# BAYESIAN AND EMPIRICAL BAYESIAN ANALYSIS FOR THE TRUNCATION PARAMETER DISTRIBUTION FAMILIES

By

YIMIN MA, B.Sc., M.Sc.

A Thesis
Submitted to the School of Graduate Studies
in Partial Fulfilment of the Requirements
for the Degree
Doctor of Philosophy

McMaster University

# BAYESIAN AND EMPIRICAL BAYESIAN ANALYSIS FOR THE TRUNCATION PARAMETER DISTRIBUTION FAMILIES

DOCTOR OF PHILOSOPHY (1998) McMaster University

(Mathematics) Hamilton, Ontario

TITLE: Bayesian and Empirical Bayesian Analysis for the Truncation Parameter Distribution Families

AUTHOR: Yimin Ma
B.Sc. (Fudan)
M.Sc. (Huazhong)
M.Sc. (McMaster)

SUPERVISOR: Professor N. Balakrishnan

NUMBER OF PAGES: viii, 79

# Abstract

Modern Bayesian analysis and empirical Bayesian analysis are dominated by the exponential distribution family; lots of research works have been done in the literature. However, for nonexponential distribution families, there are still no systematic results from Bayesian and empirical Bayesian analysis. In this Ph.D. thesis, some systematic Bayesian and empirical Bayesian analysis results are obtained for the one-parameter truncation distribution families which are nonexponential distribution families. It consists of six main chapters. In Chapter 2, the general forms of conjugate prior distributions are obtained for the two different types of truncation parameter distributions and the particular conjugate priors for the truncated exponential, Pareto and power function distributions are presented. In Chapter 3, the explicit relations between the mixing distributions and the mixture distributions are obtained and the identifiability for the mixture of these trunction parameter distributions is established. Based on these obtained relations, some procedures for estimating the mixing distributions are proposed and studied. In Chapter 4, the explicit analytical expressions of posterior moments for the two general truncation parameter likelihood functions with arbitrary priors are given by using the sufficient statistics for these truncation parameters. In particular, the explicit forms for the posterior mean and variance are presented. In Chapter 5, based on the relations between the Bayes estimators under squared error loss and the marginal distributions, the empirical Bayes estimators are proposed and the asymptotic optimalities of the proposed empirical Bayes estimators are inves-

tigated. Finally, in Chapter 6, the problems of empirical Bayes estimation for the truncated exponential distributions and the empirical Bayes rule for selecting the best of exponential populations are discussed and the convergence rates of the proposed empirical Bayes estimators and the empirical Bayes selection rule are established. And in Chapter 7, a location parameter family of gamma distribution, which is not a typical truncation parameter distribution, is considered. The empirical Bayes estimator and the empirical Bayes testing rule for the two-action problem are studied and the convergence rates for the proposed empirical Bayes estimator and the empirical Bayes testing rule are established.

# Acknowledgements

I would like to express my sincere thanks to my supervisor, Professor N. Balakrishnan, for his guidance and help in the preparation of this thesis. My sincere thanks also go to Professor Mead and Professor Min-Oo for reading this thesis and suggesting valuable options. I am especially grateful to Mrs. Debbie Iscoe for the excellent typesetting of the manuscript. My final acknowledgement is to CIDA (Canadian International Development Agency) and McMaster University for providing the financial support for my graduate studies at McMaster University.

# Contents

# Chapter 1

# Introduction

## 1.1 Truncation Parameter Distributions

In this Thesis, we consider two general types of the one-parameter truncation distribution families with density functions over any interval $(a, b)$, finite or infinite, as follows:

Type I truncation parameter density

$$f_1(x_1|\theta_1) = h_1(x_1)/k_1(\theta_1), \qquad a < \theta_1 \leq x_1 < b. \tag{1.1.1}$$

Type II truncation parameter density

$$f_2(x_2|\theta_2) = h_2(x_2)/k_2(\theta_2), \qquad a < x_2 \leq \theta_2 < b. \tag{1.1.2}$$

From the definitions of (1.1.1) and (1.1.2) as density functions, we assume the following conditions: (1) Both $h_1(x_1)$ and $h_2(x_2)$ are positive, continuous, and integrable over $(\theta_1, b)$ and $(a, \theta_2)$, respectively, for $\theta_i$ in $(a, b)$, and (2) both $k_1(\theta_1)$ and $k_2(\theta_2)$ are differentiable, and satisfying

$$k_1(\theta_1) = \int_{\theta_1}^{b} h_1(x_1)dx_1, \tag{1.1.3}$$

$$k_2(\theta_2) = \int_{a}^{\theta_2} h_2(x_2)dx_2. \tag{1.1.4}$$

for any $\theta_i$ in the interval $(a, b)$.

