

α 混合样本下稳健非参数一般权回归估计的强相合性

项 萍

(柳州工学院 数理教学部, 广西 柳州 545616)

摘要: 本文对一般权函数稳健非参数回归估计进行研究. 在 α 混合样本下, 证明了稳健条件位置函数一般权回归估计的强相合性.

关键词: α 混合序列, 一般权函数, 稳健估计, 强相合性

由于核估计法都是利用响应变量的加权平均进行估计的, 导致估计值对奇异值非常敏感, 为此提出了稳健估计, 解决了传统核估计的弱抗干扰性. 自 1953 年 Box 首先提出稳健性的概念, Huber, Tukey, Rousseeuw, Fraiman, Cox 等一批学者对参数的稳健估计进行了卓有成效的研究.

而近年来, 关于稳健条件位置函数 NW 核回归估计的研究则相对较少, 因此加强对该估计的研究很有必要. Boente 和 Fraiman (1989) 定义了一种稳健条件位置函数, 研究了独立同分布样本下稳健非参数 NW 核回归估计的相合性及渐进正态性. 之后, Boente 和 Fraiman (1989) 又给出了 α 混合样本下该估计的强相合性. Boente 和 Fraiman (1990) 在 α 混合条件下, 获得了稳健条件位置函数的 NW 核回归估计和 k -近邻估计的相合性和渐进正态性. 最近, Zhao 和 Ma (2016) 研究了稳健非参数回归函数的 G-M 核估计, 证明了均方误差, 给出了最优窗宽和渐进正态性. 项萍 (2018) 获得了 α 混合样本下稳健条件位置函数 NW 核回归估计的强相合性及渐进正态性. 本文在前人的研究基础上, 研究 α 混合样本下条件位置参数的稳健一般权函数回归估计的强相合性.

设 (X, Y) 为 $d+1$ 维随机变量, 其中 $X \in R^d$, $Y \in R$. 又设给定 X 下 Y 的条件分布函数为 $F(y|X=x)$, ψ 是 $R \rightarrow R$ 的严格递增有界连续函数. Boente 和 Fraiman (1989) 证明存在唯一的 $\sigma(X)$ 可测函数 $g(x)$ 满足下面方程

$$E \left[h(X) \psi \left(\frac{Y - g(X)}{s(X)} \right) \right] = 0$$

其中 $h(x)$ 为任意的可积函数, $s(X)$ 是条件分布尺度的稳健量, 即 $s(x) = \text{med}(|Y - m(x)|)$, 而 $m(x) = \text{med}(Y|X=x)$ 是条件分布 $F(y|X=x)$ 的中位数. 称 $g(x)$ 为稳健条件位置函数, 记为 $E^\psi(Y|X)$, 即 $g(x) = E^\psi(Y|X)$.

令混合系数

$$\alpha(n) = \sup_{k \geq 1} \{ |P(A \cap B) - P(A)P(B)| : A \in J_1^k, B \in J_{K+n}^\infty \},$$

其中 J_a^b 是由 $\{X_i, Y_i\}_a^b$ 构成的 σ 代数. 如果 $\lim_{n \rightarrow \infty} \alpha(n) = 0$, 则称 $\{X_i, Y_i : i = 1, 2, 3, \dots\}$ 为 α 混合序列.

下面给出权函数的稳健回归函数估计的定义.

定义 1.1. 稳健回归函数估计 $g_n(x) = E_{F_n}^\psi(Y|X=x)$ 定义为下面方程的解

$$\sum_{i=1}^n W_{ni}(x) \psi((Y_i - g_n(x))/s_n(x)) = 0$$

其中 $s_n(x)$ 是条件分布非参数估计 $F_n(y|X=x)$ 的稳健尺度量, 即 $s_n(x) = \text{med}(|Y - m_n(x)|)$, 而 $m_n(x) = \text{med}(Y|X=x)$ 是条件分布

$F_n(y|X=x)$ 的中位数.

一、主要结论

为了给出本文结论, 我们需要 Boente 和 Fraiman (1989) 文中的类似条件.

(A1) $\psi : R \rightarrow R$ 是一个严格递增有界的连续函数, 满足

$$\lim_{t \rightarrow +\infty} \psi(t) = a > 0 \text{ 且 } \lim_{t \rightarrow -\infty} \psi(t) = b < 0.$$

(A2) 下面两个条件之一成立:

(a) 对几乎所有的 x , $s(x)$ 是关于 x 的弱连续函数;

(b) ψ 是奇函数, 对于每一个给定的 x , 条件分布函数

$F(y|X=x)$ 关于 $g(x) = E_{F_x}^\psi(Y|X=x)$ 对称且关于 y 连续.

(H1) 随机序列 $\{X_i, Y_i : i \geq 1\}$ 与 (X, Y) 同分布且为 α 混合的, 其混合系数 $\alpha(n) = o(n^{-\lambda})$, 其中 $\lambda > 1$.

