Statistica Sinica **7**(1997), 755-769

EMPIRICAL BAYES ESTIMATION OF THE TRUNCATION PARAMETER WITH LINEX LOSS

Su-Yun Huang and TaChen Liang

Academia Sinica and Wayne State University

Abstract: This paper deals with the empirical Bayes estimation of the truncation position of the truncated family under the Linex loss. Nonparametric empirical Bayes estimator is proposed and its asymptotic optimality and rate are investigated. Under certain mild conditions without any differentiability assumption on either the prior or the marginal distribution, it is shown that the proposed empirical Bayes estimator is asymptotically optimal with convergence rate given. Simulation results on the performance of the proposed empirical Bayes estimator are also presented.

Key words and phrases: Asymptotic optimality, empirical Bayes estimator, Linex loss, rate of convergence, regret risk.

1. Introduction

Consider the truncated family with *pdf* of the following form

$$f(x|\theta) = \frac{h(x)}{A(\theta)} I_{(\theta,\infty)}(x), \tag{1.1}$$

where $h(x)$ is positive, continuous and decreasing, and where $\theta \geq 0$ is the truncation parameter of our interest. Important examples are the translated exponential distribution and the Pareto distribution. (See Mann, Schafer and Singpurwalla (1974) and Lawless (1982) for model discussions on the translated exponential distribution and applications to lifetime data analysis. See also Arnold (1983) for the Pareto distribution and applications to socio-economic data.)

In this paper, we consider empirical Bayes estimation of θ under the asymmetric Linex loss. As we know that, in the lifetime model or in the Pareto distribution for socio-economic data or in many other models, the levels of seriousness for over-estimation and under-estimation are quite different. The asymmetric Linex loss was introduced by Varian (1975) and later adopted by Zellner (1986) in the Bayesian analysis of several statistical estimation and prediction problems. For other Linex loss applications see also Parsian (1990) for a generalized Bayes estimator of a multivariate normal mean, Kuo and Dey (1990) for Poisson mean estimation, Basu and Ebrahimi (1991) for lifetime testing and reliability estimation and Huang (1995) for empirical Bayes testing procedures in a class of non-exponential families. The authors feel that the studies of statistical inferential problems with Linex loss are still too few despite its practical potentiality. We hope to make some contribution to the Linex loss literature and to draw more attention and research efforts to its development and applications.

The Linex loss function is given herewith. Let a denote an estimate of θ. The loss is

$$L(\theta, a) = \exp\{c(a-\theta)\} - c(a-\theta) - 1, \ c \in R. \tag{1.2}$$

For $c > 0$, the loss function $L(\theta, a)$ is quite asymmetric about 0 with over-estimation being more costly than under-estimation. As $|a-\theta| \to \infty$, the loss $L(\theta, a)$ increases almost exponentially when $a - \theta > 0$ and almost linearly when $a - \theta < 0$. For $c < 0$, the linearity-exponentiality phenomenon is reversed. Also, when $|a-\theta|$ is very small, $L(\theta, a)$ is near $c(a-\theta)^2/2$.

In this paper, the Linex loss is employed for estimation of the truncation parameter in model (1.1). The rest of the paper is organized as follows. In Section 2, a Bayesian framework is introduced and a Bayes estimator is derived in terms of the prior distribution. A certain nonparametric empirical Bayes (NPEB) estimator is proposed to handle the unknown prior in the Bayes estimator. In Section 3, under certain mild conditions without any differentiability assumption, the proposed NPEB estimator is shown to be asymptotically optimal with convergence rate (an upper bound) provided. A lower bound for convergence rate is established in Section 4 and the proposed NPEB estimator is shown to achieve the rate $O(n^{-2/3})$. In Section 5, some simulation studies are carried out to investigate the performance of the proposed estimator.

2. Bayes and Empirical Bayes Estimators

In this paper, we consider only $c > 0$. Discussion for the other case where $c < 0$ is similar and therefore omitted here. Assume that the parameter θ is a realization of a nonnegative random variable Θ which has an unknown prior distribution G over $[0, \infty)$. Then, the Bayes estimator of the parameter θ given $X = x$ is $\varphi_G(x)$ which minimizes $\int_{\theta=0}^{x}\{\exp[c(a-\theta)] - c(a-\theta) - 1\} dG(\theta|x)$ among all estimators a. Let $f(x) = \int_{\theta=0}^{x} f(x|\theta) dG(\theta)$, the marginal *pdf* of X, and $\psi(x) = \int_{\theta=0}^{x} (A(\theta))^{-1} dG(\theta)$. Then the marginal *pdf* $f(x)$ can be written as $f(x) = h(x)\psi(x)$.

Some straightforward computation yields $\varphi_G(x) = -c^{-1} \ln E(e^{-c\theta}|X) \equiv c^{-1} \ln \tau_G(X)$, where

$$\tau_G(x) = \frac{\int_0^x f(x|\theta) dG(\theta)}{\int_0^x e^{-c\theta} f(x|\theta) dG(\theta)} = \frac{f(x)}{e^{-cx} f(x) + K(x)} \tag{2.1}$$

with

$$K(x) = \int_0^x ce^{-ct}\left(\int_0^t f(x|\theta)dG(\theta)\right)dt = h(x)\int_{t=0}^x ce^{-ct}h^{-1}(t)f(t)dt.$$

Note that

$$1 \le \tau_G(x) = \frac{e^{cx}f(x)}{f(x)+e^{cx}K(x)} \le e^{cx}$$

for any prior distribution G (could be improper, as long as the resulting $f(x)$ is a probability density). Therefore, the Bayes estimator $\varphi_G(x)$ satisfies the condition $0 \le \varphi_G(x) \le x$.

In this paper we assume that the prior distribution G satisfies condition (A1) $G(\theta^\star) = 1$ for some predestined (i.e., known) finite positive constant $\theta^\star$.