**Example 1.1.1** The left-truncated exponential distribution

$$f_1(x_1|\theta_1) = \lambda \exp\{-\lambda(x_1 - \theta_1)\}, \qquad 0 < \theta_1 \leq x_1 < \infty, \ \lambda > 0,$$

is a type I truncation parameter distribution with $h_1(x_1) = \lambda e^{-\lambda x_1}$ and $k_1(\theta_1) = e^{-\lambda\theta_1}$.

**Example 1.1.2** The Pareto distribution

$$f_1(x_1|\theta_1) = \frac{a\theta_1^a}{x_1^{a+1}}, \qquad 0 < \theta_1 \leq x_1 < \infty, \ a > 0,$$

is also a type I truncation parameter distribution with $h_1(x_1) = a/x_1^{a+1}$ and $k_1(\theta_1) = 1/\theta_1^a$.

**Example 1.1.3** The right-truncated exponential distribution

$$f_2(x_2|\theta_2) = \lambda \exp\{-\lambda(\theta_2 - x_2)\}, \qquad -\infty < x_2 \leq \theta_2 < \infty, \ \lambda > 0,$$

is a type II truncation parameter distribution with $h_2(x_2) = \lambda e^{\lambda x_2}$ and $k_2(\theta_2) = e^{\lambda\theta_2}$.

**Example 1.1.4** The power function distribution

$$f_2(x_2|\theta_2) = \frac{a\, x_2^{a-1}}{\theta_2^a}, \qquad 0 < x_2 \leq \theta_2 < \infty, \ a > 0,$$

is also a type II truncation parameter distribution with $h_2(x_2) = a\, x_2^{a-1}$ and $k_2(\theta_2) = \theta_2^a$.

Note that in the definitions of the truncation parameter distributions (1.1.1) and (1.1.2), the ranges of the distributions depend on the parameters $\theta_i$, $i = 1, 2$, respectively. Clearly these distributions are not exponential family distributions.

A family of distributions is said to belong to the one-parameter exponential family if it is defined by density functions, with respect to some fixed measure $\mu$, of the following form

$$f(x|\theta) = c(\theta)h(x)\exp\{Q(\theta)\cdot T(x)\}$$

where $\theta$ and $x$ are in $\mathbf{R}$ and $c$, $h$, $Q$ and $T$ are real valued functions; in the particular case when

$$f(x|\theta) = c(\theta)h(x)\exp\{\theta\cdot x\}$$

the family is said to be the natural exponential family. The one-parameter exponential family includes many standard classes of distributions which arise in practice, for example, binomial, Poisson, geometric, normal, exponential and gamma distributions.

Modern Bayesian and empirical Bayesian analysis are dominated by the exponential distribution family. Lots of research works have been done in the literature; see, for example, Berger (1985), Maritz and Lwin (1989) and Robert (1994) for the general introduction. However, for non-exponential distribution families, there are still no systematic results from Bayesian and empirical Bayesian analysis. In this Ph.D. thesis, some original Bayesian and empirical Bayesian analysis results are obtained for the one-parameter truncation distribution families which are non-exponential families. All these results will be presented in the next six main chapters.

## 1.2 Prior and Posterior Distributions

The essential difference between the classical frequentist approach and the Bayesian approach is that in the latter we assume the existence of a probability distribution $G(\cdot)$ on the parameter space $\Theta$. This probability distribution describes our degrees of belief in possible parameter values prior to an observation being made.

and consequently it is called a prior distribution. That means a Bayesian statistical model is made up of a parametric statistical model, $f(x|\theta)$, and a prior distribution $G(\theta)$ on the parameter (or a prior density function $g(\theta)$). Given these two distributions, we can construct the following distributions:

(a) the joint distribution of $(\theta, x)$,

$$p(\theta, x) = f(x|\theta)g(\theta); \tag{1.2.1}$$

(b) the marginal distribution of $x$,

$$f(x) = \int f(x|\theta)g(\theta)\, d\theta; \tag{1.2.2}$$

(c) the posterior distribution of $\theta$, given $x$,

$$p(\theta|x) = \frac{f(x|\theta)g(\theta)}{f(x)}. \tag{1.2.3}$$

This conditional distribution (1.2.3) on $\Theta$ may be interpreted as describing our degrees of belief in different possible values of $\Theta$ after the observation $x$ has been made, and consequently it is called the posterior distribution of $\theta$.