(H2) 存在 $q > 2\lambda / (\lambda - 1)$ 使得

$$nh_n^{2d(1-1/q)} / \log n \rightarrow \infty,$$

且

$$\sum_{i=1}^{\infty} h_n^{d[\lambda(1-2/q)-1]} < \infty$$

(H3) 存在一串非负实常数序列 $\{c_n : n \geq 1\}$ 使得

$$\max_{1 \leq j \leq n} W_{nj}(x; X_1, X_2, \dots, X_n) \leq c_n, a.s. \forall x(P_X)$$

且当 $n \rightarrow \infty$ 时

$$c_n \log n \rightarrow 0, nc_n \rightarrow \infty.$$

定理 2.1. 假设权函数 $W_{ni}(x; X_1, X_2, \dots, X_n)$ 是概率权函数, 即

$$W_{ni}(x) \geq 0, \sum_{i=1}^n W_{ni}(x) = 1.$$

如果条件 (A1), (A2), (H3) 成立, 则有

$$g_n(x) \rightarrow g(x), a.s. \quad n \rightarrow \infty.$$

上述定理与文中定理 2.1 的比较主要有以下几点:

注 1: 条件 (A1), (A2) 与 Boente 和 Fraiman (1989) 文中的条件完全一致.

注 2: 本文定理 2.1 将结论推广到 α 混合样本下的一般权函数情形, 得出 α 混合样本下一般权函数稳健回归估计的强相合性.

定理证明

为了证明本文结论, 需要如下引理:

引理 3.1. 假设条件 (A1) 及 (A2) 成立, $F(y|X=x)$ 是一个条件分布函数, $F_n(y|X=x)$ 是一串条件分布函数序列. 如果

$$F_n(y|X=x) \rightarrow F(y|X=x), a.s. (P_x)$$

则

$$g_n(x) \rightarrow g(x), a.s. \quad n \rightarrow \infty.$$

其中 $g_n(x)$ 和 $g(x)$ 是定义 1.1 中的解.

引理 3.2. 令 $\{X_i: i \geq 1\}$ 为 α 混合序列, $|X_i| \leq b < \infty, a.s.$ 且 $E(X_i) = 0$, 假设存在 $\{K_n: n \geq 1\}$ 序列, 满足 $1 \leq K_n \leq n/2$, 则对 $\forall \varepsilon > 0$ 有

$$P\left(\left|\sum_{n=1}^{\infty} X_i\right| > n\varepsilon\right) \leq 4 \exp\left(-\frac{n\varepsilon^2}{12(3\sigma_n^2 + bK_n\varepsilon)}\right) + 36b\varepsilon^{-1}\alpha(K_n)$$

$$\text{其中 } \sigma_n^2 = n^{-1} \sum_{j=1}^{2m_n+1} E[U_j]^2, m_n = \left\lfloor \frac{n}{2K_n} \right\rfloor, U_j = \sum_{(j-1)K_n < i \leq jK_n} X_i, U_{2m_n+1} = \sum_{i=2m_nK_n+1}^n X_i$$

引理 3.3. 假设 $W_m(x)$ 是概率权函数且满足条件 (H3), 则 $\forall x \in \sigma(Y)$, 有 $\lim_{n \rightarrow \infty} P_n(A|X=x) = P(Y \in A|X=x), a.s.(P_X)$

且

$$P(\sup_y F_n(y|X=x) - F(y|X=x) \rightarrow 0) = 1$$

证明: 由 $\sum_{i=1}^n W_{ni}(x) = 1$, 有

$$\begin{aligned} P_n(A|X=x) - P(Y \in A|X=x) &= \sum_{i=1}^n \{W_{ni}(x)[I_A(Y_i) - P(Y_i \in A|X_i)]\} \\ &\quad + \sum_{i=1}^n \{W_{ni}(x)[P(Y_i \in A|X_i) - P(Y \in A|X=x)]\} \\ &= S_{1n} + S_{2n} \end{aligned}$$

记 $Z_i = I_A(Y_i) - P(Y_i \in A|X_i)$ 和 $\xi_{ni} = W_{ni}(x)Z_i$, 显然, $|Z_i| \leq 1$, 则

$$\begin{aligned} E[\xi_{ni}|X_i] &= E\{W_{ni}(x)[I_A(Y_i) - P(Y_i \in A|X_i)]|X_i\} \\ &= W_{ni}(x)[P(Y_i \in A|X_i) - P(Y_i \in A|X_i)] \\ &= 0 \end{aligned}$$

记 $U_j = \sum_{(j-1)K_n < i \leq jK_n} \xi_{ni}$, 则

$$\begin{aligned} E[U_j^2|X_i] &= E\left[\left|\sum_{(j-1)K_n < i \leq jK_n} \xi_{ni}\right|^2|X_i\right] \\ &\leq C \sum_{(j-1)K_n < i \leq jK_n} E[\xi_{ni}^4|X_i] \\ &\leq C \sum_{(j-1)K_n < i \leq jK_n} E[c_n^4|X_i] \\ &\leq Ck_n c_n^2 \end{aligned}$$

