第33卷 第3期
2013年8月
桂林理工大学学报
Journal of Guilin University of Technology
Vol. 33 No. 3
Aug. 2013

文章编号: 1674－9057(2013)03－0565－04 doi:10.3969/j.issn.1674－9057.2013.03.030

ND样本下密度核估计的相合性

苏红柳，唐国强，孙国华

（桂林理工大学 a. 理学院；b. 广西空间信息与测绘重点实验室，广西 桂林 541004）

摘　要: 设 $\{X_n, n \geqslant 1\}$ 是同分布的ND随机变量序列，$f(x)$ 为其概率密度函数，基于 $X_1, X_2, \cdots, X_n$，对密度函数 $f(x)$ 的核估计进行了讨论，根据ND序列的性质和不等式，在适当的条件下证明了其强相合和 r 阶矩相合。

关键词: ND样本；密度函数核估计；相合性

中图分类号: O212.7　　　　**文献标志码:** A

1 引言与引理

定义　称随机变量 $X_1, X_2, \cdots, X_n, n \geqslant 2$ 是ND(negatively dependent)的，若对任意的 $\forall x_1, x_2, \cdots, x_n \in \mathbf{R}$，都有

$$P(X_1 < x_1, X_2 < x_2, \cdots, X_n < x_n,) \leqslant \prod_{i=1}^{n} P(X_i < x_i),$$

及

$$P(X_1 \geqslant x_1, X_2 \geqslant x_2, \cdots, X_n \geqslant x_n,) \leqslant \prod_{i=1}^{n} P(X_i \geqslant x_i),$$

称随机变量列 $\{X_n, n \geqslant 1\}$ 是ND列，若对任意的 $X_1, X_2, \cdots, X_n, n \geqslant 2$ 是ND的。

ND序列的概念是1993年文献［1］给出的，ND相依序列比NA相依序列弱一些，这些负相依随机变量的概念在可靠性理论、渗透理论和某些多元统计分析中均有广泛的应用，因而负相依随机变量序列的概念引起了越来越多人的广泛兴趣。近年来，NA极限理论的研究发展十分迅速，文献［2－5］研究NA列的密度函数核估计的相合性、完全收敛性、强大数定律、不等式等。而关于ND的文献相对有限，但仍引人入胜：孟兵等[6]对ND阵列加权乘积和的完全收敛性的研究；陈晓林等[7]也获得了ND随机变量列的指数不等式；季洁鸥等[8]获得了同分布ND序列加权和的强大数定律等。然而对于ND序列密度函数核估计相合性研究较少，因此，笔者在前人的基础上给出ND序列的概率密度函数核估计相合性的结论和证明。本文约定文中出现的 C 总表示一个常数，在不同的地方可以代表不同的值。

引理1[9]　设 $K(\cdot)$ 为定义在 $(-\infty, +\infty)$ 上的有界密度函数，则当 $\lim\limits_{n\to\infty} h_n = 0$ 时有 $\lim\limits_{n\to\infty} \mathrm{E}f_n(x) = f(x)\ (\forall x \in c(f))$，其中 $c(f)$ 为 f 的连续点集；又若 $f(x)$ 一致连续，则有

$$\lim_{n\to\infty} \sup \left| \mathrm{E}f_n(x) - f(x) \right| = 0。$$

收稿日期: 2012－10－16

基金项目: 国家自然科学基金项目（11061012）；广西空间信息与测绘重点实验室基金项目（1103108－20）；桂林市科技局基金项目（20110120－5）

作者简介: 苏红柳（1986—），女，硕士研究生，研究方向：应用统计，175168370@qq.com。

通讯作者: 唐国强，副教授，tanggq@glut.edu.cn。

引文格式: 苏红柳，唐国强，孙国华. ND样本下密度核估计的相合性［J］. 桂林理工大学学报，2013，33（3）：565－568.