Under (A1), for $x > \theta^\star$,

$$\tau_G(x) = \frac{\int_0^{\theta^\star} f(x|\theta)dG(\theta)}{\int_0^{\theta^\star} e^{-c\theta}f(x|\theta)dG(\theta)} = \tau_G(\theta^\star) \tag{2.2}$$

and hence,

$$\varphi_G(x) = \varphi_G(\theta^\star). \tag{2.3}$$

The minimum Bayes risk, attained by the Bayes estimator φ_G, is

$$\begin{aligned}
&R(G,\varphi_G)\\
&= E\{\exp[c(\varphi_G(X)-\Theta)] - c(\varphi_G(X)-\Theta) - 1\}\\
&= \int_{x=0}^\infty \int_{\theta=0}^x \{\exp[c(\varphi_G(x)-\theta)] - c(\varphi_G(x)-\theta) - 1\}dG(\theta|x)f(x)dx\\
&= \int_{x=0}^\infty \Big\{ \exp(c\varphi_G(x))\left(\int_{\theta=0}^x \exp(-c\theta)dG(\theta|x)\right) - c\varphi_G(x)\\
&\qquad + \int_{\theta=0}^x c\theta dG(\theta|x) - 1\Big\} f(x)dx\\
&= \int_{x=0}^\infty \Big\{ \int_{\theta=0}^x c\theta dG(\theta|x) - c\varphi_G(x)\Big\} f(x)dx, \qquad (2.4)
\end{aligned}$$

as $\exp(c\varphi_G(x))\int_{\theta=0}^x \exp(-c\theta)dG(\theta|x) = 1$. Often the prior distribution G is unknown and the Bayes estimator $\varphi_G(\cdot)$ cannot be applied. In this paper, the empirical Bayes approach is employed to handle the uncertainty of G.

In the empirical Bayes framework, consider i.i.d. copies (X_1,θ_1),. . .,(X_n,θ_n) of (X,θ), where θ has a distribution G and, conditional on θ, X has a distribution with *pdf* $f(x|\theta)$. The X's are observed but the θ's are not. Let $\varphi_n(X) \equiv \varphi_n(X; X_1,\ldots,X_n)$ be an empirical Bayes estimator of the truncation parameter θ based on the past data $\underline{X}_n = (X_1,\ldots,X_n)$ and the present observation X. Let $R(G,\varphi_n|\underline{X}_n)$ denote the conditional Bayes risk of the estimator

φ_n given $\underline{X}_n$, and let $R(G, \varphi_n) = E_n[R(G, \varphi_n|\underline{X}_n)]$ denote the unconditional Bayes risk of φ_n, where the expectation E_n is taken with respect to $\underline{X}_n$. Since φ_G is the Bayes estimator, $R(G, \varphi_n|\underline{X}_n) - R(G, \varphi_G) \geq 0$ for all $\underline{X}_n$ and for all n. Therefore, $R(G, \varphi_n) - R(G, \varphi_G) \geq 0$ for all n. The nonnegative regret risk $R(G, \varphi_n) - R(G, \varphi_G)$ is used as a measure of performance of the empirical Bayes estimator φ_n. A sequence of empirical Bayes estimators $\{\varphi_n\}_{n=1}^{\infty}$ is said to be asymptotically optimal if $R(G, \varphi_n) - R(G, \varphi_G) \to 0$ as $n \to \infty$. Moreover, if $R(G, \varphi_n) - R(G, \varphi_G) = O(\alpha_n)$, where $\{\alpha_n\}_{n=1}^{\infty}$ is a sequence of positive numbers such that $\lim_{n\to\infty} \alpha_n = 0$, then the sequence $\{\varphi_n\}_{n=1}^{\infty}$ is said to be asymptotically optimal with convergence rate of order $\{\alpha_n\}_{n=1}^{\infty}$. (See Robbins (1956, 1964).)

A sequence of empirical Bayes estimators $\{\varphi_n(x)\}$ for the truncation parameter is constructed below. Let $\{b_n\}$ be a sequence of positive numbers such that $\lim_{n\to\infty} b_n = 0$ and $\lim_{n\to\infty} nb_n = \infty$. For each n and $x > 0$, let

$$f_n(x) = [F_n(x + b_n) - F_n(x)]/b_n, \tag{2.5}$$

where $F_n(x)$ is the empirical distribution based on $\underline{X}_n$. The estimator f_n is a kernel estimator with a left-sided uniform kernel. The reason for using a left-sided kernel instead of, say, a symmetric order 2 (or even higher order) kernel is based on two reasons. One is from the viewpoint of the consideration to avoid dominant bias at the left boundary. The other is that, for an order two (or higher order) kernel we have to assume continuity of the second (or higher) derivative of $f(x)$. (See Gasser and Müller (1979), Rice (1984) and Schuster (1985) for discussions on boundary behavior of kernel estimators and related theory. Also see Gasser and Müller (1979) and Silverman (1986) for general asymptotics and smoothness assumption.) The smoothness of $f(x)$ depends on the prior distribution G. In this paper, we do not impose any continuous derivative assumption on G. Let

$$K_n(x) = \frac{h(x)}{n} \sum_{j=1}^{n} \frac{ce^{-cX_j}}{h(X_j)} I_{(0,x]}(X_j). \tag{2.6}$$

Note that $E_n K_n(x) = K(x)$. Both $f_n(x)$ and $K_n(x)$ are consistent estimators for $f(x)$ and $K(x)$ respectively. Let

$$\tau_n(x) = \frac{f_n(x)}{e^{-cx} f_n(x) + K_n(x)} \vee 1,$$

where $a \vee b = \max(a, b)$ and $0/0 \equiv 0$. The proposed sequence of empirical Bayes estimators $\{\varphi_n(\cdot)\}$ is

$$\varphi_n(X) = \frac{1}{c} \ln \tau_n(X) I_{(0,\theta^\star]}(X) + \frac{1}{c} \ln \tau_n(\theta^\star) I_{(\theta^\star,\infty)}(X). \tag{2.7}$$

Note that the past data $\underline{X}_n$ is implicitly contained in the subscript n of φ_n.

