第5卷 第2期
1985年4月

数学研究与评论
JOURNAL OF MATHEMATICAL
RESEARCH AND EXPOSITION

Vol.5 No.2
April 1985

一类核估计强一致收敛的必要条件*

柴根象

（四川大学）

§1. 引言和结果

设$X_1, X_2, \cdots, X_n$是来自R^1上的$d.f.\ F(x)$的$i.i.d.$样本，$\{h_n\}$是一串正数；$K(\cdot)$是$p.d.f.$，令

$$f_n(x)=\frac{1}{nh_n}\sum_{i=1}^{n}K\left(\frac{x-X_i}{h_n}\right),\quad x\in R^1. \tag{1}$$

当F的$p.d.f.\ f$存在时，f_n是f的一类重要估计，叫做核估计．关于f_n一致强收敛于f的问题在文献中有很多讨论，所得结果一无例外地要假定f在全直线上一致连续．1969年，E. P. Schuster在[1]中提出了反面的问题：如有函数g，使得

$$\sup_x|f_n(x)-g(x)|\to 0,\quad a.s. \tag{2}$$

那末g是否必须一致连续，且g是否就是F的$p.d.f.$？他在假定$K(\cdot)$为连续、有界变差的情况下解决了此问题．但这个假定排除了一大批有意义的核估计，例如Rosenblatt估计．陈希孺教授[2]讨论了Rosenblatt估计强一致收敛的必要条件，其结果有趣之处在于它对h_n的选择没有任何限制，并有力地说明了Schuster施加于核$K(\cdot)$的假定并非必要．

本文基于[1]、[2]的方法，讨论了具有有限个间断点的一类核估计．主要结果如下：

定理 设$p.d.$核$K(\cdot)$有下述形式

$$K(u)=\sum_{j=1}^{m}c_j\chi_{A_j}(u),\quad u\in R^1, \tag{3}$$

诸$c_j\geqslant 0$，A_j为形如$(a_j, b_j]$的两两不交的有限或无限半开区间；数串$\{h_n\}$满足

1°、$0<h_n\to 0$； (4)

2°、$nh_n^2/\log n\to\infty$． (5)

则若有函数g使得(2)成立，那末g必在R^1上一致连续；而且F必是绝对连续的，g就是F的$p.d.f.$

§2 若干引理

因$K(\cdot)$是$p.d.$，故可以认为诸A_j皆是有限区间．

引理1 设数串$\{h_n\}$合于(5)，则

$$\sup_x|f_n(x)-Ef_n(x)|\to 0,\quad a.s. \tag{6}$$

证 记$X_1, X_2, \cdots, X_n$的经验分布函数为F_n，则对任给$\varepsilon>0$，有

$P(\sup_x|f_n(x)-Ef_n(x)|>\varepsilon)$

* 1982年2月12日收到．

$$= P\left(\sup_x \left|\sum_{j=1}^m c_j\{(F_n(x-h_na_j)-F_n(x-h_nb_j))-(F(x-h_na_j)-F(x-h_nb_j))\}\right| \geqslant h_n\varepsilon\right)$$

$$\leqslant P\left(\sup_y |F_n(y)-F(y)| \geqslant \frac{h_n\cdot\varepsilon}{2\cdot\sum_{j=1}^m c_j}\right)$$

$$\leqslant C\exp\left(-nh_n^2\varepsilon^2/2\left(\sum_{j=1}^m c_j\right)^2\right)[3] \tag{7}$$

由 $\{h_n\}$ 合于(5)，即得(6)成立. ▮

引理 2　若数串 $\{h_n\}$ 合于(4)、(5)，且存在 g 使得(2)成立，则 F 在 R^1 上处处连续.

证　因 K 是 p. d.，故必有 $1\leqslant j_0\leqslant m$，使得

$$c_{j_0}\cdot(b_{j_0}-a_{j_0})>0, \tag{8}$$

而且不失一般性可设 $a_{j_0}<0<b_{j_0}$.(必要时，可作一平移变换).

