}(x) &= \hat{f}_n(x) - E\hat{f}_n(x) - (n+1)^{-1}\{\hat{f}_n(x) - E\hat{f}_n(x)\} \\ &\quad + (n+1)^{-1}\{h_{n+1}^{-m} K(h_{n+1}^{-1}(x - X_{n+1})) - h_{n+1}^{-m} E(K(h_{n+1}^{-1}(x - X_{n+1})))\}. \end{aligned}$$

Thus,

$$E[Y_{n+1} | F_n] = Y_n + (n+1)^{-2} Y_n - 2(n+1)^{-1} Y_n + (n+1)^{-2} h_{n+1}^{-2m} \operatorname{Var} K(h_{n+1}^{-1}(x - X)).$$

Let $Y_n' = (n+1)^{-2} Y_n - 2(n+1)^{-1} Y_n + (n+1)^{-2} h_{n+1}^{-2m} \operatorname{Var} K(h_{n+1}^{-1}(x - X))$.

In order to use Lemma 1 we must verify

$$\sum_{n=1}^{\infty} E|Y_n'| < \infty. \tag{10}$$

But,

$$\sum_{n=1}^{\infty} (n+1)^{-1} E|Y_n| = \sum_{n=1}^{\infty} (n+1)^{-1} \operatorname{Var} \hat{f}_n(x).$$

Now, Yamato (1970), using a monotone sequence satisfying (5) and (8) has shown (Theorem 3) that

$$\lim_{n\to\infty} n h_n^m \operatorname{Var} \hat{f}_n(x) = \alpha f(x) \int_{R^m} K^2(y) dy. \tag{11}$$

Thus $n^{-1}\operatorname{Var}\hat{f}_n(x)=0(n^{-2}h_n^{-m})$ for large n. Hence, $\sum_{n=1}^{\infty}(n+1)^{-1}\operatorname{Var}\hat{f}_n(x)<\infty$ using condition (7).

The fact that $\sum_{n=1}^{\infty}(n+1)^{-2}h_{n+1}^{-2m}\operatorname{Var}K(h_{n+1}^{-1}(x-X))<\infty$ is shown to follow from condition (7) by Van Ryzin (1969).

Thus $\sum_{n=1}^{\infty}E|Y_n'|<\infty$ and the result follows by using Lemma 1 and the fact that the mean square limit and the almost sure limit coincide with probability one.

Uniform:

Let $k(t)=k(t_1,\cdots,t_m)=\int e^{it'u}K(u)du$ where $t'u=\sum_{j=1}^{m}t_ju_j$. Also let $\phi_n(t)=n^{-1}\sum_{j=1}^{n}e^{it'X_j}$ and let $\phi(t)=Ee^{it'X}$.

THEOREM 2. *If $K(u)$ satisfies conditions* (2)—(4) *and $\{h_n\}$ is a monotone decreasing sequence of real numbers satisfying* (5) *and if both $h_nh_{n+1}^{-1}\to 1$ and $nh_n^{-2m}\to\infty$ as $n\to\infty$, and if*

(10)
$$\sum_{n=1}^{\infty}(nh_n^m)^{-2}<\infty$$

(11)
$$\sum_{n=1}^{\infty}\frac{1}{nh_n^{2m-1}}\left|\frac{1}{h_{n+1}}-\frac{1}{h_n}\right|<\infty$$

and if also

(12)
$$\int|k(t)|dt<\infty$$

where $|k(u)|$ is non-decreasing on $u<0$ and non-increasing on $u\geqq 0$, then if $f(x)$ is uniformly continuous on R^m, $\sup_x|\hat{f}_n(x)-f(x)|\to 0$ with probability one as $n\to\infty$.

PROOF. (All integrals are over R^m, unless otherwise specified.) We first prove $\sup_x|\hat{f}_n(x)E\hat{f}_n(x)|\to 0$ a.s. as $n\to\infty$. Since both $k(u)$ and $K(u)$ are in L_1, using the inversion theorem for a Fourier transform

$$\sup_x|\hat{f}_n(x)-E\hat{f}_n(x)|=\sup_x\left|(2\pi)^{-1}\int n^{-1}\sum_{j=1}^{n}[e^{iu'X_j}-\phi(u)]k(h_ju)e^{-iu'x}du\right|$$
$$\leqq(2\pi)^{-m}\int\left|n^{-1}\sum_{j=1}^{n}[e^{iu'X_j}-\phi(u)]k(h_ju)\right|du$$
$$=(2\pi)^{-m}\int\left|n^{-1}\sum_{j=1}^{n}[e^{iu'X_j}-\phi(u)]\frac{k(h_ju)}{k(h_nu)}\right||k(h_nu)|du\,.$$