3. Asymptotic Optimality and Convergence Rate

The Bayes risk $R(G,\varphi_n)$ of φ_n can be written as

$$\begin{aligned}
&R(G,\varphi_n)\\
&=\int_0^\infty E_n\Big\{c\exp(\varphi_n(x))\int_{\theta=0}^{x}\exp(-c\theta)dG(\theta|x)-c\varphi_n(x)\\
&\qquad\qquad+\int_{\theta=0}^{x}c\theta dG(\theta|x)-1\Big\}f(x)dx\\
&=\int_0^\infty E_n\Big\{\exp[c(\varphi_n(x)-\varphi_G(x))]-c\varphi_n(x)+\int_{\theta=0}^{x}c\theta dG(\theta|x)-1\Big\}f(x)dx. \qquad (3.1)
\end{aligned}$$

Also note that, for $x>\theta^\star$, we have $\varphi_G(x)=\varphi_G(\theta^\star)$ and $\varphi_n(x)=\varphi_n(\theta^\star)$. Then from (2.4) and (3.1), we obtain

$$\begin{aligned}
0\le\ &R(G,\varphi_n)-R(G,\varphi_G)\\
&=\int_{x=0}^{\infty}E_n\{\exp[c(\varphi_n(x)-\varphi_G(x))]-c(\varphi_n(x)-\varphi_G(x))-1\}f(x)dx\\
&=\int_{x=0}^{\theta^\star}E_n\{\exp[c(\varphi_n(x)-\varphi_G(x))]-c(\varphi_n(x)-\varphi_G(x))-1\}f(x)dx\\
&\quad+E_n\{\exp[c(\varphi_n(\theta^\star)-\varphi_G(\theta^\star))]-c(\varphi_n(\theta^\star)-\varphi_G(\theta^\star))-1\}\{1-F(\theta^\star)\}, \qquad (3.2)
\end{aligned}$$

where F is the distribution function for f. Let ν_{Leb} be the Lebesgue measure and F^∞ be the product measure induced by $X_1,X_2,\ldots,X_n,\ldots$ By the Lebesgue dominated convergence theorem and the following limiting result $\lim_{n\to\infty}\varphi_n(x)=\varphi_G(x)$, $\nu_{\text{Leb}}\times F^\infty-a.s.$, we can pass to the limit inside the integral and expectations of (3.2) and get $\lim_{n\to\infty}R(G,\varphi_n)-R(G,\varphi_G)=0$. That is, asymptotic optimality is obtained. Below we study the rate of convergence. We need further assumptions. Let G be such a prior distribution that (A2) and (A3) are satisfied where (A2) and (A3) are as follows: (A2) As $b\to 0$,

$$\int_{\theta_0}^{\theta^\star}\frac{(f(x)-f(x+b))^2}{f(x)}\,dx=O(b^2),$$

where θ_0 is the left boundary of G's support. (A3) For any $b>0$,

$$\int_{\theta_0}^{\theta^\star}\frac{b^{-1}\int_x^{x+b}f(t)dt}{f(x)}\,dx<\infty.$$

Remarks on conditions (A2) and (A3)

1. With the translated exponential distribution $f(x|\theta)=\lambda e^{-\lambda(x-\theta)}I_{(\theta,\infty)}(x)$ for the lifetime model, a sufficient condition for both (A2) and (A3) is that the

marginal *pdf* $f(x)$ is decreasing. It is common to expect a decreasing *pdf* for lifetime data. However, if we do not expect a decreasing marginal *pdf*, then an alternative sufficient condition for (A2) and (A3) is that $G(x)$ is Lipschitz continuous.

2. For the Pareto distribution, we have $f(x|\theta) = \alpha\theta^\alpha/x^{\alpha+1}\ I_{(\theta,\infty)}$, where $\alpha > 0$ is known and where x and θ can be explained as income and minimum income respectively. Same as in the translated exponential distribution, the decreasing property assumption of $f(x)$ is sufficient for conditions (A2) and (A3). However, the decreasingness of $f(x)$, which is common in the lifetime model, may not be realistic for income data. Therefore we look for other sufficient conditions. The Lipschitz continuity assumption on G,
$$\sup_{t,x\in(0,\theta^\star]} \frac{|G(x+t)-G(x)|}{t} \le \alpha \text{ for some constant } \alpha > 0,$$
ensures that the minimum income distribution G has bounded first divided difference (or first derivative if it exists) everywhere. This assumption together with the assumption that $\theta_0 > 0$ are sufficient for conditions (A2) and (A3).

Lemma 3.1. *For $a, b \ge 0$, $e^{a-b} - (a-b) - 1 \le (e^a - e^b)^2$.*

Some calculus leads to the above inequality, and the proof is omitted.

By Lemma 3.1 and the definitions of $\varphi_n(x)$ and $\varphi_G(x)$, we have
$$\begin{aligned}
&E_n\{\exp[c(\varphi_n(x)-\varphi_G(x))] - c(\varphi_n(x)-\varphi_G(x)) - 1\}\\
&\le E_n\{\exp(c\varphi_n(x)) - \exp(c\varphi_G(x))\}^2\\
&= E_n\Big(\big(\frac{f_n(x)}{e^{-cx}f_n(x)+K_n(x)} \vee 1\big) - \frac{f(x)}{e^{-cx}f(x)+K(x)}\Big)^2\\
&\le E_n\Big(\Big|\frac{f_n(x)}{e^{-cx}f_n(x)+K_n(x)} - \frac{f(x)}{e^{-cx}f(x)+K(x)}\Big| \wedge e^{cx}\Big)^2, \qquad (3.3)
\end{aligned}$$
since $1 \le \tau_n(x) \le e^{cx}$ and $1 \le \tau_G(x) \le e^{cx}$. A lemma due to Singh (1977) is needed.

