应 用 数 学
MATHEMATICA APPLICATA
2019, 32(4): 865-871

NSD样本最近邻密度估计的强相合性

聂彩玲[1], 李永明[2,3], 应锐[4]
(1. 南昌大学理学院, 江西 南昌 330031;
2. 上饶师范学院数学与计算机科学学院, 江西 上饶 334001;
3. 上海财经大学统计与管理学院, 上海 200433; 4. 上饶师范学校, 江西 上饶 334001)

摘要: 本文研究负超可加相依样本的最近邻密度估计强相合性. 利用负超可加相依序列的不等式与性质, 获得最近邻密度估计的弱相合性、强相合性和一致强相合性.

关键词: 负超可加相依; 最近邻密度估计; 强相合性
中图分类号: O211.67 **AMS(2000)主题分类:** 62G05; 62G09; 62G20
文献标识码: A **文章编号:** 1001-9847(2019)04-0865-07

1. 引言

设总体X的分布密度函数为$f(x)$, $\{X_1,\cdots,X_n\}$是抽自该总体的负超可加相依(NSD)样本. 设$\{k_n,n\geq 1\}$为给定的正整数列, 满足$1\leq k_n\leq n$. 令$a_n(x)$为最小的正数a, 使得$[x-a,x+a]$中至少包含$X_1,X_2,\cdots,X_n$中的k_n个, 则密度函数$f(x)$的最近邻密度估计为

$$f_n(x)=\frac{k_n}{2na_n(x)}. \tag{1.1}$$

又设$F(x)$是密度函数$f(x)$的分布函数, 其对应的$F_n(x)$是样本$X_1,X_2,\cdots,X_n$的经验分布函数.

最近邻密度估计(nearest neighbor估计, 简记为NN估计)的概念是由Loftsgarden等[1]在1965年提出来的. 关于最近邻密度估计的性质, 在独立样本情形下已有许多研究结果[1−4]. 在相依样本情形下, 兰冲锋[5−6]研究了END样本最近邻密度估计的强相合速度、NQD样本最近邻密度估计的一致强相合速度, 曾翔[7]讨论了平稳φ-混合序列最近邻密度估计的相合速度, G.Boente等[8]讨论了φ-混合序列最近邻密度估计的大样本性质, 杨善朝[9]在NA下讨论了最近邻密度估计的相合性.

NSD随机变量的概念由胡太忠[10]引入, 他在文中举例说明了NSD变量不一定是NA变量. 之后, Christofides[11]证明了NA随机变量是NSD的. 鉴于NSD相依序列是NA序列的推广, 一些文献对NSD序列进行了研究. 如: 郑璐璐[12]研究了NSD随机变量加权和的强收敛性; 余云彩[13]研究了NSD序列加权和的中心极限定理及其在EV回归模型中的应用; WANG[14]讨论了NSD随机变量阵列的完全收敛性; SHEN[15]给出了NSD随机序列的Rosenthal型矩不等式的一些应用.

然而对于NSD序列样本最近邻密度估计大样本性质研究较少, 本文主要在NSD序列样本下讨论式(1.1)中给出的最近邻密度估计的相合性.

---

* **收稿日期:** 2018-10-02
基金项目: 国家自然科学基金资助项目(11461057, 11561010), 国家自然科学基金委重点项目(71331006)
作者介绍: 聂彩玲, 女, 汉族, 江西人, 研究方向: 非参数统计.
通讯作者: 李永明.

为行文方便, $C,\ C_1,\ C_2,\cdots$表示常数, 在不同地方取值可以相同也可以不同.

2. 引理与定理

定义2.1[10] 称函数$\phi:\mathbb{R}^n\to\mathbb{R}$为超可加函数, 如果对所有的向量$\mathbf{x},\mathbf{y}\in\mathbb{R}^n$, 满足

$$\phi(\mathbf{x}\vee\mathbf{y})+\phi(\mathbf{x}\wedge\mathbf{y})\geq\phi(\mathbf{x})+\phi(\mathbf{y}).$$

其中, $\vee$代表它们之间的最大值, $\wedge$代表它们之间的最小值.

