文章编号:1001-5132（2007）03-0364-04

删失场合混合样本下回归函数核估计的强相合性

余　未

（宁波大学 理学院，浙江 宁波 315211）

摘要：在删失场合下，当 (X_i, Y_i) 为同分布 φ 混合样本时，获得了回归函数 $m(x)=E(Y\mid X=x)$ 的 3 类核估计的强相合性.

关键词：回归函数；删失场合；混合样本；核估计；强相合性

中图分类号：O212.7　　　　文献标识码：A

关于回归函数非参数核估计的强相合性曾受到国内外学者的很大关注. 设 $(X_i, Y_i), i=1,2,\cdots,n$，是取自 $R^d\times R^1$ 上的随机向量 (X,Y) 中的样本，设 $E|Y|<\infty$，以 $m(x)=E(Y\mid X=x)$ 表示回归函数，以下 3 类回归函数核估计的形式是比较常见的.

$$m_n^{(1)}(x)=\sum_{i=1}^n Y_i K((X_i-x)/h_n)\Big/\sum_{j=1}^n K((X_j-x)/h_n), \tag{1}$$

$$m_n^{(2)}(x)=\sum_{i=1}^n Y_i h_i^{-d} K((X_i-x)/h_i)\Big/\sum_{j=1}^n h_j^{-d} K((X_j-x)/h_j), \tag{2}$$

$$m_n^{(3)}(x)=\sum_{i=1}^n Y_i K((X_i-x)/h_i)\Big/\sum_{j=1}^n K((X_j-x)/h_j), \tag{3}$$

其中，K 是 R^d 的概率密度，$h_n>0, h_n\to 0\ (n\to\infty)$.

当 (X_i, Y_i) 为独立同分布时，孙东初[1]研究了完全样本下上述 3 类估计的强相合性. 陈平[2]在不完全样本下，即删失场合获得了它们的强相合性. 胡玉萍等人[3]又进一步在截尾样本下对估计量进行了改良并讨论了强相合性，不过这些都是对 i.i.d. 样本所做的讨论.

当 (X_i, Y_i) 为 φ 混合样本时，最早秦永松[4]获得了完全样本下上述 3 类估计的强相合性. 最近有研究者利用截尾的方法分别对这 3 类核估计进行了改良并研究了在 φ 混合条件下它们的强一致收敛性[5,6]. 本文的目的是在假设 (X_i, Y_i) 为同分布 φ 混合序列且此时获得的样本是删失的前提下，采用文献[7]所定义的 Y_i^*，研究基于 (X_i, Y_i^*) 构造的以下 3 类估计量的强相合性.

$$m_n^{(1)*}(x)=\sum_{i=1}^n Y_i^* K((X_i-x)/h_n)\Big/\sum_{j=1}^n K((X_j-x)/h_n), \tag{4}$$

$$m_n^{(2)*}(x)=\sum_{i=1}^n Y_i^* h_i^{-d} K((X_i-x)/h_i)\Big/\sum_{j=1}^n h_j^{-d} K((X_j-x)/h_j), \tag{5}$$

$$m_n^{(3)*}(x)=\sum_{i=1}^n Y_i^* K((X_i-x)/h_i)\Big/\sum_{j=1}^n K((X_j-x)/h_j). \tag{6}$$

收稿日期：2007-04-02.　　宁波大学学报（理工版）网址：http://3xb.nbu.edu.cn
基金项目：宁波大学校科研基金（XK200454）.
作者简介：余　未（1974－），女，浙江宁波人，讲师，主要研究方向：概率与数理统计. E-mail: yuwei@nbu.edu.cn

1　若干引理

引理 1[8]　假设 (X_i, Y_i) 为同分布 φ 混合序列且此时获得的样本是删失的，则有 $EY_i^* = EY_i, E[Y_i^* \mid X_i = x] = E[Y_i \mid X_i = x]$，且 (X_i, Y_i^*) 是同分布 φ 混合.

引理 2[9]　如果 $K(x)$ 为 R^d 上的非负可测函数，且 $\int_{R^d} K(x)\mathrm{d}x = 1$，记 $\psi(x) = \sup\limits_{\|y\| \geq \|x\|} K(y)$，有 $\int_{R^d} \psi(x) \cdot \mathrm{d}x < \infty$，那么对于 R^d 上的任意非负可测函数 $f(x)$，只要有 $\int_{R^d} |f(x)|\mathrm{d}x < \infty$，就有

$$\int_{R^d} s^{-d} K\left(\frac{y-x}{s}\right) f(y)\mathrm{d}y \to f(x), s \to 0, \text{a. s.}.$$

引理 3[10]　设 $\{\xi_i, i \geq 1\}$ 是 φ 混合序列，且 $E\xi_i = 0$，存在 $0 < M < \infty$，使 $|\xi_i| \leq M$, a. s.，若混合系数满足 $\sum_{n=1}^{\infty} \varphi(n) < \infty, \varphi(n) \downarrow$，则 $\forall \varepsilon > 0$，当 n 充分大时，有

$$P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} \xi_i\right| > \varepsilon\right\} \leq C_0 \exp\{-CM^{-1}\sqrt{n}\varepsilon\},$$

其中 C_0, C 是与 ε 无关的正常数.

