

# 偏态分布截断参数的经验Bayes检验

刘蕊<sup>1,2</sup>, 谭燕<sup>1,2</sup>, 吴刘仓<sup>1,2\*</sup>

(1. 昆明理工大学理学院, 云南昆明 650504;

2. 昆明理工大学应用统计学研究中心, 云南昆明 650504)

**摘要:** 偏态分布是对称分布的一种推广, 在实际生活中应用广泛, 其中截断参数对界定偏态分布的边界具有重要意义. 文中基于经验Bayes检验方法, 研究了偏态分布截断参数的假设检验问题. 考虑普通Bayes检验中先验密度的未知性和不确定性, 利用递归核估计的密度函数代替未知的先验密度函数. 定义检验的加权线性损失函数, 从而更好地刻画了决策风险. 在给定条件下证明了所提出检验函数的渐近最优性, 同时给出确定的收敛速度. 最后通过实例验证了文中的理论结果.

**关键词:** 经验Bayes检验; 偏态分布; 递归核估计; 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:** 1000-4424(2024)01-0013-15

## §1 引言

统计分析过程中, 通常假设数据具有对称性, 而实际生活中遇到的数据往往不具有这类性质, 例如金融学, 生物医学以及社会科学等研究领域的的数据大多呈偏斜形态. 若忽略数据偏斜的特征, 仍用对称分布进行分析则可能会造成结果偏移甚至错误结论. 此时用偏态分布拟合这类数据, 能更合理地刻画其相关特征. 在实际应用中, 随机变量的定义范围直接影响偏态分布的取值, 为更准确地获得偏态分布的定义区间, 对偏态分布的截断参数进行假设检验具有现实意义.

偏态分布的研究领域内, 现已有很多关于偏斜数据非对称性质的研究. 在偏态分布的统计性质方面, Pearson<sup>[1]</sup>研究了非对称概率密度函数曲线的广义形式及拟合性质; Azzalini<sup>[2]</sup>提出偏正态分布的具体概念和模型, 并对其相关性质进行研究; 李光兴等<sup>[3]</sup>引入截断参数后给出新的偏态密度函数. 基于此, Jones等<sup>[4]</sup>提出一种偏t分布, 并对其相关性质与参数推断进行研究; Nekoukhou等<sup>[5]</sup>研究了几类Laplace分布相关偏斜分布的性质特征, 得到截断偏Laplace分布的相关性质; 杨昕<sup>[6]</sup>提出一种新的偏Logistic分布, 并讨论了其在股市风险中的应用问题; 谭佳玲等<sup>[7]</sup>基于随机缺失的情形, 研究了偏正态分布相关模型的参数估计问题. 在偏态分布的

收稿日期: 2022-10-05 修回日期: 2023-04-10

\*通讯作者, E-mail: wuliucang@163.com

基金项目: 国家自然科学基金(12261051; 11861041); 昆明理工大学学术科技创新基金(2022KJ150); 昆明理工大学哲学社会科学科研创新团队(CXTD2023005)

统计诊断方面, Nadarajah<sup>[8]</sup>针对偏Logistic分布研究了其相关的统计性质, 信息熵以及统计诊断问题; Azzalini等<sup>[9]</sup>在偏t分布中引入调节参数, 提出基于惩罚似然思想的统计诊断方法; Kahrari等<sup>[10]</sup>针对多元偏斜正态-柯西分布的尺度混合模型给出了极大似然估计的EM算法; 曹幸运等<sup>[11]</sup>在偏正态分布下众数回归模型中, 基于Pena距离研究了模型的统计诊断问题; 叶仁道等<sup>[12]</sup>研究了偏正态单向分类模型中方差分量的单边假设检验和区间估计问题.

随着检验方法的成熟, Bayes检验逐渐发展为一种热门的检验方法, 与普通Bayes检验相比, 经验Bayes方法的基本思想是利用抽样样本估计总体先验分布, 从而代替真实先验分布进行Bayes统计与决策的分析. Robbins<sup>[13]</sup>最早提出忽略参数本身是一个随机变量的经验Bayes基本思想; John<sup>[14]</sup>基于经验Bayes检验方法研究了两类连续型单参数指数分布族中参数的检验问题. 在不同分布的经验Bayes检验方面, 陈家清等<sup>[15]</sup>利用核密度估计代替概率密度函数, 给出了线性指数分布参数的经验Bayes检验研究理论和结果证明; 周雁等<sup>[16]</sup>引入加权线性损失函数对刻度指数分布参数的经验Bayes检验问题展开了研究; Tansey等<sup>[17]</sup>基于多实验研究数据提出了一种双经验Bayes检验方法. 在经验Bayes检验函数的收敛速度方面, Liang<sup>[18]</sup>研究了正态分布中均值的经验Bayes检验函数的最优收敛速度, 并得到其最优收敛速度为 $O(n^{-1})$ ; Chen<sup>[19]</sup>通过构造经验Bayes检验组间测试序列, 研究了均匀分布参数的检验问题, 并得到检验函数的收敛速度为 $O(n - 2(r + \alpha)[2(r + \alpha) + 1])$ ; 陈家清等<sup>[20]</sup>基于离散型单参数指数分布族的研究得到参数双侧经验Bayes检验函数的精确收敛速度可达 $O(n^{-1})$ ; Yuan等<sup>[21]</sup>研究Gamma分布位置参数的双侧经验Bayes问题, 得到其收敛速度 $O(n^{-\frac{1}{2}})$ .

在经验Bayes领域用抽样概率密度去估计总体概率密度的过程中, 核密度估计是最常用的方法. Parzen<sup>[22]</sup>最早提出了密度函数的核估计表达; 随着Wolverton等<sup>[23]</sup>提出了密度函数的递归核估计, 核估计的计算效率得到极大改善. 而运用递归核估计时重在选择合适的核函数和窗宽, Terrell<sup>[24]</sup>提出了过平滑窗宽的条件和准则, 该准则在后续研究中被广泛使用. 值得注意的是, 韦来生<sup>[25]</sup>研究了NA随机变量密度函数的核估计的相合性, 并给出详细证明; Alexandre<sup>[26]</sup>提出了Solve-the-Equation的方法, 其通过控制积分均方误差的大小得到最优窗宽值, 在估计过程中极大地简化了计算.

基于上述文献可知, 经验Bayes检验方法是研究偏态分布参数检验的可行方法. 但是在偏态分布截断参数的分析中, 对其进行假设检验的研究并不多, 而考虑在高维数据样本条件下, 普通核密度估计不论是在计算量还是估计精度上都存在弊端. 因此为提高密度估计的效率和精度, 本文引入递归核密度函数估计未知的先验密度函数, 利用加权线性损失函数, 有效地刻画决策的风险, 通过实例分析证明该检验的渐进最优性, 并给出收敛速度的阶数.

本文结构安排如下: §2介绍偏态分布的相关性质, 并给出截断参数 $c$ 的条件密度函数; §3给出了偏态分布中截断参数 $c$ 的Bayes检验函数; §4利用递归核估计方法得到了参数的经验Bayes检验函数; §5证明了所提出经验Bayes检验函数的渐近最优性, 并给出具体的收敛速度; §6针对双侧假设检验, 结合递归核密度估计对截断参数进行经验Bayes检验; §7通过实例验证了该检验的渐近最优性, 并得到其收敛速度.