定义2.2[10] 称随机变量$\{X_1,\cdots,X_n\}$为负超可加相依(NSD)随机变量, 如果存在相互独立的随机变量 $X_1^*,\cdots,X_n^*$, 使得对每个 i, X_i^* 与 X_i 同分布, 且

$$\mathrm{E}\phi(X_1,\cdots,X_n)\leq\mathrm{E}\phi(X_1^*,\cdots,X_n^*),$$

其中 ϕ 是超可加函数, 并且使得其期望存在.

引理2.1[10] 设$\{X_n,n\geq1\}$是NSD随机序列, $f_n(x)$关于x为非升(降)的连续函数, 则$\{f_n(X_n),n\geq1\}$仍是NSD随机序列.

引理2.2[16] 设$\{X_j,j\geq1\}$是NSD随机序列, $\mathrm{E}X_j=0,|X_j|\leq b_j$, a.s., 且$t\cdot\max\limits_{1\leq j\leq n}b_j\leq1$, 其中$t>0$.则对任意的$\varepsilon>0$,有

$$P\left(\left|\sum_{j=1}^nX_j\right|\geq\varepsilon\right)\leq2\exp(-t\varepsilon+t^2\sum_{j=1}^n\mathrm{E}X_j^2).$$

定理2.1 设$\{X_n,n\geq1\}$为同分布的NSD随机序列, 有共同的密度函数$f(x)$, 设k_n满足$k_n\to\infty,\frac{k_n}{n}\to0,n\to\infty$, 则对$f(x)$的任意连续点$x$, 有

(i) 当$\frac{k_n^2}{n}\to\infty$时, $f_n(x)\xrightarrow{P}f(x)$;

(ii) 当$\frac{k_n^2}{n\log n}\to\infty$时, $f_n(x)\xrightarrow{\text{a.s.}}f(x)$.

证 对任意的$\varepsilon>0$, 令$b_n(x)=\frac{k_n}{2n(f(x)+\varepsilon)}$, $c_n(x)=\frac{k_n}{2n(f(x)-\frac{\varepsilon}{2})}$, 则有

$$\begin{aligned}P(|f_n(x)-f(x)|>\varepsilon)&=P(f_n(x)-f(x)>\varepsilon)+P(f_n(x)-f(x)<-\varepsilon)\\&=P(f_n(x)>f(x)+\varepsilon)+P(f_n(x)<f(x)-\varepsilon)\\&=P\left(\frac{k_n}{2nf_n(x)}<\frac{k_n}{2n(f(x)+\varepsilon)}\right)+P\left(\frac{k_n}{2nf_n(x)}>\frac{k_n}{2n(f(x)-\varepsilon)}\right)\\&\leq P\left(a_n(x)<\frac{k_n}{2n(f(x)+\varepsilon)}\right)+P\left(a_n(x)>\frac{k_n}{2n(f(x)-\frac{\varepsilon}{2})}\right)\\&=P(a_n(x)<b_n(x))+P(a_n(x)>c_n(x))\hat{=}I_1+I_2.\end{aligned}$$

当$f(x)\leq\varepsilon$时, 由$f_n(x)$的非负性知, 事件$f_n(x)<f(x)-\varepsilon$是不可能事件, 其概率为零. 故对于$P(a_n(x)>c_n(x))$只考虑$f(x)>\varepsilon$时的情况.