引理 4[11]　对同分布 φ 混合序列 $\{X_i, i \geq 1\}$，记 $S_n = \sum_{i=1}^{n} X_i$，如果存在 $a_n \uparrow \infty$，$\sum_{n=1}^{\infty} \mathrm{Var}X_n / a_n^2 < \infty$，且 $\sum_{n=1}^{\infty} \varphi^{1/2}(n) < \infty$，则 $(S_n - ES_n)/a_n \to 0$, a. s..

2　主要定理及其证明

首先给出 1 组已知条件如下：

(A)　X 具有概率密度 $f(x)$，且 $f(x) > 0$, a. s.，(X_i, Y_i) 为同分布 φ 混合序列；

(B)　存在正常数 K^*，使 $0 \leq K(x) \leq K^*$，$\int_{R^d} K(x)\mathrm{d}x < \infty, \int_{R^d} \psi(x)\mathrm{d}x < \infty$，记 $\psi(x) = \sup\limits_{\|y\| \geq \|x\|} K(y)$；

(C)　$\lim\limits_{n \to \infty} h_n = 0$；

(D)　混合系数 $\varphi(n) \downarrow$，满足 $\sum_{n=1}^{\infty} \varphi(n) < \infty$.

定理 1　若条件(A)～(D)均满足，且 Y, a.s.有界，则当 $\sqrt{n} h_n^d = O(n^q)$，$q > 0$ 时，有 $m_n^{(1)*}(x) \to m(x)$, a.s..

定理 2　若条件(A)～(D)均满足，且 $EY^{*2} < \infty$，则当 $\sum_{n=1}^{\infty} (nh_n^d)^{-2} < \infty$ 时，有 $m_n^{(2)*}(x) \to m(x)$, a.s..

定理 3　若条件(A)～(D)均满足，且 $EY^{*2} < \infty$，则当 $\sum_{n=1}^{\infty} (\sum_{i=1}^{n} h_i^d)^{-2} < \infty$ 时，有 $m_n^{(3)*}(x) \to m(x)$, a.s..

以上 3 个定理的结果与文献[2]中关于删失场合独立同分布样本下的结果相类似，但在后 2 个定理中放弃了 Y 有界的条件，所付出的代价是 $EY^{*2} < \infty$，而非存在 $P \geq 1$，有 $E|Y|^P < \infty$.

定理的证明

不妨设 $\int_{R^d} K(x)\mathrm{d}x = 1$，否则令 $M_0^{-1} = \int_{R^d} K(x) \cdot \mathrm{d}x$，在 $m_n^{(i)*}(x)$ 的分子分母同乘以 M_0 即可.

首先令：

$$Z_n^{(1)*}(x) = \frac{1}{nh_n^d}\sum_{i=1}^{n} Y_i^* K\left(\frac{X_i - x}{h_n}\right),$$

$$W_n^{(1)}(x) = \frac{1}{nh_n^d}\sum_{j=1}^{n} K\left(\frac{X_j - x}{h_n}\right);$$

$$Z_n^{(2)*}(x) = \frac{1}{n}\sum_{i=1}^{n} Y_i^* h_i^{-d} K\left(\frac{X_i - x}{h_i}\right),$$

$$W_n^{(2)}(x) = \frac{1}{n}\sum_{j=1}^{n} h_j^{-d} K\left(\frac{X_j - x}{h_j}\right);$$

$$Z_n^{(3)*}(x) = \frac{1}{\sum_{i=1}^{n} h_i^d}\sum_{i=1}^{n} Y_i^* K\left(\frac{X_i - x}{h_i}\right),$$

$$W_n^{(3)}(x) = \frac{1}{\sum_{i=1}^{n} h_i^d}\sum_{j=1}^{n} K\left(\frac{X_j - x}{h_j}\right).$$

易见 $m_n^{(i)*}(x) = Z_n^{(i)*}(x) / W_n^{(i)}(x)$，$i = 1, 2, 3$.

由条件(A)可设 $Z(x) = m(x)f(x)$，若 $Z_n^{(i)*}(x) \to Z(x)$, a. s.，则令 $Y_i^* = 1$，有 $m(x) = 1$, a. s.，就有 $W_n^{(i)*}(x) \to f(x)$, a. s..