In this thesis, we always assume that $G_i(\theta_i)$ are the prior (mixing) distributions of truncation parameters $\theta_i$, $i = 1, 2$, respectively, defined in the interval $(a, b)$ with $G_i(a) = 0$ and $G_i(b) = 1$, where $G_i(\theta_i)$ might not be continuous distributions; then the marginal (mixture) distributions of $x_i$, $i = 1, 2$, are given by

$$f_{G_1}(x_1) = f_1(x_1) = \int_a^{x_1} \frac{h_1(x_1)}{k_1(\theta_1)}\, dG_1(\theta_1), \tag{1.2.4}$$

$$f_{G_2}(x_2) = f_2(x_2) = \int_{x_2}^{b} \frac{h_2(x_2)}{k_2(\theta_2)}\, dG_2(\theta_2). \tag{1.2.5}$$

In general, $f(x)$ and $p(\theta|x)$ in (1.2.2) and (1.2.3) are not easily calculable. A large part of Bayesian literature is devoted to finding prior distributions for which the posterior $p(\theta|x)$ can be easily calculated. These are the so-called conjugate priors, and were developed extensively by Raiffa and Schlaifer (1961).

**Definition 1.2.1** Let $\mathcal{F}$ denote the class of density functions $f(x|\theta)$. A class $\mathcal{P}$ of prior distributions is said to be a conjugate family for $\mathcal{F}$ if the posterior $p(\theta|x)$ is still in the class $\mathcal{P}$ for all prior $g \in \mathcal{P}$.

Many examples of conjugate priors are given in Raiffa and Schlaifer (1961); further illustrations and additional examples can be found in DeGroot (1970), Berger (1985) and Robert (1994).

**Example 1.2.1** Suppose that $x$ is distributed with the normal distribution

$$f(x|\theta) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\theta)^2}$$

where $\theta$ is unknown but $\sigma^2$ is known. Let the prior distribution $g(\theta)$ be the normal distribution $N(\mu, \tau^2)$ given by

$$g(\theta) = \frac{1}{\sqrt{2\pi}\tau} e^{-\frac{1}{2\tau^2}(\theta-\mu)^2}$$

where both $\mu$ and $\tau^2$ are known. Then

$$\begin{aligned} p(\theta|x) \propto f(x|\theta)g(\theta) &= \frac{1}{2\pi\sigma\tau} \exp\left\{-\frac{1}{2}\left[\frac{(\theta-\mu)^2}{\tau^2} + \frac{(x-\theta)^2}{\sigma^2}\right]\right\} \\ &= \frac{1}{2\pi\sigma\tau} \exp\left\{-\frac{1}{2}\rho\left[\theta - \frac{1}{\rho}\left(\frac{\mu}{\tau^2} + \frac{x}{\sigma^2}\right)\right]^2\right\} \\ &\quad \cdot \exp\left\{-\frac{(\mu-x)^2}{2(\sigma^2+\tau^2)}\right\}. \end{aligned}$$

It follows that the posterior distribution of $\theta$ given $x$ is normal $N(\mu(x), \rho^{-1})$, where

$$\begin{aligned} \rho &= \tau^{-2} + \sigma^{-2}, \\ \mu(x) &= \rho^{-1}\left(\frac{\mu}{\tau^2} + \frac{x}{\sigma^2}\right) = \frac{\sigma^2}{\sigma^2+\tau^2}\mu + \frac{\tau^2}{\sigma^2+\tau^2}x. \end{aligned}$$

So the prior distribution $g(\theta)$ is conjugate.