令 $p_n=\int_{x-b_n(x)}^{x+b_n(x)}f(t)\mathrm{d}t$. 因为$x$是$f(x)$的连续点, 由$\frac{k_n}{n}\to0$得$b_n(x)\to0$, 故当$n$充分大时有

$$p_n\leq\frac{f(x)+\varepsilon}{f(x)+\frac{\varepsilon}{2}}f(x)2b_n(x)=\frac{f(x)+\varepsilon}{f(x)+\frac{\varepsilon}{2}}f(x)2\frac{k_n}{2n(f(x)+\varepsilon)}=\frac{k_n}{n}\frac{f(x)}{f(x)+\frac{\varepsilon}{2}}\leq\frac{k_n}{n}.\tag{2.1}$$

故

$$k_n-np_n\geq k_n-n\frac{k_n}{n}\frac{f(x)}{f(x)+\frac{\varepsilon}{2}}=k_n\frac{\varepsilon}{2f(x)+\varepsilon}>0.\tag{2.2}$$

记$\eta_i=I(X_i\leq x+b_n(x))$, $\zeta_i=I(X_i\leq x-b_n(x))$, $\xi_i=I(x-b_n(x)<X_i\leq x+b_n(x))$, $i=1,2,\cdots,n$. 则 $\xi_i=\eta_i-\zeta_i$,且$\mathrm{E}\eta_i-\mathrm{E}\zeta_i=p_n$, 由$a_n(x)$的定义, 有

$$I_1=P(a_n(x)<b_n(x))=P(\sum_{i=1}^n\xi_i\geq k_n)=P(\sum_{i=1}^n(\eta_i-\zeta_i)\geq k_n)$$

$$
\begin{aligned}
&= P\left(\left[\sum_{i=1}^{n}(\eta_i - \mathrm{E}\eta_i) - \sum_{i=1}^{n}(\zeta_i - \mathrm{E}\zeta_i)\right] \geq k_n - \sum_{i=1}^{n}(\mathrm{E}\eta_i - \mathrm{E}\zeta_i)\right) \\
&= P\left(\left[\sum_{i=1}^{n}(\eta_i - \mathrm{E}\eta_i) - \sum_{i=1}^{n}(\zeta_i - \mathrm{E}\zeta_i)\right] \geq k_n - np_n\right) \\
&\leq P\left(\left|\sum_{i=1}^{n}(\eta_i - \mathrm{E}\eta_i)\right| \geq \frac{k_n - np_n}{2}\right) + P\left(\left|\sum_{i=1}^{n}(\zeta_i - \mathrm{E}\zeta_i)\right| \geq \frac{k_n - np_n}{2}\right) \\
&\hat{=} I_{11} + I_{12}.
\end{aligned}
$$

令$Z_i = \eta_i - \mathrm{E}\eta_i, T_i = \zeta_i - \mathrm{E}\zeta_i$, 由引理2.1知$\{Z_i, 1 \leq i \leq n\}$和$\{T_i, 1 \leq i \leq n\}$仍为NSD随机序列. 且

$$
\mathrm{E}Z_i = \mathrm{E}T_i = 0, |Z_i| \leq 1, |T_i| \leq 1, \sum_{i=1}^{n}\mathrm{E}(Z_i)^2 \leq n, \sum_{i=1}^{n}\mathrm{E}(T_i)^2 \leq n.
$$

令 $t = \frac{k_n - np_n}{4n}$, 由引理2.2知

$$
\begin{aligned}
I_{11} &= P\left(\left|\sum_{i=1}^{n} Z_i\right| \geq \frac{k_n - np_n}{2}\right) \leq 2\exp\left\{-t\frac{k_n - np_n}{2} + t^2 n\right\} \\
&= 2\exp\left\{-\frac{(k_n - np_n)^2}{8n} + \frac{(k_n - np_n)^2}{16n}\right\} = 2\exp\left\{-\frac{(k_n - np_n)^2}{16n}\right\} \\
&\leq 2\exp\left\{-\frac{k_n^2}{n}\frac{\varepsilon^2}{16(2f(x)+\varepsilon)^2}\right\}.
\end{aligned}
$$

