第 34 卷　第 4 期　　华侨大学学报（自然科学版）　　Vol. 34　No. 4
2013 年 7 月　　Journal of Huaqiao University (Natural Science)　　Jul. 2013

文章编号：1000-5013(2013)04-0477-04　　**doi**：10.11830/ISSN.1000-5013.2013.04.0477

α-混合序列下核密度估计量的 r 阶平均相合性

曾翔，吴群英

（桂林理工大学 理学院，广西 桂林 541004）

摘要：　在 α-混合序列情形下，讨论核密度估计量的 r 阶平均相合性，并将该条件弱化为 $nh_n\to\infty$，给出更好的 r 阶最优相合速度 $O(n^{-2r/5})$．结果表明：所要求的条件弱于韦来生关于负相关(NA)样本概率密度函数核估计的 r 阶平均相合性所要求的条件，且相合速度更优，推广了韦来生的结果．

关键词：　核密度估计；α-混合；平均相合；相合速度；负相关

中图分类号：　O 211.4；O 212.7　　　　**文献标志码**：　A

核密度估计是一类重要的非参数分布密度估计．设 $X_1,X_2,\cdots,X_n$ 为总体 X 的样本，X 具有未知的密度函数 $f(x)$，其核密度估计量为

$$f_n(x)=\frac{1}{nh_n}\sum_{i=1}^{n}K\left(\frac{x-X_t}{h_n}\right). \tag{1}$$

式(1)中：$K(\cdot)$为核函数；h_n 为窗宽．

令$\{X_n;n\geqslant 1\}$为概率空间上实值随机变量序列，如记 $F_{-\infty}^{j}=\sigma(X_i,i\leqslant j)$，$F_j^{\infty}=\sigma(X_i,i\geqslant j)$，$\alpha(n)=\sup\limits_j\sup\limits_{A\in F_{-\infty}^{j},B\in F_{n+j}^{\infty}}|P(AB)-P(A)P(B)|$．若当 $n\to\infty$时，$\alpha(n)\to 0$，则称$\{X_n\}$为 α-混合或强混合，$\alpha(n)$称为强混合系数．

$f_n(x)$的极限性质在独立样本下已取得较深入研究[1]．由于相依样本下广泛存在，有不少学者讨论了 $f_n(x)$的极限性质．如文献[2]对 φ-混合序列讨论 $f_n(x)$的均方相合性和渐近正态性，文献[3]进一步弱化了文献[2]中所给的条件，对 $f_n(x)$的渐近正态性进行讨论．目前，基于 α-混合相依结构的模型引起很多统计、金融、经济等方面学者的关注．Chanda[4]首先提出对线性随机过程是 α-混合的随机变量序列；Gorodeskii[5]，Pham 等[6]对此作了进一步的讨论．文献[7]给出了 α-混合序列下的核密度估计量的强相合性与一致强相合性．叶大相等[8]讨论了 α-混合随机变量几乎处处中心极限定理．韦来生[9]讨论了 NA 样本概率密度函数核估计的 r 阶相合性，用到条件 $nh_n^4\to\infty$．本文在 α-混合序列下讨论概率密度函数核估计的 r 阶相合性，并弱化该条件，给出更好的 r 阶最优相合速度．

1　引理及定理

假设核函数 $K(\cdot)$为 Borel 可测有界密度函数，且满足以下 4 个基本条件：

ⅰ) $\int_{-\infty}^{\infty}uK(u)\mathrm{d}u=0$；

ⅱ) $\int_{-\infty}^{\infty}u^2K(u)\mathrm{d}u<\infty$；

ⅲ) $\sup\limits_n\sum\limits_{i=1}^{n}K\left(\frac{x-X_i}{h_n}\right)<\infty$；

ⅳ) $\lim\limits_{|n|\to\infty}uK(u)=0$．

收稿日期：　2012-11-19

通信作者：　曾翔(1977-)，男，讲师，主要从事概率统计的研究．E-mail：168zxzx@163.com．

基金项目：　国家自然科学基金资助项目(11061012)；广西自然科学基金资助项目(2012GXNSFAA053010)

文中 C 是与 n 无关的正常数,即便在同一个式子中也可以表示不同的值.

