1982年　第2卷　第3期
Vol.2, No.3. 1982

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

# 密度核估计的一致收敛速度*

陈希孺

（中国科技大学）

设$X_1,\cdots,X_n$为从具密度$f$的$m$维总体中抽出的独立随机样本．要由它去估计$f$．二十多年来，这个问题吸引了不少统计学家的注意．在提出的种种方法中，如下形式的所谓核估计

$$f_n(x)=f_n(X_1,\cdots,X_n;x)=\frac{1}{nh_n^m}\sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right) \tag{1}$$

得到最深入的研究，主要是在大样本方面．在(1)式中，$K$是一个选定的定义于$R^m$上的函数，称为核(函数)，$\{h_n\}$是一串选定的数，$0<h_n\to 0$．

尽管学者们对核估计作了大量的研究，但关于这种估计是否有什么优良性的问题，则所知极少．1972年，Farrell在[1]中研究了上面的问题：指定$R^m$中的一点$x$，例如$x=0$．假定未知的密度$f$在某个族C内．基于样本$X_1,\cdots,X_n$在C内去估计$f(0)$，最好的收敛速度能达到多少？他证明了：若$k$为自然数，$a>0$为常数，以$C_{ka}$记满足如下条件的密度函数$f$的族：

$$|\partial^k f(x)/\partial x_1^{k_1}\cdots\partial x_m^{k_m}|\leqslant a \tag{2}$$

对任何$x=(x_1,\cdots x_m)\in R^m$．非负整数$k_1,\cdots k_m$，$\sum_{i=1}^m k_i=k$．设$\xi_n(X_1,\cdots,X_n)$为$f(0)$的任一估计量，如果对某一串常数$\{a_n\}$有

$$\lim_{n\to\infty}\left\{\inf_{f\in C_{ka}} P_f(|\xi_n(X_1,\cdots,X_n)-f(0)|\leqslant a_n)\right\}=1$$

则必然有

$$\lim_{n\to\infty} a_n n^{k/(2k+m)}=\infty$$

这个结果可解释为，若要在族$C_{ka}$内去估计$f(0)$，则任何估计的收敛速度都不能比$O(n^{-k/(2k+m)})$高．另一方面，如Farrell在[1]中指出的，由Parzen[2]中的结果可推出：在$m=1$的情况，存在适当的核估计，达到这个速度．Farrell猜测这个事实对$m>1$也成立（这一猜测是正确的，我们最近的工作附带证明了这一点）．

Farrell的结果，在考虑估计密度在一个点处之值的情况下，确立了核估计在大样本意义下的一个最优性质．更有意义且更为困难的问题，是关于整个密度函数的估计，在一

* 1981年9月17日收到．

致收敛的意义下，可能达到的收敛速速．这个收敛速度显然不能高于 Farrell 结果中的界限 $O(n^{-k/(2k+m)})$．有兴趣的是：我们最近的工作证明了，即使在一致收敛的意义下，收敛速度的主要部分并未降低．确切地说，有：

**定理 1** 可找到适当的核估计 $f_n(x)=f_n(X_1,\cdots,X_n;x)$，使

$$\lim_{a\to\infty}\lim_{n\to\infty}\left\{\inf_{f\in C_{ka}}P_f\left(\sup_x|f_n(x)-f(x)|\leqslant a\sqrt{\log n}\ n^{-k/(2k+m)}\right)\right\}=1 \tag{3}$$

（3）式指出的收敛速度在主要部分上与可能的最佳速度一致．但是，其中的因子$\sqrt{\log n}$能否去掉或降低，则尚待进一步研究，这问题估计很困难．

在得出定理 1 的过程中，附带证明了以下的结果：

**定理 2** 可找到适当的核估计 $f_n$，使对任何密度函数 $f$，只要其一切 $k$ 阶混合偏导数都在 $R^m$ 上有界，则有

$$\lim_{n\to\infty}\left\{\frac{1}{a_n}(\log)^{-1/2}\ n^{k/(2k+m)}\sup_x|f(x)-f(x)|\right\}=0,\qquad a.s. \tag{4}$$