令 $C_1 = \frac{\varepsilon^2}{16(2f(x)+\varepsilon)^2}$, 则

$$
I_{11} \leq 2\mathrm{e}^{-C_1\frac{k_n^2}{n}}. \tag{2.3}
$$

同理可证

$$
I_{12} \leq 2\mathrm{e}^{-C_1\frac{k_n^2}{n}}. \tag{2.4}
$$

令 $q_n = \int_{x-c_n(x)}^{x+c_n(x)} f(t)\mathrm{d}t$, 类似于式(2.1)和(2.2),有

$$
q_n \geq \frac{f(x) - \frac{\varepsilon}{2}}{f(x) - \frac{\varepsilon}{4}} f(x) 2c_n(x) = \frac{k_n}{n}\frac{f(x)}{f(x) - \frac{\varepsilon}{4}} > \frac{k_n}{n}. \tag{2.5}
$$

$$
nq_n - k_n \geq n\frac{k_n}{n}\frac{f(x)}{f(x) - \frac{\varepsilon}{4}} - k_n = k_n\frac{\varepsilon}{4f(x) - \varepsilon} > 0. \tag{2.6}
$$

记 $\eta_i^{'} = I(X_i \leq x + c_n(x))$, $\zeta_i^{'} = I(X_i \leq x - c_n(x))$, $\xi_i^{'} = I(x - c_n(x) < X_i \leq x + c_n(x))$, $i = 1, 2, \cdots, n$. 则 $\xi_i^{'} = \eta_i^{'} - \zeta_i^{'}$, 且 $\mathrm{E}\eta_i^{'} - \mathrm{E}\zeta_i^{'} = q_n$, 由 $a_n(x)$ 的定义, 有

$$
\begin{aligned}
I_2 &= P(a_n(x) > c_n(x)) = P(\sum_{i=1}^{n}\xi_i^{'} \leq k_n) = P(\sum_{i=1}^{n}(\eta_i^{'} - \zeta_i^{'}) \leq k_n) \\
&= P\left(\left[\sum_{i=1}^{n}(\eta_i^{'} - \mathrm{E}\eta_i^{'}) - \sum_{i=1}^{n}(\zeta_i^{'} - \mathrm{E}\zeta_i^{'})\right] \leq k_n - \sum_{i=1}^{n}(\mathrm{E}\eta_i^{'} - \mathrm{E}\zeta_i^{'})\right) \\
&= P\left(\left[\sum_{i=1}^{n}(\eta_i^{'} - \mathrm{E}\eta_i^{'}) - \sum_{i=1}^{n}(\zeta_i^{'} - \mathrm{E}\zeta_i^{'})\right] \leq k_n - nq_n < 0\right) \\
&\leq P\left(\left|\sum_{i=1}^{n}(\eta_i^{'} - \mathrm{E}\eta_i^{'}) - \sum_{i=1}^{n}(\zeta_i^{'} - \mathrm{E}\zeta_i^{'})\right| \geq nq_n - k_n\right)
\end{aligned}
$$

$$\le P\left(\left|\sum_{i=1}^{n}(\eta_i^{'}-\mathrm{E}\eta_i^{'})\right|\ge\frac{nq_n-k_n}{2}\right)+P\left(\left|\sum_{i=1}^{n}(\zeta_i^{'}-\mathrm{E}\zeta_i^{'})\right|\ge\frac{nq_n-k_n}{2}\right)$$
$$\doteq I_{21}+I_{22}.$$

令$Z_i^{'}=\eta_i^{'}-\mathrm{E}\eta_i^{'}$, $T_i^{'}=\zeta_i^{'}-\mathrm{E}\zeta_i^{'}$, 由引理2.1知 $\{Z_i^{'},1\le i\le n\}$和$\{T_i^{'},1\le i\le n\}$ 仍为NSD随机序列. 且

$$\mathrm{E}Z_i^{'}=\mathrm{E}T_i^{'}=0,\ |Z_i^{'}|\le 1,\ |T_i^{'}|\le 1,\ \sum_{i=1}^{n}\mathrm{E}(Z_i^{'})^2\le n,\ \sum_{i=1}^{n}\mathrm{E}(T_i^{'})^2\le n.$$

