International Journal of Pure and Applied Mathematics

Volume 27 No. 1 2006, 11-20

EMPIRICAL BAYES TEST FOR ONE-SIDE TRUNCATION PARAMETERS WITH ASYMMETRIC LOSS FUNCTIONS USING NA SAMPLES

Jun-Feng Zhao[1], De-Qin Chen[2], Yong Xu[1 §], Yi-Min Shi[1]

[1,3,4]Department of Applied Mathematics
Northwestern Polytechnical University
Xi'an, 710072, P.R. CHINA
[1]e-mall: zhaojf@nwpu.edu.cn
[3]e-mail: hsux3@263.net
[4]e-mail: ymshi@nwpu.edu.cn
[2]Department of Mathematics
Sichuan University of Science and Engineering
Zigong, Sichuan, 643000, P.R. CHINA
e-mail: chendeqin-66@163.com

Abstract: In the case of negatively associated (NA) samples, this paper is to investigate empirical Bayes test of parameters in one-side truncated distribution family under the asymmetric loss function of the form $L(\theta,\theta_0) = k_1(\theta-\theta_0)^2 I_{(\theta<\theta_0)} + [k_1(\theta-\theta_0)^2 + k_2(\theta-\theta_0)]\, I_{(\theta\geq\theta_0)}, k_i \geq 0$, $i = 1, 2$. The kernel estimation of probability density function is used to construct EB test function, and its asymptotical optimality is obtained.

AMS Subject Classification: 62C12
Key Words: negatively associated (NA), asymmetric loss functions, empirical Bayes (EB) test, asymptotical optimality

1. Introduction

Empirical Bayes tests for parameters of truncation distribution families and its asymptotical optimality under the linear loss or linear weighted loss func-

Received: December 16, 2005

§Correspondence author

tions have been examined by many investigators [13], [8], [6], [7], however in some practical problems such as reliability and predictions linear loss or linear weighted loss functions may be inappropriate [10], [17], [1], [3], [4], [9], [14], [11]. Varian [10] introduced a kind of asymmetric loss function called linex loss function, that has been applied greatly in practical problems such as reliability and predictable problems (for instance, see Zellner [17] for the Bayes analysis of several statistical estimation and prediction problems, Basu and Ebrahimi [1] applied the linex loss in lifetime testing and reliability estimation, Huang [3] for empirical Bayes testing procedures in a class of nonexponential families, Huang and Liang [4] for the empirical Bayes estimation of the truncation parameter with linex loss, Shi and Xu [9], [14] for EB estimation of two-side truncated parameters in the case of NA samples with linex loss, and Xu [16] for EB estimation for one side-truncation distribution under the Canfield asymmetric loss function [2]). These investigations describe the application of asymmetric loss functions are of importance in practical science.

However, in reliability, penetration theory and multivariable analysis the random samples are not always iid samples but negatively associated (NA). Few articles investigated on EBT or EBE are reported [11], especially using asymmetric loss function in the case of NA samples [14], [15], [12].

In this paper, we construct the EBT with asymptotic optimality in one-side truncation distribution family with a kind of asymmetric loss functions introduced by Canfield [2] under NA samples. Firstly, we introduce the definition of NA sequence by Joag-Dev and Proschan [5].

Definition 1.1. Random variables $X_1 X_2 \cdots X_n (n \geq 2)$ are said to be NA if for every pair of disjoint subsets T_1 and T_2 of $\{1, 2, 3, \cdots, n\}$, $\mathrm{Cov}(f_1(X_i, i \in T_1), f_2(X_j, j \in T_2)) \leq 0$, where f_1 and f_2 are increasing or decreasing for every variable such that this covariance exists. Random variables sequence $\{X_i, i \in N\}$ are said to be NA if for every natural number $n \geq 2$, $X_1 X_2 \cdots X_n$ are negatively associated.