引理 2[1] 设$\{X_n, n \geqslant 1\}$是 ND 随机变量列,$f_n(x)$关于x为单调增(减)函数列,则$\{f_n(X_n); n \geqslant 1\}$仍为 ND 随机变量列。

引理 3[8] 令$\{X_n, n \geqslant 1\}$为一列满足$\delta_k^2 = EX_k^2 < \infty$,$EX_k = 0(k = 1,2,\cdots,n)$的 ND 随机变量,且存在$H > 0$,整数$m \geqslant 2$,使得$|EX_k^m| \leqslant \dfrac{m!}{2} \cdot \delta_k^2 \cdot H^{m-2}(k = 1,2,\cdots,n)$,则

$$P\Big(\sum_{i=1}^n X_i > x\Big) \leqslant \exp\Big(-x^2/(4\sum_{i=1}^n \delta_i^2)\Big), 0 \leqslant x \leqslant \sum_{i=1}^n \frac{\delta_i^2}{H}。$$

引理 4[7] 设$\{X_n, n \geqslant 1\}$是 ND 序列,$EX_n = 0$,$EX_n^r < \infty$,$S_j(k) = \sum\limits_{i=j+1}^{j+k} X_i$,$j \geqslant 0$,则有

当$r \leqslant 2$时,有

$$E|S_n|^r \leqslant C_r \sum_{i=1}^n E|X_i|^r;$$

当$r > 2$时,有

$$E|S_n|^r \leqslant C_r\Big\{\sum_{i=1}^n E|X_i|^r + \Big(\sum_{i=1}^n EX_i^2\Big)^{r/2}\Big\}。$$

2 结果及其证明

定理 1 (逐点强相合性)设$\{X_n, n \geqslant 1\}$为同分布的 ND 的随机变量列,有共同的密度$f(x)$,设$K(u)$为有界变差密度函数,如果$EX_k = 0$,$\lim\limits_{n\to\infty} h_n = 0$,$\lim\limits_{n\to\infty} \dfrac{nh_n^2}{\log n} = +\infty$,则

$$f_n(x) - f(x) \to 0, \text{a. s. }, \forall x \in c(f)。$$

证明 由于

$$\Big|f_n(x) - f(x)\Big| \leqslant \Big|f_n(x) - Ef_n(x)\Big| + \Big|Ef_n(x) - f(x)\Big| = I_1 + I_2, \tag{1}$$

由引理 1 知$\lim\limits_{n\to\infty} Ef_n(x) = f(x)$,故

$$I_2 = \Big|Ef_n(x) - f(x)\Big| \to 0,\ n \to \infty。 \tag{2}$$

下面只需证明$I_1 \to 0$,a. s. ,由于$K(x)$为$(-\infty, +\infty)$上的有界变差函数,因此存在单调有界变差函数$K_1(x)$和$K_2(x)$,使得$K(x) = K_1(x) - K_2(x)$。所以

$$\begin{aligned} f_n(x) - Ef_n(x) &= \frac{1}{nh_n}\sum_{j=1}^n\Big[K\Big(\frac{x-X_j}{h_n}\Big) - EK\Big(\frac{x-X_j}{h_n}\Big)\Big] \\ &= \frac{1}{nh_n}\Big\{\sum_{j=1}^n\Big[K_1\Big(\frac{x-X_j}{h_n}\Big) - EK_1\Big(\frac{x-X_j}{h_n}\Big)\Big] - \sum_{j=1}^n\Big[K_2\Big(\frac{x-X_j}{h_n}\Big) - EK_2\Big(\frac{x-X_j}{h_n}\Big)\Big]\Big\} \\ &= \frac{1}{nh_n}\Big[\sum_{j=i}^n Y_j - \sum_{j=1}^n Z_j\Big] = \frac{1}{nh_n}[S_{n_1} - S_{n_2}]。\end{aligned}$$

此处:

$$Y_j = K_1\Big(\frac{x-X_j}{h_n}\Big) - EK_1\Big(\frac{x-X_j}{h_n}\Big);\quad S_{n_1} = \sum_{j=1}^n Y_j;$$

$$Z_j = K_2\Big(\frac{x-X_j}{h_n}\Big) - EK_2\Big(\frac{x-X_j}{h_n}\Big);\quad S_{n_2} = \sum_{j=1}^n Z_j。$$

因为$K_1(x)$和$K_2(x)$均为有界单调变差函数,所以$Y_j = r(X_j)$和$Z_j = s(X_j)$分别为X_j的单调函数,由引理 2 可知Y_j、Z_j也是 ND 变量。所以,

$$P\Big(\Big|f_n(x) - Ef_n(x)\Big| \geqslant \varepsilon\Big) = P\Big(\Big|\frac{1}{nh_n}(S_{n_1} - S_{n_2}) \geqslant \varepsilon\Big|\Big) \leqslant P\Big(|S_{n_1}| \geqslant \frac{\varepsilon}{2}nh_n\Big) + P\Big(|S_{n_2}| \geqslant \frac{\varepsilon}{2}nh_n\Big)$$

$$= P\left(\left|\sum_{i=1}^{n} Y_i\right| \geqslant \frac{\varepsilon}{2} n h_n\right) + P\left(\left|\sum_{i=1}^{n} Z_i\right| \geqslant \frac{\varepsilon}{2} n h_n\right) = I_{11} + I_{12}。$$