## §2 偏态分布

根据文献[3]给出的偏态密度函数, 随机变量 $X$ 服从截断参数为 $c$ , 尺度参数为 $\sigma^2$ 的偏态分布,

考虑其概率密度函数为

$$f(x|c) = \frac{2}{\sigma}(c-x) \exp\left\{-\frac{(c-x)^2}{\sigma}\right\}, \quad (1)$$

其中,  $\sigma > 0$ , 截断参数  $c$  可在  $(0, +\infty)$  上取值, 随机变量  $x$  的取值区间为  $(-\infty, c)$ . 偏态的形态主要由截断参数  $c$  决定, 且截断参数  $c$  决定了偏态分布的边界定义. 如何确定偏斜数据的总体分布区间便与参数  $c$  有密切关系.

采用矩母函数的方法能得到(1)式中偏态分布的期望和方差分别为

$$E(X) = c - \frac{\sqrt{\sigma\pi}}{2}; \quad \text{Var}(X) = \sigma - \frac{\sigma\pi}{4}.$$

综合上述偏态分布的性质分析可知, 密度函数  $f(x)$  在  $(-\infty, c)$  上仅存在一个极大值点, 其极大值点为  $x_0 = c - \sqrt{\frac{\sigma}{2}}$ , 最大值为  $f(x_0) = \sqrt{\frac{2}{\sigma}} \exp(-\frac{1}{2})$ . 又由于  $E(X) = c - \frac{\sqrt{\sigma\pi}}{2} < c - \sqrt{\frac{\sigma}{2}}$ , 可以得到该偏态分布的密度函数  $f(x)$  呈现一种单峰的右偏态. 故通过偏态分布均值和方差得到

$$\sigma = \frac{\text{Var}(X)}{1 - \frac{\pi}{4}}; \quad c = E(X) + \frac{\sqrt{\sigma\pi}}{2}.$$

现给出此类偏态密度函数的曲线图像, 如下图1和图2所示.

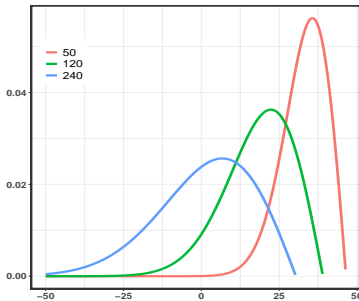


图1 期望相等时不同方差的密度函数曲线

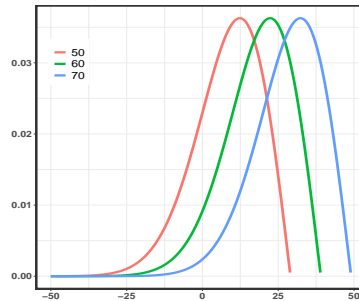


图2 方差相等时不同期望的密度函数曲线

图1为考虑期望  $E(X) = 60$ , 方差  $\text{Var}(X)$  分别取 50, 120, 240 时, 三条离散程度不等的密度函数  $f(x)$  的曲线. 随着方差  $\text{Var}(X)$  的增大, 参数  $\sigma$  的值急剧增大, 截断参数  $c$  的值则轻微减小; 其中, 密度函数曲线随方差  $\text{Var}(X)$  的增大而趋于平缓. 图2为考虑方差  $\text{Var}(X) = 120$ , 期望  $E(X)$  分别取 50, 60, 70 时, 三条同一水平上但位置不同的密度函数  $f(x)$  的曲线. 由于具有相同的方差  $\text{Var}(X)$  和参数  $\sigma$ , 三条曲线的极大值均相等; 而随着期望  $E(X)$  的增大, 截断参数  $c$  的值逐渐增大, 密度函数曲线也随期望  $E(X)$  的增大而向右平移.

为此(1)式中概率密度函数的形式有助于完成对偏态分布边界的精确定义, 故本文在该偏态分布密度函数的基础上对截断参数  $c$  进行检验分析.

### §3 Bayes检验函数

对于给定随机变量  $X$  服从偏正态分布时的截断参数  $c$ , 假设  $\sigma$  为已知常数, 有条件密度函数

$$f(x|c) = \frac{2}{\sigma}(c-x) \exp\left\{-\frac{(c-x)^2}{\sigma}\right\}. \quad (2)$$

令其边缘密度函数为

$$f(x) = \int_C f(x|c)dG(c) = \int_C \frac{2}{\sigma}(c-x) \exp\left\{-\frac{(c-x)^2}{\sigma}\right\} dG(c),$$

其中  $x \in \Omega : (0, c)$ ,  $c \in C : (0, +\infty)$ ;  $G(c)$  为参数  $c$  未知的非退化先验分布, 且  $\int_{\Omega} f(x|c)dx = 1$ ,  $dG(c) = g(c)dc$ .

现考虑在模型(2)中对参数  $c$  的单侧假设检验问题展开研究, 对于给定的常数  $c_0$  有

$$H_0 : c \leq c_0 \quad vs \quad H_1 : c > c_0.$$

针对该检验问题, 取其加权线性损失函数为

$$L_m(c, d_m) = (1-m)q(c-c_0)I(c > c_0) + mq(c_0-c)I(c \leq c_0), \quad m = 0, 1,$$

其中,  $D = \{d_0, d_1\}$  表示判决空间;  $d_i$  为接受  $H_i$  的决策,  $i = 0, 1$ ;  $I(A)$  表示  $A$  发生的示性函数.

设随机化判决函数为  $\delta_G(x) = P\{\text{接受}H_0 | X = x\}$ , 则在先验分布  $G(c)$  下随机化判决函数  $\delta(x)$  的风险函数为

$$\begin{aligned} R(\delta_G(x), G(c)) &= \int_C \int_{\Omega} [L_0(c, d_0) f(x|c) \delta_G(x) + L_1(c, d_1) f(x|c)(1 - \delta_G(x))] dx dG(c) \\ &= \int_C \int_{\Omega} [L_0(c, d_0) - L_1(c, d_1)] f(x|c) \delta_G(x) dx dG(c) + \int_C \int_{\Omega} L_1(c, d_1) f(x|c) dx dG(c) \\ &= q \int_{\Omega} \beta(x) \delta_G(x) dx + C_G, \end{aligned} \tag{3}$$

其中  $\beta(x) = \int_C (c - c_0) f(x|c) dG(c)$ ,  $C_G = \int_C L_1(c, d_1) dG(c)$ .

记  $f^{(i)}(x)$  为  $f(x)$  关于  $x$  的  $i$  阶导数, 则  $f(x)$  的一阶导数为

$$f^{(1)}(x) = \int_C \frac{2}{\sigma} \exp\left\{-\frac{(c-x)^2}{\sigma}\right\} \left[\frac{2(c-x)^2}{\sigma} - 1\right] dG(c).$$

由(3)式结果经计算有

$$\beta(x) = \int_C (c - c_0) f(x|c) dG(c) = [\gamma(x) - c_0] f(x), \tag{4}$$

记  $\gamma(x) = [\int_C c f(x|c) dG(c)] / [\int_C f(x|c) dG(c)]$ .

令  $P_G(x) = \int_C \exp\left\{-\frac{(c-x)^2}{\sigma}\right\} dG(c)$ , 又

$$P_G^{(1)}(x) = \int_C \frac{2}{\sigma} (c-x) \exp\left\{-\frac{(c-x)^2}{\sigma}\right\} dG(c) = f(x),$$

故  $\int_{-\infty}^x f(y) dy = P_G(x)$ . 进一步可将  $\beta(x)$  表示为

$$\beta(x) = N_1 P_G(x) - c_0 f(x) + N_2 f^{(1)}(x), \tag{5}$$

其中  $N_1 = \frac{c_0}{c_0-x}$ ,  $N_2 = \frac{c_0 \sigma}{2(c_0-x)}$ .