Singh's Lemma. *Let $y, z \ne 0$ and $L > 0$ be real numbers. If Y and Z are two random variables, then for every $r > 0$,*
$$\begin{aligned}
&E\left(|\ y/z - Y/Z\ | \wedge L\right)^r\\
&\le 2^{r+(r-1)\vee 0}|z|^{-r}\left\{E|y-Y|^r + \left(|y/z|^r + 2^{-(r-1)\vee 0}L^r\right)E|z-Z|^r\right\},
\end{aligned}$$
where the expectation E is taken with respect to the joint distribution of (Y, Z).

Proceeding from (3.3), we have

$$\begin{aligned}
&E_n\{\exp[c(\varphi_n(x)-\varphi_G(x))]-c(\varphi_n(x)-\varphi_G(x))-1\}\\
&\le \frac{8}{[e^{-cx}f(x)+K(x)]^2}E_n[f_n(x)-f(x)]^2\\
&\quad+\frac{12e^{2cx}E_n[e^{-cx}f_n(x)+K_n(x)-e^{-cx}f(x)-K(x)]^2}{[e^{-cx}f(x)+K(x)]^2}\\
&=\frac{20[E_nf_n(x)-f(x)]^2+32\mathrm{Var}\,(f_n(x))+24e^{2cx}\mathrm{Var}\,K_n(x)}{[e^{-cx}f(x)+K(x)]^2},
\end{aligned}\tag{3.4}$$

as $\mathrm{Var}\,(X+Y)\le 2(\mathrm{Var}\ X+\mathrm{Var}\ Y)$ for random variables X and Y.

Lemma 3.2. *For $x>0$, we have*
(a) $\mathrm{Var}\ (f_n(x))\le b_n^{-1}\int_x^{x+b_n}f(t)dt/(nb_n)$,
(b) $\mathrm{Var}\ K_n(x)\le c^2\psi(x)h(x)(1-e^{-2cx})/(2n)$.

Proof. (a)

$$\begin{aligned}
\mathrm{Var}\,(f_n(x)) &= \frac{[F(x+b_n)-F(x)][1-F(x+b_n)+F(x)]}{nb_n^2}\\
&\le \frac{F(x+b_n)-F(x)}{nb_n^2}=\frac{b_n^{-1}\int_x^{x+b_n}f(t)dt}{nb_n}.
\end{aligned}$$

(b) Begin with

$$\begin{aligned}
&\mathrm{Var}\,(K_n(x))=\frac{c^2h^2(x)}{n}\mathrm{Var}\,(e^{-cX_1}h^{-1}(X_1)I_{(0,x]}(X_1))\\
&\le \frac{c^2h^2(x)}{n}E_n[e^{-2cX_1}(h^{-1}(X_1))^2I_{(0,x]}(X_1)]\\
&=\frac{c^2h^2(x)}{n}\int_0^x e^{-2ct}h^{-1}(t)\psi(t)dt\le\frac{c^2\psi(x)h(x)(1-e^{-2cx})}{2n}.
\end{aligned}\tag{3.5}$$

End of proof.

From (3.4) and Lemma 3.2,

$$\begin{aligned}
&\int_0^{\theta^\star}E_n\{\exp[c(\varphi_n(x)-\varphi_G(x))]-c(\varphi_n(x)-\varphi_G(x))-1\}f(x)dx\\
&\le\int_{x=0}^{\theta^\star}\frac{32f(x)b_n^{-1}\int_x^{x+b_n}f(t)dt}{[e^{-cx}f(x)+K(x)]^2nb_n}dx+\int_{x=0}^{\theta^\star}\frac{20f(x)(E_nf_n(x)-f(x))^2}{[e^{-x}f(x)+K(x)]^2}dx\\
&\quad+\int_{x=0}^{\theta^\star}\frac{12c^2f(x)(e^{2cx}-1)h(x)\psi(x)}{[e^{-cx}f(x)+K(x)]^2n}dx\\
&\le\int_{x=\theta_0}^{\theta^\star}\frac{32e^{2cx}b_n^{-1}\int_x^{x+b_n}f(t)dt}{nb_nf(x)}dx+\int_{x=\theta_0}^{\theta^\star}\frac{20e^{2cx}(E_nf_n(x)-f(x))^2}{f(x)}dx\\
&\quad+\int_{x=\theta_0}^{\theta^\star}\frac{12c^2e^{2cx}(e^{2cx}-1)}{n}dx,
\end{aligned}\tag{3.6}$$

as $f^2(x)/[e^{-cx}f(x)+K(x)]^2 \le e^{2cx}$. By conditions (A2) and (A3), we have

$$(3.6) = O(\frac{1}{nb_n}) + O(b_n^2) + O(\frac{1}{n}). \tag{3.7}$$

Also from Lemma 3.2 and conditions (A2) and (A3), we can easily get

$$\begin{aligned} &E_n\{\exp[c(\varphi_n(\theta^\star) - \varphi_G(\theta)^\star)] - c(\varphi_n(\theta^\star) - \varphi_G(\theta)^\star) - 1\} \\ &= O(\frac{1}{nb_n}) + O(b_n^2) + O(\frac{1}{n}). \end{aligned}$$

The results obtained above are summarized in the theorem below.

