数 学 年 刊
2016, 37A(4):463–472
DOI: 10.16205/j.cnki.cama.2016.0038

强混合样本下最近邻密度估计的渐近正态性*

秦永松[1]　雷庆祝[1]

提要 研究了 α-混合样本下最近邻密度估计的渐近性质，证明了估计的渐近正态性并且给出了其渐近方差的显式表达式，由此构造了 α-混合样本下概率密度的渐近置信区间.

关键词 α-混合样本，最近邻密度估计，渐近正态性
MR (2000) 主题分类 62G05, 62E20
中图法分类 O212.7
文献标志码 A
文章编号 1000-8314(2016)04-0463-10

1 引言

设总体 $X \in \mathbb{R}$ 有概率密度函数 (简记为 p.d.f.) f 且 $X_1, X_2, \cdots, X_n$ 为 X 的样本，对给定的 $x \in \mathbb{R}$, Loftsgarden 和 Quesenberry[1] 首次提出了 $f(x)$ 的如下形式的最近邻估计：

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

其中 $k = k_n$ 为 1 到 n 之间的整数且

$$a_n(x) = \min\{a \mid [x-a, x+a] \text{ 至少包含 } X_1, X_2, \cdots, X_n \text{ 的 } k_n \text{ 个样本点}\}.$$

非参数密度估计是考察数据结构的一个重要工具，而最近邻密度估计是被广泛应用的概率密度估计方法，自 $f_n(x)$ 问世以来，对其大样本性质的研究引起了学者的较大兴趣，在此列举一些相关研究成果：对于独立样本情形，Quesenberry[1], Wagner[2] 和 Chen[3] 分别证明了 $f_n(x)$ 的弱相合性、强相合性和渐近正态性；对于 α-混合样本情形， Liu 和 Zhang[4] 研究了 $f_n(x)$ 的渐近正态性，但没有给出渐近方差的显式表达式.

本文继续研究 α-混合样本情形 $f_n(x)$ 的渐近正态性，我们得到了渐近方差的显式表达式，由此构造了 α-混合样本下概率密度 $f(x)$ 的渐近置信区间.

在结束本节前，我们首先给出 α-混合序列的定义.

定义 1.1 设 $\{\eta_i, i \geqslant 1\}$ 为随机变量序列 (简记为 r.v.s.), $\mathcal{F}_s^t$ 为 $\{\eta_i, s \leqslant i \leqslant t\}$ $(s \leqslant t)$ 生成的 σ-代数. 如果当 $n \to \infty$ 时,

$$\alpha(n) = \sup_{k \geqslant 1} \sup_{\substack{A \in \mathcal{F}_1^k \\ B \in \mathcal{F}_{k+n}^\infty}} |P(AB) - P(A)P(B)| \to 0,$$

本文 2013 年 11 月 13 日收到， 2015 年 6 月 13 日收到修改稿.
[1]广西师范大学数学系，广西　桂林　541004. E-mail: ysqin@gxnu.edu.cn; qzlei@gxnu.edu.cn
*本文受到国家自然科学基金 (No. 11671102, No. 11271088, No. 11361011) 和广西省自然科学基金 (No. 2013GXNSF-AA019004, No. 2013GXNSFAA019007) 的资助.

则称随机变量序列 $\{\eta_i, i \geqslant 1\}$ 为强混合或者 α-混合随机变量序列[5], $\alpha(n)$ 称为 α-混合系数.

α-混合是常见的一些混合结构中最广泛的一类混合结构，它包含 ϕ, ρ 和 β-混合结构，混合结构的较全面的讨论见文 [6]. 另外， Chen 和 Tang[7] 给出了强混合序列在经济等领域的应用的一个较好的综述.

本文其余部分的结构如下：第 2 节给出了本文的主要结果，第 3 节给出了本文提出的 $f(x)$ 的渐近置信区间在有限样本下的效果的模拟结果，第 4 节给出了证明主要结果所需要的一些引理，第 5 节给出了主要结果的证明.

2 主要结果

为了得到 $f_n(x)$ 的渐近分布，我们需要如下正则条件.