Consider a kind of one-side truncation distribution family with probability density function (pdf) of the form:

$$f(x\,|\theta) = u(x)\varphi(\theta) I_{(a<\theta<x<b)}, \tag{1.1}$$

where $u(x)$ is positive, integral on (a, b), $\varphi(\theta) = [\int_\theta^b u(x)dx]^{-1}, -\infty \leq a < \theta < b \leq +\infty$. In this paper we only consider (1.1) and the other form $f_1(x\,|\theta) = u_1(x)\varphi_1(\theta) I_{(a<x<\theta<b)}$ can be similarly investigated, where $u_1(x)$ is positive, integral on$(a, b), \varphi_1(\theta) = [\int_a^\theta u_1(x)dx]^{-1}, -\infty \leq a < \theta < b \leq +\infty$.

Let pdf of r.v.X be the form of (1.1) and the prior distribution function of θ is assumed to be $G(\theta)$ with pdf $g(\theta)$. Then the marginal density function of

X can be written as,

$$f(x) = \int_a^x f(x\,|\theta)dG(\theta) = \int_a^x u(x)\varphi(\theta)dG(\theta). \tag{1.2}$$

The hypothesis to be tested is

$$H_0 : \theta < \theta_0 \leftrightarrow H_1 : \theta \geq \theta_0. \tag{1.3}$$

Now let the asymmetric loss function be

$$L(\theta, d) = L(\theta, d_0) + L(\theta, d_1),$$

where $L(\theta, d_0) = k_1(\theta - \theta_0)^2 I_{(\theta<\theta_0)}$, $L(\theta, d_1) = [k_1(\theta - \theta_0)^2 + k_2(\theta - \theta_0)]I_{(\theta\geq\theta_0)}$; $k_j > 0, j = 1, 2$, d_i indicates accepting H_i $(i = 0, 1)$ and $D = \{d_0, d_1\}$ is the decision space.

Suppose that $\delta(x) = P(\text{accept} H_0\,|X = x)$ is the randomized decision function, then one can obtain the Bayes risk of $\delta(x)$ as,

$$\begin{aligned} R(G, \delta) &= \iint \{L(\theta, d_0)\delta(x) + L(\theta, d_1)(1 - \delta(x))\}f(x\,|\theta)dG(\theta)dx \\ &= \iint \{L(\theta, d_0) - L(\theta, d_1)\}\delta(x)f(x\,|\theta)dG(\theta)dx \\ &\quad + \iint L(\theta, d_1)f(x\,|\theta)dG(\theta)dx = \int r(x)\delta(x)dx + C_G, \end{aligned} \tag{1.4}$$

where $C_G = \iint L(\theta, d_1)f(x\,|\theta)dG(\theta)dx$,

$$\begin{aligned} r(x) &= \int \{L(\theta, d_0) - L(\theta, d_1)\}f(x\,|\theta)dG(\theta) \\ &= \int L(\theta, d_0)f(x\,|\theta)dG(\theta) - \int L(\theta, d_0)f(x\,|\theta)dG(\theta) \\ &= \int k_1(\theta - \theta_0)^2 f(x\,|\theta)dG(\theta) - \int [k_1(\theta - \theta_0)^2 + k_2(\theta - \theta_0)]f(x\,|\theta)dG(\theta) \\ &= k_1 \int \theta^2 f(\theta\,|x)dG(\theta) + (k_2 - 2k_2\theta_0) \int \theta f(\theta\,|x)dG(\theta) + (k_1\theta_0^2 - k_2\theta_0)f(x)\,. \end{aligned}$$

Note that

$$\begin{aligned} \int_a^x \frac{u(x)}{u(t)}dF(t) &= \int_a^x \int_a^t \frac{u(x)}{u(t)}f(t\,|\theta)dG(\theta)dt = \int_a^x \int_a^t u(x)\varphi(\theta)dG(\theta)dt \\ &= xf(x) - E(\theta\,|x)f(x) = xf(x) - \int \theta f(\theta\,|x)dG(\theta), \end{aligned}$$

$$\int_a^x t\frac{u(x)}{u(t)}dF(t) = \int_a^x \int_a^t t\frac{u(x)}{u(t)}f(t\,|\theta)dG(\theta)dt = \int_a^x \int_a^t tu(x)\varphi(\theta)dG(\theta)dt$$
$$= \frac{x^2}{2}f(x) - \frac{1}{2}E(\theta^2\,|x)f(x) = \frac{x^2}{2}f(x) - \int \frac{\theta^2}{2}f(\theta\,|x)dG(\theta),$$