又由有界性知，$\exists M > 0$，使得 $|K_1(x)| \leqslant M$，从而有

$$\delta^2 = \mathrm{E}Y_i^2 = \mathrm{E}\left[K_1\left(\frac{x - X_1}{h_n}\right)\right]^2 = \int_{-\infty}^{+\infty} K_1^2\left(\frac{x-y}{h_n}\right) f(y)\,\mathrm{d}y \leqslant M^2。$$

并且，又因为$\lim\limits_{n\to\infty} h_n = 0$，当取 $H = \dfrac{2M^2}{\varepsilon h_n}$ 时，有

$$|\mathrm{E}Y_i^m| = \mathrm{E}\left[K_1\left(\frac{x-X_1}{h_n}\right)\right]^m = \int_{-\infty}^{+\infty} K_1^m\left(\frac{x-y}{h_n}\right) f(y)\,\mathrm{d}y \leqslant M^m \leqslant \frac{m!}{2}\cdot\delta_k^2\cdot H^{m-2} = \frac{C_m\delta_i^2}{h_n^{m-2}}\cdot M^m。$$

因此，由定理 3，可得

$$I_{11} \leqslant \exp\left\{-\frac{\left(\frac{\varepsilon}{2}nh_n\right)^2}{4\sum\limits_{i=1}^{n}\delta_i^2}\right\} \leqslant \exp\left\{-\frac{Cn^2h_n^2}{nM^2}\right\} \leqslant \exp\{-Cnh_n^2\}。$$

因为$\lim\limits_{n\to\infty}\dfrac{nh_n^2}{\log n} = +\infty$，所以，当 n 充分大时，有 $nh_n^2 \geqslant \dfrac{2}{C}\log n$，因此$\sum\limits_{n=1}^{\infty} I_{11} < \infty$，同理$\sum\limits_{n=1}^{\infty} I_{12} < \infty$，所以

$$P\left(\left|f_n(x) - \mathrm{E}f_n(x)\right| \geqslant \varepsilon\right) < \infty。$$

由 Borel-Cantlelli 引理，得

$$I_1 = \left|f_n(x) - \mathrm{E}f_n(x)\right| \to 0, \text{a. s.}\ (n\to\infty)。\tag{3}$$

故由式(1) ~ (3)，得知：$f_n(x) - f(x) \to 0$，a. s.，证毕。

定理 2　设$\{X_n, n\geqslant 1\}$为同分布的 ND 的随机变量序列，设 $K(u)$ 是有界变差密度函数，$\lim\limits_{n\to\infty} h_n = 0$，$\lim\limits_{n\to\infty} nh_n^2 = +\infty$，则有

$$\lim_{n\to\infty}\mathrm{E}|f_n(x) - f(x)|^r = 0, r\in(0, +\infty)$$

成立。

证明　由 C_r - 不等式，得

$$\mathrm{E}|f_n(x) - f(x)|^r \leqslant C_1\mathrm{E}|f_n(x) - \mathrm{E}f_n(x)|^r + C_2|\mathrm{E}f_n(x) - f(x)|^r = J_1 + J_2。$$

由式(2)知，$J_2 \to 0$，下面只需证明 $J_1 \to 0$。

这里仍沿用定理 1 证明中的记号，如 $K_1(\,)$，$K_2(\,)$，S_{n_1}，S_{n_2}，Y_j，Z_j 等记号，现分 $0 < r \leqslant 2$ 和 $r > 2$ 两种情况讨论。

(1) 当 $0 < r \leqslant 2$ 时，由定理 1 的证明、C_r - 不等式和引理 4，得

$$\mathrm{E}[f_n(x) - \mathrm{E}f_n(x)]^2 = \frac{1}{n^2h_n^2}\mathrm{E}[S_{n_1} - S_{n_2}]^2 \leqslant \frac{C}{n^2h_n^2}(\mathrm{E}S_{n_1}^2 + \mathrm{E}S_{n_2}^2) \leqslant \frac{C}{n^2h_n^2}\left\{\sum_{i=1}^{n}\mathrm{E}Y_i^2 + \sum_{i=1}^{n}\mathrm{E}Z_i^2\right\}$$

$$\leqslant \frac{C}{nh_n^2}\{\mathrm{E}Y_i^2 + \mathrm{E}Z_i^2\} \leqslant \frac{C}{nh_n^2},$$