假设 2.1 (A1) (i) 随机变量序列 $X_1, X_2, \cdots, X_n$ 为平稳序列且 f 为 X_1 的密度函数 (关于勒贝格测度).

(ii) $\{X_i, 1 \leqslant i \leqslant n\}$ 为 α-混合序列，其混合系数为 $\alpha(\cdot)$.

(iii) 对任意 $j \geqslant 1$, X_1, X_{j+1} 的联合密度函数 $f_{1,j}$ 存在且在 $(x,x)'$ 的某个邻域内满足 $|f_{1,j}(u,v) - f(u)f(v)| \leqslant C$.

(iv) $f(x) > 0$, 存在整数 $r \geqslant 2$, 使得 $f^{(r)}$ 在 x 的某个邻域内存在，且 $f^{(2)}(x) = f^{(3)}(x) = \cdots = f^{(r-1)}(x) = 0$ (当 $r = 2$ 时，条件 $f^{(2)}(x) = f^{(3)}(x) = \cdots = f^{(r-1)}(x) = 0$ 取消).

(A2) (i) $k \to \infty$ 且 $k = o\big(n^{\frac{2r}{2r+1}}\big)$.

(ii) 设 $p = p(n)$ 和 $q = q(n)$ 均为正整数， $p + q \leqslant n$, $\frac{q}{p} \to 0$, $\frac{pk}{n} \to 0$, $\frac{p^2}{k} \to 0$, $\big(\frac{n}{k}\big) \sum\limits_{s=q}^{\infty} \alpha(s) \to 0$ 且 $\widetilde{k}\alpha(q) \to 0$, 其中 $\widetilde{k} = \big[\frac{n}{p+q}\big]$, 此处 $[t]$ 表示 t 的整数部分.

注 2.1 由 (A2) 知， $\frac{p}{n} \to 0$, $\frac{p\widetilde{k}}{n} \to 1$ 且 $\big(\frac{n}{k}\big) \sum\limits_{s=p}^{\infty} \alpha(s) \to 0$, 这些结论将在后文中不加声明地引用.

注 2.2 条件 (A1)(iv) 和 (A2)(i) 在独立样本情形也被引用[3].

注 2.3 令 $p = [n^a]$, $q = [n^b]$, $k = [n^c]$, 其中 $0 < a, b, c < 1$. 设存在 $\tau_0 > 1$, 使得

$$\alpha(m) \leqslant C m^{-\tau_0},$$

则

$$\sum_{s=q}^{\infty} \alpha(s) \leqslant C q^{-(\tau_0 - 1)}.$$

易知下述条件 (2.1)–(2.2) 为 (A2) 的充分条件：

$$\max\{2a, 1 - b(\tau_0 - 1)\} < c < \min\Big\{\frac{2r}{2r+1}, 1 - a\Big\}, \tag{2.1}$$

$$a > \max\{b, 1 - b\tau_0\}, \quad b > 0. \tag{2.2}$$

在下述条件下可得 (2.1) 中的 c: a 和 b 满足 $b > \frac{1}{(\tau_0-1)(2r+1)}$ 且

$$\max\{b, 1-b\tau_0\} < a < \min\Big\{\frac{1}{3}, \frac{r}{2r+1}, b(\tau_0-1)\Big\}. \tag{2.3}$$

注意到 $r \geqslant 2$, 在下述条件下可得 (2.3) 中的 a:

$$\max\Big\{\frac{1}{(\tau_0-1)(2r+1)}, \frac{2}{3\tau_0}, \frac{1}{2\tau_0-1}\Big\} < b < \frac{1}{3}. \tag{2.4}$$

再次注意到 $r \geqslant 2$, 故在 $\tau_0 > 2$ 的条件下可得上式中的 b, 条件 $\tau_0 > 2$ 明显弱于被 Boenet 和 Fraiman[8] 及 Chen 和 Tang[7] 引用的几何 α-混合过程. 换言之，如果存在 $\tau_0 > 2$, 使得 $\alpha(m) \leqslant Cm^{-\tau_0}$ 满足，则可以适当选取 p, q 和 k, 使得条件 (A2) 满足.

现在给出本文的主要结果.