对任何常数列 $a_n\to\infty$．

本定理表示，$\sup\limits_x|f_n(x)-f(x)|$ 趋于 0 的速度，大体上可达到$n^{-k/(2k+m)}\sqrt{\log n}$ 的数量级．这个数量级的研究在核估计大样本理论中占有重要地位．1969 年，Schuster[3] 对 $m=1$ 的情况，在核 K 于 $R^1$ 连续有界变差及$\int_{-\infty}^{\infty}|xK(x)|dx<\infty$ 的条件下，证明了：对任给 $\varepsilon>0$ 适当选择 $h_n$，可达到

$$\sup_x|f_n(x)-f(x)|=O(n^{-1/4+\varepsilon}),\qquad a.s.$$

1977年，Singh [4]对 $m=1$ 的情况．仍假定核 $K$ 在 $R^1$ 连续有界变差，并假定

$$\sup_x|Ef_n(x)-f(x)|=O(h_n^k), \tag{5}$$

证明了（$k$ 为自然数）：

$$\sup_x|f_n(x)-f(x)|=O(^{-k/(2k+2)}\sqrt{\log\log n})\qquad a.s. \tag{6}$$

应当注意：条件（5）是一个不便于验证的综合性条件．最容易设想的满足条件（5）的情况是：$f$ 的任何 $k$ 阶混合偏导数在 $R^m$ 上有界，而核函数 $K$ 满足：

$$\begin{cases}\int_{R^m}K(x)dx=1,\\ \int_{R^m}|K(x)|\,\|x\|^k dx<\infty,\\ \int_{R^m}K(x)\prod\limits_{i=1}^m x_i^{k_i}dx=0\end{cases} \tag{7}$$

对任何非负整数 $k_1,\cdots,k_m$，$0<\sum\limits_{i=1}^m k_i<k$，此处 $\|x\|^2=x_1^2+\cdots+x_m^2$．

以上结果都远未达到（4）式给出的数量级（即使限于 $m=1$ 的情况）．1975 年，Silverman 在[5]中，对 $m=1$，证明了：若核函数 $k$ 满足[8]中所规定的一大堆条件 $(C1)$和$(C2)$．且未知密度函数 $f$ 在 $R^1$ 一致连续，则有

$$\sup_x |Ef_n(x)-f(x)| = O((nh_n)^{-1/2}\sqrt{-\log|h_n|}),\quad a.s. \tag{8}$$

由(8)出发所能得出的关于$\sup_x|f_n(x)-f(x)|$的数量级，不超过由(4)给出的。为达到这样的数量级，还必须有(5)。而这在最常见的情况下需要核函数$K$满足条件(7)。但条件(7)与[5]中的条件($C1$)和($C2$)是否相容尚未可知，至少并不明显。

除核估计以外，受到较多注意的还有所谓"正交级数估计"。Schwarz 在[6]中，在准则

$$\mathrm{MISE}(f_n) = \int_{\mathrm{R}^m} E(f_n(x)-f(x))^2 dx$$

之下讨论了这种估计的收敛速度。他自己得出的结论是：与核估计比较，当维数$m$低时，对比对核估计有利；而当$m$上升时，则正交级数估计愈来愈有利，因此，正交级数估计适用于高维密度估计。然而，从 Farrell 和我们的结果显示，不论维数$m$大小如何，看不出正交级数估计有什么优于核估计的地方，且不谈计算问题。

处理$\sup_x|f_n(x)-f_n(x)|$的方法，一般总是从

$$\sup_x|f_n(x)-f(x)| \leqslant \sup_x|f_n(x)-Ef_n(x)| + \sup_x|Ef_n(x)-f(x)| \tag{9}$$

然后分别去处理(9)式右边的两项。第二项是非随机的，它的处理属于纯分析范围。拿本文所报导的工作来说，这一项的处理需要这样一个事实：存在只与$a, k, m$有关的常数$B$，使对任何$f\in C_{ka}$，$x\in\mathrm{R}^m$，及$0\leqslant l\leqslant k$，非负整数$l_1,\cdots,l_m$，$\sum_{i=1}^m l_i = l$，有$|f^{(l)}(l_1,\cdots,l_m;x)|\leqslant B$。这个事实的证明很复杂。(9)式右边第一项$\sup_x|f_n(x)-Ef_n(x)|$的处理，得力于[7]中最近得到的一个有力不等式(见[7]，(2.5)式)。看来，要改进$\sup_x|f_n(x)-f(x)|$的结果，和研究在种种情况下这个量的收敛速度问题，关键在于这种不等式的改进。

最后我们提到一下，我们的工作中提到的"适当的核估计"，就是使(3)、(4)成立的核估计，其核函数$K$满足条件(7)，以及

$$|\partial K(x)/\partial x_i| \leqslant M,\quad i=1,\cdots,m, \text{对任何 } x=(x_1,\cdots,x_m)\in\mathrm{R}^m. \tag{10}$$

这里$M$为一常数，不难指出满足条件(7)、(10)的核函数$K$的例子。

## 参考文献


[1] Farrell, R.H., On the best obtainable asymptotic rates of convergence in estimation of a density function at a point. *Ann. Math. Statist*, 43 (1972), pp. 170—180.