针对 $\gamma(x)$ , 可求得其关于 $x$ 的一阶导数为

$$\begin{aligned}\gamma^{(1)}(x) &= \left[ \frac{\int_C c f(x|c) dG(c)}{\int_C f(x|c) dG(c)} \right]' \\ &= \frac{\int_C \frac{2c}{\sigma} \exp\left\{-\frac{(c-x)^2}{\sigma}\right\} \left[\frac{2(c-x)^2}{\sigma} - 1\right] dG(c) \int_C \frac{2}{\sigma} (c-x) \exp\left\{-\frac{(c-x)^2}{\sigma}\right\} dG(c)}{\left[\int_C f(x|c) dG(c)\right]^2} \\ &= \frac{\int_C \frac{2c}{\sigma} (c-x) \exp\left\{-\frac{(c-x)^2}{\sigma}\right\} dG(c) \int_C \frac{2}{\sigma} \exp\left\{-\frac{(c-x)^2}{\sigma}\right\} \left[\frac{2(c-x)^2}{\sigma} - 1\right] dG(c)}{\left[\int_C f(x|c) dG(c)\right]^2}.\end{aligned}$$

由Cauchy-Schwarz不等式可得:  $\gamma^{(1)}(x) \leq 0$ , 所以 $\gamma(x)$ 在 $\Omega$ 上是连续的减函数. 现假设 $\gamma(x)$ 满足

**条件I**  $\lim_{x \rightarrow c^-} \gamma(x) < c_0 < \lim_{x \rightarrow 0^+} \gamma(x)$ .

则由介值定理可知, 存在 $a_G \in (0, c)$ , 使得 $\gamma(a_G) = c_0$ , 又由(4)式得

$$\begin{aligned}\beta(x) \leq 0 &\Leftrightarrow \gamma(x) - c_0 \leq 0 \Leftrightarrow \gamma(x) \leq c_0 \Leftrightarrow x \geq a_G, \\ \beta(x) > 0 &\Leftrightarrow \gamma(x) - c_0 > 0 \Leftrightarrow \gamma(x) > c_0 \Leftrightarrow x < a_G.\end{aligned}\tag{6}$$

由(6)式可知Bayes判决函数为

$$\delta_G(x) = \begin{cases} 1, & \beta(x) \leq 0 \\ 0, & \beta(x) > 0 \end{cases} = \begin{cases} 1, & x \geq a_G, \\ 0, & x < a_G.\end{cases}\tag{7}$$

则 $\delta_G(x)$ 的Bayes风险为

$$R_G(\delta_G, G) = \inf_{\delta_G(x)} R(\delta_G(x), G(c)) = q \int_{\Omega} \beta(x) \delta_G(x) dx + C_G,\tag{8}$$

由于这里参数 $c$ 的先验分布 $G(c)$ 未知, 故导致Bayes检验函数在实际问题中没有运用价值, 因而 $\delta_G(x)$ 没法应用, 于是考虑引入下一章的经验Bayes方法.

## §4 经验Bayes检验函数

本节将在以下框架的基础上构造经验Bayes函数: 设 $X_1, X_2, \dots, X_n, X_{n+1}$ 为独立同分布的随机变量序列, 其共同的边缘概率密度函数为 $f_G(x)$ ,  $X_1, X_2, \dots, X_n$ 为历史样本,  $X_{n+1}$ 是当前样本. 本文中还需假定以下条件.

**条件II**  $f_G(x) \in C_{k,\beta}$ ,  $X \in R$ .  $C_{k,\beta}$ 是 $R$ 上的一族密度函数, 它的 $k$ 阶导数存在且连续.

统计学研究中用概率密度函数描述分布函数的特征, 而当概率密度函数未知时, 通常考虑非参数估计的方法. 故为构造经验Bayes检验函数, 本文将引入如下递归核估计的方法. 首先需要假设先验分布和核函数满足条件.

**条件III** 假设先验分布 $G(c)$ 属于以下分布族, 即

$$G(c) \in \varphi = \left\{ G \mid \int_C c dG(c) < \infty, A_1 < a_G < A_2 \right\},$$

其中 $A_1$ 和 $A_2$ 为给定常数, 且满足 $0 < A_1 < A_2 < c$ .

**条件IV** 令核函数 $D_i(x)$  ( $i = 0, 1$ ) 是 $R$ 上有界的Borel可测函数, 并在 $R$ 上可微, 在区间 $(0, 1)$ 之外其取值为零, 且 $\frac{1}{h} \int_0^1 x^t D_i(x) dx = \begin{cases} (-1)^t, & t = i; \\ 0, & t \neq i, \end{cases}$  其中 $t = 0, 1, \dots, k-1$ .

于是根据Wolverton等<sup>[23]</sup>于1969年所提出的核估计方法定义 $f^{(i)}(x)$ 的递归核密度估计为

$$f_n^{(i)}(x) = \frac{1}{n} \sum_{j=1}^n \frac{1}{h_j^{i+1}} D_i\left(\frac{X_j - x}{h_j}\right), \quad (9)$$

其中 $h_n$ 为正整数序列, 且 $\lim_{n \rightarrow \infty} h_n = 0$ . 该递归核密度估计与普通核密度估计最大的区别在于其递归性质, 表示为

$$f_n^{(i)}(x) = \frac{n-1}{n} f_{n-1}^{(i)}(x) + \frac{1}{nh_j^{i+1}} D_i\left(\frac{X_i - x}{h_j}\right).$$

由这种递归性质可知, 在估计未知密度函数 $f^{(i)}(x)$ 的过程中,  $f_n^{(i)}(x)$ 可以在原始样本的核密度估计基础上, 添加新增样本的信息, 而无需再重新对全体样本进行核密度估计. 这在大样本数据的条件下或是计算机编程的计算中, 对计算效率都有明显的改进. 除此之外, 递归核密度估计针对不同区间内样本点选择了不同的窗宽 $h_j$ . 在实际应用中, 即可根据条件的变化定义适应窗宽 $h_j$ 的选取准则, 从而实现避免过平滑和过锐化的作用.

本文采用递归核密度估计进行检验, 参考文献<sup>[26]</sup>的方法, 能够找到合适上述条件III和条件IV的核函数 $D_i(x)$ . 由于 $P_G(x) = \int_{-\infty}^x f(y) dy = E\{I(X_i \leq x)\}$ , 因此不失一般性, 定义 $P_G(x)$ 的估计量为

$$P_n(x) = \frac{1}{n} \sum_{i=1}^n I(X_i \leq x). \quad (10)$$

于是可定义 $\beta(x)$ 的估计量为

$$\beta_n(x) = N_1 P_n(x) - c_0 f_n(x) + N_2 f_n^{(1)}(x). \quad (11)$$

与(7)式同理, 当 $x \in (A_1, A_2)$ 时, 可构造参数 $c$ 的经验Bayes判决函数为

$$\delta_n(x) = \begin{cases} 1, & x \geq A_1 \text{ 或 } \beta_n(x) \leq 0, \\ 0, & x < A_2 \text{ 或 } \beta_n(x) > 0. \end{cases}$$

则 $\delta_n(x)$ 的全面Bayes风险为

$$R_n(\delta_n, G) = q \int_{\Omega} \beta_n(x) E_n[\delta_n(x)] dx + C_G', \quad (12)$$

其中 $E_n$ 表示联合分布 $(X_1, X_2, \dots, X_n)$ 的均值.

