Statistics and Application 统计学与应用, 2016, 5(3), 225-231
Published Online September 2016 in Hans. http://www.hanspub.org/journal/sa
http://dx.doi.org/10.12677/sa.2016.53022

The Empirical Bayes Two-Sided Tests for the Parameter of Linear Exponential Distribution under Longitudinal Data

Xiuhui Luo, Chengdong Wei*, Jinli Li

School of Mathematical and Statistics Sciences, Guangxi Teachers Education University, Nanning Guangxi
Email: *wcdbbh@163.com

Received: Aug. 19th, 2016; accepted: Sep. 4th, 2016; published: Sep. 9th, 2016

Abstract

In the case of longitudinal data, this paper studies two-side test problem of linear exponential distribution parameters under square loss function. By applying Markov inequality, the EB test rules for parameter of the linear exponential distribution are constructed and the asymptotically optimal property is obtained. Finally, we obtain the convergence rate of the proposed EB test under suitable conditions.

Keywords

Longitudinal Data, Linear Exponential Distribution, Empirical Bayes Test, Asymptotic Optimality, Convergence Rate

纵向数据下线性指数分布参数的经验Bayes双边检验

罗修辉，韦程东*，李进立

广西师范学院数学与统计科学学院，广西 南宁
Email: *wcdbbh@163.com

*通讯作者。

文章引用: 罗修辉, 韦程东, 李进立. 纵向数据下线性指数分布参数的经验 Bayes 双边检验[J]. 统计学与应用, 2016, 5(3): 225-231. http://dx.doi.org/10.12677/sa.2016.53022

收稿日期：2016年8月19日；录用日期：2016年9月4日；发布日期：2016年9月9日

摘　要

基于纵向数据下，本文讨论了在平方损失函数下线性指数分布参数的双边检验问题，利用Markov不等式证明了构造的经验贝叶斯双边检验函数具有渐近最优性，并获得了其收敛速度。

关键词

纵向数据，线性指数分布，经验Bayes，渐近最优性，收敛速度

1. 引言

自 John 和 Ryzin [1] [2]分别对独立同分布($i.i.d.$)样本情形的离散型和连续型指数族提出经验贝叶斯(EB)检验方法以来，经过国内外学者的致力研究，目前对于分布族参数的经验贝叶斯双边检验问题的研究已有了一定的成果[3]-[8]。但大多数文献都是针对无重复 $i.i.d.$ 样本进行讨论的，而很少讨论重复样本下贝叶斯的双边检验问题。然而，在实际研究中常常需要对随机变量进行大量重复的观测，如在生物学，医学和经济学等领域的研究中常常对观测个体进行重复的大量观察，获得一系列的纵向数据。鉴于此，本文将在纵向数据情形下，构造线性指数分布族参数的检验函数，并对其双边检验问题进行讨论。

考虑如下的线性指数分布族

$$f\left(x|\theta\right)=\left(\mu x+\theta\right)\mathrm{e}^{-\theta x-\frac{\mu x^2}{2}} \tag{1}$$

其中，θ 为未知参数，$\mu>0$ 为常数且已知。参数空间为 $\Theta=\left\{\theta>0:\int_{\Omega} f\left(x|\theta\right)\mathrm{d}x=1\right\}$。样本空间为 $\Omega=\left\{x|x>0\right\}$。

本文讨论下列双边检验问题

$$H_0:\theta_1\le\theta\le\theta_2 \Leftrightarrow H_1\theta\le\theta_1 \text{ 或 } \theta>\theta_2, \tag{2}$$

此处 θ_1,θ_2 为给定的常数，如果取 $\theta_0=\dfrac{\theta_1+\theta_2}{2}$ 和 $\varepsilon_0=\dfrac{\theta_2-\theta_1}{2}$，则(2)等价于

$$H_0^*:\left|\theta-\theta_0\right|\le\theta_2 \Leftrightarrow H_1^*:\left|\theta-\theta_0\right|>\varepsilon_0 \tag{3}$$