**Example 1.2.2** Suppose that $x$ is distributed with the Poisson distribution

$$f(x|\theta) = \frac{\theta^x e^{-\theta}}{x!}, \qquad x = 0, 1, \ldots .$$

Let the prior distribution be gamma $\mathcal{G}(\alpha, \beta)$ with

$$g(\theta) = \frac{1}{\Gamma(\alpha)\beta^\alpha}\, \theta^{\alpha-1} e^{-\theta/\beta}, \qquad \theta > 0.$$

Then

$$\begin{aligned} p(\theta|x) \propto f(x|\theta)g(\theta) &= \frac{\theta^x e^{-\theta}}{x!} \frac{\theta^{\alpha-1} e^{-\theta/\beta}}{\Gamma(\alpha)\beta^\alpha} \\ &= \frac{\theta^{(\alpha+x-1} e^{-(1+1/\beta)\theta}}{x!\Gamma(\alpha)\beta^\alpha}, \quad \theta > 0. \end{aligned}$$

So the posterior distribution is $\mathcal{G}(\alpha + x, (1 + 1/\beta)^{-1})$ and the prior distribution $g(\theta)$ is conjugate.

We now present a general result about conjugate prior families for the exponential distribution family in the following proposition.

**Proposition 1.2.1** *Consider natural exponential families distributed as*

$$f(x|\theta) = h(x)\; e^{\theta \cdot x - \psi(\theta)};$$

*then the conjugate prior family for* $f(x|\theta)$ *is given by*

$$g(\theta|\mu, \lambda) = k(\mu, \lambda)\; e^{\theta\cdot\mu - \lambda\psi(\theta)},$$

*and the posterior distribution is* $g(\theta|\mu + x, \lambda + 1)$.

**Proof.** See Proposition 3.3 in Robert (1994).

We will present the general conjugate prior forms for the one-parameter truncation distribution families in Chapter 2.

In order to calculate the posterior distribution, it is often helpful to use the sufficient statistics.

**Definition 1.2.2** Let $x$ be a random variable with distribution $f(x|\theta)$. A statistic $t = t(x)$ is said to be sufficient for the parameter $\theta$ if the conditional distribution of $x$ given $t$ does not depend on $\theta$.

**Proposition 1.2.2** *A necessary and sufficient condition for a statistic $t(x)$ to be sufficient is that $f(x|\theta)$ can be expressed in the form*

$$f(x|\theta) = u(t(x), \theta)v(x)$$

*where $u$ and $v$ are measurable functions and $v$ does not depend on $\theta$.*

**Proof.** See Theorem 2.8 and Corollary 2.1 in Lehmann (1986).

This is the so-called factorization theorem: we can use it to derive the following useful result for us to determine the posterior distribution.

**Proposition 1.2.3** *Suppose that $t(x)$ is the sufficient statistic for parameter $\theta$ and the factorization theorem holds: then for $t(x) = t$,*

$$p(\theta|x) = p(\theta|t) = \frac{f(t|\theta)g(\theta)}{f(t)} .$$

**Proof.** It is straightforward by using the factorization theorem.

The reason to determine $p(\theta|t)$ rather than $p(\theta|x)$ is that $f(t|\theta)$ and $f(t)$ are usually easier to handle. On the other hand, given a sample of $n$ independent observations $y = (x_1, \ldots, x_n)$, we know that $f(y|\theta) = \prod_{i=1}^{n} f(x_i|\theta)$. If we can find a statistic $t$ which is sufficient for $\theta$, then we may work with $p(\theta|t)$, instead of $p(\theta|y)$ which depends on the entire data set $y$, thus reducing the dimensionality of the problem.

## 1.3 Bayes and Empirical Bayes Methods

After specifying the prior distribution for the unknown random parameter $\theta$ and the form of the likelihood function for the observations, to set up the general Bayesian decision framework we need the following further components: A loss function, a class of allowable actions and decision rules.

The loss function $L(\theta, a)$ gives the loss incurred when $\theta$ is the true state of nature and we take action $a$. The decision rule $d$ maps the observed data $x$ into an action $a$.

The Bayesian outlook on the problem of selecting a decision rule is as follows: if the posterior distribution is given, we should choose the action that minimizes the posterior risk

$$\rho(g, d) = E_{\theta|x}[L(\theta, d)] = \int L(\theta, d(x))p(\theta|x)\, d\theta. \qquad (1.3.1)$$

To find the Bayesian solution to a decision problem in another way, we define the Bayes risk as

$$r(g, d) = E_\theta E_{x|\theta}[L(\theta, d)] = \int g(\theta)d\theta \int L[\theta, d(x)]f(x|\theta)\, dx. \qquad (1.3.2)$$

Since $g(\theta)f(x|\theta) = f(x)p(\theta|x)$, we can obtain the alternate computational form

$$r(g, d) = E_x E_{\theta|x}[L(\theta, d)] = \int f(x)dx \int L[\theta, d(x)]p(\theta|x)\, d\theta; \qquad (1.3.3)$$

then we could choose the action that minimizes the Bayes risk. Under very broad conditions, these two operations are virtually equivalent. Because $f(x)$ is nonnegative we minimize the double integral by minimizing the inner integral for each fixed $x$. In other words, we choose $d(x)$ for fixed $x$ to be that decision which minimizes the expected posterior loss.