设 $F$ 是参数 $c$ 的先验分布族, 对 $\forall G(c) \in F$ , 有 $\lim_{n \rightarrow \infty} R_n(\delta_n, G) = R_G(\delta_G, G)$ , 则称 $\delta_n(x)$ 为渐近最优的经验Bayes函数; 若 $R_n(\delta_n, G) - R_G(\delta_G, G) = O(n^{-p})$  ( $p > 0$ ), 那么称 $\delta_n$ 的收敛速度的阶数为 $O(n^{-p})$ .

下面为验证本文所提 $\delta_n$ 的渐近最优性和收敛速度, 首先给出以下三条引理.

引理1 令 $R_G, R_n$ 分别由(8)式和(12)式定义, 则

$$0 \leq R_n(\delta_n, G) - R_G(\delta_G, G) \leq q \int_{\Omega} |\beta(x)| P[|\beta_n(x) - \beta(x)| \geq |\beta(x)|] dx.$$

证 参见文献[14]中有关引理1的证明.

引理2  $f_n^{(i)}(x)$  ( $i = 0, 1$ )由(9)式定义,  $X_1, X_2, \dots, X_n$ 是独立同分布的随机样本序列, 假设条件II及条件III成立, 当 $0 < \theta < 1$ ,  $h_n = n^{-\frac{1}{2(k-1)}}$ 时, 存在正整数 $M$ , 使得

$$E_n |f_n^{(i)}(x) - f^{(i)}(x)|^{2\theta} \leq M n^{-\frac{\theta(k-i-1.5)}{k-1}}$$

成立.

证 由C-R不等式和Jensen不等式得

$$E_n |f_n^{(i)}(x) - f^{(i)}(x)|^{2\theta} \leq 2|E_n f_n^{(i)}(x) - f^{(i)}(x)|^{2\theta} + 2[\text{Var} f^{(i)}(x)]^\theta = 2(I_{1i}^{2\theta} + I_{2i}^\theta). \quad (13)$$

由假定条件IV和核密度估计的性质可得

$$\begin{aligned} E_n f_n^{(i)}(x) &= \frac{1}{n} \sum_{j=1}^n \frac{1}{h_j^{i+1}} E_n [D_i(\frac{X_j - x}{h_j})] = \frac{1}{n} \sum_{j=1}^n \frac{1}{h_j^{i+1}} \int_x^{x+h_j} D_i(\frac{y-x}{h_j}) f(y) dy \\ &= \frac{1}{n} \sum_{j=1}^n \frac{1}{h_j^i} \int_0^1 D_i(v) f(x + h_j v) dv, \end{aligned} \quad (14)$$

其中将 $f(x + h_j v)$ 在 $x$ 处作Taylor展开到第 $k+1$ 项有

$$f(x + h_j v) = f(x) + \frac{f^{(1)}(x)}{1!} (h_j v) + \dots + \frac{f^{(k)}(x + \xi h_j v)}{k!} (h_j v)^k, \quad (15)$$

将(15)式带入(14)式有

$$\begin{aligned} E_n f_n^{(i)}(x) &= \frac{1}{n} \sum_{j=1}^n \frac{1}{h_j^i} [h_j^i f^{(i)}(x) + \int_0^1 D_i(v) \frac{f^{(k)}(x + \xi h_j v)}{k!} (h_j v)^k dv] \\ &= \frac{1}{n} \sum_{j=1}^n \frac{1}{h_j^i} [h_j^i f^{(i)}(x) + h_j^k \int_0^1 D_i(v) \frac{f^{(k)}(x + \xi h_j v)}{k!} v^k dv] \\ &= f^{(i)}(x) + \frac{1}{n} \sum_{j=1}^n h_j^{k-i} \int_0^1 D_i(v) \frac{f^{(k)}(x + \xi h_j v)}{k!} v^k dv. \end{aligned}$$

又由 $f(x) \in C_{k,\beta}$ 及核函数的有界性可知

$$\begin{aligned} I_{1i} &= |E_n f_n^{(i)}(x) - f^{(i)}(x)| = \left| \frac{1}{n} \sum_{j=1}^n h_j^{k-i} \int_0^1 D_i(v) \frac{f^{(k)}(x + \xi h_j v)}{k!} v^k dv \right| \\ &\leq \frac{1}{n} \sum_{j=1}^n h_j^{k-i} \int_0^1 |D_i(v)| v^k \left| \frac{f^{(k)}(x + \xi h_j v)}{k!} \right| dv \leq M \frac{1}{n} \sum_{j=1}^n h_j^{k-i} \leq M h_n^{k-i}. \end{aligned} \quad (16)$$

同理计算可得

$$\begin{aligned} I_{2r} &= \text{Var}[f_n^{(i)}(x)] = \frac{1}{n^2} \sum_{j=1}^n \frac{1}{h_j^{2(i+1)}} \text{E}_n[D_i(\frac{X_j - x}{h_j})]^2 = \frac{1}{n^2} \sum_{j=1}^n \frac{1}{h_j^{2(i+1)}} \int_x^{x+h_j} D_i^2(\frac{y-x}{h_j}) f(x) dx \\ &= \frac{1}{n^2} \sum_{j=1}^n \frac{1}{h_j^{2i+1}} \int_0^1 D_i^2(v) f(x+h_j v) dv. \end{aligned} \quad (17)$$

再由  $f(x) \in C_{k,\beta}$  及核函数的有界性可知

$$I_{2r} \leq M \frac{1}{n^2} \sum_{j=1}^n \frac{1}{h_j^{2i+1}} \leq M \frac{1}{nh_n^{2i+1}} = M(nh_n^{2i+1})^{-1}. \quad (18)$$

因此当取  $h_n = n^{-\frac{1}{2(k-1)}}$  时, 代入(16)式和(18)式有

$$I_{1i} \leq M h_n^{k-i} = M n^{-\frac{k-i}{2(k-1)}} \leq M n^{-\frac{1}{2}}, \quad I_{2i} \leq M(nh_n^{2i+1})^{-1} \leq M n^{-\frac{k-i-1.5}{k-1}}.$$

故有

$$I_{1i}^{2\theta} \leq M n^{-\theta}, \quad (19)$$

$$I_{2i}^\theta \leq M n^{-\frac{\theta(k-i-1.5)}{k-1}}. \quad (20)$$

最后将(19)式和(20)式带入(13)式, 结论即成立.

**引理3** 设  $P_G(x)$  和  $P_n(x)$  分别由(5)式和(10)式定义, 其中  $X_1, X_2, \dots, X_n$  为独立同分布的随机样本序列, 则当  $0 < \theta \leq 1$  时, 有  $\text{E}_n |P_n(x) - P_G(x)|^{2\theta} \leq n^{-\theta}$ .

**证** 由于

$$\text{E}_n P_n(x) = \text{E}\{I(X_1 \leq x)\} = \int_{-\infty}^x f(y) dy = P_G(x),$$

故  $P_n(x)$  为  $P_G(x)$  的无偏估计. 由 Jensen 不等式可知

$$\text{E}_n [P_n(x) - P_G(x)]^{2\theta} = \text{E}_n \{ [P_n(x) - P_G(x)]^2 \}^\theta \leq \{ \text{Var}(P_n(x)) \}^\theta, \quad (21)$$

其中

$$\begin{aligned} \text{Var}[P_n(x)] &= \text{E}\left\{ \frac{1}{n} \sum_{i=1}^n [I(X_i \leq x) - P_G(x)] \right\}^{2\theta} = \frac{1}{n^2} \sum_{i=1}^n \text{Var}[I(X_i \leq x)] \\ &= \frac{1}{n} \text{Var}[P_G(X_1)] \leq \frac{1}{n} \text{E}[P_G(X_1)]^2 = \frac{1}{n} \int_{\mathcal{T}} f_G(t) dt = \frac{1}{n}. \end{aligned} \quad (22)$$

最后将(22)式带入(21)式, 引理3即得证.