则

$$\sigma_n^2 = \frac{1}{n} \sum_{j=1}^{2m_n+1} E[U_j^2|X_i] \leq C \frac{1}{n} k_n c_n^2$$

取 $k_n = c_n^{-(1-\frac{2}{q})}$, 由引理 3.2, 对 $\forall \varepsilon > 0$, 有

$$\begin{aligned} P\left(\left|\sum_{i=1}^n \xi_{ni} > \varepsilon\right|\right) &\leq 4 \exp\left\{-\frac{\varepsilon^2}{12(3\sigma_n^2 + c_n k_n \varepsilon)}\right\} + 36b\varepsilon^{-1}\alpha(k_n) \\ &\leq 4 \exp\left\{-\frac{\varepsilon^2}{12(Ck_n c_n + c_n k_n \varepsilon)}\right\} + c_n k_n^{-\lambda} \\ &\leq 4 \exp\left\{-\frac{C\varepsilon^2 \log n}{c_n \log n}\right\} + Cc_n^{[\lambda(1-2/q)+1]} \cdot c_n^{-2} \\ &\leq 4 \exp\{-2 \log n\} + Cc_n^{[\lambda(1-2/q)-1]} \\ &\leq 4n^{-2} + Cc_n^{[\lambda(1-2/q)-1]} \end{aligned}$$

又, 设 $d=1$ 和 $h_n = cn^{-1/5}$, 则对任意的 $q > 2$ 都有

$$nh_n^{2d(1-1/d)} / \log n = Cn^{1-2(1-1/q)/5} / \log n = Cn^{3/5+2/(5q)} / \log n \rightarrow \infty$$

又由于 $\lambda > 6$, 所以可以取 q 充分大, 使 $(\lambda-1)/5 - 2\lambda/(5q) > 1$,

从而

$$\begin{aligned} \sum_{i=1}^{\infty} h_n^{d[\lambda(1-2/q)-1]} &= C \sum_{i=1}^{\infty} n^{-[\lambda(1-2/q)-1]/5} \\ &= C \sum_{i=1}^{\infty} n^{-[(\lambda-1)/5 - 2\lambda/(5q)]} \\ &< \infty \end{aligned}$$

故, 这里取 $c_n = Cn^{-1/5}$, 令 $\lambda > 6$, 可证 $\sum_{i=1}^n c_n^{[\lambda(1-2/q)-1]} < \infty$,

又 $\sum_{i=1}^n n^{-2} < \infty$, 所以

$$\sum_{n=1}^{\infty} P\left(\left|\sum_{i=1}^n \xi_{ni}\right| > \varepsilon\right) < \infty \quad (1)$$

由 Borel-cantelli 引理有

$$\sum_{i=1}^n \xi_{ni} \rightarrow 0, a.s. \quad (2)$$

即 $S_{1n} \rightarrow 0$.

下证 $S_{2n} \rightarrow 0, n \rightarrow \infty$. 记 $G_i = P(Y_i \in A|X_i) - P(Y \in A|X=x)$,

则

$$S_{2n} = \sum_{i=1}^n W_{ni}(x)(G_i - EG_i) + \sum_{i=1}^n W_{ni}(x)EG_i$$

由 L.Devroye (1981) 的引理 2.1, 有

$$E\left[\sum_{i=1}^n W_{ni}(x)|f(X_i) - f(X)|^p\right] \rightarrow 0, n \rightarrow \infty.$$

当 $P=1$, 可知 $E\left(\sum_{i=1}^n W_{ni}(x)G_i\right) \rightarrow 0, n \rightarrow \infty$, 故只需证

$$\sum_{i=1}^n W_{ni}(x)(G_i - EG_i) \rightarrow 0, n \rightarrow \infty.$$

$$\text{记 } U_i = W_{ni}(x)(G_i - EG_i), W_j = \sum_{(j-1)K_n < i \leq jK_n} U_i$$

显然 $EU_i = 0, |U_i| \leq c_n$, 由 $0 < W_{ni}(x) \leq \max_{1 \leq j \leq n} W_j \leq c_n$, 有

$$E|W_j|^2 = E\left|\sum_{(j-1)K_n < i \leq jK_n} U_i\right|^2 \leq C \sum_{(j-1)K_n < i \leq jK_n} (E|U_i|^q)^{2/q} \leq Ck_n c_n^2$$

则

$$\sigma_n^2 = n^{-1} \sum_{j=1}^{2m_n+1} E|W_j|^2 \leq n^{-1} Ck_n c_n^2$$

同理对 $\sum_{i=1}^n U_i$ 可得 (1) 和 (2) 式. 因此有 $S_{2n} \rightarrow 0$. 证毕.

证明定理 2.1. 由引理 3.1 和引理 3.3 立即可得定理 2.1 的结论, 证毕.

参考文献:

[1] 项萍. 混合样本下稳健非参数 NW 核回归估计 [J]. 广西师范大学, 2019 (01).

基金项目: 柳州工学院科学基金项目 (2019LSKY07)