定理 1 假设 $f_n(x)$ 由式(1)定义,$X_1,X_2,\cdots,X_n,\cdots$ 为同分布的 α 混合序列,$f(x)$ 的二阶导数有界,$f(x)$ 在 x 点连续,若条件ⅰ)~ⅳ)成立且

$$\sum_{i=1}^{n}\alpha(i)<\infty, \tag{3}$$

则对 $0<r\leqslant 2$,当 $h_n\to 0$,$nh_n\to\infty$ 时,有

$$\lim_{n\to\infty}|f_n(x)-f(x)|^r=0.$$

推论 1 在定理 1 的条件下,当 $h_n=n^{-1/5}$ 时,有

$$E|f_n(x)-f(x)|^r=O(n^{-2r/5}).$$

引理 1 $X_1,X_2,\cdots,X_n,\cdots$ 为同分布的 α-混合序列,$K(\cdot)$ 满足条件ⅰ)~ⅱ),则有

$$|Ef_n(x)-f(x)|\leqslant Ch_n^2. \tag{4}$$

证明 将 $f(x-h_nu)$ 作泰勒展开,即有

$$f(x-h_nu)=f(x)-f'(x)h_nu+\frac{f''(x-\zeta h_nu)}{2!}(h_nu)^2,\qquad 0<\zeta<1. \tag{5}$$

利用条件ⅰ)~ⅱ),式(5)和 $f(x)$ 的二阶导数有界,以及 $X_1,X_2,\cdots,X_n$ 的同分布性,可得

$$\begin{aligned}|Ef_n(x)-f(x)|\leqslant Ch_n^2=&\left|\frac{1}{nh_n}\sum_{i=1}^{n}EK\left(\frac{x-X_i}{h_n}\right)-f(x)\right|=\\&\left|\frac{1}{nh_n}\sum_{i=1}^{n}EK\left(\frac{x-X_i}{h_n}\right)-f(x)\right|=\\&\left|\frac{1}{h_n}\int_{-\infty}^{\infty}K\left(\frac{x-y}{h_n}\right)\cdot f(y)\mathrm{d}y-f(x)\right|=\\&\left|\int_{-\infty}^{\infty}K(u)[f(x-h_nu)-f(x)]\mathrm{d}u\right|=\\&\left|\int_{-\infty}^{\infty}K(u)\frac{f''(x-\xi h_nu)}{2}h_n^2u^2\mathrm{d}u\right|\leqslant Ch_n^2.\end{aligned} \tag{6}$$

由式(6)可知式(4)成立.

引理 2[10] 设 $\xi\in\mathscr{F}_{-\infty}^{m}$,$\eta\in\mathscr{F}_{m+n}^{\infty}$,若 $\xi\leqslant C_1$,$\eta\leqslant C_2$,则有

$$|E\xi\eta-E\xi E\eta|\leqslant 4C_1C_2\alpha(n).$$

引理 3[11] 设 $X_1,X_2,\cdots,X_n,\cdots$ 为有共同分布函数 F 的平稳 α-混合序列,且对某 $v>3$,有 $\alpha(n)=O(n^{-v})$,又设 F_n 是基于样本 $X_1,X_2,\cdots,X_n$ 的经验分布函数,则有

$$\sup_{x\in R}|F_n(x)-F(x)|=O\left(\left(\frac{\log\log n}{n}\right)^{1/2}\right)\qquad \text{a. s.}$$

引理 4[12] 设 $K(\cdot)$ 及 $g(\cdot)$ 均为 R 上的 Borel 可测函数,并且满足下述条件:1) K 有界;2) $\int_{-\infty}^{\infty}|K(u)|\mathrm{d}u<\infty$;3) $\lim\limits_{|u|\to\infty}uK(u)=0$ 或 g 有界;4) $\int_{-\infty}^{\infty}|g(u)|\mathrm{d}u<\infty$.

常数序列 $\{h_n\}$ 满足 $\lim\limits_{u\to\infty}h_n=0$. 令 $g_n(x)=\frac{1}{h_n}\int K\left(\frac{u}{h_n}\right)g(x-u)\mathrm{d}u$,则有 $\lim\limits_{u\to\infty}g_n(x)=g(x)\int K(u)\mathrm{d}u$,任意 $x\in C(g)$. 其中:$C(g)$ 为 g 的连续点集.