## §5 经验 Bayes 检验函数的渐近最优性和收敛速度

**定理1** 定义  $X_1, X_2, \dots, X_n$  为独立同分布的随机样本序列, 其中  $R_G$  和  $R_n$  分别由(8)式和(12)式定义. 假定条件 III 和条件 IV 均满足, 则对  $\forall x \in \Omega$ , 若

(i)  $h_n$  为正数序列, 且  $\lim_{n \rightarrow \infty} h_n = 0$ ,  $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n h_j = 0$ ,  $\lim_{n \rightarrow \infty} n h_n^3 = 0$ ;



(ii)  $E(c) = \int_C cdG(c) < \infty$ ;

(iii)  $f^{(2)}(x)$ 为 $x$ 的连续函数.

则有  $\lim_{n \rightarrow \infty} R_n(\delta_n, G) = R_G(\delta_G, G)$ .

证 由引理1可知

$$0 \leq R_n(\delta_n, G) - R_G(\delta_G, G) \leq q \int_{\Omega} |\beta(x)| P[|\beta_n(x) - \beta(x)| \geq |\beta(x)|] dx,$$

令  $Q_n(x) = |\beta(x)| P[|\beta_n(x) - \beta(x)| \geq |\beta(x)|]$ , 有  $Q_n(x) \leq |\beta(x)|$ , 故由(4)式和Fubini定理可知

$$\begin{aligned} \int_{\Omega} |\beta(x)| dx &= \int_{\Omega} \left| \int_C f(x|c)(c - c_0) dG(c) \right| dx \\ &\leq \int_{\Omega} \int_C cf(x|c) dG(c) dx + |c_0| \int_{\Omega} \int_C f(x|c) dG(c) dx = |c_0| + \int_C cdG(c) < \infty. \end{aligned}$$

由控制收敛定理得

$$0 \leq \lim_{n \rightarrow \infty} [R_n(\delta_n, G) - R_G(\delta_G, G)] \leq \int_{\Omega} \lim_{n \rightarrow \infty} Q_n(x) dx.$$

要使定理成立, 则只需证明  $\lim_{n \rightarrow \infty} Q_n(x) \stackrel{a.s.}{=} 0$  成立即可.

根据Markov不等式和Jensen不等式可得

$$\begin{aligned} 0 \leq Q_n(x) &\leq E_n |\beta_n(x) - \beta(x)| \\ &\leq |N_1| E_n |P_n(x) - P_G(x)| + |c_0| E_n |f_n(x) - f(x)| + |N_2| E_n |f_n^{(1)}(x) - f^{(1)}(x)| \\ &\leq |N_1| [E_n |P_n(x) - P_G(x)|^2]^{\frac{1}{2}} + |c_0| [E_n |f_n(x) - f(x)|^2]^{\frac{1}{2}} + |N_2| [E_n |f_n^{(1)}(x) - f^{(1)}(x)|^2]^{\frac{1}{2}}. \end{aligned}$$

又由引理2可知, 当 $\theta = 1$ 时, 有  $E_n |f_n^{(i)}(x) - f^{(i)}(x)|^2 \leq Mn^{-\frac{k-i-1.5}{k-1}}$ . 其中

当 $i = 0$ 时, 有  $E_n |f_n(x) - f(x)|^2 \leq Mn^{-\frac{k-1.5}{k-1}}$ ;

当 $i = 1$ 时, 有  $E_n |f_n^{(1)}(x) - f^{(1)}(x)|^2 \leq Mn^{-\frac{k-2.5}{k-1}}$ .

故在  $\lim_{n \rightarrow \infty} h_n = n^{-\frac{1}{2(k-1)}} = 0$ ,  $\lim_{n \rightarrow \infty} nh_n^{2i+1} = \infty$  成立时, 有

$$\lim_{n \rightarrow \infty} Mn^{-\frac{k-1.5}{(k-1)}} = \lim_{n \rightarrow \infty} Mh_n^{2(k-1.5)} = 0, \quad \lim_{n \rightarrow \infty} Mn^{-\frac{k-2.5}{(k-1)}} = \lim_{n \rightarrow \infty} Mh_n^{2(k-2.5)} = 0.$$

所以对 $\forall x \in \Omega$ 有

当 $i = 0$ 时,  $\lim_{n \rightarrow \infty} E_n |f_n(x) - f(x)|^2 = 0$ ;

当 $i = 1$ 时,  $\lim_{n \rightarrow \infty} E_n |f_n^{(1)}(x) - f^{(1)}(x)|^2 = 0$ .

再由引理3可知, 当 $\theta = 1$ 时, 有  $0 \leq E_n |P_n(x) - P_G(x)|^2 \leq n^{-1}$ . 所以对 $\forall x \in \Omega$ , 有

$$\lim_{n \rightarrow \infty} E_n |P_n(x) - P_G(x)|^2 = 0.$$

因此有

$$\begin{aligned} \lim_{n \rightarrow \infty} Q_n(x) &\leq |N_1| \lim_{n \rightarrow \infty} [E_n |P_n(x) - P_G(x)|^2]^{\frac{1}{2}} + |c_0| \lim_{n \rightarrow \infty} [E_n |f_n(x) - f(x)|^2]^{\frac{1}{2}} \\ &\quad + |N_2| \lim_{n \rightarrow \infty} [E_n |f_n^{(1)}(x) - f^{(1)}(x)|^2]^{\frac{1}{2}} = 0. \end{aligned}$$

所以有

$$0 \leq \lim_{n \rightarrow \infty} [R_n(\delta_n, G) - R_G(\delta_G, G)] \leq \int_0^\infty \lim_{n \rightarrow \infty} Q_n(x) dx \leq 0.$$

综上所述,  $\lim_{n \rightarrow \infty} R_n(\delta_n, G) = R_G(\delta_G, G)$  得证.

**定理2**  $X_1, X_2, \dots, X_n$  为独立同分布的随机样本序列,  $R_G$  和  $R_n$  由(8)式和(12)式定义. 假定条件III和条件IV成立, 当  $0 < \theta < 1, h_n = n^{-\frac{1}{2(k-1)}}$  时有

$$R_n(\delta_n, G) - R_G(\delta_G, G) = O(n^{-q}), \quad q = \frac{\theta(k-2.5)}{k-1}.$$

**证** 由(8)式和(11)式可得

$$\begin{aligned} 0 &\leq R_n(\delta_n, G) - R_G(\delta_G, G) \leq q \int_{\Omega} \beta(x) \{E_n[\delta_n(x)] - \delta_G(x)\} dx \\ &= q \int_0^{A_1} \beta(x) \{E_n[\delta_n(x)] - \delta_G(x)\} dx + q \int_{A_1}^{a_G} \beta(x) \{E_n[\delta_n(x)] - \delta_G(x)\} dx + \\ & q \int_{a_G}^{A_2} \beta(x) \{E_n[\delta_n(x)] - \delta_G(x)\} dx + q \int_{A_2}^c \beta(x) \{E_n[\delta_n(x)] - \delta_G(x)\} dx \\ &= q(K_1 + K_2 + K_3 + K_4). \end{aligned} \quad (23)$$

又由(5)式和(10)式及C-R不等式和引理2, 引理3得

$$\begin{aligned} &E_n |\beta_n(x) - \beta(x)|^{2\theta} \\ &= E_n |N_1 [P_n(x) - P_G(x)] + c_0 [f_n(x) - f(x)] + N_2 [f_n^{(1)}(x) - f^{(1)}(x)]|^{2\theta} \\ &\leq C_1 |N_1|^{2\theta} E_n |P_n(x) - P_G(x)|^{2\theta} + C_2 |c_0|^{2\theta} E_n |f_n(x) - f(x)|^{2\theta} + \\ & C_3 |N_2|^{2\theta} E_n |f_n^{(1)}(y) - f^{(1)}(y)|^{2\theta} \\ &\leq [C_4 |N_1|^{2\theta} + C_5] n^{\frac{\theta(k-2.5)}{k-1}}. \end{aligned} \quad (24)$$

当  $x \in (0, A_1]$  时, 有  $\delta_G(x) = 1, \delta_n(x) = 1$ ;

当  $x \in [A_2, c)$  时, 有  $\delta_G(x) = 0, \delta_n(x) = 0$ .