Theorem 3.1. *Let $\{\varphi_n\}_{n=1}^\infty$ be the sequence of empirical Bayes estimators constructed in section* 2. *Suppose conditions* (A1), (A2) *and* (A3) *hold. Then $\{\varphi_n\}_{n=1}^\infty$ is asymptotically optimal and, as $n \to \infty$,*

$$R(G,\varphi_n) - R(G,\varphi_G) = O(\frac{1}{nb_n}) + O(b_n^2) + O(\frac{1}{n}).$$

4. A Lower Bound for $R(G,\varphi_n) - R(G,\varphi_G)$

Consider the translated exponential distribution with $h(x) = e^{-x}$. We assume that the prior distribution G is as given below.

$$G(\theta) = \begin{cases} 0, & \text{if } \theta < 0, \\ \frac{1}{2} + \frac{1}{2}\theta, & \text{if } 0 \le \theta \le 1, \\ 1, & \text{if } \theta > 1. \end{cases} \tag{4.1}$$

Then,

$$f(x) = \begin{cases} \frac{1}{2}, & \text{if } 0 < x \le 1, \\ \frac{1}{2}e^{1-x}, & \text{if } x > 1. \end{cases} \tag{4.2}$$

Conditions (A1), (A2) and (A3) are satisfied. Also,

$$K(x) = \begin{cases} \frac{c(e^{-x}-e^{-cx})}{2(c-1)}, & \text{if } 0 < x \le 1, \\ \frac{e^{-x}}{2}(\frac{c}{c-1} - \frac{e^{1-c}}{c-1} - e^{1-cx}), & \text{if } x > 1, \end{cases} \tag{4.3}$$

and

$$\tau_G(x) = \begin{cases} \frac{c-1}{ce^{-x}-e^{-cx}}, & \text{if } 0 < x \le 1, \\ \frac{c-1}{ce^{-1}-e^{-c}}, & \text{if } x > 1. \end{cases} \tag{4.4}$$

(When $c = 1$, the above expressions for $K(x)$ and $\tau_G(x)$ are still valid by taking $\lim_{c\to1}$.) For the prior distribution G given in (4.1), we claim the following result.

Theorem 4.1. *Let* $\{\varphi_n\}_{n=1}^{\infty}$ *be the sequence of empirical Bayes estimators constructed in section 2. Assume that* $\limsup_{n\to\infty} nb_n^3 = \beta$ *with* $0 \le \beta < \infty$. *If the prior distribution* G *is the one given in* (4.1), *then*

$$R(G,\varphi_n) - R(G,\varphi_G) \ge \gamma_1 (nb_n)^{-1} + \gamma_2 b_n^2$$

for some constants $\gamma_1 > 0$ *and* $\gamma_2 > 0$.

The lower bound rate in Theorem 4.1 is established below. Note that

$$\begin{aligned}
&R(G,\varphi_n) - R(G,\varphi_G)\\
&= \int_{x=0}^{1} E_n\{\exp[c(\varphi_n(x)-\varphi_G(x))] - c(\varphi_n(x)-\varphi_G(x)) - 1\} f(x)dx\\
&\quad + E_n\{\exp[c(\varphi_n(1)-\varphi_G(1))] - c(\varphi_n(1)-\varphi_G(1)) - 1\}[1-F(1)]\\
&= \int_{x=0}^{1} E_n\left(\frac{\tau_n(x)}{\tau_G(x)} - \ln\frac{\tau_n(x)}{\tau_G(x)} - 1\right) f(x)dx\\
&\quad + E_n\left(\frac{\tau_n(1)}{\tau_G(1)} - \ln\frac{\tau_n(1)}{\tau_G(1)} - 1\right)[1-F(1)]. \qquad (4.5)
\end{aligned}$$

The following two inequalities are needed. (They can be obtained by elementary calculus.)

Lemma 4.1.

$$\begin{aligned}
\frac{(t-1)^2}{2} &\le t - \ln t - 1, \quad \textit{for}\, 0 < t \le 1,\\
\frac{(1-t^{-1})^2}{2} &\le t - \ln t - 1, \quad \textit{for}\, t > 1.
\end{aligned}$$

For every $x \in (0,1]$,

$$\begin{aligned}
&E_n\left(\frac{\tau_n(x)}{\tau_G(x)} - \ln\frac{\tau_n(x)}{\tau_G(x)} - 1\right)\\
&\ge \frac{1}{2}E_n\left((\frac{\tau_n(x)}{\tau_G(x)}-1)^2 I(\frac{\tau_n(x)}{\tau_G(x)} \le 1)\right) + \frac{1}{2}E_n\left((\frac{\tau_G(x)}{\tau_n(x)}-1)^2 I(\frac{\tau_n(x)}{\tau_G(x)} > 1)\right)\\
&\ge \frac{e^{-2cx}}{2}E_n[(\tau_n(x)-\tau_G(x))^2 I(\tau_n(x) \le \tau_G(x))]\\
&\quad + \frac{e^{-2cx}}{2}E_n[(\tau_n(x)-\tau_G(x))^2 I(\tau_n(x) > \tau_G(x))], \text{ as } \tau_n(x), \tau_G(x) \le e^{cx}\\
&= \frac{e^{-2cx}}{2}E_n[\tau_n(x)-\tau_G(x)]^2. \qquad (4.6)
\end{aligned}$$

From (4.5) and (4.6),

$$\begin{aligned}
R(G,\varphi_n) - R(G,\varphi_G) \ge \frac{e^{-2c}}{2}\int_{x=0}^{1} E_n[\tau_n(x)-\tau_G(x)]^2 f(x)dx&\\
+\frac{e^{-2c}}{2}E_n[\tau_n(1)-\tau_G(1)]^2[1-F(1)].& \qquad (4.7)
\end{aligned}$$

Let $0 < \delta < 1/2$ be an arbitrary positive number. For $x \in [\delta, 1-\delta]$, let

$$B_n(x) = I\Big(\frac{f_n(x)}{e^{-cx} f_n(x) + K_n(x)} \geq 1\Big) \text{ and } B_n^c(x) = 1 - B_n(x)$$

Lemma 4.2. *For $\delta \leq x \leq 1-\delta$, we have*

$$\lim_{n\to\infty} E_n[B_n^c(x)] = 0 \text{ and } \lim_{n\to\infty} E_n[B_n(x)] = 1.$$

Proof. Assume n is sufficiently large such that $b_n < \delta$ to ensure $x + b_n$ inside (0,1). For $x \in [\delta, 1-\delta]$,

$$\begin{aligned} E_n[B_n^c(x)] &= P\left\{\frac{f_n(x)}{e^{-cx} f_n(x) + K_n(x)} < 1\right\} \\ = P\Big\{[1-e^{-cx}][f_n(x) - f(x)] &- [K_n(x) - K(x)] < -(1-e^{-cx})f(x) + K(x)\Big\}. \end{aligned}$$