令 $t=\frac{nq_n-k_n}{4n}$, 由引理2.2知

$$\begin{aligned}I_{21}&=P\left(\left|\sum_{i=1}^{n}Z_i^{'}\right|\ge\frac{nq_n-k_n}{2}\right)\le 2\exp\left\{-t\frac{nq_n-k_n}{2}+t^2n\right\}\\&=2\exp\left\{-\frac{(nq_n-k_n)^2}{8n}+\frac{(nq_n-k_n)^2}{16n}\right\}=2\exp\left\{-\frac{(nq_n-k_n)^2}{16n}\right\}\\&\le 2\exp\left\{-\frac{k_n^2}{n}\frac{\varepsilon^2}{16(4f(x)-\varepsilon)^2}\right\}.\end{aligned}$$

令 $C_2=\frac{\varepsilon^2}{16(4f(x)-\varepsilon)^2}$，则

$$I_{21}\le 2\mathrm{e}^{-C_2\frac{k_n^2}{n}}. \tag{2.7}$$

同理可证

$$I_{22}\le 2\mathrm{e}^{-C_2\frac{k_n^2}{n}}. \tag{2.8}$$

取常数 $C=\min\{C_1,C_2\}$, 由式(2.3), (2.4), (2.7)和(2.8), 可得

$$P(|f_n(x)-f(x)|>\varepsilon)\le 8\mathrm{e}^{-C\frac{k_n^2}{n}}.$$

又当 $n\to\infty$时, $\frac{k_n^2}{n}\to\infty$, 故

$$P(|f_n(x)-f(x)|>\varepsilon)\to 0,$$

即

$$f_n(x)\xrightarrow{P}f(x).$$

当 $\frac{k_n^2}{n\log n}\to\infty$时, 对充分大的$n$, 有 $\frac{k_n^2}{n}>\frac{2}{C}\log n$ 成立, 故

$$\sum_{i=1}^{n}P(|f_n(x)-f(x)|>\varepsilon)\le\sum_{i=1}^{n}8\mathrm{e}^{-C\frac{k_n^2}{n}}\le\sum_{i=1}^{n}8\mathrm{e}^{-C\frac{2}{C}\log n}=\sum_{i=1}^{n}\frac{8}{n^2}<\infty.$$

根据Borel-Cantelli引理,

$$P(|f_n(x)-f(x)|>\varepsilon,\mathrm{i.o.})=0,$$

即

$$f_n(x)\xrightarrow{\mathrm{a.s.}}f(x).$$

证毕.

注2.1 文[9]定理1中的(2)假设的是$\forall\lambda>0$, $\sum_{n=1}^{\infty}\exp(-\lambda k_n^2/n)<\infty$, 而本文定理2.1的(ii)假设为$\frac{k_n^2}{n\log n}\to\infty$, 得到了与NA样本情形下相同的结论.

引理2.3[9] 设$F(x)$是连续分布函数, 其经验分布函数为$F_n(x)$. 令$x_{n,k}$满足$F(x_{n,k})=k/n$, 其中$n\ge 3$, $k=1,2,\cdots,n-1$, 则

$$\sup_{-\infty<x<+\infty}|F_n(x)-F(x)|\le\max_{1\le k\le n-1}|F_n(x_{n,k})-F(x_{n,k})|+\frac{2}{n}.$$

定理2.2(一致强相合性)　设$\{X_n, n \geq 1\}$为同分布的NSD随机序列, 有共同的密度函数$f(x)$, $\sup\limits_x |f(x)| < \infty$且$f(x)$一致连续, 设$k_n \to \infty$, $\frac{k_n}{n} \to 0$, $\frac{k_n^2}{n \log n} \to \infty$, 则

$$\sup_x |f_n(x) - f(x)| \to 0, \quad \text{a.s.}$$

证

$$\begin{aligned} A_x &= \{|f_n(x) - f(x)| > \varepsilon\} \subset \{a_n(x) < b_n(x)\} \cup \{a_n(x) > c_n(x)\} \\ &\subset \left\{F_n(x + b_n(x)) - F_n(x - b_n(x)) \geq \frac{k_n}{n}\right\} \cup \left\{F_n(x + c_n(x)) - F_n(x - c_n(x)) \leq \frac{k_n}{n}\right\} \\ &\hat{=} A_{1x} \cup A_{2x}, \end{aligned} \tag{2.9}$$

其中$F_n(\cdot)$表示样本的经验分布函数.