Using Schwartz's inequality, this is less than

$$(2\pi)^{-m}\left\{\int\left|k(h_nu)|du\right\}^{\frac{1}{2}}\left\{\int\left|n^{-1}\sum_{j=1}^{n}[e^{iu'X_j}-\phi(u)]\frac{k(h_ju)}{k(h_nu)}\right|^2|k(h_nu)|du\right\}$$
$$=(2\pi)^{-m}\left\{\int\left|h_n^{-m}|k(u)|du\right\}^{\frac{1}{2}}\left\{\int\left|n^{-1}\sum_{j=1}^{n}[e^{iu'X_j}-\phi(u)]\frac{k(h_jn)}{k(h_nu)}\right|^2|k(h_nu)|du\right\}.$$

Since $\int|k(u)|du<\infty$ by assumption (12), we need only consider the term

$$\int h_n^{-m}\left|n^{-1}\sum_{j=1}^{n}[e^{iu'X_j}-\phi(u)]\frac{k(h_ju)}{k(h_nu)}\right|^2|k(h_nu)|du$$

which we will denote by Y_n. Taking expectations we get

$$EY_n = \int h_n^{-m} E|n^{-1}\sum_{j=1}^{n}[e^{iu'X_j}-\phi(u)]\frac{k(h_j u)}{k(h_n u)}\Big|^2 |k(h_n u)|\,du\,.$$

But,

$$E|e^{iu'X_j}-\phi(u)|^2 = E[e^{iu'X_j}-\phi(u)][e^{-iu'X_j}-\overline{\phi(u)}] = 1-|\phi(u)|^2$$

so that, we obtain

$$EY_n = n^{-2}h_n^{-m}\int(1-|\phi(u)|^2)\sum_{j=1}^{n}\left|\frac{k(h_j u)}{k(h_n u)}\right|^2 |k(h_n u)|\,du\,. \tag{13}$$

However, since $|k(u)|$ is non-decreasing on $u<0$ and non-increasing on $u\geqq 0$, we have

$$\left|\frac{k(h_j u)}{k(h_n u)}\right| \leqq 1\,, \qquad j\leqq n\,. \tag{14}$$

Thus, using inequality (14) in equation (13) we get

$$EY_n \leqq n^{-1}h_n^{-m}\int |k(h_n u)du| = n^{-1}h_n^{-2m}\int |k(u)du\,.$$

Condition (12) and the assumption $nh_n^{2m}\to\infty$ ensure that EY_n tends to zero as $n\to\infty$. We now prove Y_n converges with probability one. Write,

$$\begin{aligned} Y_{n+1} &= h_{n+1}^{-m}\int (n+1)^{-2}\left|\sum_{j=1}^{n+1}[e^{iu'X_j}-\phi(u)]\frac{k(h_j u)}{k(h_{n+1}u)}\right|^2 |k(h_{n+1}u)|\,du \\ &= (n+1)^{-2}h_{n+1}^{-m}\int\left|\sum_{j=1}^{n}[e^{iu'X_j}-\phi(u)]\frac{k(h_j u)}{k(h_n u)} + e^{iu'X_{n+1}}-\phi(u)\right|^2 |k(h_{n+1}u)|\,du\,. \end{aligned}$$

Let F_n be defined as in Theorem 1. We now expand the first term in the integral using the fact that something of the form $|z|^2 = z\bar{z}$ and that

$$E(e^{iu'X_{n+1}}-\phi(u)|F_n)=0 \quad\text{and}\quad E(e^{-iu'X_{n+1}}-\overline{\phi(u)}|F_n)=0\,.$$

Thus,

$$\begin{aligned} E[Y_{n+1}|F_n] &= (n+1)^{-2}h_{n+1}^{-m}\int\left|\sum_{j=1}^{n}[e^{iu'X_j}-\phi(u)]\frac{k(h_j u)}{k(h_n u)}\right|^2 |k(h_{n+1}u)|\,du \\ &\quad + h_{n+1}^{-m}(n+1)^{-2}\int(1-|\phi(u)|^2)|k(h_{n+1}u)|\,du\,. \end{aligned}$$

But $(1-|\phi(u)|^2)\leqq 1$ and $|k(u)|\leqq 1$ since we assume $K(u)$ is a density function. Thus,

$$\begin{aligned} E[Y_{n+1}|F_n] &\leqq n^{-2}h_{n+1}^{-m}\int\left|\sum_{j=1}^{n}[e^{iu'X_j}-\phi(u)]\frac{k(h_j u)}{k(h_n u)}\right|^2 |k(h_n u)|\,du \\ &\quad + n^{-2}h_{n+1}^{-2m}\int |k(u)|\,du\,. \end{aligned}$$