Since $f(x)$ is constant for $x \in (0,1]$ as appeared in (4.2), $Ef_n(x) = f(x)$. Also $E_n K_n(x) = K(x)$. Therefore,

$$E_n\left\{(1-e^{-cx})[f_n(x) - f(x)] - [K_n(x) - K(x)]\right\} = 0.$$

Also note that, from (4.2) and (4.3),

$$(1-e^{-cx})f(x) - K(x) = \frac{1}{2}\Big(1 - \frac{ce^{-x}}{c-1} + \frac{e^{-cx}}{c-1}\Big) > 0$$

for $0 < x \leq 1$. By Chebychev's inequality, Lemma 3.2 and equation (3.6),

$$\begin{aligned} &E_n[B_n^c(x)] \\ &\leq 4\mathrm{Var}\,[(1-e^{-cx})f_n(x) - K_n(x)]\Big(1 - \frac{ce^{-x}}{c-1} + \frac{e^{-cx}}{c-1}\Big)^{-2} \\ &\leq 8\Big\{\mathrm{Var}\,[(1-e^{-cx})f_n(x)] + \mathrm{Var}\,[K_n(x)]\Big\}\Big(1 - \frac{ce^{-x}}{c-1} + \frac{e^{-cx}}{c-1}\Big)^{-2} \\ &\leq 8\Big(1 - \frac{ce^{-x}}{c-1} + \frac{e^{-cx}}{c-1}\Big)^{-2}\Big(\frac{h(x)\psi(x+b_n)}{nb_n} + \frac{c^2 h(x)\psi(x)}{2n}\Big), \end{aligned}$$

which tends to 0 as $n \to \infty$. Therefore, we have $\lim_{n\to\infty} E_n[B_n^c(x)] = 0$ and $\lim_{n\to\infty} E_n[B_n(x)] = 1$. End of proof.

Lemma 4.3. *We have $\int_0^1 E_n[\tau_n(x) - \tau_G(x)]^2 f(x)dx \geq \gamma_1 (nb_n)^{-1}$ for some constant $\gamma_1 > 0$.*

Proof. Start with

$$\int_0^1 E_n[\tau_n(x) - \tau_G(x)]^2 f(x)dx \geq \int_\delta^{1-\delta} E_n[(\tau_n(x) - \tau_G(x))^2 B_n(x)]f(x)dx.$$

For $x \in [\delta, 1-\delta]$, we have

$$\begin{aligned} &E_n[(\tau_n(x) - \tau_G(x))^2 B_n(x)] \\ &= E_n\left((\frac{f_n(x)}{e^{-cx} f_n(x) + K_n(x)} - \frac{c-1}{ce^{-x} - e^{-cx}})^2 B_n(x) \right), \end{aligned} \tag{4.8}$$

$E_n f_n(x) = f(x)$, $E_n K_n(x) = K(x)$, $\mathrm{Var}\,(f_n(x)) = f(x)(1 - b_n f(x))/(nb_n)$, and $\mathrm{Var}\,(K_n(x)) = O(1/n)$. For $x > 0$ and as $n \to \infty$,

$$\begin{aligned} \sqrt{nb_n}\,[f_n(x) - f(x)] &\xrightarrow{d} N(0, f(x)), \\ \sqrt{nb_n}\,(K_n(x) - K(x)) &\xrightarrow{p} 0, \quad \text{and} \\ [e^{-cx} f_n(x) + K_n(x)] \xrightarrow{p} [e^{-cx} f(x) + K(x)] &= \frac{ce^{-x} - e^{-cx}}{2(c-1)}. \end{aligned}$$

Hence, by Slutsky's theorem,

$$\begin{aligned} &\sqrt{nb_n}\left(\frac{f_n(x)}{e^{-cx} f_n(x) + K_n(x)} - \frac{c-1}{ce^{-x} - e^{-cx}} \right) B_n(x) \\ &\xrightarrow{d} N\left(0, \frac{f(x)}{[e^{-cx} f(x) + K(x)]^2}\right). \end{aligned} \tag{4.9}$$

Denote the above variance by $\sigma^2(x)$. The following convergence theorem is needed: If $U_n \to U$ in distribution, then $EU^2 \leq \liminf EU_n^2$. The above theorem can be easily obtained from Skorohod's theorem and Fatou's lemma. (For reference, see Theorem 25.6 (and possibly also Theorem 25.11 for similar proof) in Billingsley (1986).) From (4.8) and (4.9), we have

$$\liminf E_n[nb_n(\tau_n(x) - \tau_G(x))^2 B_n(x)] \geq \sigma^2(x) \tag{4.10}$$

for $x \in [\delta, 1-\delta]$. By Fatou's lemma again,

$$\begin{aligned} &\liminf \int_\delta^{1-\delta} E_n[nb_n(\tau_n(x) - \tau_G(x))^2 B_n(x)] f(x) dx \\ &\geq \int_\delta^{1-\delta} \liminf E_n[nb_n(\tau_n(x) - \tau_G(x))^2 B_n(x)] f(x) dx \\ &\geq \int_\delta^{1-\delta} \frac{\sigma^2(x)}{2} dx > 0. \end{aligned}$$

Hence there exists a constant $\gamma_1 > 0$ such that $\int_0^1 E_n[\tau_n(x) - \tau_G(x)]^2 f(x) dx \geq \gamma_1 (nb_n)^{-1}$. End of proof.