In the estimation problem, the most common loss functions are the squared error loss and the absolute error loss. Here, we consider more general losses as follows:

$$L_2(\theta, d) = w(\theta)(\theta - d)^2 \qquad (1.3.4)$$

where $w(\theta)$ is a nonnegative function, and

$$L_1(\theta, d) = \begin{cases} k_1(\theta - d), & \text{if } d < \theta, \\ k_2(d - \theta), & \text{if } d \geq \theta. \end{cases} \qquad (1.3.5)$$

Then the Bayes estimators under $L_i$, $i = 1, 2$, are given, respectively, by Propositions 1.3.1 and 1.3.2.

**Proposition 1.3.1** *The Bayes estimator $d_2$ associated with prior $g$ and with the weighted quadratic loss $L_2$ is*

$$d_2(x) = \frac{E(w(\theta)\theta|x)}{E(w(\theta)|x)} .$$

*In particular, when $w(\theta) = 1$, $L_2(\theta, d) = (\theta - d)^2$, then $d_2(x) = E(\theta|x)$.*

**Proof.** See Corollary 2.19 in Robert (1994).

**Proposition 1.3.2** *The Bayes estimator $d_1$ associated with prior $g$ and with the linear loss $L_1$ is the $(k_1/(k_1 + k_2))$ fractile of the posterior $p(\theta|x)$. In particular, if $k_1 = k_2$, in the case of the absolute error loss, $L_1(\theta, d) = |\theta - d|$, $d_1(x)$ is the posterior median.*

**Proof.** See Corollary 2.18 in Robert (1994).

For the one-parameter exponential distribution family, the Bayes estimator under squared error loss can be expressed in terms of the marginal distribution of $x$ as follows.

**Proposition 1.3.3** *Consider the natural exponential family given by*

$$f(x|\theta) = h(x)e^{\theta \cdot x - \psi(\theta)}.$$

*Then for every prior distribution $g$, the posterior mean of $\theta$ is given by*

$$d_g(x) = \frac{d}{dx} \log f(x) - \frac{d}{dx} \log h(x)$$

*where $f(x)$ is the marginal distribution of $x$ associated with $g$,*

$$f(x) = \int f(x|\theta)g(\theta)d\theta.$$

**Proof.** See Lemma 4.1 in Robert (1994).

Note that this result is satisfied for every prior $g$. We will present similar results about the posterior mean for the truncation parameter distributions in Chapter 4.

In practical situations, however, the prior distribution $G$ is rarely known even if it is believed to exist. In the empirical Bayes framework, we assume that $(x_i, \theta_i)$, $i = 1, 2, \ldots$, is a sequence of pairs of random variables, where the $x_i$'s are observable but the $\theta_i$'s are not and conditional on $\theta_i = \theta$, $x_i$ has probability distribution $f(x|\theta)$. It is also assumed that the $\theta_i$'s are i.i.d. having unknown distribution $G$. Therefore, the pairs $(x_i, \theta_i)$ are i.i.d. Let $(x_1, \ldots, x_n)$ denote the $n$ past observations and let $x_{n+1} = x$ denote the current observation whose observed value is $x$. We want to make a decision about $\theta_{n+1}$ with loss $L$, where $G$ is assumed unknown and $x_1, \ldots, x_n$ is a random sample from

$$f_G(x) = f(x) = \int f(x|\theta)\, dG(\theta).$$

We expect that $x_1, \ldots, x_n$ do contain some information about the prior $G$ and we want to use such information to define $t_n(\cdot) = t_n(x_1, \ldots, x_n; \cdot)$, a decision rule for use in the $(n+1)$th decision problem to decide about $\theta_{n+1}$. We then incur an expected loss at $(n+1)$th stage given by

$$E(r(t_n, G)) = \int\int E[L(t_n(x), \theta)] f(x|\theta)\, dx dG(\theta).$$

Since the Bayes rule $d_G$ achieves the minimum Bayes risk $r(G)$, then $E\{r(t_n, G)\} - r(G) \geq 0$. This nonnegative difference is often used as a measure of the optimality of the emprical Bayes rule $\{t_n\}$.