定理 2.1　设条件 (A1) 和 (A2) 满足，则

$$\sqrt{k}\frac{f_n(x)-f(x)}{f(x)} \xrightarrow{d} N(0,1). \tag{2.5}$$

注 2.4　在文 [4] 中，除条件 (A1) 中的 (i), (ii), (iv) 外，下述条件也被用到:

(A3) 存在 $\frac{1}{2} < a_0 < \frac{2r+1}{2r+2}$, 使得 $k = O(n^{a_0})$ 且存在 $\delta > 0$, 使得 $\sum\limits_{n=1}^{\infty} \alpha^{\frac{\delta}{2+\delta}}(n) < \infty$.

易知，条件 (A3) 较 (A2) 弱，但文 [4] 应用 α-混合过程的一个中心极限定理来证明 $f_n(x)$ 的渐近正态性, 故未给出渐近方差的显式表达式, 本文则利用分块技术来证明 $f_n(x)$ 的渐近正态性，此方法可较好地利用估计量的特性，由此得到了渐近方差的表达式.

对任意 $0 < \alpha < 1$, 设 u_α 满足

$$P(|N(0,1)| \leqslant u_\alpha) = 1-\alpha,$$

则由 (2.5) 知， $\theta = f(x)$ 的渐近覆盖概率为 $1-\alpha$ 的基于正态逼近的置信区间为

$$\Big[f_n(x) - \frac{u_\alpha f_n(x)}{\sqrt{k}}, f_n(x) + \frac{u_\alpha f_n(x)}{\sqrt{k}}\Big]. \tag{2.6}$$

3　模拟结果

本节通过模拟评估置信区间 (2.6) 在有限样本下的优劣，为此，我们选取如下两种 AR 模型:

$$X_t = \frac{1}{2}X_{t-1} + \epsilon_t \tag{3.1}$$

以及

$$X_t = \frac{3}{4}X_{t-1} - \frac{1}{8}X_{t-2} + \epsilon_t, \tag{3.2}$$

其中 $\{\epsilon_t\}$ 为独立同分布随机变量序列. 易知, 上述两种随机过程均为平稳过程, 模型 (3.1) 中 $X_1 \sim N\big(0, \frac{4}{3}\big)$, 模型 (3.2) 中 $X_1 \sim N\big(0, \frac{32}{17}\big)$.

我们从上述模型中产生样本容量分别为 $n = 100, 150, 200, 350, 500$ 和 $1,000$ 的随机样本 $\{X_i,\ i = 1, \cdots, n\}$，重复次数为 $2,000$ 次． k 的选取是一个至今没有很好解决的重要课题，有待未来进行深入研究．依据独立样本情形[3, 注2], k 可取为 $[n^{\frac{4}{5}}]$. 在我们的模拟

研究中，取 $k=[n^{\frac{3}{5}}]$ 和 $[n^{\frac{7}{10}}]$, 这样选取的 k 在 $[n^{\frac{4}{5}}]$ 附近但比 $[n^{\frac{4}{5}}]$ 小．我们发现，在这样的 k 之下，置信区间的效果比 $k=[n^{\frac{4}{5}}]$ 稍好．用 f 记 X_1 的概率密度，选取名义覆盖水平为 $1-\alpha=0.95$, 利用模拟样本，我们分别考察 $f(x)$ 在 $x=0$ 和 1 处的置信区间的覆盖概率 (简记为 CP) 和平均区间长度 (简记为 AL). 表 1 和表 2 分别给出了模型 (3.1) 和 (3.2) 下的模拟结果，同时，我们还模拟研究了独立同分布样本下的置信区间的情况，用于与 α-混合情况下的结果进行比较，这些模拟结果也在表 1–2 中.

从模拟结果可以看出，随着样本容量的增加，覆盖概率逐渐接近名义覆盖水平，且置信区间的平均覆盖长度逐步减小．另外，在样本容量较小时，独立样本情形的置信区间的效果好于 α-混合样本，但这种差异随着样本容量的增加而越来越不明显.