对于(3)式的双边假设检验问题，可设其平方损失函数为

$$L_0\left(\theta,d_0\right)=a\left[\left(\theta-\theta_0\right)^2-\varepsilon_0^2\right]I_{\left[\left|\theta-\theta_0\right|>\varepsilon_0\right]},\ L_1\left(\theta,d_1\right)=a\left[\varepsilon_0^2-\left(\theta-\theta_0\right)^2\right]I_{\left[\left|\theta-\theta_0\right|\le\varepsilon_0\right]}$$

其中，$a>0$ 为常数，$d=\left\{d_0,d_1\right\}$ 表示行动空间；d_0 表示接受 H_0^*，d_1 则表示否定 H_0^*；此处假定参数 θ 的先验分布为 $G\left(\theta\right)$，且 $\mathrm{d}G\left(\theta\right)=g\left(\theta\right)\mathrm{d}\theta$，其中先验分布族为 $\left\{G:\int_{\Theta}\left|\theta\right|^{\delta}<\infty,\delta\ge1\right\}$。而且 $G\left(\theta\right)$ 未知。

设随机判决函数为

$$\delta\left(x\right)=P\left(\text{接受 } H_0^*\left|X=x\right.\right), \tag{4}$$

则 $\delta\left(x\right)$ 的风险函数为

$$R\left(\delta(x),G(\theta)\right)=\int_{\Theta}\int_{\Omega}\left[L_0(\theta,d_0)f(x|\theta)\delta(x)+L_1(\theta,d_1)f(x|\theta)\left(1-\delta(x)\right)\right]\mathrm{d}x\mathrm{d}G(\theta)=a\int_{\Omega}\beta(x)\delta(x)\mathrm{d}x+C_G \quad (5)$$

此处 $C_G=\int_{\Theta}L_1(\theta,d_1)\mathrm{d}G(\theta)$，

$$\beta(x)=\int_{\Theta}\left[(\theta-\theta_0)^2-\varepsilon_0^2\right]f(x|\theta)\mathrm{d}G(\theta) \quad (6)$$

令 r,v,X 的边缘分布为

$$f(x)=\int_{\Theta}f(x|\theta)\mathrm{d}G(\theta)=\int_{\Theta}(\mu x+\theta)\mathrm{e}^{-\theta x-\frac{1}{2}\mu x^2}\mathrm{d}G(\theta)$$

令 $\rho(x)=\int_{\Theta}\mathrm{e}^{-\theta x-\frac{1}{2}\mu x^2}\mathrm{d}G(\theta)$

又 $\rho^{(1)}(x)=-\int_{\Theta}(\mu x+\theta)\mathrm{e}^{-\theta x-\frac{1}{2}\mu x^2}\mathrm{d}G(\theta)=-f(x)$，故有 $\rho(x)=\int_x^{\infty}f(x)\mathrm{d}x$

所以由(6)得

$$\beta(x)=A(x)\rho(x)-B(x)f(x)-C(x)f^{(1)}(x)-f^{(2)}(x), \quad (7)$$

其中 $f^{(1)}(x)$, $f^{(2)}(x)$ 分别表示 $f(x)$ 的一阶和二阶导数，且 $A(x)=-2\mu^2x-2\mu\theta_0$，$B(x)=\varepsilon_0-3\mu-\mu^2x^2-2\mu x\theta_0-\theta_0^2$， $C(x)=-2\mu x-2\theta_0$

由(6)可知贝叶斯检验函数为

$$\delta_G(x)=\begin{cases}1, & 若\beta(x)\le 0,\\ 0, & 若\beta(x)>0.\end{cases} \quad (8)$$

其贝叶斯风险为

$$R(G)=\inf_{\delta}R(\delta,G)=R(\delta_G,G)=a\int_{\Omega}\beta(x)\delta_G(x)\mathrm{d}x+C_G. \quad (9)$$

若先验分布 $G(\theta)$ 已知且 $\delta(x)$ 等于 $\delta_G(x)$ 时， $R(G)$ 是可以精确达到的。然而 $G(\theta)$ 是未知的，因此 $\delta_G(x)$ 对我们而言是不适用的，于是需要引入 EB 方法。

