

α -混合样本下 Pareto 分布参数的 经验 Bayes 双边检验*

黄娟¹, 李乃医¹, 韦盛学², 刘华祥¹

(1. 广东海洋大学理学院, 广东 湛江 524088;

2. 中山大学卫生统计学院 广东 广州 510080)

摘要: 在弱平稳 α -混合样本下, 利用核估计构造了 Pareto 分布参数的经验 Bayes 检验函数, 在适当的条件下证明了所提出的经验 Bayes 检验函数的渐近最优 (a. o.) 性, 并获得了它的收敛速度。

关键词: α -混合样本; 经验 Bayes 检验; 渐近最优性; 收敛速度

中图分类号: O212.1 **文献标识码:** A **文章编号:** 0529-6579 (2009) 05-0024-3

Empirical Bayes Two-sided Test for the Parameter of Pareto Distribution under α -Mixing Random Samples

HUANG Juan¹, LI Naiyi¹, WEI Shengxue², LIU Huaxiang¹

(1. School of Sciences, Guangdong Ocean University, Zhanjiang 524088, China;

2. School of Public Health, Sun Yat-sen University, Guangzhou 510080, China)

Abstract: By the kernel estimation under weakly stationary α -mixing random samples, the empirical Bayes test rule for the parameter of Pareto distribution is constructed. The asymptotically optimal property and convergence rates for the proposes EB test are obtained under suitable conditions.

Key words: α -mixing samples; empirical Bayes test; asymptotic optimality; convergence rates

EB 问题在文献中已研究的相当深入了^[1-2], 但几乎所有这些文献中的 EB 方法都是针对独立同分布 (iid) 样本而考虑的。本文利用弱平稳 α -混合样本^[3], 讨论 Pareto 分布参数的经验 Bayes 双边检验问题。

考虑如下模型^[4]: 设随机变量 X 的条件

$$f(x|\theta) = \frac{\theta\alpha^\theta}{x^{\theta+1}} \quad (1)$$

其中 θ 为参数, α 已知。样本空间为 $\Omega = \{x | x > \alpha\}$, 参数空间为 $\Theta = \{\theta | \theta > 0\}$ 。

讨论双侧检验问题: $H_0: \theta_1 \leq \theta \leq \theta_2 \Leftrightarrow H_1: \theta < \theta_1 \text{ or } \theta > \theta_2$ (2)

其中 θ_1 和 θ_2 为给定的常数, 如果取 $\theta_0 = \frac{\theta_1 + \theta_2}{2}$ 和 $\gamma_0 = \frac{\theta_2 - \theta_1}{2}$, 则双侧问题 (2) 等价于

$$H_0^*: |\theta - \theta_0| \leq \gamma_0 \Leftrightarrow H_1^*: |\theta - \theta_0| > \gamma_0 \quad (3)$$

关于假设检验问题 (3), 对 $i = 0, 1$, 设损失函数为

$$L_i(\theta, d_i) = (1-i)a[(\theta - \theta_0)^2 - \gamma_0^2]I_{[|\theta - \theta_0| > \gamma_0]} + ia[\gamma_0^2 - (\theta - \theta_0)^2]I_{[|\theta - \theta_0| \leq \gamma_0]}$$

此处 a 是正的常数, $d = \{d_0, d_1\}$ 是行动空间, d_0 表示接受 H_0^* , d_1 表示否定 H_0^* 。

假设参数 θ 的先验分布 $G(\theta)$ 且 $G(\theta)$ 未知。设随机化判决函数为

$$\delta(x) = P(\text{accept } H_0^* | X = x) \quad (4)$$

则 $\delta(x)$ 的风险函数为

$$R(\delta(x), G(\theta)) = \int_{\theta} \int_{\Omega} [L_0(\theta, d_0)f(x|\theta)\delta(x) + L_1(\theta, d_1)f(x|\theta)(1 - \delta(x))] dx dG(\theta) =$$

* 收稿日期: 2008-09-23

基金项目: 国家自然科学基金资助项目 (10661003); 广东海洋大学科研资助项目 (0612163)

作者简介: 黄娟 (1978 年生), 女, 讲师; E-mail: huangjuan401178@163.com

$$a \int \beta(x) \delta(x) dx + C_G \quad (5)$$

$$\text{此处 } C_G = \int_{\Theta} L_1(\theta, d_1) dG(\theta), \quad \beta(x) = \int_{\Theta} [(\theta - \theta_0)^2 - \gamma_0^2] f(x|\theta) dG(\theta) \quad (6)$$