表 1 模型 (3.1) 下 $f(x)$ 的置信区间的覆盖概率 (简记为 CP) 和平均区间长度 (简记为 AL) (括号中的数字为独立同分布样本下的模拟结果)

		CP		AL	
x	n	$k=[n^{\frac{3}{5}}]$	$k=[n^{\frac{7}{10}}]$	$k=[n^{\frac{3}{5}}]$	$k=[n^{\frac{7}{10}}]$
$x=0$	100	0.9090	0.9146	0.3716	0.2771
		(0.9650)	(0.9684)	(0.3725)	(0.2772)
	150	0.9164	0.9256	0.3163	0.2398
		(0.9648)	(0.9664)	(0.3169)	(0.2402)
	200	0.9346	0.9360	0.2874	0.2169
		(0.9634)	(0.9628)	(0.2874)	(0.2175)
	250	0.9590	0.9576	0.2692	0.2002
		(0.9583)	(0.9568)	(0.2689)	(0.1998)
	350	0.9584	0.9570	0.2421	0.1764
		(0.9577)	(0.9564)	(0.2415)	(0.1762)
	500	0.9582	0.9548	0.2159	0.1553
		(0.9578)	(0.9547)	(0.2158)	(0.1549)
	1000	0.9520	0.9530	0.1728	0.1218
		(0.9510)	(0.9535)	(0.1726)	(0.1215)
$x=1$	100	0.9112	0.9236	0.2550	0.1917
		(0.9690)	(0.9734)	(0.2550)	(0.1914)
	150	0.9360	0.9310	0.2183	0.1656
		(0.9616)	(0.9716)	(0.2185)	(0.1660)
	200	0.9380	0.9370	0.1978	0.1502
		(0.9606)	(0.9666)	(0.1964)	(0.1496)
	250	0.9402	0.9384	0.1846	0.1376
		(0.9594)	(0.9622)	(0.1865)	(0.1383)
	350	0.9430	0.9386	0.1663	0.1214
		(0.9574)	(0.9612)	(0.1667)	(0.1217)
	500	0.9431	0.9390	0.1485	0.1070
		(0.9528)	(0.9600)	(0.1488)	(0.1068)
	1000	0.9444	0.9434	0.1186	0.0836
		(0.9544)	(0.9538)	(0.1189)	(0.0837)

表 2　模型 (3.2) 下 $f(x)$ 的置信区间的覆盖概率 (简记为 CP) 和平均区间长度 (简记为 AL) (括号中的数字为独立同分布样本下的模拟结果)

		CP		AL	
x	n	$k=[n^{\frac{3}{5}}]$	$k=[n^{\frac{7}{10}}]$	$k=[n^{\frac{3}{5}}]$	$k=[n^{\frac{7}{10}}]$
$x=0$	100	0.9098	0.9126	0.3172	0.2370
		(0.9708)	(0.9726)	(0.3144)	(0.2341)
	150	0.9188	0.9336	0.2710	0.2051
		(0.9626)	(0.9678)	(0.2653)	(0.2019)
	200	0.9324	0.9604	0.2451	0.1855
		(0.9620)	(0.9670)	(0.2417)	(0.1825)
	250	0.9610	0.9596	0.2297	0.1707
		(0.9606)	(0.9650)	(0.2275)	(0.1686)
	350	0.9596	0.9534	0.2097	0.1505
		(0.9568)	(0.9618)	(0.2026)	(0.1478)
	500	0.9586	0.9528	0.1845	0.1330
		(0.9565)	(0.9606)	(0.1816)	(0.1307)
	1000	0.9540	0.9511	0.1472	0.1036
		(0.9541)	(0.9530)	(0.1449)	(0.1019)
$x=1$	100	0.9296	0.9142	0.2436	0.1822
		(0.9632)	(0.9706)	(0.2395)	(0.1796)
	150	0.9368	0.9260	0.2062	0.1564
		(0.9626)	(0.9676)	(0.2047)	(0.1554)
	200	0.9370	0.9265	0.1862	0.1414
		(0.9614)	(0.9654)	(0.1863)	(0.1405)
	250	0.9418	0.9380	0.1753	0.1299
		(0.9606)	(0.9614)	(0.1735)	(0.1291)
	350	0.9426	0.9386	0.1577	0.1150
		(0.9604)	(0.9606)	(0.1566)	(0.1140)
	500	0.9442	0.9408	0.1402	0.1007
		(0.9594)	(0.9580)	(0.1394)	(0.1001)
	1000	0.9468	0.9444	0.1122	0.0788
		(0.9538)	(0.9536)	(0.1117)	(0.0785)

4 引　理

我们用 C 表示不依赖于 n 的正常数，其在不同的场合可以表示不同的值，对任意整数 $s\geqslant 1$ 和满足 $E|\xi|^s<\infty$ 的随机变量，记 $\|\xi\|_s=\{E|\xi|^s\}^{\frac{1}{s}}$. 为证明本文的主要结果，我们需要如下引理.