2. EB 检验函数的构造

本节主要目的是构造纵向数据下的 EB 检验函数。设 $i.i.d.$ 序列 $X_1,X_2,\cdots,X_n,X$ 共同的边际概率密度函数为 $f(x)$， $X_1,X_2,\cdots,X_n$ 表示历史样本， X 表示当前样本。记纵向数据 X_{ij} 为对历史样本 X_i 进行第 j 次观测所得的样本 $(i=1,\cdots,n;j=1,\cdots,n_i)$， $X_{11},X_{12},\cdots,X_{1n_1},\cdots,X_{21},X_{22},\cdots,X_{2n_2},\cdots,X_{n1},X_{n2},\cdots,X_{nn_n}$ 相互独立，且 $f(x)$ 为其共同的边际概率密度函数。记 $n=\sum_{i=1}^{n}n_i$， 假定 $f(x)\in C_{s,\alpha},x\in R^1$， 其中 $C_{s,\alpha}=\left\{g(x)\middle|g^{(s)}(x)\right\}$ 表示 R^1 中的一族概率密度函数，其有 s 阶导函数，连续且有 $\left\{\left|g^{(s)}(x)\right|\le\alpha,s\ge 4,s\in N^+,\alpha>0\right\}$。接下来首先构造 $\beta(x)$ 的估计量。

令 $K_r(x)(r=0,1,\cdots,s-1)$ 为有界的 Borel 可测函数，其在区间 $(0,1)$ 之外取值为零且满足条件 $(A1)$。

$(A1)$ $\frac{1}{t!}\int_0^1u^tK_r(u)\mathrm{d}u=\begin{cases}1, & t=r\\ 0, & t\ne r,t=0,1,\cdots,s-1.\end{cases}$

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

$$f_n^{(r)}(X)=\frac{1}{nh_n^{r+1}}\sum_{i=1}^{n}\sum_{j=1}^{n_i}\int K\left(\frac{X-X_{ij}}{h_n}\right) \tag{10}$$

其中$\{h_n\}$为正数序列，且$\lim_{n\to\infty}h_n=0$。由于$\rho(x)=\int_x^\infty f(x)\mathrm{d}x=E\left\{I_{(X_i>x)}\right\}$，故可定义$\rho(x)$的估计量为

$$\rho_n(x)=\frac{1}{n}\sum_{i=1}^{n}I_{(X_i>x)}$$

所以$\beta(x)$的估计量定义为

$$\beta(x)=A(x)\rho_n(x)-B(x)f_n(x)-C(x)f_n^{(1)}(x)-f_n^{(2)}(x) \tag{11}$$

故 EB 检验函数定义为

$$\delta_n(x)=\begin{cases}1, & 若\ \beta_n(x)\le 0,\\ 0, & 若\ \beta_n(x)>0.\end{cases} \tag{12}$$

设E_n为$X_{11},X_{12},\cdots,X_{1n_1},\cdots,X_{21},X_{22},\cdots,X_{2n_2},\cdots,X_{n1},X_{n2},\cdots,X_{nn_n}$的联合分布均值，若有$\lim_{n\to\infty}R(\delta_n,G)=R(\delta_G,G)$，其中$R(\delta_n,G)$为 EB 检验函数$\delta_n(x)$的全面 Bayes 风险

$$R(\delta_n,G)=a\int_\Omega\beta(x)E_n\left[\delta_n(x)\right]\mathrm{d}x+C_G. \tag{13}$$

则称$\{\delta_n\}$为渐近最优的 EB 检验函数。

若有$R(\delta_n,G)-R(\delta_G,G)=O(n^{-q}),q>0$，则称 EB 检验函数$\{\delta_n\}$的收敛速度为$O(n^{-q})$。接下来将导出$\{\delta_n\}$的渐近最优性及其收敛速度。

3. 几个主要的引理

引理 3.1 [5]：设$f_n^{(r)}(x)$由(10)式定义，其中$X_{11},X_{12},\cdots,X_{1n_1},\cdots,X_{21},X_{22},\cdots,X_{2n_2},\cdots$为独立同分布随机变量序列。假定条件(A1)成立，对$\forall x\in\Omega$。