2 定理 1 的证明

由于 $0<r\leqslant 2$,故由 C_r-不等式和 Jensen 不等式可知

$$\begin{aligned}E|f_n(x)-f(x)|^r\leqslant &E|f_n(x)-Ef_n(x)+Ef_n(x)-f(x)|^r\leqslant\\&C\{E|f_n(x)-Ef_n(x)|^r+E|Ef_n(x)-f(x)|^r\}\leqslant\\&C[\mathrm{var}(f_n(x))]^{r/2}+C|Ef_n(x)-f(x)|^r\triangleq I_1+I_2.\end{aligned} \tag{7}$$

由引理 1 可知

$$I_2 = C \mid Ef_n(x) - f(x) \mid^r \leqslant Ch_n^{2r} \to 0, \qquad n \to \infty. \tag{8}$$

记 $g_n(x,X_i) = \frac{1}{n^2}\sum_{i=1}^{n} K(\frac{x-X_i}{h_n})$，由引理 2 可得

$$\begin{aligned}\operatorname{var}(f_n(x)) =& \frac{1}{n^2}E[\sum_{i=1}^{n}(g_n(x,X_i) - Eg_n(x,X_i))]^2 = \\ & \frac{1}{n^2}E[\sum_{i=1}^{n}(g_n(x,X_i) - Eg_n(x,X_i))^2] + \\ & \frac{2}{n^2}\sum_{1\leqslant i<j\leqslant n}\operatorname{cov}[g_n(x,X_i), g_n(x,X_j)] \triangleq I_{11} + I_{12}.\end{aligned} \tag{9}$$

由 $X_1, X_2, \cdots, X_n$ 的同分布性和引理 4，可知

$$\begin{aligned}I_{11} =& \frac{1}{n^2}\sum_{i=1}^{n}\operatorname{var}(g_n(x,X_i)) \leqslant \\ & \frac{1}{n}E[g_n^2(x,X_1)] = \frac{1}{nh_n^2}\int_{-\infty}^{\infty}K^2(\frac{x-y}{h_n})f(y)\mathrm{d}y = \\ & \frac{1}{nh_n}\int_{-\infty}^{\infty}K^2(u)f(x-h_nu)\mathrm{d}u \leqslant \frac{C}{nh_n}\int_{-\infty}^{\infty}K(u)f(x-h_nu)\mathrm{d}u \leqslant \\ & Cf(x)(nh_n)^{-1} \leqslant C(nh_n)^{-1}.\end{aligned} \tag{10}$$

由条件 ⅲ)，式(3)以及引理 2，可知

$$\begin{aligned}I_{12} =& \frac{2}{n^2}\sum_{1\leqslant i<j\leqslant n}\{E[g_n(x,X_i)g_n(x,X_j)] - Eg_n(x,X_i)Eg_n(x,X_j)\} \leqslant \\ & \frac{2}{n^2}\sum_{i<j}[4\frac{1}{h_n}\sum_{i=1}^{n}K(\frac{x-X_i}{h_n}) \cdot \frac{1}{h_n}\sum_{i=1}^{n}K(\frac{x-X_j}{h_n}) \cdot \alpha(j-i)] \leqslant \\ & \frac{8}{n^2h_n^2}\sum_{i<j}[\sum_{i=1}^{n}K(\frac{x-X_i}{h_n}) \cdot \sum_{i=1}^{n}K(\frac{x-X_j}{h_n}) \cdot \alpha(j-i)] \leqslant \\ & \frac{8}{n^2h_n^2}\sum_{j=1}\alpha(j) \leqslant C(nh_n)^{-2}.\end{aligned} \tag{11}$$

将式(10)和式(11)代入式(9)，可得

$$I_1 = C[\operatorname{var}(f_n(x))]^{r/2} \leqslant C(nh_n)^{-r/2}, \qquad n \to \infty. \tag{12}$$

将式(8)和式(12)代入式(7)，可知

$$0 \leqslant \lim_{u\to\infty} E \mid f_n(x) - f(x) \mid^r \leqslant \lim_{u\to\infty} C(nh_n)^{-r/2} + \lim_{u\to\infty} Ch_n^{2r} = 0. \tag{13}$$

定理证毕.

推论 1 的证明　当取 $h_n = n^{-1/5}$，$E|f_n(x) - f(x)|^r$ 达到最优相合速度，将式(8)和式(12)代入式(7)，可得

$$E \mid f_n(x) - f(x) \mid^r \leqslant C(nh_n)^{-r/2} + Ch_n^{2r} \leqslant Cn^{-2r/5},$$

即 $E|f_n(x) - f(x)|^r = O(n^{-2r/5})$. 推论 1 得证.