then the expression $r(x)$ can be rewritten as:

$$r(x) = m(x)f(x) - m_1w_1(x) - m_2w(x), \tag{1.5}$$

where $m(x) = k_2x^2 + k_2x - 2k_2\theta_0x + 2k_1\theta_0^2 - k_2\theta_0$, $m_1 = 2k_1$, $m_2 = k_2 - 2k_2\theta_0$, $w_1(x) = \int_a^x t\frac{u(x)}{u(t)}dF(t)$, $w(x) = \int_a^x \frac{u(x)}{u(t)}dF(t)$, and $F(t)$denotes the distribution funcation of X.

From (1.4) one can obtain the Bayes test function as:

$$\delta_G(x) = \begin{cases} 1, & r(x) \le 0, \\ 0, & r(x) > 0. \end{cases} \tag{1.6}$$

Then the Bayes risk of $\delta_G(x)$ can be written as:

$$R_G = \inf_\delta R(\delta, G) = R(\delta_G, G) = \int_a^b r(x)\delta_G(x)dx + C_G. \tag{1.7}$$

Note that in (1.7) Bayes risk $R(G, \delta_G)$ can be obtained accurately when prior distribution $G(\theta)$ is known and $\delta(x) = \delta_G(x)$. However, in practical problems usually $G(\theta)$ is unknown and then $\delta_G(x)$ could not be applied. Then the empirical Bayes approach is introduced to construct the EB test function of the parameter θ in (1.1) using the kernel estimation of density functions of identical NA samples and the asymptotical optimality is to be obtained.

Let $(X_1, \theta_1), (X_2, \theta_2), \cdots, (X_n, \theta_n)$ and (X, θ) be weakly stationary NA samples with the identical pdf $f(x|\theta)$ of the form (1.2), θ_i and θ have the identical prior distribution $G(\theta)(i = 1, 2, 3, \cdots)$. $X_1, X_2, \cdots, X_n$ are said to be past samples and X the present sample with the identical marginal pdf $f(x)$ of the form (1.2), and $f(x) \in C_{s,\alpha},\ x \in R^1$, where $C_{s,\alpha}$ denotes a family of probability density function with s times derivatives and its absolute value not exceeding α, $\alpha \in Z^+$, Z^+ is the set of the positive integers, $s > 1$, $s \in N$. It is assumed in this paper that NA sequence has the property $\sum\limits_{i,j=1}^{\infty} [-\mathrm{Cov}(X_i, X_j)] < \infty$.

Let $k(x)$ be Borel measurable and bounded function on $(0, 1)$, such that:

(1) $\int_0^1 y^t k\,(y)\,dy = \begin{cases} 1\,, & t = 0, \\ 0\,, & t \ne 0,\ t = 1, 2 \cdots, s - 1. \end{cases}$

(2) $k(x)$ is differentiable on R^1 not including finite set E_0 and

$$\sup_{x \in R^1 \setminus E_0} \left| k'(x) \right| \leq c < \infty.$$

Therefore the estimation of $f(x)$ is defined to be

$$f_n(x) = \frac{1}{nh} \sum_{i=1}^{n} k(\frac{x - x_i}{h}), \tag{1.8}$$

with $h = h_n \to 0\,(n \to \infty)$. The EB test of θ can be defined to be

$$\delta_n(x) = \begin{cases} 1 & r_n(x) \leq 0, \\ 0 & r_n(x) > 0, \end{cases} \tag{1.9}$$

where

$$r_n(x) = m(x) f_n(x) - m_1 w_{1n}(x) - m_2 w_n(x),$$

$$w_{1n}(x) = \frac{1}{n} \sum_{i=1}^{n} X_i \frac{u(x)}{u(X_i)} I_{(a<X_i<x)}(X_i), w_n(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{u(x)}{u(X_i)} I_{(a<X_i<x)}(X_i).$$

The all Bayes risk of $\delta_n(x)$ can be obtained as:

$$R(G, \delta_n) = E_n \int_a^b r(x) \delta_n(x) dx + C_G = \int_a^b r(x) E_n[\delta_n(x)] dx + C_G, \tag{1.10}$$

where E_n indicates the expectation with respect to joint distribution of $(X_1, \cdots, X_n)$.