(I) 当$f^{(r)}(x)$关于x连续，当$\lim_{n\to\infty}h_n=0$且$\lim_{n\to\infty}nh_n^{2r+1}=0$时，则有

$$\lim_{n\to\infty}E\left|f_n^{(r)}(x)-f^{(r)}(x)\right|^2=0$$

(II) 当$f(x)\in C_{S,a}$，当取$h_n=n^{-\frac{1}{2+s}}$时，对于$0<\lambda\le 1$有

$$E\left|f_n^{(r)}(x)-f^{(r)}(x)\right|^{2\lambda}\le c\cdot n^{-\frac{\lambda(s-2\lambda+1)}{2+s}}.$$

引理 3.2 [2]：令$R(\delta_G,G)$和$R(\delta_n,G)$分别由(9)和(13)式给出，则

$$0\le R(\delta_n,G)-R(\delta_G,G)\le a\int_\Omega|\beta(x)|\,p\left(|\beta_n(x)-\beta(x)|\ge|\beta(x)|\right)\mathrm{d}x.$$

引理 3.3 [6]：设$\rho(x)=\int_\Theta \mathrm{e}^{-\theta x-\frac{1}{2}\mu x^2}\mathrm{d}G(\theta)$和$\rho_n(x)=\frac{1}{n}\sum_{i=1}^{n}I_{(X_i>x)}$，则对$0<\lambda\le 1$有

$$E_n\left|\rho_n(x)-\rho(x)\right|^\lambda\le n^{-\lambda/2}$$

4. EB 检验函数的性质

本节主要讨论线性指数分布函数族的$\{\delta_n\}$的性质并导出其收敛速度为$O\left(n^{-\frac{\lambda(s-2)}{2s+3}}\right)$。

定理 4.1：设 $\delta_n(x)$ 由(12)式定义， $X_{11}, X_{12}, \cdots, X_{1n_1}, \cdots, X_{21}, X_{22}, \cdots, X_{2n_2}, \cdots$ 为独立同分布随机变量序列。假定条件(A1)成立，若

(i) $\{h_n\}$ 为正数序列，且满足 $\lim\limits_{n\to\infty} h_n = 0$, $\lim\limits_{n\to\infty} nh_n^5 = \infty$ ；

(ii) $\int_\Theta \theta^2 \mathrm{d}G(\theta) < \infty$ ；

(iii) $f^{(3)}(x)$ 为 x 的连续函数。

则有

$$\lim_{n\to\infty} R(\delta_n, G) = R(\delta_G, G) \tag{14}$$

证明：由引理 3.2 得

$$0 \le R(\delta_n, G) - R(\delta_G, G) \le a\int_\Omega |\beta(x)| p\left(|\beta_n(x) - \beta(x)| \ge |\beta(x)|\right) \mathrm{d}x \tag{15}$$

记 $Q_n(x) = |\beta(x)| P\left(|\beta_n(x) - \beta(x)| \ge |\beta(x)|\right)$，则有 $Q_n(x) \le \beta(x)$。又由式(6)和 Fubini 定理得

$$\begin{aligned}\int_\Omega \beta(x)\mathrm{d}x &\le \left|\theta_0^2 - \varepsilon_0^2\right| \int_\Omega f(x)\mathrm{d}x + \int_\Omega \int_\Theta \left(\theta^2 + 2\theta|\theta_0|\right) f(x|\theta)\mathrm{d}G(\theta)\mathrm{d}x \\ &= \left|\theta_0^2 - \varepsilon_0^2\right| + \int_\Theta \left(\theta^2 + 2\theta|\theta_0|\right) \int_\Omega f(x|\theta)\mathrm{d}x\mathrm{d}G(\theta) \\ &= \left|\theta_0^2 - \varepsilon_0^2\right| + \int_\Theta \theta^2 \mathrm{d}G(\theta) + 2|\theta_0| \int_\Omega \theta \mathrm{d}G(\theta) < \infty\end{aligned}$$