**Definition 1.3.1** $T = \{t_n\}$ is said to be asymptotically optimal relative to the prior $G$ if

$$\lim_{n\to\infty} E\{r(t_n, G)\} - r(G) = 0.$$

## 1.4 Organization of This Thesis

Modern Bayesian analysis and empirical Bayesian analysis are dominated by the exponential distribution family. Lots of results are available in the literature for this family; see, for example, Berger (1985) and Maritz and Lwin (1989) for the general introduction. However, for non-exponential distribution families, there are still no systematic results from Bayesian and empirical Bayesian analysis. This thesis considers truncation parameter distribution families (1.1.1) and (1.1.2), which are non-exponential distribution models, by Bayesian and empirical Bayesian methods. It consists of six main chapters. In Chapter 2, the general forms of conjugate prior distributions are obtained for the two truncation parameter distributions and the particular conjugate prior distributions for truncated exponential, Pareto and power function distributions are presented. In Chapter 3, the explicit relations between the mixing distributions and the mixture distributions are obtained and the identifiability for the mixtures of these truncation parameter distributions is established. Based on the obtained relations, some procedures for estimating the mixing distribution are presented and studied. In Chapter 4, the explicit analytical expressions of posterior moments for the two general truncation parameter likelihood functions with arbitrary priors are obtained by using the sufficient statistics for these truncation parameters. In particular, the explicit forms for the posterior mean and variance are given. In Chapter 5, based on the relations between the Bayes estimators under squared error loss and the marginal distributions of $x_i$, $i = 1, 2$, the empirical Bayes estimators are proposed and the asymptotic optimality of the proposed empirical Bayes estimators is investigated. Finally, in Chapter 6, the problems of empirical Bayes estimation for the truncated exponential distributions and the empirical Bayes rule for selecting the best of exponential populations are discussed and the convergence rates of the proposed empirical Bayes estimators and the empirical Bayes selection rule are obtained. And in Chapter 7, the location parameter family of the gamma distribution, which is not the typical truncation parameter distribution, is considered. The empirical Bayes estimator and the empirical Bayes testing rule for the two-action problem are studied and the convergence rates for the proposed empirical Bayes estimator and the empirical Bayes testing rule are established.

# Chapter 2

# Conjugate Prior Distributions

## 2.1 Likelihood Functions

Let $x_{i1}, \ldots, x_{in}$ be independent and identically distributed samples of size $n$ from truncation parameter distributions $f_i(x_i|\theta_i)$, $i = 1, 2$, given by (1.1.1) and (1.1.2), respectively; then the likelihood functions of samples $y_i = (x_{i1}, \ldots, x_{in})$ are given by

$$\ell_1(y_1|\theta_1) = \frac{\prod_{j=1}^{n} h_1(x_{1j})}{\{k_1(\theta_1)\}^n}, \quad a < \theta_1 \le x_{1j} < b, \ j = 1, \ldots, n. \tag{2.1.1}$$

and

$$\ell_2(y_2|\theta_2) = \frac{\prod_{j=1}^{n} h_2(x_{2j})}{\{k_2(\theta_2)\}^n}, \quad a < x_{2j} \le \theta_2 < b, \ j = 1, \ldots, n. \tag{2.1.2}$$

These likelihood functions can be rewritten in the following forms

$$\ell_1(y_1|\theta_1) = \frac{\prod_{j=1}^{n} h_1(x_{1j})}{\{k_1(\theta_1)\}^n}, \quad a < \theta_1 \le t_1 < b, \tag{2.1.3}$$

and

$$\ell_2(y_2|\theta_2) = \frac{\prod_{j=1}^{n} h_2(x_{2j})}{\{k_2(\theta_2)\}^n}, \quad a < t_2 \le \theta_2 < b, \tag{2.1.4}$$

where $t_1 = \min\{x_{11}, \ldots, x_{1n}\}$ and $t_2 = \max\{x_{21}, \ldots, x_{2n}\}$. It is obvious from the factorization criterion theorem that $t_1$ and $t_2$ are sufficient statistics for truncation parameters $\theta_1$ and $\theta_2$, respectively.