引理 4.1　设 $\{\eta_i, i\geqslant 1\}$ 为 α-混合过程，对 $s\leqslant t$, 用 $\mathcal{F}_s^t$ 表示由 $\{\eta_i, s\leqslant i\leqslant t\}$ 生成的 σ-代数. 设 ξ 和 η 分别为 $\mathcal{F}_1^k$ 和 $\mathcal{F}_{k+m}^{\infty}$ 可测的随机变量，且 $\|\xi\|_{p_1}<\infty$, $\|\eta\|_{q_1}<\infty$,

其中 $p_1, q_1 > 1$, $p_1^{-1} + q_1^{-1} < 1$, 则

$$|E(\xi\eta) - E(\xi)E(\eta)| \leqslant 10\{\alpha(m)\}^{1-p_1^{-1}-q_1^{-1}} \|\xi\|_{p_1} \|\eta\|_{q_1}.$$

进一步地，如果 ξ 和 η 分别为有界随机变量，则

$$|E(\xi\eta) - E(\xi)E(\eta)| \leqslant C\alpha(m),$$

其中 C 为正常数.

证 见文 [9] 中的引理 1.

引理 4.2 设 $\{\eta_i, i \geqslant 1\}$ 为 α-混合过程，对 $s \leqslant t$, 用 $\mathcal{F}_s^t$ 表示由 $\{\eta_i, s \leqslant i \leqslant t\}$ 生成的 σ-代数. 设 $\{\xi_i, 1 \leqslant i \leqslant n\}$ 分别为 $\mathcal{F}_{i_1}^{j_1}, \cdots, \mathcal{F}_{i_n}^{j_n}$ 可测的，其中 $1 \leqslant i_1 < j_1 < i_2 \cdots < j_n$, $i_{l+1} - j_l \geqslant m$, 且 $|\xi_l| \leqslant 1$ $(l = 1, \cdots, n)$, 则

$$\Big|E\Big(\prod_{l=1}^n \xi_l\Big) - \prod_{l=1}^n E(\xi_l)\Big| \leqslant 16(n-1)\alpha(m). \tag{4.1}$$

证 见文 [10] 中的引理 1.1. 注意到 (4.1) 对复随机变量 $\{\xi_i, 1 \leqslant i \leqslant n\}$ 亦成立，其中 $|a|$ 用复数 a 的模代替.

对任意固定的 $x, y \in \mathbb{R}$, 下文将用到如下记号：

$$b_n = \frac{k}{2nf(x)(1 + yk^{-\frac{1}{2}})}, \quad p_n = \int_{x-b_n}^{x+b_n} f(t)\mathrm{d}t, \quad q_n = 1 - p_n, \tag{4.2}$$

$$\xi_{ni} = \begin{cases} 1, & 若\ X_i \in [x - b_n, x + b_n], \\ 0, & 其它, \end{cases} \tag{4.3}$$

$$\eta_{ni} = \frac{\xi_{ni} - p_n}{\sqrt{np_nq_n}}, \quad i = 1, 2, \cdots, n. \tag{4.4}$$

引理 4.3 在定理 2.1 的条件下，当 $n \to \infty$ 时，有下述结论成立：

$$\sum_{i=1}^n \eta_{ni} \xrightarrow{d} N(0, 1). \tag{4.5}$$

证 我们采用分组方法证明引理 4.3. 采用条件 (A2) 中同样的 p, q 和 $\widetilde{k}$, 令 $Z_{ni} = \xi_{ni} - p_n$, $1 \leqslant i \leqslant n$, 且

$$S_n = \sum_{i=1}^n \eta_{ni} = \frac{1}{\sqrt{np_nq_n}} \sum_{i=1}^n Z_{ni},$$

则 S_n 可被拆分成如下形式：

$$S_n = S_n' + S_n'' + S_n''',$$

其中

$$S_n' = \sum_{m=1}^{\widetilde{k}} e_{nm}, \quad S_n'' = \sum_{m=1}^{\widetilde{k}} e_{nm}', \quad S_n''' = e_{n,\widetilde{k}+1}',$$