由微分中值定理知, 存在$\theta_1 \in (x - b_n(x), x + b_n(x)), \theta_2 \in (x - c_n(x), x + c_n(x))$, 使得

$$F(x + b_n(x)) - F(x - b_n(x)) = 2b_n(x) f(\theta_1). \tag{2.10}$$

$$F(x + c_n(x)) - F(x - c_n(x)) = 2c_n(x) f(\theta_2). \tag{2.11}$$

因为$f(x)$一致连续, 故对于任意的$\varepsilon > 0$, 存在$\delta > 0$, 使得当$|x - y| < \delta$时，有

$$|f(x) - f(y)| < \frac{\varepsilon}{4}. \tag{2.12}$$

由$\frac{k_n}{n} \to 0$知, 存在正整数$N > 1$, 使得当$n > N$时, 有$\frac{k_n}{\varepsilon n} < \delta$, 从而有

$$b_n(x) = \frac{k_n}{2n(f(x) + \varepsilon)} \leq \frac{k_n}{2n\varepsilon} < \delta$$

关于x一致成立, 且由$f(x) > \varepsilon$,

$$c_n(x) = \frac{k_n}{2n(f(x) - \frac{\varepsilon}{2})} \leq \frac{k_n}{2n(\varepsilon - \frac{\varepsilon}{2})} < \delta$$

关于x一致成立.

因为$\theta_1 \in (x - b_n(x), x + b_n(x))$, $\theta_2 \in (x - c_n(x), x + c_n(x))$, 所以$|x - \theta_1| < \delta$, $|x - \theta_2| < \delta$. 故由式(2.12)得

$$|f(x) - f(\theta_1)| < \frac{\varepsilon}{4}, \ |f(x) - f(\theta_2)| < \frac{\varepsilon}{4}. \tag{2.13}$$

记$M = \sup\limits_x f(x) < \infty$. 由式(2.9)-(2.11)和(2.13)可得

$$\begin{aligned} &F_n(x + b_n(x)) - F_n(x - b_n(x)) - F(x + b_n(x)) + F(x - b_n(x)) \\ \geq &\frac{k_n}{n} - 2b_n(x) f(\theta_1) = \frac{k_n}{n} \frac{f(x) - f(\theta_1) + \varepsilon}{f(x) + \varepsilon} \geq \frac{k_n}{n} \frac{-\frac{\varepsilon}{4} + \varepsilon}{f(x) + \varepsilon} \geq \frac{k_n}{n} \frac{\varepsilon}{4(M + \varepsilon)}, \end{aligned}$$

以及

$$\begin{aligned} &F_n(x + c_n(x)) - F_n(x - c_n(x)) - F(x + c_n(x)) + F(x - c_n(x)) \\ \leq &\frac{k_n}{n} - 2c_n(x) f(\theta_2) = \frac{k_n}{n} \frac{f(x) - f(\theta_2) - \frac{\varepsilon}{2}}{f(x) - \frac{\varepsilon}{2}} \leq -\frac{k_n}{n} \frac{\frac{\varepsilon}{2} - \frac{\varepsilon}{4}}{f(x) - \frac{\varepsilon}{2}} \leq -\frac{k_n}{n} \frac{\varepsilon}{4(M + \varepsilon)}. \end{aligned}$$