随机变量 X 的边缘分布为

$$f_G(x) = \int_{\Theta} f(x|\theta) dG(\theta) = \int_{\Theta} \theta \alpha^{\theta} x^{-(\theta+1)} dG(\theta)$$

所以由 (5) 式经计算可得 $\beta(x) = A(x)f_G^{(2)}(x) + B(x)f_G^{(1)}(x) + Cf_G(x)$

其中 $A(x) = x^2, B(x) = (2\theta_0 + 3)x, C = (\theta_0 + 1)^2 - \gamma_0^2, f_G^{(1)}(x)$ 和 $f_G^{(2)}(x)$ 分别表示 $f_G(x)$ 的一、二阶导数。由 (4) 可知, Bayes 检验函数为:

$$\delta_G(x) = \begin{cases} 1, & \beta(x) \leq 0 \\ 0, & \beta(x) > 0 \end{cases} \quad (7)$$

$$\text{其 Bayes 风险为 } R(G) = \inf_{\delta} R(\delta, G) = R(\delta_G, G) = a \int \beta(x) \delta_G(x) dx + C_G \quad (8)$$

上述风险当先验分布 $G(\theta)$ 已知, 且 $\delta(x) = \delta_G(x)$ 是可以到达的。但此处 $G(\theta)$ 未知, 因而 $\delta_G(x)$ 无使用价值, 这就导致引入 EB 方法。

1 α -混合样本下 EB 检验函数的构造

本节在下列框架下构造了 EB 检验函数: 设 $X_1, X_2, \dots, X_n, X$ 为有共同的边缘概率密度函数为 $f_G(x)$, 设 $X_1, X_2, \dots, X_n$ 为弱平稳的 α -混合随机变量序列且 $X_1, X_2, \dots, X_n$ 与 X 相互独立, 其中 $X_1, X_2, \dots, X_n$ 为历史样本, X 为当前样本。本文中假定 $f_G(x) \in C_{s,\alpha}, x \in \mathbb{R}^1$, 其中 $C_{s,\alpha}$ 表示 $\mathbb{R}^1$ 中的一族概率密度函数, 其 s 阶导数存在, 连续且绝对值不超过 $\alpha, s \geq 3$ 为正整数。首先构造 $\beta(x)$ 的估计量。

令 $K_r(x)$ ($r = 0, 1, \dots, s-1$) 是有界 Borel 可测的函数, 在区间 $(0, 1)$ 之外为零且满足下列条件:

$$(A_1) \quad \frac{1}{t!} \int_0^1 v^t K_r(v) dv = \begin{cases} 1, & t = r \\ 0, & t \neq r, t = 0, 1, \dots, s-1 \end{cases}$$

本文中 α -混合序列的混合速度满足:

$$(A_2) \quad \sum_{i=1}^{\infty} \alpha(i) < \infty.$$

记 $f_G^{(r)}(x)$ 表示 $f_G(x)$ 的第 r 阶导数。对 $r = 0, 1, \dots, s-1$ 。定义 $f_G^{(r)}(x)$ 的核估计为:

$$f_n^{(r)}(x) = \frac{1}{nh_n^{1+r}} \sum_{j=1}^n K_r\left(\frac{x - X_j}{h_n}\right), \quad (9)$$

其中 $\{h_n\}$ 为正数序列, 且 $\lim_{n \rightarrow \infty} h_n = 0$ 。

所以 $\beta(x)$ 的估计量为 $\beta_n(x) = A(x)f_n^{(2)}(x) + B(x)f_n^{(1)}(x) + Cf_n(x)$ 。

故 EB 检验函数定义为

$$\delta_n(x) = \begin{cases} 1, & \beta_n(x) \leq 0 \\ 0, & \beta_n(x) > 0 \end{cases} \quad (10)$$

令 E_n 表示对随机变量 $X_1, X_2, \dots, X_n$ 的联合分布求均值, 则 $\delta_n(x)$ 的全面风险为

$$R(\delta_n(x), G) = a \int \beta(x) E_n[\delta_n(x)] dx + C_G \quad (11)$$

若 $\lim_{n \rightarrow \infty} R(\delta_n, G) = R(\delta_G, G)$, 称 $\{\delta_n(x)\}$ 为 a. o. 的 EB 检验函数; 若 $R(\delta_n, G) - R(\delta_G, G) = O(n^{-q}), q > 0$, 则称 EB 检验函数 $\{\delta_n(x)\}$ 的收敛速度为 $O(n^{-q})$ 。