其中

$$e_{nm} = \frac{1}{\sqrt{np_nq_n}} \sum_{i=r_m}^{r_m+p-1} Z_{ni}, \quad e'_{nm} = \frac{1}{\sqrt{np_nq_n}} \sum_{i=l_m}^{l_m+q-1} Z_{ni},$$

$$e'_{n,\widetilde{k}+1} = \frac{1}{\sqrt{np_nq_n}} \sum_{i=\widetilde{k}(p+q)+1}^{n} Z_{ni},$$

$$r_m = (\widetilde{k}-1)(p+q)+1, \quad l_m = (\widetilde{k}-1)(p+q)+p+1, \quad m = 1, \cdots, \widetilde{k}.$$

如果可以证明下面的 3 个结论成立，则 (4.5) 成立：

$$S'_n \xrightarrow{d} N(0,1), \tag{4.6}$$

$$S''_n = o_p(1), \tag{4.7}$$

$$S'''_n = o_p(1). \tag{4.8}$$

作为准备，我们首先证明

$$\mathrm{Var}(S'_n) = 1 + o(1). \tag{4.9}$$

由 (4.2), $f(x) > 0$ 和 f 在 x 处连续，可知

$$p_n \to 0, \quad np_nq_n = k + o(k). \tag{4.10}$$

由平稳性，条件 (A1)(iii), $b_n = O\left(\frac{k}{n}\right)$, (4.10), $\frac{\widetilde{k}p}{n} \to 1$, $\frac{pk}{n} \to 0$ 以及 $\mathrm{Var}(Z_{n1}) = p_nq_n$, 可知

$$\begin{aligned}
&\frac{\widetilde{k}}{np_nq_n} \sum_{1\leqslant i<j\leqslant p} |\mathrm{Cov}(Z_{ni}, Z_{nj})| \\
&= \frac{\widetilde{k}}{np_nq_n} \sum_{1\leqslant i<j\leqslant p} \Big| \int_{x-b_n}^{x+b_n} \int_{x-b_n}^{x+b_n} \{f_{1,j-i}(u,v) - f(u)f(v)\} \mathrm{d}u\mathrm{d}v \Big| \\
&\leqslant \frac{C\widetilde{k}p^2b_n^2}{np_nq_n} \to 0
\end{aligned}$$

且

$$\sum_{m=1}^{\widetilde{k}} \mathrm{Var}(e_{nm}) = \frac{\widetilde{k}p}{np_nq_n} \mathrm{Var}(Z_{n1}) + \frac{2\widetilde{k}}{np_nq_n} \sum_{1\leqslant i<j\leqslant p} \mathrm{Cov}(Z_{ni}, Z_{nj}) \to 1. \tag{4.11}$$

由平稳性可以证明

$$\begin{aligned}
\sum_{1\leqslant i<j\leqslant \widetilde{k}} |\mathrm{Cov}(e_{ni}, e_{nj})| &= \sum_{l=1}^{\widetilde{k}-1} (\widetilde{k}-l) |\mathrm{Cov}(e_{n1}, e_{n,l+1})| \leqslant \widetilde{k} \sum_{l=1}^{\widetilde{k}-1} |\mathrm{Cov}(e_{n1}, e_{n,l+1})| \\
&\leqslant \frac{\widetilde{k}p}{np_nq_n} \sum_{l=1}^{\widetilde{k}-1} \sum_{s=l(p+q)-p}^{l(p+q)+p} |\mathrm{Cov}(Z_{n1}, Z_{n,s+1})|.
\end{aligned}$$

故由引理 4.1, (4.10), $\frac{\widetilde{k}p}{n} \to 1$ 和 $\left(\frac{n}{k}\right) \sum\limits_{s=q}^{\infty} \alpha(s) \to 0$, 可知

$$\sum_{1\leqslant i<j\leqslant \widetilde{k}} \mathrm{Cov}(e_{ni}, e_{nj}) \leqslant \frac{C\widetilde{k}p}{np_nq_n} \sum_{s=q}^{\infty} \alpha(s) \to 0. \tag{4.12}$$

由 (4.11)–(4.12) 知 (4.9) 成立.