3　应用举例

失效率是工作到某时刻尚未失效的产品，在该时刻后单位时间内发生失效的概率，一般记为 λ. 它也是时间 t 的函数，故也记为 $\lambda(t)$，称为失效率函数，有时也称为故障率函数或风险函数. 按上述定义，失效率是在时刻 t 尚未失效产品在 $t+\Delta t$ 的单位时间内发生失效的条件概率，即反映 t 时刻失效的速率，也称为瞬时失效率. 设在可靠性问题中 r. v. X 的分布函数记为 $F(x)$，密度函数为 $f(x)$，生存函数和失效率函数分别定义为 $\bar{F}(x) = 1 - F(x) = P(X \geqslant x)$ 和 $\lambda(t) = f(x)/\bar{F}(x)$.

设 $X_1, X_2, \cdots, X_n$ 为从总体 $F(x)$ 中抽取的样本，记 $\bar{F}_n(x) = 1 - F_n(x)$，则生存函数和失效率函数的自然估计为

$$\bar{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n}I(X_i \geqslant x), \qquad \lambda(t) = \frac{\hat{f}_n(x)}{\bar{F}_n(x)}.$$

推论 2 设 $X_1,X_2,\cdots,X_n$ 为平稳 α-混合样本，对某 $v>3$，混合系数 $\alpha(n)=O(n^{-v})$，且满足定理的条件，则有 $\lim\limits_{u\to\infty}E|\lambda_n(t)-\lambda(t)|^r=0$.

证明 由引理 3 及式(13)可知

$$E\mid\lambda_n(t)-\lambda(t)\mid^r=E\mid\frac{f_n(x)}{\bar{F}_n(x)}-\frac{f(x)}{\bar{F}_n(x)}\mid^r=E\mid\frac{f_n(x)\bar{F}(X)-\bar{F}_n(x)f(x)}{\bar{F}_n(x)\bar{F}(X)}\mid^r\leqslant$$
$$CE(\mid f_n(x)-f(x)\mid\bar{F}(X)+f(x)E\mid\bar{F}_n(x)-\bar{F}(X)\mid)^r\leqslant$$
$$CE\mid f_n(x)-f(x)\mid^r+CE\mid\bar{F}_n(x)-\bar{F}(X)\mid^r\to 0.$$

证毕.

参考文献:

[1] 陈希孺，方兆本，李国英，等. 非参数统计[M]. 上海：上海科学技术出版社，1989：284-305.

[2] 林正炎. 相依样本情形时密度的核估计[J]. 科学通报，1983，28(12)：709-713.

[3] 杨善朝. φ-混合样本密度估计的渐近正态性[J]. 广西师范大学学报：自然科学版，1996，14(1)：20-21.

[4] CHANDA K C. Strong mixing properties of linear stochastic processes[J]. Journal of Applied Probability，1974，11(2)：401-408.

[5] GORODESKII V V. On the strong mixing property for linear sequences[J]. Theory of Probability and Their Applications，1977，22(2)：421-423.

[6] PHAM T D，TRAN L T. Some mixing properties of time series models[J]. Stoch Process Appl，1985，19(2)：297-303.

[7] GEORGE G. Roussas，asymptotic normality of the kernel estimate of a probability density function under association[J]. Statistics and Prob Lett，2000，50(1)：1-12.

[8] YE Da-xiang，WU Qun-ying. Almost sure central limit theorem for product of partial sums of strongly mixing random variables[J]. Journal of Inequalities and Applications，2011，2011(1)：576301-576309.

[9] 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.

[10] ROUSSAS G G，IOANNIDES D. Moment inequalities for mixing sequences of random variables[J]. Stochastic Analysis and Applications，1987，5(1)：61-120.

[11] CAI Zong-wu. Asymptotic properties of Kaplan-Meier estimator for censored dependent data[J]. Statistics and Probability Letters，1988，37(4)：381-389.

[12] PARZEN E. On estimation of a probability density function and mode[J]. Ann Math Statist，1962，33(3)：1065-1076.

Consistence r-th Mean of Density Kernal Estimators for α-Mixing Sequences

ZENG Xiang，WU Qun-ying

(College of Science，Guilin University of Technology，Guilin 541004，China)

Abstract: Under α-mixing random variable sequences，we study the consistence in r-th mean of the kernel density estimate. The condition is weakened as $nh_n\to\infty$，and r-th optimal convergence rate $O(n^{-2r/5})$ is better. The result shows that the required condition is weaker than the condition proposed by WEI Lai-sheng on consistence in r-th mean of the kernel density estimator for NA samples，and we convergence rate is obtained. Our results improve the one made by WEI Lai-sheng.

Keywords: kernel density estimate；α-mixing sequence；mean consistency；convergence rate；negatively associated

(**责任编辑**：钱筠 **英文审校**：黄心中)