## 2.2 Conjugate Prior Distribution

In this section, we will present the conjugate prior distributions for the samples $y_i = (x_{i1}, \ldots, x_{in})$ from truncation parameter distributions (1.1.1) and (1.1.2) in the following Theorems 2.2.1 and 2.2.2.

**Theorem 2.2.1** *Let $x_{11}, \ldots, x_{1n}$ be a random sample of size $n$ from the type I truncation parameter distribution. Suppose that the prior distribution of truncation parameter $\theta_1$ is given by*

$$g_1(\theta_1) = \frac{\alpha\{k_1(\theta_0)\}^{\alpha}}{\{k_1(\theta_1)\}^{\alpha+1}} h_1(\theta_1), \qquad a < \theta_1 < \theta_0, \ \alpha > 0, \tag{2.2.1}$$

*with parameter $a < \theta_0 < b$ and $\alpha > 0$; if*

$$\lim_{\theta_1 \to a} k_1(\theta_1) = \lim_{\theta_1 \to a} \int_{\theta_1}^{b} h_1(x_1)dx_1 = \infty$$

*then the posterior distribution of $\theta_1$ given $X_{1j} = x_{1j}, \ j = 1, \ldots, n$, is*

$$\begin{aligned} p_1(\theta_1|x_{11}, \ldots, x_{1n}) &= \frac{(\alpha + n)\{k_1(\theta_0^*)\}^{\alpha+n}}{\{k_1(\theta_1)\}^{\alpha+n+1}} h_1(\theta_1), \\ & \qquad a < \theta_1 < \theta_0^*, \ \alpha > 0, \end{aligned} \tag{2.2.2}$$

*with parameters $\theta_0^* = \min\{x_{11}, \ldots, x_{1n}, \theta_0\}$ and $\alpha + n$.*

**Proof.** Since

$$g_1(\theta_1) \propto \begin{cases} \frac{h_1(\theta_1)}{\{k_1(\theta_1)\}^{\alpha+1}}, & a < \theta_1 < \theta_0 \\ 0, & \text{otherwise} \end{cases}$$

and

$$\ell_1(y_1|\theta_1) \propto \begin{cases} \frac{1}{\{k_1(\theta_1)\}^{n}}, & a < \theta_1 < \min\{x_{11}, \ldots, x_{1n}\} \\ 0, & \text{otherwise} \end{cases}$$

then we have

$$p_1(\theta_1|x_{11}, \ldots, x_{1n}) \propto \begin{cases} \frac{h_1(\theta_1)}{\{k_1(\theta_1)\}^{\alpha+n+1}}, & a < \theta_1 < \theta_0^* \\ 0, & \text{otherwise.} \end{cases} \tag{2.2.3}$$

It is seen from the relation (2.2.3) that the posterior distribution of $\theta_1$ must be given by (2.2.2) with parameters $\theta_0^* = \min\{x_{11}, \ldots, x_{1n}, \theta_0\}$ and $\alpha + n$.

**Theorem 2.2.2** *Let $x_{21}, \ldots, x_{2n}$ be a random sample of size n from the type II truncation parameter distribution. Suppose that the prior distribution of truncation parameter $\theta_2$ is given by*

$$g_2(\theta_2) = \frac{\alpha\{k_2(\theta_0)\}^\alpha}{\{k_2(\theta_2)\}^{\alpha+1}}\, h_2(\theta_2), \qquad \theta_0 < \theta_2 < b,\ \alpha > 0, \tag{2.2.4}$$

*with parameter $a < \theta_0 < b$ and $\alpha > 0$; if*

$$\lim_{\theta_2 \to b} k_2(\theta_2) = \lim_{\theta_2 \to b} \int_a^{\theta_2} h_2(x_2)dx_2 = \infty$$

*then the posterior distribution of $\theta_2$ given $X_{2j} = x_{2j}$, $j = 1, \ldots, n$, is*

$$\begin{aligned} p_2(\theta_2|x_{21}, \ldots, x_{2n}) &= \frac{(\alpha + n)\{k_2(\theta_0^*)\}^{\alpha+n}}{\{k_2(\theta_2)\}^{\alpha+n+1}}\, h_2(\theta_2), \\ &\qquad \theta_0^* < \theta_2 < b,\ \alpha > 0, \end{aligned} \tag{2.2.5}$$