为导出 $\{\delta_n(x)\}$ 的 a. o. 性和收敛速度, 首先证明以下引理。

本文中令 $c, c_1, c_2, c_3, \dots, c_6$ 表示不同的常数, 即使在同一表达式中它们也可取不同的值。

引理 1^[5] 设 $\{X_i, i \geq 1\}$ 是 α -混合随机变量序列, $|X_i| < c_1, |X_j| < c_2, i, j \geq 1$, 则

$$|EX_i X_{i+j} - EX_i EX_{i+j}| \leq 4c_1 c_2 \alpha(j)$$

引理 2^[6] 设 $f_n^{(r)}(x)$ 由 (9) 式定义, 其中 $X_1, X_2, \dots, X_n$ 为弱平稳 α -混合样本序列。假定条件 (A1) - (A2) 成立, 对 $\forall x \in \Omega$,

(1) 当 $f_n^{(r)}(x)$ 关于 x 连续, 则当 $\lim_{n \rightarrow \infty} h_n = 0$ 且 $\lim_{n \rightarrow \infty} nh_n^{2r+2} = \infty$ 时, 有

$$\lim_{n \rightarrow \infty} E |f_n^{(r)}(x) - f_G^{(r)}(x)|^2 = 0$$

(2) 当 $f_G(x) \in C_{s,\alpha}$, 当取 $h_n = n^{-\frac{1}{2s+2}}$ 时, 对于 $0 < \lambda \leq 1$ 有

$$E |f_n^{(r)}(x) - f_G^{(r)}(x)|^{2\lambda} \leq c \cdot n^{-\frac{\lambda(s-r)}{1+s}}$$

注由引理 1 和类似文献 [4] 中引理 2.2 的证明方法可得。

引理 3^[7] 令 $R(\delta_G, G)$ 和 $R(\delta_n, G)$ 分别由 (8) 和 (11) 式给出, 则

$$0 \leq R(\delta_n, G) - R(\delta_G, G) \leq a \int_{\Omega} |\beta(x)| p(|\beta_n(x) - \beta(x)| \geq |\beta(x)|) dx$$

2 EB 检验函数的渐近性及其收敛速度

定理 1 设 $\delta_n(x)$ 由 (10) 式定义, 假定条件

(A1) - (A2) 成立。若 (1) $\{h_n\}$ 为正数序列, 且满足 $\lim_{n \rightarrow \infty} h_n = 0$, $\lim_{n \rightarrow \infty} nh_n^6 = \infty$, (2) $\int_{\theta} \theta^2 dG(\theta) < \infty$, (3) $f_G^{(2)}(x)$ 为 x 的连续函数, 则有 $\lim_{n \rightarrow \infty} R(\delta_n, G) = R(\delta_G, G)$ 。

证明 由引理 3 得

$$0 \leq R(\delta_n, G) - R(\delta_G, G) \leq a \int_{\Omega} |\beta(x)| p(|\beta_n(x) - \beta(x)| \geq |\beta(x)|) dx$$

记 $\Psi_n(x) = |\beta(x)| p(|\beta_n(x) - \beta(x)| \geq |\beta(x)|)$, 则有 $\Psi_n(x) \leq |\beta(x)|$ 。

又由 (5) 式和 Fubini 定理得

$$\int_{\Omega} |\beta(x)| dx \leq |\theta_0^2 - \gamma_0^2| + \int_{\theta} \theta^2 dG(\theta) + 2|\theta_0| \int_{\theta} \theta dG(\theta) < \infty$$

故由控制收敛定理可知 $0 \leq \lim_{n \rightarrow \infty} R(\delta_n, G) -$

$$R(\delta_G, G) \leq \int_{\Omega} [\lim_{n \rightarrow \infty} \Psi_n(x)] dx \quad (11)$$

所以要使定理 1 成立, 要证明 $\lim_{n \rightarrow \infty} \Psi_n(x) = 0$ 对 a. s. x 成立即可。

由 Markov 不等式和 Jensen 不等式可得

$$\begin{aligned} \Psi_n(x) &\leq |A(x)| [E|f_n^{(2)}(x) - f_G^{(2)}(x)|^2]^{1/2} + \\ &|B(x)| [E|f_n^{(1)}(x) - f_G^{(1)}(x)|^2]^{1/2} + \\ &|C| [E|f_n(x) - f_G(x)|^2]^{1/2} \quad (12) \end{aligned}$$

再由引理 2 (1) 和引理 3 可知, 对任意固定的 $x \in \Omega$, 当 $r = 0, 1, 2$ 和 $\lambda = 1$ 有

$$\begin{aligned} 0 &\leq \lim_{n \rightarrow \infty} \Psi_n(x) \leq |A(x)| \lim_{n \rightarrow \infty} [E|f_n^{(2)}(x) - f_G^{(2)}(x)|^2]^{1/2} + \\ &|B(x)| \lim_{n \rightarrow \infty} [E|f_n^{(1)}(x) - f_G^{(1)}(x)|^2]^{1/2} + |C| \\ &[\lim_{n \rightarrow \infty} E|f_n(x) - f_G(x)|^2]^{1/2} = 0 \end{aligned}$$