此即 Jessen 不等式，得

$$J_1 = \mathrm{E}|f_n(x) - \mathrm{E}f_n(x)|^r \leqslant \{\mathrm{E}[f_n(x) - \mathrm{E}f_n(x)]^2\}^{r/2} \leqslant C\left(\frac{1}{nh_n^2}\right)^{r/2}。$$

又因为由条件知道 $nh_n^2 \to \infty$，则 $J_1 \to 0$。

(2) 当 $r > 2$ 时，由 C_r - 不等式可知，

$$J_1 = \mathrm{E}|f_n(x) - \mathrm{E}f_n(x)|^r = \frac{1}{n^rh_n^r}\mathrm{E}[S_{n_1} - S_{n_2}]^r \leqslant \frac{C}{n^rh_n^r}(\mathrm{E}|S_{n_1}|^r + \mathrm{E}|S_{n_2}|^r)。\tag{4}$$

由引理 4 可得

$$\mathrm{E}|S_{n_1}|^r \leqslant C_r\{\sum_{i=1}^{n}\mathrm{E}|Y_i|^r + (\sum_{i=1}^{n}\mathrm{E}Y_i^2)^{r/2}\} \leqslant C_r(n + n^{r/2}) \leqslant Mn^{r/2}。 \tag{5}$$

$$\mathrm{E}|S_{n_2}|^r \leqslant C_r\{\sum_{i=1}^{n}\mathrm{E}|Z_i|^r + (\sum_{i=1}^{n}\mathrm{E}Z_i^2)^{r/2}\} \leqslant C_r(n + n^{r/2}) \leqslant Mn^{r/2}。 \tag{6}$$

联合式(4)~(6),有:

$$J_1 \leqslant \frac{C}{n^r h_n^r}[n^{r/2} + n^{r/2}] \leqslant \frac{C}{n^{r/2} h_n^r} = C\left(\frac{1}{nh_n^2}\right)^{r/2} \to 0。$$

即 $J_1 \to 0$,定理 2 证毕。

参考文献:

[1] Bozorgnia A, Patterson R F, Taylor R L. Limit theorems for ND r. v. , s [R]. Athens: University of Georgia, 1993.
[2] 韦来生. NA 样本概率密度函数核估计的相合性 [J]. 系统科学与数学, 2001, 21 (1): 79 - 87.
[3] 文志成, 杨善朝. NA、PA 样本下密度核估计相合性 [J]. 数学研究, 2002, 35 (3): 309 - 319.
[4] 吴群英. 两两 NQD 列的收敛性质 [J]. 数学学报, 2002, 45 (3): 617 - 624.
[5] 王岳宝, 苏淳. 关于两两 NQD 列的若干极限性质 [J]. 应用数学报, 1998, 21 (3): 404 - 414.
[6] 孟兵, 吴群英. ND 阵列加权乘积和的完全收敛性 [J]. 纯粹数学和应用数学, 2010, 26 (2): 84 - 90.
[7] 陈晓林, 吴群英, 周德宏. ND 随机变量列的指数不等式 [J]. 浙江大学学报: 理学版, 2011, 38 (1): 31 - 33.
[8] 季洁鸥, 林正炎. 同分布 ND 序列加权和的强大数定律 [J]. 浙江大学学报: 理学版, 2007, 34 (5): 499 - 504.
[9] 陈希孺, 陶波, 方兆本, 等. 非参数统计 [M]. 北京: 科学技术出版社, 1989.

Consistencies for Kernel-Type Density Estimation in ND Samples

SU Hong-liu, TANG Guo-qiang, SUN Guo-hua
(a. *College of Science*; b. *Guangxi Key Laboratory of Spatial Information and Gemomatics*, *Guilin Univercity of Technology*, *Guilin* 541004 , *China*)

Abstract: Let $\{X_n, n \geqslant 1\}$ be a sequence of identically distributed and negatively dependent random variables with probability density function $f(x)$. Based on ND samples, the kernel-type density estimation is studied in this paper, According to the nature and the inequality of ND, strong consistency and the moment consistency are proved under suitable conditions.

Key words: ND samples; kernel-type density estimation; strong consistency; moment consistence