故

$$\begin{aligned} A_{1x} &= \{F_n(x + b_n(x)) - F_n(x - b_n(x)) \geq \frac{k_n}{n}\} \\ &= \{F_n(x + b_n(x)) - F_n(x - b_n(x)) - F(x + b_n(x)) + F(x - b_n(x)) \geq \frac{k_n}{n} - 2b_n(x) f(\theta_1)\} \\ &\subset \{F_n(x + b_n(x)) - F_n(x - b_n(x)) - F(x + b_n(x)) + F(x - b_n(x)) \geq 2\frac{k_n}{n} u\} \end{aligned}$$

$$\subset \{|F_n(x+b_n(x))-F(x+b_n(x))| \geq \frac{k_n}{n}u\} \cup \{|F_n(x-b_n(x))-F(x-b_n(x))| \geq \frac{k_n}{n}u\}$$

$$\subset D.$$

$$\begin{aligned}A_{2x} &= \{F_n(x+c_n(x))-F_n(x-c_n(x)) \leq \frac{k_n}{n}\}\\ &= \{F_n(x+c_n(x))-F_n(x-c_n(x))-F(x+c_n(x))+F(x-c_n(x)) \leq \frac{k_n}{n}-2c_n(x)f(\theta_2)\}\\ &\subset \{F_n(x+c_n(x))-F_n(x-c_n(x))-F(x+c_n(x))+F(x-c_n(x)) \leq -2\frac{k_n}{n}u\}\\ &\subset \{|F_n(x+c_n(x))-F(x+c_n(x))| \geq \frac{k_n}{n}u\} \cup \{|F_n(x-c_n(x))-F(x-c_n(x))| \geq \frac{k_n}{n}u\}\\ &\subset D,\end{aligned}$$

其中

$$u = \frac{\varepsilon}{8(M+\varepsilon)},\ D = \{\sup_x |F_n(x)-F(x)| \geq \frac{k_n}{n}u\}.$$

因此

$$A_x \subset A_{1x} \cup A_{2x} \subset D.$$

由 $k_n \to \infty$, 可知 $\frac{2}{n} < \frac{k_n}{2n}u$. 故由引理2.3, 得

$$\begin{aligned}&P(\sup_x |f_n(x)-f(x)| > \varepsilon)\\ =&P(\bigcup_x A_x) \leq P(D) = P(\sup_x |F_n(x)-F(x)| \geq \frac{k_n}{n}u)\\ \leq&P(\max_{1\leq k\leq n-1} |F_n(x_{n,k})-F(x_{n,k})| + \frac{2}{n} \geq \frac{k_n}{n}u) \leq P(\max_{1\leq k\leq n-1} |F_n(x_{n,k})-F(x_{n,k})| \geq \frac{k_n}{2n}u)\\ \leq&\sum_{k=1}^{n-1} P(|F_n(x_{n,k})-F(x_{n,k})| \geq \frac{k_n}{2n}u) \leq \sum_{k=1}^{n-1} P\left\{\left|\sum_{i=1}^{n} X_i^{(nk)}\right| \geq \frac{k_n}{2}u\right\},\end{aligned}$$

这里 $X_i^{(nk)} = I(X_i < x_{n,k}) - \mathrm{E}I(X_i < x_{n,k}),\ 1 \leq i \leq n$. 由引理2.1可知$\{X_i^{(nk)},\ 1 \leq i \leq n\}$仍为NSD随机序列, 且

$$\mathrm{E}X_i^{(nk)} = 0,\ |X_i^{(nk)}| \leq 1,\ \sum_{i=1}^{n} \mathrm{E}(X_i^{(nk)})^2 \leq n.$$

令 $t = \frac{k_n}{4n}u$,当 $n \to \infty$时, 由 $\frac{k_n}{n} \to 0$, 可知 $t \to 0$, 故满足引理2.2的条件. 由 $\frac{k_n^2}{n\log n} \to \infty$, 可知当 n 足够大时, 有 $\frac{k_n^2}{n} > \frac{3\log n}{C}$. 因此

$$\begin{aligned}P(\sup_x |f_n(x)-f(x)| > \varepsilon) &\leq \sum_{k=1}^{n-1} P\left\{\left|\sum_{i=1}^{n} X_i^{(nk)}\right| \geq \frac{k_n}{2}u\right\}\\ &\leq \sum_{k=1}^{n-1} 2\exp\left\{-\frac{tk_n u}{2} + t^2 n\right\} = \sum_{k=1}^{n-1} 2\exp\left\{-\frac{k_n^2}{n}\frac{u^2}{16}\right\}\\ &\leq 2(n-1)\exp\{-3\log n\} \leq 2n^{-2}.\end{aligned}$$