类似地，可证

$$\begin{aligned}\sum_{m=1}^{\widetilde{k}}\mathrm{Var}(e'_{nm})&=\frac{\widetilde{k}q}{np_nq_n}\mathrm{Var}(Z_{n1})+\frac{2\widetilde{k}}{np_nq_n}\sum_{1\leqslant i<j\leqslant q}\mathrm{Cov}(Z_{ni},Z_{nj})\\&\leqslant\frac{C\widetilde{k}q}{n}+\frac{C\widetilde{k}q^2b_n^2}{np_nq_n}\to 0\end{aligned}\tag{4.13}$$

及

$$\begin{aligned}\sum_{1\leqslant i<j\leqslant\widetilde{k}}|\mathrm{Cov}(e'_{ni},e'_{nj})|&=\sum_{l=1}^{\widetilde{k}-1}(\widetilde{k}-l)|\mathrm{Cov}(e'_{n1},e'_{n,l+1})|\leqslant\widetilde{k}\sum_{l=1}^{\widetilde{k}-1}|\mathrm{Cov}(e'_{n1},e'_{n,l+1})|\\&\leqslant\widetilde{k}q\sum_{l=1}^{\widetilde{k}-1}\sum_{r=l(p+q)-(q-1)}^{l(p+q)+(q-1)}|\mathrm{Cov}(Z_{n1},Z_{n,r+1})|,\end{aligned}$$

从而

$$E(S''_n)^2\leqslant\frac{C\widetilde{k}q}{n}+\frac{C\widetilde{k}q^2b_n^2}{np_nq_n}+\frac{C\widetilde{k}q}{np_nq_n}\sum_{s=p}^{\infty}\alpha(s)\to 0,\tag{4.14}$$

故 (4.7) 成立.

注意到 $n-\widetilde{k}(p+q)\leqslant p$, 可证

$$E(S'''_n)^2\leqslant\frac{Cp}{n}+\frac{Cp^2b_n^2}{np_nq_n}\to 0,\tag{4.15}$$

故 (4.8) 成立.

下证 (4.6). 由引理 4.2 知

$$\Big|E\mathrm{e}^{\mathrm{i}t\sum\limits_{m=1}^{\widetilde{k}}e_{nm}}-\prod_{m=1}^{\widetilde{k}}E\mathrm{e}^{\mathrm{i}te_{nm}}\Big|\leqslant C(\widetilde{k}-1)\alpha(q)\to 0,\tag{4.16}$$

故 $\{e_{nm},1\leqslant m\leqslant\widetilde{k}\}$ 渐近独立. 令

$$X_{nm}=\frac{e_{nm}}{\tau_n},\quad \tau_n^2=\sum_{m=1}^{\widetilde{k}}\mathrm{Var}(e_{nm}).$$

由 $|X_{n1}|\leqslant\frac{Cp}{\sqrt{np_nq_n}}\tau_n$, a.s., $\mathrm{Var}(X_{n1})=\frac{\mathrm{Var}(e_{n1})}{\tau_n^2}=\frac{1}{\widetilde{k}}$ 知，对任意 $\varepsilon>0$, 有

$$\begin{aligned}\sum_{m=1}^{\widetilde{k}}EX_{nm}^2I_{\{|X_{nm}|\geqslant\varepsilon\}}&\leqslant\frac{Cp^2\widetilde{k}}{np_nq_n\tau_n^2}P(|X_{n1}|\geqslant\varepsilon)\leqslant\frac{Cp^2\widetilde{k}}{np_nq_n\tau_n^2}\frac{\mathrm{Var}(X_{n1})}{\varepsilon^2}\\&=\frac{C}{\varepsilon^2\tau_n^2}\frac{p^2}{np_nq_n}\leqslant\frac{C}{\varepsilon^2\tau_n^2}\frac{p^2}{k}\to 0,\end{aligned}$$

此处用到了如下事实： $\tau_n^2\to 1$, (4.10)–(4.11) 且 $\frac{p^2}{k}\to 0$. 故由 Feller-Lindeberg 中心极限定理知 (4.6) 成立.