故在  $(0, A_1]$  和  $[A_2, c)$  这两个区间内:  $E_n[\delta_n(x)] - \delta_G(x) = 0$ . 所以有

$$\begin{aligned} K_1 &= \int_0^{A_1} \beta_n(x) \{E_n[\delta_n(x)] - \delta_G(x)\} dx = 0, \\ K_4 &= \int_{A_2}^c \beta_n(x) \{E_n[\delta_n(x)] - \delta_G(x)\} dx = 0. \end{aligned}$$

当  $x \in [A_1, a_G)$  时, 有  $\delta_G(x) = 1, E_n[\delta_n(x)] = P(\beta_n(x)) \leq 0$ , 则由(24)式可知

$$\begin{aligned} K_2 &= \int_{A_1}^{a_G} \beta(x) \{E_n[\delta_n(x)] - \delta_G(x)\} dx = \int_{A_1}^{a_G} \beta(x) \{1 - P[\beta_n(x) \leq 0]\} dx \\ &\leq \int_{A_1}^{a_G} |\beta(x)| P[|\beta_n(x) - \beta(x)| \geq |\beta(x)|] dx \leq \int_{A_1}^{a_G} |\beta(x)|^{1-2\theta} E_n |\beta_n(x) - \beta(x)|^{2\theta} dx \\ &\leq n^{-\frac{\theta(k-2.5)}{k-1}} [C_2 \int_{A_1}^{a_G} |\beta(x)|^{1-2\theta} |N_1|^{2\theta} dx + C_3 \int_{A_1}^{a_G} |\beta(x)|^{1-2\theta} dx]. \end{aligned}$$

当  $0 < \theta < \frac{1}{2}$  时, 由于  $\beta(x)$  和  $N_1$  是闭区间上的连续函数, 由闭区间上连续函数的有界性可知  $\int_{A_1}^{a_G} |\beta(x)|^{1-2\theta} dx < \infty$ ,  $\int_{A_1}^{a_G} |\beta(x)|^{1-2\theta} |N_1|^{2\theta} dx < \infty$ , 故有:  $K_2 \leq Cn^{-\frac{\theta(k-2.5)}{k-1}}$ .

当  $0 < \theta < 1$  时, 由于  $\lim_{x \rightarrow a_G} |\beta(x)| = \lim_{x \rightarrow a_G} f(x) |\gamma(x) - c_0| = 0$ , 所以有

$$\int_{A_1}^{a_G} |\beta(x)|^{1-2\theta} dx = \int_{A_1}^{a_G} \frac{1}{|\beta(x)|^{2\theta-1}} dx$$

是第二类型广义积分, 奇点为  $x = a_G$ . 故由第二类型广义积分的比较判别法可得

$$\begin{aligned} \lim_{x \rightarrow a_G} \frac{(a_G - x)^{2\theta-1}}{|\beta(x)|^{2\theta-1}} &= \lim_{x \rightarrow a_G} \left[ \frac{a_G - x}{f(x)[\gamma(x) - c_0]} \right]^{2\theta-1} = \left[ \frac{1}{f(a_G)} \lim_{x \rightarrow a_G} \frac{-1}{\gamma^{(1)}(x)} \right]^{2\theta-1} \\ &= \left[ \frac{-1}{f(a_G)\gamma^{(1)}(a_G)} \right]^{2\theta-1} \geq 0. \end{aligned}$$

所以  $\int_{A_1}^{a_G} \frac{1}{(a_G-x)^{2\theta-1}} dx$  是收敛的. 又由于  $\int_{A_1}^{a_G} \frac{1}{|\beta(x)|^{2\theta-1}} dx$  和  $\int_{A_1}^{a_G} \frac{1}{(a_G-x)^{2\theta-1}} dx$  具有相同的敛散性, 故  $\int_{A_1}^{a_G} \frac{1}{|\beta(x)|^{2\theta-1}} dx$  也是收敛的. 因此由连续函数的有界性可知  $|N_1|^{2\theta} \leq M$ , ( $M$  为正实数). 故有

$$K_2 \leq n^{-\frac{\theta(k-2.5)}{k-1}} \left[ C_2 \int_{A_1}^{a_G} \frac{1}{(a_G-x)^{2\theta-1}} dx + C_3 M \int_{A_1}^{a_G} \frac{1}{(a_G-x)^{2\theta-1}} dx \right] \leq C_4 n^{-\frac{\theta(k-2.5)}{k-1}}.$$

同理可得

$$\begin{aligned} K_3 &= \int_{a_G}^{A_2} \beta(x) \{E_n[\delta_n(x)] - \delta_G(x)\} dx = \int_{a_G}^{A_2} \beta(x) [P(\beta_n(x)) \leq 0] dx \\ &\leq \int_{a_G}^{A_2} |\beta(x)|^{1-2\theta} E_n |\beta_n(x) - \beta(x)|^{2\theta} dx \leq C_5 n^{-\frac{\theta(k-2.5)}{k-1}}. \end{aligned}$$

最后把  $K_1, K_2, K_3, K_4$  带入(23)式, 即可得到

$$R_n(\delta_n, G) - R_G(\delta_G, G) = q(K_1 + K_2 + K_3 + K_4) = O(n^{-\frac{\theta(k-2.5)}{k-1}}).$$

注1 本文所得到判决函数的收敛阶数为  $O(n^{-\frac{\theta(k-2.5)}{k-1}})$ , 当  $k \rightarrow \infty, \theta \rightarrow 1$  时, 有

$$O(n^{-\frac{\theta(k-2.5)}{k-1}}) \rightarrow O(n^{-1}).$$

## §6 双侧检验问题

现考虑参数  $c$  的双侧检验问题

$$H_0 : c_1 \leq c \leq c_2 \quad vs \quad H_1 : c < c_1, c > c_2,$$

其中  $c_1$  和  $c_2$  为给定的常数. 如果取  $c_0 = \frac{c_1+c_2}{2}$  和  $\lambda_0 = \frac{c_2-c_1}{2}$ , 那么双侧检验问题就等价于

$$H_0^* : |c - c_0| \leq \lambda_0 \quad vs \quad H_1^* : |c - c_0| > \lambda_0.$$

取其加权线性损失函数为

$$L_m^*(c, d_m) = (1-m)q[(c-c_0)^2 - \lambda_0^2]I(|c-c_0| > \lambda_0) + mq[\lambda_0^2 - (c-c_0)^2]I(|c-c_0| \leq \lambda_0),$$

其中  $m = 0, 1$ ;  $D^* = \{d_0, d_1\}$  为双侧检验问题的判决空间;  $d_i$  为接受  $H_i^*$  的决策,  $i = 0, 1$ ;  $I(A)$  表示事件  $A$  发生的示性函数.