将 (12) 代入 (11) 式定理得证。

定理 2 设 $\delta_n(x)$ 由 (10) 式定义, 假定条件 (A₁) - (A₂) 成立, 且 $f_G(x) \in C_{s,a}$, 若 $0 < \lambda \leq 1$, 当取 $h_n = n^{-\frac{1}{2s+2}}$ 时, 有 (4) $\int_{\Omega} x^{m\lambda} |\beta(x)|^{1-\lambda} dx < \infty$, $m = 0, 1, 2$ 。则有

$$R(\delta_n, G) - R(\delta_G, G) = O(n^{-\frac{\lambda(s-2)}{2s+2}}), \text{ 其中 } s \geq 3 \text{ 为正整数。}$$

证明 由引理 3 和 Markov 不等式得

$$0 \leq R(\delta_n, G) - R(\delta_G, G) \leq c_1 \int_{\Omega} |\beta(x)|^{1-\lambda} |A(x)|^{\lambda} E|f_n^{(2)}(x) - f_G^{(2)}(x)|^{\lambda} dx +$$

$$\begin{aligned} &c_2 \int_{\Omega} |\beta(x)|^{1-\lambda} |B(x)|^{\lambda} E|f_n^{(1)}(x) - f_G^{(1)}(x)|^{\lambda} dx + \\ &c_3 \int_{\Omega} |\beta(x)|^{1-\lambda} |C|^{\lambda} E|f_n(x) - f_G(x)|^{\lambda} dx = \\ &A_n + B_n + C_n \quad (13) \end{aligned}$$

由引理 2 和条件 (4) 知

$$\begin{aligned} A_n &\leq c_1 n^{-\frac{\lambda(s-2)}{2s+2}} \int_{\Omega} |\beta(x)|^{1-\lambda} |A(x)|^{\lambda} dx \\ &\leq c_4 n^{-\frac{\lambda(s-2)}{2s+2}} \quad (14) \end{aligned}$$

$$\begin{aligned} B_n &\leq c_2 n^{-\frac{\lambda(s-1)}{2s+2}} \int_{\Omega} x^{\lambda} |\beta(x)| |B(x)|^{\lambda} dx \\ &\leq c_5 n^{-\frac{\lambda(s-1)}{2s+2}} \quad (15) \end{aligned}$$

$$\begin{aligned} C_n &\leq c_3 n^{-\frac{\lambda s}{2s+2}} \int_{\Omega} |\beta(x)|^{1-\lambda} |C|^{\lambda} dx \\ &\leq c_6 n^{-\frac{\lambda s}{2s+2}} \quad (16) \end{aligned}$$

$$\text{将 (14) - (16) 代入 (13) 得 } R(\delta_n, G) - R(\delta_G, G) = O(n^{-\frac{\lambda(s-2)}{2s+2}})$$

注 1 当 $\lambda \rightarrow 1$ 时, $s \rightarrow \infty$ 时, $O(n^{-\frac{\lambda(s-2)}{2s+2}})$ 可任意接近 $O(n^{-\frac{1}{2}})$ 。

参考文献:

- [1] KARUNAMUNI R J, SINGH R S, SHANG S. On empirical Bayes estimation in the location family[J]. Journal of Nonparametric Statistics, 2002, 14(4): 435 - 448.
- [2] LIANG T. An improved empirical Bayes test for positive exponential families[J]. Statistica neerlandica, 2002, 56(3): 346 - 361.
- [3] CHAI G X, WU Y Q. The nearest neighbor estimation in contaminated linear model under dependent samples[J]. Acta Mathematicae Applicatae Sinica, 2006, 29(3): 542 - 554.
- [4] PARETO V. Cour economic politique [M]. Lausanne and pairs, Rouge and Cie, 1987.
- [5] LU Z R. Limit theorems for sums of dependent random variables[J]. Chinese Journal of Appl Prob and Stat, 1993, 9(1): 94 - 103.
- [6] LI N Y, HUANG J. Empirical Bayes two-sided test problem for the parameter of linear exponential distribution; in the case of ϕ -mixing random samples[J]. Appl Math J Chinese Univ, 2008, 23(2): 213 - 218.
- [7] JOHNS M V Jr, VAN RYZIN J. Convergence rates in empirical Bayes two-action problems 2: continuous case [J]. Ann Math Statist, 1972, 43: 934 - 937.