由引理 1 可知，若有 g 使得(2)成立，则必有 $Ef_n(x)\to g(x), n\to\infty$，对每一 $x\in\mathrm{R}^1$.　(9)

若引理的结论不真，则必存在一点 x_0，使得 $\Delta=F(x_0+0)-F(x_0-0)>0$. 注意到

$$EK\left(\frac{x_0-X}{h_n}\right)=\sum_{j=1}^m c_j[F(x_0-h_na_j)-F(x_0-h_nb_j)]$$
$$\geqslant c_{j_0}[F(x_0-h_na_{j_0})-F(x_0-h_nb_{j_0})]$$
$$\to c_{j_0}\Delta>0,\quad n\to\infty.$$

从而 $Ef_n(x_0)=\frac{1}{h_n}EK\left(\frac{x_0-X}{h_n}\right)\to\infty$，自与(9)式成立矛盾. ▮

引理 3　在引理 2 的假定下，则 g 必在 R^1 上一致连续.

证　由引理 1 可知，当 $n\to\infty$ 时，$Ef_n(x)\to g(x)$，关于 $x\in\mathrm{R}^1$ 一致成立. 从而对任给 $\varepsilon>0$，存在 n 致

$$\sup_x |Ef_n(x)-g(x)|<\varepsilon/4, \tag{10}$$

固定此 n，由引理 2 知 F 在 R^1 上一致连续，于是存在 $\delta>0$，使得

$$|x-y|<\delta \Rightarrow |F(x)-F(y)|<\frac{1}{4\sum_{j=1}^m c_j}\cdot\varepsilon\cdot h_n. \tag{11}$$

因此，当 $|x-y|<\delta$，就有

$$|Ef_n(x)-Ef_n(y)|\leqslant \sum_{j=1}^m c_j\left\{\frac{|F(x-h_na_j)-F(y-h_na_j)|}{h_n}+\frac{|F(x-h_nb_j)-F(y-h_nb_j)|}{h_n}\right\}$$

$$\leqslant\left(\frac{\varepsilon}{4\sum_{j=1}^m c_j}+\frac{\varepsilon}{4\sum_{j=1}^m c_j}\right)\cdot\left(\sum_{j=1}^m c_j\right)=\varepsilon/2. \tag{12}$$

由(10)、(12)可得，当 $|x-y|<\delta$，恒有

$$|g(x)-g(y)|\leqslant|g(x)-Ef_n(x)|+|Ef_n(x)-Ef_n(y)|+|Ef_n(y)-g(y)|$$
$$<\frac{\varepsilon}{4}+\frac{\varepsilon}{2}+\frac{\varepsilon}{4}=\varepsilon.$$ ▮

引理 4　在引理 2 的假定下，令

$$A=\{x:\ x\in\mathrm{R}^1,\ F'(x)\text{不存在或 }F'(x)\neq g(x)\}, \tag{13}$$

则 $\lambda(A)=0$（λ 是 R^1 上的勒贝格测度）。

证　因 $\sup_x|Ef_n(x)-g(x)|\to 0$，故只须证使 $F'(x)$ 存在的 x 处有 $\lim_n Ef_n(x)=F'(x)$ 即可，但

$$Ef_n(x)=\frac{1}{h_n}\sum_{j=1}^{m}c_j[F(x-h_na_j)-F(x-h_nb_j)],\text{ 故而}$$

$$\begin{aligned}\lim_n Ef_n(x)&=\lim_n\sum_{j=1}^{m}c_j\left\{a_j\left(\frac{F(x-h_na_j)-F(x)}{h_na_j}\right)+b_j\left(\frac{F(x)-F(x-h_nb_j)}{h_nb_j}\right)\right\}\\&=\sum_{j=1}^{m}c_j(b_j-a_j)\cdot F'(x)\\&=F'(x).\end{aligned}$$

∎

依引理2，F 有下述分解：

$$F=F_{AC}+F_S,\tag{14}$$

其中，F_{AC}、F_S 分别为 F 的绝对连续及奇异部分。由引理4应有

$$F'_{AC}(x)=g(x),\quad a.\ e.\ [\lambda].\tag{15}$$

故有

$$F_{AC}(x)=F_{AC}(0)+\int_0^x g(t)dt,\tag{16}$$

但由引理 3，g 在 R^1 处处连续，从而

$$F'_{AC}(x)=g(x),\text{ 对每一}x\in R^1\text{成立.}\tag{17}$$

引理 5　在引理2的条件下，则在每一 $x\in R^1$ 处，F_S 有一导出数为 0。

证　因 $f_n(x)\xrightarrow{p}g(x)$，关于 $x\in R^1$ 一致成立，故对任一 $x_0\in R^1$，取 $x_n=x_0+h_na_{j_0}$，则由 g 的连续性易知