设随机化判决函数为  $\delta_G^*(x) = P\{\text{接受 } H_0^* \mid X = x\}$ , 则在先验分布  $G(c)$  未知的条件下, 与(3)式同理, 有  $\delta^*(x)$  的 Bayes 风险为

$$\begin{aligned} R(\delta_G^*(x), G(c)) &= \int_C \int_{\Omega} [L_0^*(c, d_0) f(x|c) \delta_G^*(x) + L_1^*(c, d_1) f(x|c) (1 - \delta_G^*(x))] dx dG(c) \\ &= q \int_{\Omega} \beta^*(x) \delta_G^*(x) dx + C_G^*. \end{aligned} \quad (25)$$

记  $\gamma^*(x) = \frac{\int_C (c^2 - 2c_0 c) f(x|c) dG(c)}{\int_C f(x|c) dG(c)}$ , 则上式中  $\beta^*(x)$  和  $C_G^*$  分别表示

$$\begin{cases} C_G^* = \int_C L_1^*(c, d_1) dG(c), \\ \beta^*(x) = \int_C [(c - c_0)^2 - \lambda_0^2] f(x|c) dG(c) = [c_0^2 - \lambda_0^2 + \gamma^*(x)] f(x). \end{cases}$$

经计算有

$$\beta^*(x) = H_1 P_G(x) + H_2 f(x) + H_3 f^{(1)}(x) + H_4 f^{(2)}(x), \quad (26)$$

其中  $H_1 = x$ ,  $H_2 = c_0^2 - \lambda_0^2 + \frac{3}{2}\sigma$ ,  $H_3 = \frac{\sigma}{2}x$ ,  $H_4 = \frac{\sigma^2}{4}$ .

由(26)式可知 Bayes 双侧检验的判决函数为

$$\delta_G^*(x) = \begin{cases} 1, & \beta^*(x) \leq 0, \\ 0, & \beta^*(x) > 0. \end{cases} \quad (29)$$

则  $\delta_G(x)$  的 Bayes 风险为

$$R_G^*(\delta_G^*, G) = \inf R^*(\delta_G^*(x), G(c)) = q \int_{\Omega} \beta^*(x) \delta_G^*(x) dx + C_G^*. \quad (30)$$

现考虑使用经验 Bayes 检验函数, 利用(9)式的递归核密度形式, 可以得到  $f^{(2)}(x)$ ,  $f^{(1)}(x)$  和  $f(x)$  的核密度估计为  $f_n^{(2)}(x)$ ,  $f_n^{(1)}(x)$ ,  $f_n(x)$ . 由(26)式可定义  $\beta^*(x)$  的估计量为

$$\beta_n^*(x) = H_1 P_n(x) + H_2 f_n(x) + H_3 f_n^{(1)}(x) + H_4 f_n^{(2)}(x).$$

再由(27)式定义参数  $c$  的双侧经验 Bayes 检验的随机化判决函数为

$$\delta_n^*(x) = \begin{cases} 1, & \beta_n^*(x) \leq 0, \\ 0, & \beta_n^*(x) > 0. \end{cases}$$

则  $\delta_n^*(x)$  的全面 Bayes 风险为

$$R_n^*(\delta_n^*, G) = q \int_{\Omega} \beta_n^*(x) E_n[\delta_n^*(x)] dx + C_G^{*'} \quad (29)$$

**定理3** 定义  $X_1, X_2, \dots, X_n$  为独立同分布的随机样本序列, 其中  $R_G^*$  和  $R_n^*$  分别由(28)式和(29)式所定义. 假定条件 III 和条件 IV 均满足, 则对  $\forall x \in \Omega$ , 若

(i)  $\{h_n\}$  为正数序列, 且  $\lim_{n \rightarrow \infty} h_n = 0$ ,  $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n h_j = 0$ ,  $\lim_{n \rightarrow \infty} nh_n^5 = \infty$ ;

(ii)  $E(c^2) = \int_C c^2 dG(c) < \infty$ ;

(iii)  $f_n^3(x)$  为  $x$  的连续函数;

则有  $\lim_{n \rightarrow \infty} R_n^*(\delta_n^*, G) = R_G^*(\delta_G^*, G)$ .

**定理4**  $X_1, X_2, \dots, X_n$  为独立同分布的随机样本序列,  $R_G^*$  和  $R_n^*$  由(8)式和(12)式定义. 假定条件III和条件IV均成立, 则当  $0 < \theta < 1$ ,  $h_n = n^{-\frac{1}{2(k-1)}}$  时, 有

$$R_n^*(\delta_n^*, G) - R_G^*(\delta_G^*, G) = O(n^{-q}), \quad q = \frac{\theta(k-3.5)}{k-1},$$

其中当  $k \rightarrow \infty$ ,  $\theta \rightarrow 1$  时, 该收敛速度可任意地接近  $O(n^{-\frac{1}{2}})$ .

定理3和定理4的证明方法和过程与定理1和定理2类同, 此处省略证明过程.

## §7 实例

本节将给出一个满足前述性质的实例. 当给定  $\sigma = 1$  时, 随机变量  $X$  在参数  $c$  给定条件下的条件密度函数为

$$f(x|c) = 2(c-x)\exp\{-(c-x)^2\},$$

取参数  $c$  的先验分布为  $g(c) = \frac{1}{\tau(r)} c^{r-1} \exp\{-c\} I_{(c>0)}$ , 故得到  $x$  的边缘密度函数

$$\begin{aligned} f(x) &= \int_C f(x|c)g(c)dc = \int_C \frac{2(c-x)c^{r-1}}{\tau(r)} \exp\{-c^2 + (2x-1)c - x^2\} dc \\ &= \frac{2\exp\{-x^2\}}{\tau(r)} \int_C (c^r - c^{r-1}x) \exp\{-c^2 + (2x-1)c\} dc < \frac{2\exp\{-x^2\}}{\tau(r)}. \end{aligned}$$

由此可知  $f(x)$  可导, 其导函数连续且有界, 即满足  $f_G(x) \in C_{k,\beta}$ , 条件II成立. 于是假定条件IV成立, 只需验证条件I成立即可. 那么将  $f(x)$  及  $f(x|c)$  带入(4)式即得

$$\gamma(x) = \frac{\int_C c f(x|c) dG(c)}{\int_C f(x|c) dG(c)} = \frac{\int_C (c^{r+1} - c^r x) \exp\{-c^2 + (2x-1)c\} dc}{\int_C (c^r - c^{r-1}x) \exp\{-c^2 + (2x-1)c\} dc},$$

有  $\gamma(x)$  的一阶导数

$$\begin{aligned} \gamma^{(1)}(x) &= \frac{\int_C (2c^{r+2} - 2c^{r+1}x - c^r) \exp\{-c^2 + (2x-1)c\} dc \int_C (c^r - c^{r-1}x) \exp\{-c^2 + (2x-1)c\} dc}{[\int_C (c^r - c^{r-1}x) \exp\{-c^2 + (2x-1)c\} dc]^2} \\ &\quad - \frac{\int_C (c^{r+1} - c^r x) \exp\{-c^2 + (2x-1)c\} dc \int_C c^{r-1} (2c^2 - 2cx - 1) \exp\{-c^2 + (2x-1)c\} dc}{[\int_C (c^r - c^{r-1}x) \exp\{-c^2 + (2x-1)c\} dc]^2} \\ &\leq 0. \end{aligned}$$