The EB test $\{\delta_n(x)\}$ is said to be asymptotically optimal if $\lim\limits_{n \to \infty} R_n = R_G$ for every $G(\theta) \in F$, where F is the prior distribution family of θ.

Now we claim the following theorem:

Theorem 1. *Let $\delta_n(x)$ be defined by (1.9), $\sup\limits_x |f(x)| < \infty$, $\sup\limits_x \left| f^{(s)}(x) \right| < \infty$, for every $0 < \delta < 1$, such that:*

(1) $E(\theta) < \infty, E(\theta^2) < \infty$,

(2) $f(x)$ is continuous and $u(x)$ is nondecreasing for x.

Then $\lim\limits_{n \to \infty} R_n = R_G$ as $nh \to \infty$.

2. Some Lemmas

In this section it is assumed that M is a positive constant in different cases even in the same expression.

Now we introduce the following lemmas to complete the proof of Theorem 1.

Lemma 2.1. *Let X and Y be NA variables with finite covariance for differentiable functions g_1 and g_2 ,*

$$|\mathrm{Cov}(g_1(X), g_2(X))| \leq \sup_X \left|g_1'(X)\right| \sup_Y \left|g_Y'(Y)\right| [-\mathrm{Cov}(X, Y)],$$

As g_1 and g_2 are not differentiable on finite or countable set E_0^1 and E_0^2, we have

$$|\mathrm{Cov}(g_1(X), g_2(X))| \leq \sup_{X\in R^1\setminus E_0^1} \left|g_1'(X)\right| \sup_{Y\in R^1\setminus E_0^2} \left|g_Y'(Y)\right| [-\mathrm{Cov}(X, Y)].$$

Proof. See Lemma 2.1 in [11]. □

Lemma 2.2. *Let R_n and R_G be defined above, then*

$$0 \leq R_n - R_G \leq a \int |\alpha(x)| P\left(|\alpha_n(x) - \alpha(x)| \geq |\alpha(x)|\right) dx\,.$$

Proof. See Lemma 1 in Johns and Van Ryzin [6]. □

Lemma 2.3. *Let $f_n(x)$ be defined by (1.8), $\sup_x |f(x)| < \infty$, $\sup_x \left|f^{(s)}(x)\right| < \infty$, then for every $0 < \delta < 1$, one has*

$$E_n\,|f_n(x) - f(x)| \to 0 \ (n \to \infty) \text{ as } nh^4 \to \infty.$$

Proof. $E_n\,|f_n(x) - f(x)|^2 \leq (\mathrm{Var}[f_n(x)]) + |E_n f_n(x) - f(x)|^2$. But

$$\begin{aligned}\mathrm{Var}[f_n(x)] &= (nh)^{-2}\mathrm{Var}[\sum_{i=1}^n k(\frac{X_i - x}{h})]\\ &= (nh)^{-2}\{\sum_{i=1}^n \mathrm{Var}[k(\frac{X_i - x}{h})] + 2\sum_{i<j} \mathrm{Cov}(k(\frac{X_i - x}{h}), k(\frac{X_j - x}{h}))\}\\ &\leq (nh)^{-2}\{nE[k(\frac{X_i - x}{h})]^2 + 2\sum_{i<j} \mathrm{Cov}(k(\frac{X_i - x}{h}), k(\frac{X_j - x}{h}))\},\end{aligned}$$

where $(nh)^{-2}nE[k(\frac{X_i-x}{h})]^2 = (nh)^{-1}\int_0^1 f(x+hu)k^2(u)du \le M(nh)^{-1}$.

From Lemma 2.1 we get $(nh)^{-2}\sum\limits_{i<j}\mathrm{Cov}(k(\frac{X_i-x}{h}),k(\frac{X_j-x}{h})) \le M(nh^4)^{-1}$.

Then $\mathrm{Var}[f_n(x)] \to 0,\ nh^4 \to \infty$.

Note that

$$|E_n f_n(x) - f(x)| = \frac{h^s}{s!}\int_0^1 f^{(s)}(x+hu\xi)k(u)u^s du \le Mh^s$$

and $h_n \to 0\,(n\to\infty)$, then $|E_n f_n(x) - f(x)| \to 0\ (n\to\infty)$.