$$f_n(x_n)\xrightarrow{p}g(x_0),\quad n\to\infty.\tag{18}$$

但另一方面，依引理1有

$$f_n(x_n)-Ef_n(x_n)\xrightarrow{p}0,$$

于是

$$\lim_n Ef_n(x_n)=g(x_0).\tag{19}$$

$$\begin{aligned}Ef_n(x_n)=&\sum_{j=1}^{m}\left[\frac{F_{AC}(x_0+h_n(a_{j_0}-a_j))-F_{AC}(x_0-h_n(b_j-a_{j_0}))}{(b_j-a_j)h_n}\right]\cdot(b_j-a_j)\cdot c_j\\&+\sum_{j=1}^{m}\left[\frac{F_S(x_0+h_n(a_{j_0}-a_j))-F_S(x_0-h_n(b_j-a_{j_0}))}{(b_j-a_j)h_n}\right]\cdot(b_j-a_j)\cdot c_j,\end{aligned}$$

由(17)、(19)可知

$$\sum_{j=1}^{m}\lim_n\left[\frac{F_S(x_0+h_n(a_{j_0}-a_j))-F_S(x_0-h_n(b_j-a_{j_0}))}{(b_j-a_j)h_n}\right]\cdot(b_j-a_j)\cdot c_j=0,$$

注意到 j_0 的选取合于(8)，从而

$$\lim_n\left[\frac{F_S(x_0)-F_S(x_0-h_n(b_{j_0}-a_{j_0}))}{(b_{j_0}-a_{j_0})h_n}\right]=0.$$

∎

§3　定理的证明

令

$$\Phi(x)=F_S(x)+x,\quad x\in R^1,\tag{20}$$

则 Φ 连续、严增，且由引理 5 知 Φ 在每一 $x\in R^1$ 处有一导出数为1。于是有[4]

$\Phi(b)-\Phi(a)\leqslant b-a$，对每一$-\infty<a<b<\infty$成立，此即

$F_S(b)-F_S(a)\leqslant 0$，对每一$-\infty<a<b<\infty$成立，但F_S非降，只好$F_S\equiv$常数，可取此常数为0，即得$F_S\equiv 0$。于是$F=F_{AC}$。最后，由(17)有

$$F'(x)=g(x),\ \text{对每一}x\in\mathbf{R}^1\text{成立},\tag{21}$$

由是定理证毕。

注 对于形如(3)的非$p.\ d.$核$K(\cdot)$，只要

$$\infty>C\triangleq\int_{-\infty}^{\infty}K(u)du>0,$$

定理的第一个结论仍成立，而且在相差常数C的意义下，g为F的p. d. f. 证明的明显改动之处是(17)式改为

$$F'_{AC}(x)=\frac{1}{C}\cdot g(x),\ \text{对每一}\ x\in\mathbf{R}^1.\tag{22}$$

老师陈希孺教授、敖顾昌教授都曾看过本文的初稿，并提出许多宝贵意见，在此表示衷心的谢意。

参 考 文 献

[1] Schuster, E. P., Estimation of a probability density function and its derivatives, *Ann. Math, Statist.* (1969), PP, 1187—1195.

[2] 陈希孺，Rosenblatt估计一致收敛的必要条件，《数学研究与评论》第二卷（1982)第四期，85—92.

[3] Kiefer, J. and Wolfowitz, J. On the deviations of the empiric distribution function of vector chance variables, *Trans, Am, Math. Soc.*, 87, (1958), PP, 173～186.

[4] 复旦大学主编，实变数函数论与泛函分析概要，上海科学技术出版社（1963)，P. 187.

A Necessary Condition of Strong Uniform Convergence of One Class Kernel Estimates

Chai Gen-xiang

Abstract

The object of this article is to discuss the necessary condition of strong uniform convergence of the kernel estimation, when the kernel in the problem poses form as follow:

$$K(u)=\sum_{j=1}^{m}c_j\chi_{A_j}(u),\ \ u\in\mathbf{R}^1.$$

Let $X_1,X_2,\cdots X_n$ be i. i. d. samples from distribution function F, and f_n be kernel estimate with kernel K based on $X_1,X_2,\cdots X_n$.

If there is function g such that

$$\sup_x|f_n(x)-g(x)|\to 0,\ \ a.\ s.$$

then g is the uniformly continuous derivative of F.