故  $\gamma(x)$  在  $x \in \Omega$  上是连续的单调减函数. 令  $\gamma(a_G) = c_0$ , 并取  $A_1 = 10^{-10}$ ,  $A_2 = 10^{10}$ . 经计算有  $0 < A_1 < a_G < A_2 < c$ , 故能得到  $\lim_{x \rightarrow c^-} \gamma(x) < \gamma(a_G) = c_0 < \lim_{x \rightarrow 0^+} \gamma(x)$ , 即有条件I成立.

$$E(c) = \int_C c g(c) dc = \int_C \frac{c^r}{\tau(r)} \exp\{-c\} dc < \infty,$$

故有条件III 成立. 又

$$\begin{aligned}\beta(x) &= f(x) [\gamma(x) - c_0] \\ &= \frac{2 \exp\{-x^2\}}{\tau(r)} [c_0 \int_C (c^r - c^{r-1}x) \exp\{-c^2 + (2x-1)c\} dc \\ &\quad - \int_C (c^{r+1} - c^r x) \exp\{-c^2 + (2x-1)c\} dc] \leq W \frac{2 \exp\{-x^2\}}{\tau(r)},\end{aligned}$$

其中  $W \in \mathbf{R}$ , 故在  $x \in \Omega$  上, 函数  $\beta(x)$  是有界函数.

综上证明可得: 定理1, 定理2及定理3的条件在本例中均满足, 即本文所给经验Bayes检验的结论均成立. 那么代入定理2和定理3的理论结果, 可验证得到  $\lim_{n \rightarrow \infty} R_n(\delta_n, G) = R_G(\delta_G, G)$ , 且收敛速度的阶数满足  $O(n^{-\frac{\theta(k-2.5)}{k-1}})$ .

#### 参考文献:

- [1] Pearson K. Asymmetrical frequency curves[J]. Nature, 1893, 48(1252): 615-616.
- [2] Azzalini A. A Class of distribution which include the normal ones[J]. Scandinavian Journal of Statistics, 1985, 12(2): 171-178.
- [3] 李光兴, 余洪琼. 考试成绩转换成偏态分布量化成绩的算法[J]. 高师理科学刊, 2020, 40(1): 22-26.
- [4] Jones M C, Faddy M J. A skew extension of the t-distribution, with applications[J]. Journal of the Royal Statistical Society: Series B, 2003, 65(1): 159-174.
- [5] Nekoukhou V, Alamatsaz M H. A family of skew-symmetric-Laplace distributions[J]. Statistical Papers, 2012, 53(3): 685-696.
- [6] 杨昕. 对数收益率的偏斜Logistic分布与VaR估计[J]. 数理统计与管理, 2011, 30(3): 548-553.
- [7] 谭佳玲, 曾鑫, 吴刘仓. 带缺失数据的偏正态众数回归模型的参数估计[J]. 高校应用数学学报, 2022, 37(1): 24-34.
- [8] Nadarajah S. The skew logistic distribution[J]. AStA Advances in Statistical Analysis, 2009, 93(2): 187-203.
- [9] Azzalini A, Arellano-Valle R B. Maximum penalized likelihood estimation for skew-normal and skew-t distributions[J]. Journal of Statistical Planning and Inference, 2013, 143(2): 419-433.
- [10] Kahrari F, Arellano-Valle R B, Rezaei M, Yousefzadeh F. Scale mixtures of skew-normal-Cauchy distributions[J]. Statistics and Probability Letters, 2017, 126: 1-6.
- [11] 曹幸运, 曾鑫, 吴刘仓. 偏正态数据下众数回归模型的统计诊断[J]. 高校应用数学学报, 2021, 36(1): 9-20.
- [12] 叶仁道, 戚戩. 偏正态单向分类随机效应模型中方差分量函数的Bootstrap推断[J]. 数理统计与管理, 2022, 41(2): 309-321.
- [13] Robbins H. The empirical Bayes approach to statistical decision problems[J]. Annals of Mathematical Statistics, 1964, 35(1): 1-20.
- [14] Johns M V, Ryzin A. Convergence rates for empirical Bayes two-action problems I. Discrete Case[J]. Annals of Mathematical Statistics, 1971, 42(5): 1521-1539.
- [15] 陈家清, 刘次华. 线性指数分布参数的经验Bayes检验问题[J]. 系统科学与数学, 2008, 28(5): 616-626.

- [16] 周雁, 韦来生. 刻度指数族参数的经验Bayes检验函数收敛速度的改进[J]. 高校应用数学学报, 2008, 23(2): 219-226.
- [17] Tansey W, Wang Yixin, Rabadan R, Blei D. Double empirical Bayes testing[J]. International Statistical Review, 2020, 88(S1): 91-113.
- [18] Liang T C. On optimal convergence rate of empirical Bayes tests[J]. Stats and Probability Letters, 2004, 68(2): 189-198.
- [19] Chen L S. Empirical Bayes testing for uniform distributions: non identical components case[J]. Communications in Statistics-Theory and Methods, 2011, 40(17): 3181-3197.
- [20] 陈家清, 陈志强, 金倩聿, 等. 线性指数分布参数的经验贝叶斯检验收敛速度研究[J]. 统计与决策, 2013, 382(10): 27-30.
- [21] Yuan Min, Wei Laisheng. Two-sided empirical Bayes test for location parameter in the gamma distribution[J]. Communications in Statistics - Theory and Methods, 2017, 46(9): 4215-4225.
- [22] Parzen E. On estimation of a probability density function and mode[J]. The Annals of Mathematical Statistics, 1962, 33(3): 1065-1076.
- [23] Wolverson C T, Wagner T J. Asymptotically optimal discriminant functions for pattern classification[J]. IEEE Transactions on Information Theory, 1969, 15(2): 258-265.
- [24] Terrell G R. The maximal smoothing principle in density estimation[J]. Journal of the American Statistical Association, 1990, 85(410): 470-477.
- [25] 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.
- [26] Alexandre L A. A solve-the-equation approach for unidimensional data kernel bandwidth selection [R/OL]. [2008-11-29] [2008-01-01]. Beira Interior, Portugal: University of Beira Interior. <http://www.di.ubi.pt/lfbaa/entnetsPubs/bandwidth.pdf>
- [27] 韦博成. 参数统计教程[M]. 北京: 高等教育出版社, 2006.

## Empirical Bayes test for the truncation parameter of skew distribution

LIU Rui<sup>1,2</sup>, TAN Yan<sup>1,2</sup>, WU Liu-cang<sup>1,2</sup>

- (1. Faculty of Science, Kunming University of Science and Technology, Kunming 650504, China;  
2. Center of Applied Statistics, Kunming University of Science and Technology, Kunming 650504, China)

**Abstract:** Skew distribution is a kind of generalization of the symmetric distribution and is widely applied in real life, where the truncation parameter to define the boundary of the skew distribution is of great significance. In this study, the empirical Bayes test is discussed for the truncation parameter in the skew distribution. Considering the unpredictability and uncertainty of prior density function in the common Bayes test, the unknown prior density function is estimated by the recursive kernel density function. In order to better characterize the decision risk, the weighted linear loss function of the test is defined. The asymptotic optimality of the proposed test function is proved under given conditions, and the determined convergence rates is given. Finally, the theoretical results are verified by a real example.

**Keywords:** empirical Bayes test; skew distribution; recursive kernel estimation; asymptotic optimality; convergence rates

**MR Subject Classification:** 62C12; 62F05