Therefore $E_n\,|f_n(x)-f(x)| \to 0 \quad (n\to\infty)$ as $nh^4 \to \infty$. □

Lemma 2.4. *Assume that $u(x)$ is nondecreasing for x, then:*

$$(1)\qquad E_n\,|w_n(x)-w(x)|^{2\delta} \le n^{-\delta}[\int_a^x (\frac{u(x)}{u(t)})^2 dF(t)]^\delta;$$

$$(2)\qquad E_n\,|w_{1n}(x)-w_1(x)|^{2\delta} \le n^{-\delta}[\int_a^x (t\frac{u(x)}{u(t)})^2 dF(t)]^\delta.$$

Proof. Here we only prove (1), and (2) can be obtained similarly.

Since $u(x)$ is nondecreasing for x, $I_{(a<X_i<x)}$ is decreasing for X_i, we get

$$\frac{u(x)}{u(X_i)}I_{(a<X_i<x)}$$

is decreasing for X_i. By the definition of NA sequence, $\{\frac{u(x)}{u(X_i)}I_{(a<X_i<x)},\ i = 1,2,3,...\}$ is still NA, i.e. $\mathrm{Cov}(\frac{u(x)}{u(X_i)}I_{(a<X_i<x)}, \frac{u(x)}{u(X_j)}I_{(a<X_j<x)}) \le 0$. Then

$$\sum_{i<j}\mathrm{Cov}(\frac{u(x)}{u(X_i)}I_{(a<X_i<x)}, \frac{u(x)}{u(X_j)}I_{(a<X_j<x)}) \le 0.$$

Hence

$$E_n\,|w_n(x)-w(x)|^2 = \mathrm{Var}(w_n(x)) = n^{-2}\left\{\sum_{i=1}^n \mathrm{Var}[\frac{u(x)}{u(X_i)}I_{(a<X_i<x)}]\right.$$

$$\left. +2\sum_{i<j}\mathrm{Cov}(\frac{u(x)}{u(X_i)}I_{(a<X_i<x)}, \frac{u(x)}{u(X_j)}I_{(a<X_j<x)})\right\}$$

$$= n^{-2}\left\{n\int_a^x (\frac{u(x)}{u(t)})^2 dF(t) + 2\sum_{i<j}\mathrm{Cov}(\frac{u(x)}{u(X_i)}I_{(a<X_i<x)}, \frac{u(x)}{u(X_j)}I_{(a<X_j<x)})\right\}$$

$$\leq n^{-1} \int_a^x (\frac{u(x)}{u(t)})^2 dF(t).$$

From Jensen inequality one can easily complete the proof of Lemma 2.4. □

3. Proof of Theorem 1

By Lemma 2.2, one arrives at

$$0 \leq R_n - R_G \leq \int |r(x)| P\left(|r_n(x) - \alpha(x)| \geq |\alpha(x)|\right) dx.$$

Let $\beta_n(x) = |r(x)|\, P\left(|r_n(x) - r(x)| \geq |r(x)|\right)$, then $\beta_n(x) \leq |r(x)|$. Begin with

$$\int |r(x)|\, dx \leq k_1 \int \left| \int \theta^2 f(\theta\,|x) dG(\theta) \right| dx$$
$$+ |(k_2 - 2k_2\theta_0)| \int \left| \int \theta f(\theta\,|x) dG(\theta) \right| dx + \left|(k_1\theta_0^2 - k_2\theta_0) f(x)\right| dx$$
$$\leq k_1 E(\theta^2) + |(k_2 - 2k_2\theta_0)E(\theta)| + \left|k_1\theta_0^2 - k_2\theta_0\right| \int f(x) dx$$
$$= k_1 E(\theta^2) + |(k_2 - 2k_2\theta_0)E(\theta)| + \left|k_1\theta_0^2 - k_2\theta_0\right| < \infty\,.$$

By Markov inequality, one can get

$$\beta_n(x) \leq E_n\,|r_n(x) - r(x)|\ \leq m_1(x) E_n\,|f_n(x) - f(x)|$$
$$+ m_1 E_n\,|w_{1n}(x) - w_1(x)| + m_2 E_n\,|w_n(x) - w(x)|\ .$$

From Lemma 2.3 and Lemma 2.4, we obtain

$$\lim_{n\to\infty} \beta_n(x) = 0$$

for every fixed $x, a < x < b$.