[2] Parzen, E., On the estimation of a probability density function and its mode. *Ann. Math. Statist*, 33 (1962), pp. 1065—1076.

[3] Schuster, E.F., Estimation of a probability density function and its derivatives, *Ann. Math. Statist*. 40 (1969), pp. 1187—1195.

[4] Singh, R.S., Improvement on some known nonparametric uniformly consistent estimators of derivatives of a density, *Ann. Statist*. 5 (1977), pp. 394—399.

[5] Silverman, B. W., Weak and strong consistency of the kernel estimate of a density and its derivatives. Ann, Statist, 3 (1975), pp. 177—184.

[6] Schwartz, S. C., Estimation of a probability density by an orthogonal series, *Ann, Math. Statist*. 38 (1967), pp. 1261—1265.

[7] Devroye, L. P. and Wagner, T. J., The strong uniform consistency of kernel density estimates, *Proc. Symp. Multi. Analysis* V (1980), pp. 59—77.

# Uniform Convergence Rates of Kernel Density Function Estimates

By Chen Xiru (陈希孺)


## Abstract

Suppose that $X_1, \cdots, X_n$ are samples drawn from a m-dimensional population with probability density function f belonging to a family $C_{k\alpha}$(where $k$ is a given positive integer. and $\alpha$ is a given positive number) defined as follows: $f \in C_{k\alpha}$ if and only if

$$\left| \frac{\partial^k f(x)}{\partial x_1^{k_1} \cdots \partial x_m^{k_m}} \right| \leqslant a$$

for any $x = (x_1, \cdots, x_m) \in R^m$, $k_i \geqslant 0$, $i = 1, \cdots, m$ and $\sum_1^m k_i = k$. In 1972, Farrell proved that for any estimator $\xi_n(X_1, \cdots, X_n)$ of $f(0)$, if for some sequence $\{a_n\}$ of positive constants we have

$$\lim_{n\to\infty}\left\{ \inf_{f \in C_{k\alpha}} P_f(|\xi_n(X_1, \cdots, X_n) - f(0)| \leqslant a_n) \right\} = 1$$

then $\lim\limits_{n\to\infty} a_n n^{k/(2k+m)} = \infty$

This result can roughly be interpreted as follows: Within the whole class $C_{k\alpha}$ the accuracy of any estimator (for estimating f(0)) can never exceed $O(n^{-k/(2k+m)})$. As pointed out by Farrell, in the case of $m = 1$ this accuracy can be reached by some properly chosen kernel estimate,

Stimulated by Farrell's work, we proved recently the following results:

**Theorem 1** There exists kernel estimate $f_n(x) = f_n(X_1, \cdots, X_n; x)$ such that

$$\lim_{a\to\infty}\lim_{n\to\infty}\left\{ \inf_{f \in C_{k\alpha}} P_f\left(\sup_x |f_n(x) - f(x)| \leqslant a\sqrt{\log n}\, n^{-k/(2k+m)}\right)\right\} = 1$$

**Theorem 2** There exists kernel estimate $f_n(x) = f_n(X_1, \cdots X_n; x)$ such that for any density function $f$ with all its k-th order partial derivatives bounded on $R^m$, we have

$$\lim_{n\to\infty}\left\{ \frac{1}{a_n} \frac{1}{\sqrt{\log n}} n^{k/(2k+m)} \sup_x |f_n(x) - f(x)| \right\} = 0, \quad a.s.$$

for any $a_n \to \infty$.

From this and Farrell's result we see that, for the main part, the best rate attainable for estimating the whole density function $f$ is the same as that of estimating the function $f$ at a fixed point, i. e., $O(n^{-k/(2k+m)})$. As a by-product, we proved that in the case of estimating $f(0)$, the accuracy $O(n^{-k/(2k+m)})$ can be reached by some property chosen kernel estimate even when $m > 1$, as conjectured by Farrell.

We also note that the a. s. uniform convergence rate given by Theorem 2 improves substantially various previous results in this respect.