故由控制收敛定理我们可以得到

$$0 \le \lim_{n\to\infty} R(\delta_n, G) - R(\delta_G, G) \le \int_\Omega \left[\lim_{n\to\infty} Q_n(x)\right] \mathrm{d}x$$

所以要证明定理 1 成立，只需证明 $\lim\limits_{n\to\infty} Q_n(x) = 0$ $a.s.$ 成立即可。

由马尔科夫不等式和 Jansen 不等式有

$$\begin{aligned}Q_n(x) &\le E_n\left|\beta_n(x) - \beta(x)\right| \\ &\le |A(x)| E_n\left|\rho_n(x) - \rho(x)\right| + |B(x)| E_n\left|f_n(x) - f(x)\right| \\ &\quad + |C(x)| E_n\left|f_n^{(1)}(x) - f^{(1)}(x)\right| + E_n\left|f_n^{(2)}(x) - f^{(2)}(x)\right| \\ &\le |A(x)|\left[E_n\left|\rho_n(x) - \rho(x)\right|^2\right]^{\frac{1}{2}} + |B(x)|\left[E_n\left|f_n(x) - f(x)\right|^2\right]^{\frac{1}{2}} \\ &\quad + |C(x)|\left[E_n\left|f_n^{(1)}(x) - f^{(1)}(x)\right|^2\right]^{\frac{1}{2}} + \left[E_n\left|f_n^{(2)}(x) - f^{(2)}(x)\right|^2\right]^{\frac{1}{2}}\end{aligned}$$

再由引理 3.1 中的(I)知， $\forall x \in \Omega$， 当 $r = 0,1,2,3$ 时， 有

$$\begin{aligned}\lim_{n\to\infty} Q_n(x) &\le |A(x)|\left[\lim_{n\to\infty} E_n\left|\rho_n(x) - \rho(x)\right|^2\right]^{\frac{1}{2}} + |B(x)|\left[\lim_{n\to\infty} E_n\left|f_n(x) - f(x)\right|^2\right]^{\frac{1}{2}} \\ &\quad + |C(x)|\left[\lim_{n\to\infty} E_n\left|f_n^{(1)}(x) - f^{(1)}(x)\right|^2\right]^{\frac{1}{2}} + \left[\lim_{n\to\infty} E_n\left|f_n^{(2)}(x) - f^{(2)}(x)\right|^2\right]^{\frac{1}{2}} = 0\end{aligned} \tag{16}$$

将式(16)带入式(15)，定理得证。

定理 4.2：设 $\delta_n(x)$ 由(12)式定义，其中 $X_{11}, X_{12}, \cdots, X_{1n_1}, \cdots, X_{21}, X_{22}, \cdots, X_{2n_2}, \cdots$ 为独立同分布随机变量序列。假定条件 $(A1)$ 成立，且 $f(x) \in C_{S,a}$， 若 $0 < \lambda \le 1$ 有

(iv) $\int_\Omega x^{m\lambda} |\beta(x)|^{1-\lambda} \mathrm{d}x < \infty, m = 0,1,2.$

则当取 $h_n = n^{-\frac{1}{2s+4}}$ 时，有

$$R(\delta_n, G) - R(\delta_G, G) = O\left(n^{-\frac{\lambda(s-3)}{2s+4}}\right), \tag{17}$$

其中 $s \ge 4$ 为正整数。

证明：由引理 3.2，引理 3.3 和 Markov 不等式得

$$\begin{aligned}
0 &\le R(\delta_n, G) - R(\delta_G, G) \le a\int_\Omega |\beta(x)|^{1-\lambda} E|\beta_n(x) - \beta(x)|^\lambda \,\mathrm{d}x \\
&\le c_1 \int_\Omega |\beta(x)|^{1-\lambda} |A(x)|^\lambda \, E|\rho_n(x) - \rho(x)|^\lambda \,\mathrm{d}x \\
&\quad + c_2 \int_\Omega |\beta(x)|^{1-\lambda} |B(x)|^\lambda \, E|f_n(x) - f(x)|^\lambda \,\mathrm{d}x \\
&\quad + c_3 \int_\Omega |\beta(x)|^{1-\lambda} |C(x)|^\lambda \, E\left|f_n^{(1)}(x) - f^{(1)}(x)\right|^\lambda \,\mathrm{d}x \\
&\quad + c_4 \int_\Omega |\beta(x)|^{1-\lambda} |B(x)|^\lambda \, E\left|f_n^{(2)}(x) - f^{(2)}(x)\right|^\lambda \,\mathrm{d}x \\
&= A_n + B_n + C_n + D_n.
\end{aligned} \tag{18}$$