因此

$$\sum_{n=1}^{\infty} P(\sup_x |f_n(x)-f(x)| > \varepsilon) \leq 2\sum_{n=1}^{\infty} n^{-2} < \infty.$$

由Borel-Cantelli引理得

$$\sup_x |f_n(x)-f(x)| \to 0, \quad \text{a.s.}$$

参考文献:

[1] LOFTSGAARDEN D O, QUESENBERRY C P. A nonparametric estimate of a multivariate density function[J]. Annals of Mathematical Statistics, 1965, 36(3): 1049-1051.

[2] MOORE D S, HENRICHON E G. Uniform consistency of some estimates of a density function[J]. Annals of Mathematical Statistics, 1969, 40(4): 1499-1502.

[3] 陈希孺. 最近邻密度估计的收敛速度[J]. 中国科学, 1981(12): 3-12.

[4] DEVROYE L P, WAGNER T J. The strong uniform consistency of nearest neighbor density sstimates[J]. Annals of Statistics, 1977, 5(3): 536-540.

[5] 兰冲锋, 吴群英. END样本最近邻密度估计的强相合速度[J]. 数学杂志, 2015, 35(3): 665-671.

[6] 兰冲锋. NQD样本最近邻密度估计的一致强相合速度[J]. 阜阳师范学院学报(自科版), 2013, 33(4): 5-8.

[7] 曾翔. 平稳(Φ)-混合序列最近邻密度估计的相合速度[J]. 桂林理工大学学报, 2013, 33(1): 174-177.

[8] BOENTE G, FRAIMAN R. Consistency of a nonparametric estimate of a density function for dependent variables[J]. Journal of Multivariate Analysis, 1988, 25(1): 90-99.

[9] 杨善朝. NA样本最近邻密度估计的相合性[J]. 应用数学学报, 2003, 26(3): 385-395.

[10] HU T Z. Negatively superadditive dependence of random variables with application [J]. Chinese Journal of Applied Probability and Statistics, 2000, 16(2): 133-144.

[11] CHRISTOFIDES T C, VAGGELATOU E. A connection between supermodular ordering and positive/ negative association [J]. Journal of Multivariate Analysis, 2014, 88: 138-151.

[12] 郑璐璐, 王嫱, 王学军. NSD随机变量加权和的强收敛性[J]. 中国科学技术大学学报, 2015(2): 101-106.

[13] 余云彩, 胡宏昌. NSD序列加权和的中心极限定理及其在EV回归模型中的应用[J]. 纯粹数学与应用数学, 2016, 32(5): 525-535.

[14] WANG X J, DENG X, ZHENG L L, et al. Complete convergence for arrays of rowwise negatively superadditive dependent random variables and its applications[J]. Statistics, 2014, 48: 834-850.

[15] SHEN A T, ZHANG Y, VOLODIN A. Applications of the Rosenthal-type inequality for negatively superadditive dependent random variables[J]. Metrika, 2015, 78(3): 295-311.

[16] SHEN A T, WANG X J. Kaplan-Meier estimator and hazard estimator for censored negatively superadditive dependent data[J]. Statistics, 2016, 50(2): 377-388.

Strong Consistency of Nearest Neighbor Estimator of Density Function for NSD Sequences

NIE Cailing [1], *LI Yongming* [2,3], *YING Rui* [4]

(1. College of Sciences, Nanchang University, Nangchang 330031, China; 2. School of Mathematics and Computer Science, Shangrao Normal University, Shangrao 334001, China; 3. School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai 200433, China; 4. Shangrao Normal School, Shangrao 334001,China)

Abstract: In this paper, we study strong consistency of the nearest neighbor density estimator for NSD samples. By using the inequality and property of NSD, we obtian the weak consistency, strong consistency and uniformly strong consistency of the nearest neighbor estimator for unknown probability density function.

Key words: NSD sample; Nearest neighbor density estimator; Strong consistency