Lastly we apply the dominated convergence theorem to get

$$0 \leq \lim_{n\to\infty} (R_n - R_G) \leq \int \lim_{n\to\infty} \beta_n(x) dx.$$

This completes the proof of Theorem 1. □

Remark. From Theorem 1, one can find that the proposed EB test is asymptotical optimal, and its convergence rate will be obtained in our future work.

Acknowledgments

This work was supported by NSF of P.R. China under Grant No. 70471057 and Foundation of the Educational Committee of Shaanxi Province under Grant no. 03JK065.

References

[1] A.P. Basu, N. Ebrahimi, Bayesian approach to life testing and reliability estimation using asymmetric loss function, *J. Statist. Plann. Inference*, **29** (1991), 21-31.

[2] R.V. Canfield, A Bayesian approach to reliability estimation using a loss function, *IEEE Trans. Reliab.*, **19** (1970), 13-16.

[3] S.Y. Huang, Empirical Bayes testing procedures in some nonexponential families using asymmetric Linex loss, *J. Statist. Plann. Inference*, **46** (1995), 293-309.

[4] S.Y. Huang, T.C. Liang, Empirical Bayes estimation of the truncation parameter with Linex loss, *Statistica Sinica*, **7** (1997), 755-769.

[5] K. Joag-Dev, F. Proschan, Negative association of random variables with application, *Ann. Stasist.*, **11** (1983), 286-295.

[6] M.V. Johns Jr., J. Van Ryzin, Convergence rates in empirical Bayes two-action problem II: Continuous case, *Ann. Math. Statist.*, **43** (1972), 934-947.

[7] R.J. Karunamuni, Optimal rates of convergence of empirical Bayes tests for the continuous one-parameter exponential family, *Ann. Stasist.*, **24** (1996), 212-231.

[8] Y.M. Ma, N. Balakarishnan, Empirical Bayes estimation for truncation parameters, *J. Statist. Plann. Inference*, **84** (2000), 111-120.

[9] Y.M. Shi, Empirical Bayes estimation for parameter of two-sided truncated distribution under Linex loss function, *Appl. Math. J. Chinese Univ. Ser. A*, **15**, No. 4 (2000), 475-483, In Chinese.

[10] H.R. Varian, A Bayesian approach to real estate assessment, In: *Studies in Bayesian Economics and Statistics in Honour of L.S. Savage* (Ed-s: S.E. Feinberg, A. Zellner), North-Holland (1975), 195-208.

[11] L.S. Wei, The empirical Bayes test problem for scale exponential family: in the case of NA samples, *Acta Mathematicae Applicatae Sinica*, **23**, No. 3 (2000), 403-412, In Chinese.

[12] Y. Xu, Y.M. Shi, Empirical Bayes test for truncation parameter with NA samples, *Mathematicae Applicatae*, **14**, No. 4 (2001), 98-102.

[13] Y. Xu, Y.M. Shi, Empirical bayes test for truncation parameters using linex loss, *Bulletin of The Institute of Mathematics*, Academia Sinica, **3** (2004), 207-220.

[14] Y. Xu, X.L. Shi, Y.M. Shi, The EB estimation of parameter in truncated family with Linex loss using NA Samples, *Chinese Journal of Mathematics*, **2** (2004), 124-130.

[15] Y. Xu, S.L. Wang, Y.M. Shi, Empirical Bayes estimation for truncation parameters using NA samples, *Chinese Quarterly Journal of Mathematics*, **17**, No. 4 (2002), 43-47.

[16] Y.S. Xu, Y. Xu, Y.M. Shi, Convergence rates of empirical Bayes test for one-side truncation parameters with asymmetric loss functions, *International Journal of Pure and Applied Mathematics*, To Appear.

[17] A. Zellner, Bayesian estimation and prediction using asymmetric loss functions, *J. Amer. Stat. Assoc.*, **81** (1986), 446-451.