由引理 3.2 (II)和条件(iv)知

$$A_n \le c_1 n^{-\lambda/2} \int_\Omega |A(x)|^\lambda \left|\beta(x)^{1-\lambda}\right| \mathrm{d}x \le c_5 n^{\frac{-\lambda(s-3)}{2s+4}} \tag{19}$$

$$B_n \le c_2 n^{\frac{-\lambda(s+1)}{2s+3}} \int_\Omega |B(x)|^\lambda |\beta(x)|^{1-\lambda} \mathrm{d}x \le c_6 n^{\frac{-\lambda(s+1)}{2s+3}} \tag{20}$$

$$C_n \le c_3 n^{\frac{-\lambda(s-1)}{2s+4}} \int_\Omega C|(x)|^\lambda |\beta(x)|^{1-\lambda} \mathrm{d}x \le c_7 n^{\frac{-\lambda(s-1)}{2s+4}} \tag{21}$$

$$D_n \le c_4 n^{\frac{-\lambda(s-3)}{2s+4}} \int_\Omega |D(x)|^\lambda |\beta(x)|^{1-\lambda} \mathrm{d}x \le c_8 n^{\frac{-\lambda(s-3)}{2s+4}} \tag{22}$$

其中 $c, c_1, c_2, \cdots, c_8$ 表示常数。

将式(19)~(22)带入(18)得

$$R(\delta_n, G) - R(\delta_G, G) = O\left(n^{-\frac{\lambda(s-3)}{2s+4}}\right).$$

定理得证。

从定理 4.2 可以看出，当 $\lambda \to 1, s \to \infty$ 时，$O\left(n^{-\frac{\lambda(s-3)}{2s+4}}\right)$ 收敛于 $O\left(n^{-\frac{1}{2}}\right)$。

基金项目

国家自然科学基金项目(A011103)；数据科学广西高等学校重点实验室资助项目。

参考文献 (References)

[1] John Jr., M.V. and Van Ryzin, J. (1971) Convergence Rates in Empirical Bayes Two-Action Problems 1: Discrete Case. *The Annals of Mathematical Statistics*, **42**, 1521-1539. http://dx.doi.org/10.1214/aoms/1177693151

[2] John Jr., M.V. and Van Ryzin, J. (1972) Convergence Rates in Empirical Bayes Two-Action Problems 2: Continuous Case. *The Annals of Mathematical Statistics*, **43**, 934-947. http://dx.doi.org/10.1214/aoms/1177692557

[3] Shang, S.G. and Li, J. (2005) One Empirical Bayes Procedures for Selecting Good Populations in a Positive Exponen-

tial Family. *Journal of Statistical Planning and Inference*, **129**, 3-18. http://dx.doi.org/10.1016/j.jspi.2004.06.036

[4] 韦来生. 刻度指数族参数的经验 Bayes 检验问题: NA 样本情形[J]. 应用数学学报, 2000, 23(3): 403-412.

[5] 李乃医. 纵向数据下一类可靠性分布参数的经验贝叶斯统计分析[J]. 中北大学学报(自然科学版), 2010, 31(5): 438-442.

[6] 陈家清, 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(10): 27-29.

[7] 王莉. 不完全数据下不同统计模型的估计和检验[D]: [硕士学位论文]. 西安: 西北大学, 2012: 14-19.

[8] 王海建, 赵跃生. 删失数据密度函数估计及其强一致收敛速度[J]. 应用概率统计, 1999, 15(3): 240-244.