Lemma 4.4. *Suppose that* $\limsup_{n\to\infty} nb_n^3 = \beta$ *with* $0 \leq \beta < \infty$. *Then there exists a constant* $\gamma_2 > 0$ *such that*

$$E_n[\tau_n(1) - \tau_G(1)]^2 \geq \gamma_2 b_n^2.$$

Proof. Let $A_n = I(|\tau_n(1) - \tau_G(1)| \geq b_n)$. Then

$$\begin{aligned}
&E_n[\tau_n(1) - \tau_G(1)]^2 \geq b_n^2 E_n[A_n B_n(1)] \\
&\geq b_n^2 E_n I(\tau_n(1) - \tau_G(1) \geq b_n) B_n(1) \\
&= b_n^2 P\left\{ \frac{f_n(1)}{e^{-c} f_n(1) + K_n(1)} - \frac{c-1}{ce^{-1} - e^{-c}} \geq b_n \right\} \\
&= b_n^2 P\left\{ \sqrt{nb_n} \left(\frac{f_n(1)}{e^{-c} f_n(1) + K_n(1)} - \frac{c-1}{ce^{-1} - e^{-c}} \right) \geq \sqrt{nb_n^3} \right\} \\
&\geq \gamma_2 b_n^2, \text{ (for } n \text{ large enough)}
\end{aligned}$$

where γ_2 is any number satisfying $0 < \gamma_2 < 1 - \Phi\left(\sqrt{\beta}\,/\sigma(1)\right)$ with Φ the cdf of the standard normal. End of proof.

Proof of Theorem 4.1. Theorem 4.1 is a direct result of (4.7), Lemmas 4.3 and 4.4. End of proof.

From Theorems 3.1 and 4.1, there is a prior distribution G satisfying conditions (A1), (A2) and (A3) such that

$$\gamma_1 (nb_n)^{-1} + \gamma_2 b_n^2 \leq R(G, \varphi_n) - R(G, \varphi_G) = O(\frac{1}{nb_n}) + O(b_n^2).$$

With $b_n = O(n^{-1/3})$ the optimal convergence rate of the proposed estimator is of order $O(n^{-2/3})$.

5. Simulation Studies

Monte Carlo studies have been carried out to investigate the performance of the proposed empirical Bayes estimator φ_n. In this simulation study, we consider the translated exponential distribution with $h(x) = e^{-x}$. Two prior distributions are studied. They are:

$$G_1(\theta) = \begin{cases} 0, & \text{if } \theta < 0, \\ \frac{1}{2} + \frac{1}{2}\theta, & \text{if } 0 \leq \theta \leq 1, \\ 1, & \text{if } \theta > 1. \end{cases} \qquad G_2(\theta) = \begin{cases} 0, & \text{if } \theta < 0, \\ \frac{e}{2e-1} + \frac{e}{2e-1}[1 - e^{-\theta}], & \text{if } 0 \leq \theta \leq 1, \\ 1, & \text{if } \theta > 1. \end{cases}$$

For each prior distribution G_i, let $f_i(x)$ denote the corresponding marginal *pdf* of the random variable X. Then,

$$f_1(x) = \begin{cases} 0, & \text{if } x < 0, \\ \frac{1}{2}, & \text{if } 0 \leq x \leq 1, \\ \frac{1}{2} e^{1-x}, & \text{if } x > 1. \end{cases} \qquad f_2(x) = \begin{cases} 0, & \text{if } x < 0, \\ \frac{e}{2e-1}[e^{-x} + xe^{-x}], & \text{if } 0 \leq x \leq 1, \\ \frac{2e}{2e-1} e^{-x}, & \text{if } x > 1. \end{cases}$$

We take $c = 1$. At state $n+1$, let $\underline{X}_n = (X_1, \ldots, X_n)$ denote the past n random observations and X_{n+1} denote the present random observation. Define

$$D_n(X_{n+1}) = \exp(\varphi_n(X_{n+1}) - \varphi_G(X_{n+1})) - (\varphi_n(X_{n+1}) - \varphi_G(X_{n+1})) - 1.$$

Consider the expectation of $D_n(X_{n+1})$ taken with respect to $(\underline{X}_n, X_{n+1})$. Then,

$$\begin{aligned} ED_n(X_{n+1}) &= E_n\{E[D_n(X_{n+1})|\underline{X}_n]\} \\ &= E_n \int_{x=0}^{\infty} [\exp(\varphi_n(x) - \varphi_G(x)) - (\varphi_n(x) - \varphi_G(x)) - 1] f(x)dx \\ &= R(G, \varphi_n) - R(G, \varphi_G). \end{aligned} \tag{5.1}$$

That is, $D_n(X_{n+1})$ is an unbiased estimator of the regret risk of the empirical Bayes estimator φ_n. Our simulation study is based on (5.1). The simulation schemes used in this paper are described below.

(a) Generate random variables $X_1, \ldots, X_{n+1}$ according to a distribution function $F_i(x)$ having $f_i(x)$ as its *pdf*.
(b) Compute $\varphi_n(X_{n+1}), \varphi_G(X_{n+1})$ and $D_n(X_{n+1})$ using $b_n = n^{-1/3}$.
(c) The above process was repeated 500 times. Denote the jth run sampled $D_n(X_{n+1})$ value by $D_{nj}, j = 1, \ldots, 500$. Let $\bar{D}_n = \frac{1}{500}\sum_{j=1}^{500} D_{nj}$. We use $\bar{D}_n$ as an estimator of the regret risk $R(G, \varphi_n) - R(G, \varphi_G)$. The associated standard deviation of the sample mean $\bar{D}_n$ is estimated by $SE(\bar{D}_n)$, where $SE^2(\bar{D}_n) = \frac{1}{499\times 500}\sum_{j=1}^{500}(D_{nj} - \bar{D}_n)^2$.

Tables 1 and 2 are simulation results with prior distributions G_1 and G_2, respectively. Values of $n^{2/3}\bar{D}_n$ are provided. They are becoming stable when n is getting large. This phenomenon is consistent with the optimal rate found in Section 4.