于是就得证 $m_n^{(i)*}(x) \to m(x)$, a. s.，$i = 1, 2, 3$.

记：

$$\phi(x, s) = E\left\{s^{-d} Y_i^* K\left(\frac{X_i - x}{s}\right)\right\} =$$

$$E\left\{s^{-d} E\left[Y_i^* \mid X_i\right] K\left(\frac{X_i - x}{s}\right)\right\} =$$

$$E\left\{s^{-d}m(X_i)K\left(\frac{X_i-x}{s}\right)\right\}=$$

$$\int_{R^d}s^{-d}K\left(\frac{y-x}{s}\right)m(y)f(y)\mathrm{d}y=$$

$$\int_{R^d}s^{-d}K\left(\frac{y-x}{s}\right)z(y)\mathrm{d}y\to Z(x)\quad \text{a. s.}.$$

以上分别利用了引理 1 和引理 2 .

因为 $h_n\to 0$，所以

$$EZ_n^{(1)*}(x)=\frac{1}{n}\sum_{i=1}^{n}E\left\{h_n^{-d}Y_i^*K\left(\frac{X_i-x}{h_n}\right)\right\}=$$

$$\phi(x,h_n)\xrightarrow{\text{a.s.}}Z(x).$$

由 Toeplitz 引理得

$$EZ_n^{(2)*}(x)=\frac{1}{n}\sum_{i=1}^{n}E\left\{h_i^{-d}Y_i^*K\left(\frac{X_i-x}{h_i}\right)\right\}=$$

$$\frac{1}{n}\sum_{i=1}^{n}\phi(x,h_i)\xrightarrow{\text{a.s.}}Z(x),$$

$$EZ_n^{(3)*}(x)=\left(\sum_{i=1}^{n}h_i^d\right)^{-1}\sum_{i=1}^{n}E\left\{Y_i^*K\left(\frac{X_i-x}{h_i}\right)\right\}=$$

$$\left(\sum_{i=1}^{n}h_i^d\right)^{-1}\sum_{j=1}^{n}h_j^d\phi(x,h_j)\xrightarrow{\text{a.s.}}Z(x).$$

又因为 $\left|Z_n^{(i)*}(x)-Z(x)\right|\le\left|Z_n^{(i)*}(x)-EZ_n^{(i)*}(x)\right|+\left|EZ_n^{(i)*}(x)-Z(x)\right|$，

所以只要证 $Z_n^{(i)*}(x)-EZ_n^{(i)*}(x)\xrightarrow{\text{a.s.}}0$.

令 $T_i(s)=s^{-d}Y_i^*K\left(\left(X_i-x\right)/s\right)$， $i=1,2,\cdots$

则 $Z_n^{(1)*}(x)-EZ_n^{(1)*}(x)=\dfrac{1}{n}\sum_{i=1}^{n}\left\{T_i(h_n)-ET_i(h_n)\right\}$.

易知 $\left\{T_i(h_n)-ET_i(h_n)\right\}$ 是均值为 0 的同分布 φ 混合序列.

因为 Y, a.s. 有界，由 Y_i^* 的定义[7]知 $\left|Y_i^*\right|\le M<\infty$, a. s..

$$\left|T_i(h_n)-ET_i(h_n)\right|=\left|h_n^{-d}Y_i^*K\left(\frac{X_i-x}{h_n}\right)-E\left(h_n^{-d}Y_i^*K\left(\frac{X_i-x}{h_n}\right)\right)\right|\le$$

$$2h_n^{-d}MK^*\stackrel{\Delta}{=}C_1h_n^{-d},\ \text{a. s.}.$$

由引理 3 得

$$P\{|Z_n^{(1)*}(x)-EZ_n^{(1)*}(x)|>\varepsilon\}=$$

$$P\left\{\frac{1}{n}\left|\sum_{i=1}^{n}(T_i(h_n)-ET_i(h_n))\right|>\varepsilon\right\}\le$$

$$C_0\exp\{-Ch_n^dC_1^{-1}\sqrt{n}\varepsilon\}\le C_0\exp\{-C\sqrt{n}h_n^d\}.$$

又因为 $\sqrt{n}h_n^d=O(n^q)$, $q>0$，所以 $\sum_{n=1}^{\infty}P\{|Z_n^{(1)*}(x)-EZ_n^{(1)*}(x)|>\varepsilon\}<\infty$.

由 Borel-Cantelli 引理知 $Z_n^{(1)*}(x)-EZ_n^{(1)*}(x)\to 0$, a. s.，于是定理 1 得证.