由 (4.6)–(4.8) 和 Cramer-Wold 定理知， (4.5) 成立，引理 4.3 得证.

5 定理 2.1 的证明

利用 (4.2)–(4.4) 中的记号，对任意固定的 $y \in \mathbb{R}$, 有

$$P\Big(k\frac{f_n(x)-f(x)}{f(x)}<y\Big)=P(a_n(x)>b_n)=P\Big(\sum_{i=1}^{n}\eta_{ni}<\frac{k-np_n}{\sqrt{np_nq_n}}\Big). \tag{5.1}$$

由 (A1)(iv) 和泰勒展开，易知

$$p_n=2f(x)b_n+\frac{2}{(r+1)!}\{f^{(r)}(x)+o(1)\}b_n^{r+1}. \tag{5.2}$$

由 (5.2) 和 $k=o(n^{\frac{2r}{2r+1}})$, 利用文 [3] 中的结论，可知

$$k-np_n=y\sqrt{k}+o(\sqrt{k}). \tag{5.3}$$

由 (5.3) 和 $np_nq_n=k+o(k)$, 可推出

$$\frac{k-np_n}{\sqrt{np_nq_n}}=y+o(1). \tag{5.4}$$

由 (5.1), (5.4) 和引理 4.3 知定理 2.1 成立.

致谢 作者衷心感谢审稿人提出的宝贵修改意见.

参 考 文 献

[1] Loftsgarden D O, Quesenberry C D. A nonparametric estimator of a multivariate density function [J]. *Ann Statist*, 1965, 36:1049–1051.

[2] Wagner T J. Strong consistency of a nonparametric estimate of a density function [J]. *IEEE Trans Systems Man Cybernet*, 1973, 3:289–290.

[3] Chen X R. Convergence rates for nearest neighbor density estimator [J]. *Sci China, Ser A*, 1980, 12:1419–1428.

[4] Liu Y, Zhang Y. The consistency and asymptotic normality of nearest neighbor density estimator under α-mixing condition [J]. *Acta Math Sci, Ser B*, 2010, 30:733–738.

[5] Rosenblatt M. A central limit theorem and a strong mixing condition [J]. *Proc Natl Acad Sci USA*, 1956, 42:43–47.

[6] Doukhan P. Mixing: Properties and examples [M]. New York: Springer-Verlag, 1994.

[7] Chen S X, Tang C Y. Nonparametric inference of value at risk for dependent financial returns [J]. *J Finan Econom*, 2005, 3:227–255.

[8] Boente G, Fraiman R. Consistency of a nonparametric estimate of a density function for dependent variables [J]. *J Multi Anal*, 1988, 25:90–99.

[9] Deo C M. A note on empirical processes of strong-mixing sequences [J]. *Ann Prob*, 1973, 1:870–875.

[10] Volkonskii V A, Rozanov Y A. Some limit theorems for random functions I [J]. *Theor Prob Appl*, 1959, 4:178–197.

Asymptotic Normality of the Nearest Neighbor Density Estimators under Strong Mixing Samples

QIN Yongsong[1] LEI Qingzhu[1]

[1]Department of Mathematics, Guangxi Normal University, Guilin 541004, Guangxi, China. E-mail: ysqin@gxnu.edu.cn; qzlei@gxnu.edu.cn

Abstract In this paper, the asymptotic normality of a nearest neighbor density estimator under an α-mixing sample is studied. The authors obtain the asymptotic normality of the estimator with an explicit form of the asymptotic variance, which is used to construct the asymptotic confidence interval for a probability density function under an α-mixing sample.

Keywords α-Mixing sample, Nearest neighbor density estimator, Asymptotic normality

2000 MR Subject Classification 62G05, 62E20

The English translation of this paper will be published in

Chinese Journal of Contemporary Mathematics, Vol. 37 No. 4, 2016

by ALLERTON PRESS, INC., USA