Adding and subtracting Y_n we obtain

$$\begin{aligned} E[Y_{n+1}|F_n] &\leqq Y_n - n^{-2}(h_n^{-m}-h_{n+1}^{-m})\int\left|\sum_{j=1}^{n}[e^{iu'X_j}-\phi(u)]\frac{k(h_j u)}{k(h_n u)}\right|^2 |k(h_n u)|\,du \\ &\quad + n^{-2}h^{-2m}\int |k(u)|\,du\,. \end{aligned}$$

Let,

$$V_n = (nh_{n+1}^{2m})^{-1}\int |k(u)|\,du \tag{15}$$

and

(16) $$U_n = n^{-2}(h_n^{-m} - h_{n+1}^{-m}) \int \left| \sum_{j=1}^{n} [e^{iu'X_j} - \phi(u)] \frac{k(h_j u)}{k(h_n u)} \right|^2 |k(h_n u)| du .$$

Note that V_n is not a random variable. Hence $EV_n = V_n$, and, using condition (10), we have $\sum_{n=1}^{\infty} E|V_n| = \sum_{n=1}^{\infty} |V_n| < \infty$. Thus, in order to use Lemma 1 we need to show now that $\sum_{n=1}^{\infty} E|U_n| < \infty$. From (16) we have, taking expectations,

(17) $$E|U_n| = n^{-2}|h_n^{-m} - h_{n+1}^{-m}| \int (1 - |\phi(u)|^2) \sum_{j=1}^{n} \left| \frac{k(h_j u)}{(kh_n u)} \right|^2 |k(h_n u)| du$$
$$\leqq n^{-2}|h_n^{-m} - h_{n+1}^{-m}| h_n^{-m} \int n |k(u)| du .$$

By using the fact that $(1-t^m) = \left(\sum_{j=1}^{m} t^{j-1}\right)(1-t)$ and $h_n h_{n+1}^{-1} \to 1$, Van Ryzin (1969) shows

$$|h_{n+1}^{-m} - h_n^{-m}| \sim m h_n^{m-1} |h_{n+1}^{-1} - h_n^{-1}| .$$

Thus the upper bound in (17) is asymptotically equivalent to $\frac{m}{n h_n^{2m-1}} |h_{n+1}^{-1} - h_n^{-1}| \times \int |k(u)| du$. Hence, using condition (11), $\sum_{n=1}^{\infty} E|U_n| < \infty$. Thus conditions (iii) and (iv) of Lemma 1 are satisfied, where $Y_n' = U_n + V_n$.

The conditions of Lemma 1 are satisfied for Y_n and so Y_n tends to zero almost everywhere. Thus,

$$\sup_x |\hat{f}_{n+1}(x) - E\hat{f}_{n+1}(x)| \to 0 \qquad \text{a. s.} \qquad \text{as} \quad n \to \infty .$$

Now,

$$\sup_x |E\hat{f}_n(x) - f(x)| = \sup_x \left| \int \left(n^{-2} \sum_{j=1}^{n} K_j(u) f(x-u) - n^{-1} \sum_{j=1}^{n} K_j(u) f(x) \right) du \right|$$
$$\leqq \sup_x \int n^{-1} \left| \sum_{j=1}^{n} K_j(u) \right| |f(x-u) - f(x)| du$$
$$\leqq \sup_x \int_{\|u\| \leqq \delta} n^{-1} \left| \sum_{j=1}^{n} K_j(u) \right| |f(x-u) - f(x)| du$$
$$+ \sup_x \int_{\|u\| > \delta} |f(x-u) - f(x)| n^{-1} \sum_{j=1}^{n} K_j(u) du$$
$$\leqq \sup_x \sup_{\|u\| \leqq \delta} |f(x-u) - f(x)| + 2 \sup_x n^{-1} f(x) \sum_{j=1}^{n} \int_{\|u\| > \delta/h_j} K(u) du$$

Since $f(x)$ is uniformly continuous, the first term can be made arbitrarily small by choosing δ sufficiently small. In the second term, notice that for this fixed δ

$$f(x) \int_{\|u\| \geqq \delta/h_n} K(u) du \to 0$$

as $n \to \infty$ at all points x. Hence the Cesàro sum approaches zero and the proof is complete.

Acknowledgement.

I wish to thank Professor E. J. Wegman for suggesting the problem and for his valuable guidance.

References

[1] VAN RYZIN, J. (1969), *"On strong consistency of density estimates"*, Ann. Math. Statist., **40**, 1765-1772.

[2] YAMATO, H. (1971), *"Sequential estimation of a continuous probability density function and mode"*, Bull. Math. Statist., 14, No. 3~4, 1-12.