因 $Z_n^{(2)*}(x)-EZ_n^{(2)*}(x)=\dfrac{1}{n}\sum_{i=1}^{n}\left\{T_i(h_i)-ET_i(h_i)\right\}$ ，

$$E|T_i(s)|^2=E\left|s^{-d}Y_i^*K\left(\frac{X_i-x}{s}\right)\right|^2\le$$

$$K^{*2}s^{-2d}EY_i^{*2}\text{，}$$

且 $\{T_i(h_i)\}$ 是同分布 φ 混合序列，则当 $\sum_{n=1}^{\infty}(nh_n^d)^{-2}<\infty$ 时，有

$$\sum_{n=1}^{\infty}\frac{E|T_n(h_n)|^2}{n^2}\le K^{*2}EY_i^{*2}\sum_{n=1}^{\infty}(nh_n^d)^{-2}<\infty.$$

由引理 4，取 $a_n=n$，得 $Z_n^{(2)*}(x)-EZ_n^{(2)*}(x)\to 0$ a. s.，于是定理 2 得证.

又因为

$$Z_n^{(3)*}(x)-EZ_n^{(3)*}(x)=\left(\sum_{i=1}^{n}h_i^d\right)^{-1}\sum_{i=1}^{n}\left\{h_i^dT_i(h_i)-h_i^dET_i(h_i)\right\}\text{，}$$

当 $\sum_{n=1}^{\infty}(\sum_{i=1}^{n}h_i^d)^{-2}<\infty$ 时，$\lim_{n\to\infty}\sum_{i=1}^{n}h_i^d=\infty$ ，且

$$\sum_{n=1}^{\infty}E|h_n^dT_n(h_n)|^2/(\sum_{i=1}^{n}h_i^d)^2\le$$

$$K^{*2}EY^{*2}\sum_{n=1}^{\infty}(\sum_{i=1}^{n}h_i^d)^{-2}<\infty.$$

同样可利用引理 4，取 $a_n=\sum_{i=1}^{n}h_i^d$，得 $Z_n^{(3)*}(x)-EZ_n^{(3)*}(x)\to 0$, a. s.，于是定理 3 也得证.

最近赵霞等人[12]在相对较弱的条件下讨论了当 (X_i,Y_i) 为 ρ 混合时回归函数改良核估计的强相合性. 张日权[13]则研究了当 (X_i,Y_i) 为 α 混合下一般函数核估计的强一致收敛速度. 因此也值得继续关注在删失场合，ρ 混合或 α 混合样本下回归函数核估计的收敛问题.

参考文献：

[1] 孙东初. 回归函数核估计的强相合性[J]. 数学年刊, 1985, 6A(4):481-486.

[2] 陈平. 随机删失场合回归函数的核估计及强相合性[J]. 系统科学与数学, 1992, 12(4):350-355.

[3] 胡玉萍, 薛留根. 截尾样本下回归函数改良核估计的强相合性[J]. 应用概率统计, 2000, 16(4):379-390.

[4] 秦永松. 相依样本情形下回归函数核估计的强相合性[J]. 工程数学学报, 1990, 7(1):61-66.

[5] Qian Weimin, Mammitzsch Volker. Strong uniform convergence for the estimator of the regression function under φ-mixing conditions[J]. Metrika, 2000, 52(1):45-61.

[6] Wang Li, Liang Hanying. Strong uniform convergence of the recursive regression estimator under φ-mixing conditions[J]. Metrika, 2004, 59(3):245-261.

[7] Zhengzu Kang. A class of estimators of the parameters in linear regression with censored data[J]. Acta Math Appl Sinica, 1987, 3:231-241.

[8] 王志江, 陈彩琴. 删失场合下回归函数估计的强相合性[J]. 杭州大学学报, 1996, 23(3):212-218.

[9] Stein E M. Singular integral and differentiability properties of functions[M]. Princeton N J: Princeton Univ Press, 1970.

[10] 柴根象. 平稳序列最近邻密度估计的相合性[J]. 数学学报, 1989, 32(3):423-432.

[11] Iosifescu M, Theodorescu R. Random processes and learning[M]. New York: Springer-Verlag Press, 1969.

[12] 赵霞, 李学芳. 相依样本下回归函数核估计的强相合性[J]. 曲阜师范大学学报, 2005, 31(1):11-14.

[13] 张日权. 相依数据下一般函数核估计的强一致收敛速度[J]. 应用概率统计, 2005, 21(1):88-96.

Strong Consistency of Kernel Estimators for φ-mixed-samples-based Regression Function with Censored Data

YU Wei

(Faculty of Science, Ningbo University, Ningbo 315211, China)

Abstract: Given random censored data, the strong consistency of three kernel estimators for regression function $m(x)=E(Y\mid X=x)$ is investigated in the case of (X_i,Y_i) being in identical distribution with φ-mixing samples.

Key words: regression function; censored data; mixed sample; kernel estimator; strong consistency

CLC number: O212.7　　　　**Document code:** A

（责任编辑　史小丽）