Table 1. Small sample performance of φ_n using prior G_1.

n	40	80	120	160	200
$\bar{D}_n$	12.892 (-3)	10.723 (-3)	8.599 (-3)	7.918 (-3)	6.919 (-3)
$SE(\bar{D}_n)$	6.831 (-4)	6.170 (-4)	5.538 (-4)	5.310 (-4)	4.825 (-4)
$n^{2/3}\bar{D}_n$	0.1509	0.1992	0.2092	0.2334	0.2366
n	240	280	320	360	400
$\bar{D}_n$	7.145 (-3)	5.956 (-3)	5.432 (-3)	4.812 (-3)	4.690 (-3)
$SE(\bar{D}_n)$	4.838 (-4)	4.250 (-4)	3.924 (-4)	3.825 (-4)	3.444 (-4)
$n^{2/3}\bar{D}_n$	0.2759	0.2549	0.2541	0.2435	0.2550
n	560	720	880	1040	1200
$\bar{D}_n$	3.920 (-3)	3.387 (-3)	2.775 (-3)	2.767 (-3)	2.121 (-3)
$SE(\bar{D}_n)$	2.982 (-4)	2.783 (-4)	2.462 (-4)	2.105 (-4)	1.720 (-4)
$n^{2/3}\bar{D}_n$	0.2663	0.2721	0.2503	0.2840	0.2395
n	1360	1520	1680	1840	2000
$\bar{D}_n$	2.083 (-3)	1.921 (-3)	1.896 (-3)	1.763 (-3)	1.594 (-3)
$SE(\bar{D}_n)$	1.730 (-4)	1.730 (-4)	1.6844 (-4)	1.592 (-4)	1.377 (-4)
$n^{2/3}D_n$	0.2557	0.2540	0.2679	0.2647	0.2530

⋆ The entry 1.342 (-3) means 1.342×10^{-3}.

Table 2. Small sample performance of φ_n using prior G_2.

n	40	80	120	160	200
$\bar{D}_n$	10.797 (-3)	7.564 (-3)	6.728 (-3)	5.317 (-3)	5.192 (-3)
$SE(\bar{D}_n)$	5.926 (-4)	4.219 (-4)	3.709 (-4)	3.045 (-4)	2.983 (-4)
$n^{2/3}\bar{D}_n$	0.1263	0.1404	0.1637	0.1567	0.1776
n	240	280	320	360	400
$\bar{D}_n$	4.281 (-3)	4.976 (-3)	4.433 (-3)	4.171 (-3)	4.178 (-3)
$SE(\bar{D}_n)$	2.651 (-4)	2.761 (-4)	2.696 (-4)	2.642 (-4)	2.629 (-4)
$n^{2/3}\bar{D}_n$	0.1653	0.2130	0.2074	0.2111	0.2268
n	560	720	880	1040	1200
$\bar{D}_n$	3.006 (-3)	2.957 (-3)	2.302 (-3)	2.276 (-3)	2.498 (-3)
$SE(\bar{D}_n)$	2.129 (-4)	2.101 (-4)	1.777 (-4)	1.674 (-4)	1.824 (-4)
$n^{2/3}\bar{D}_n$	0.2043	0.2375	0.2114	0.2336	0.2820
n	1360	1520	1680	1840	2000
$\bar{D}_n$	2.073 (-3)	1.821 (-3)	1.752 (-3)	1.710 (-3)	1.544 (-3)
$SE(\bar{D}_n)$	1.600 (-4)	1.400 (-4)	1.418 (-4)	1.524 (-4)	1.251 (-4)
$n^{2/3}D_n$	0.2544	0.2408	0.2475	0.2924	0.2452

* The entry 1.342 (-3) means 1.342×10^{-3}.

References

Arnold, B. C. (1983). *Pareto Distributions.* International co-operative publishing house, Fairland, Maryland.

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

Billingsley, P. (1986). *Probability and Measure.* Wiley, New York.

Gasser, T. and Müller, H.-G. (1979). Kernel estimation of regression functions. Smoothing techniques for curve estimation. In *Lecture Notes in Mathematics* 757 (Edited by Th. Gasser and M. Rosenblatt), 23-68. Springer-Verlag.

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

Kuo, L. and Dey, D. K. (1990). On the admissibility of the linear estimators of the Poisson mean using Linex loss functions. *Statist. Decisions* **8**, 201-210.

Lawless, J. F. (1982). *Statistical Models and Methods for Lifetime Data.* Wiley, New York.

Mann, N. R., Schafer, R. E. and Singpurwalla, N. D. (1974). *Methods for Statistical Analysis of Reliability and Lifetime Data.* Wiley, New York.

Parsian, A. (1990). On the admissibility of an estimator of a normal mean vector under a Linex loss function. *Ann. Inst. Statist. Math.* **42**, 657-669.

Rice, J. (1984). Boundary modification for kernel regression. *Comm. Statist. Theory Methods* **13**, 893-900.

Robbins, H. (1956). An empirical Bayes approach to statistics. *Proc. Third Berkeley Symp. Math. Statist. Probab. I*, 157-163. University of California Press.

Robbins, H. (1964). The empirical Bayes approach to statistical decision problems. *Ann. Math. Statist.* **35**, 1-20.

Schuster, E. F. (1985). Incorporating support constraints into nonparametric estimators of densities. *Comm. Statist. Theory Methods* **14**, 1123-1136.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, London.

Singh, R. S. (1977). Applications of estimators of a density and its derivatives to certain statistical problems. *J. Roy. Statist. Soc. Ser. B* **39**, 357-363.

Varian, H. R. (1975). A Bayesian approach to real estate assessment. In *Studies in Bayesian Econometrics and Statistics in Honor of L. J. Savage* (Edited by S. E. Feinberg and A. Zellner), 195-208. North-Holland, Amsterdam.

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

Institute of Statistical Science, Academia Sinica, Taipei 115.
E-mail: syhuang@stat.sinica.edu.tw

Department of Mathematics, Wayne State University, Detroit, MI 48202, U.S.A.
E-mail: liang@math.wayne.edu

(Received December 1994; accepted August 1996)