*with parameters $\theta_0^* = \max\{x_{21}, \ldots, x_{2n}, \theta_0\}$ and $\alpha + n$.*

**Proof.** Since

$$g_2(\theta_2) \propto \begin{cases} \frac{h_2(\theta_2)}{\{k_2(\theta_2)\}^{\alpha+1}}, & \theta_0 < \theta_2 < b \\ 0, & \text{otherwise} \end{cases}$$

and

$$\ell_2(y_2|\theta_2) \propto \begin{cases} \frac{1}{\{k_2(\theta_2)\}^n}, & \max\{x_{21}, \ldots, x_{2n}\} < \theta_2 < b \\ 0, & \text{otherwise} \end{cases}$$

then we have

$$p_2(\theta_2|x_{21}, \ldots, x_{2n}) \propto \begin{cases} \frac{h_2(\theta_2)}{\{k_2(\theta_2)\}^{\alpha+n+1}}, & \theta_0^* < \theta_2 < b \\ 0, & \text{otherwise.} \end{cases} \tag{2.2.6}$$

It is easily seen from the relation (2.2.6) that the posterior distribution of $\theta_2$ must be given by (2.2.5) with parameters $\theta_0^* = \max\{x_{21}, \ldots, x_{2n}, \theta_0\}$ and $\alpha + n$.

**Example 2.2.1** Let $x_{11}, \ldots, x_{1n}$ be an independent and identically distributed sample from the Pareto distribution as follows

$$f_1(x_1|\theta_1) = \frac{a\,\theta_1^a}{x_1^{a+1}}, \qquad 0 < \theta_1 \le x_1 < \infty,\ a > 0.$$

This is a type I truncation parameter density with $h_1(x_1) = a/x_1^{a+1}$ and $k_1(\theta_1) = 1/\theta_1^a$. Then the conjugate prior distribution is given by

$$g_1(\theta_1) = \frac{\alpha\{k_1(\theta_0)\}^\alpha}{\{k_1(\theta_1)\}^{\alpha+1}} h_1(\theta_1), \qquad 0 < \theta_1 \le \theta_0 < \infty, \ \alpha > 0,$$

$$= \frac{\alpha\{\theta_0^{-a}\}^\alpha}{\{\theta_1^{-a}\}^{\alpha+1}} \left(\frac{a}{\theta_1^{a+1}}\right), \qquad 0 < \theta_1 \le \theta_0 < \infty, \ \alpha > 0,$$

$$= \frac{\alpha a\, \theta^{\alpha a-1}}{\theta_0^{\alpha a}} = \frac{\beta\theta_1^{\beta-1}}{\theta_0^{\beta}}, \qquad 0 < \theta_1 \le \theta_0 < \infty, \ \beta > 0,$$

which is a power function distribution. In this case, the posterior distribution is

$$p_1(\theta_1|x_{11}, \ldots, x_{1n}) = \frac{(\alpha+n)\{k_1(\theta_0^*)\}^{\alpha+n}}{\{k_1(\theta_1)\}^{\alpha+n+1}} h_1(\theta_1)$$

$$= \frac{(\alpha+n)\{\theta_0^{*-a}\}^{\alpha+n}}{\{\theta_1^{-a}\}^{\alpha+n+1}} \left(\frac{a}{\theta_1^{a+1}}\right)$$

$$= \frac{(\alpha+n)a\ \theta_1^{(\alpha+n)a-1}}{\theta_0^{*(\alpha+n)a}}, \qquad 0 < \theta_1 \le \theta_0^* < \infty,$$

which is still a power function distribution.

**Example 2.2.2** Let $x_{21}, \ldots, x_{2n}$ be an independent and identically distributed sample from the right-truncated exponential distribution as follows

$$f_2(x_2|\theta_2) = \lambda \exp\{-\lambda(\theta_2 - x_2)\}, \qquad -\infty < x_2 \le \theta_2 < \infty, \ \lambda > 0.$$

This is a type II truncation parameter density with $h_2(x_2) = \lambda\ e^{\lambda x_2}$ and $k_2(\theta_2) = e^{\lambda\theta_2}$. Then the conjugate prior distribution is given by

$$g_2(\theta_2) = \frac{\alpha\{k_2(\theta_0)\}^\alpha}{\{k_2(\theta_2)\}^{\alpha+1}} h_2(\theta_2), \qquad -\infty < \theta_0 \le \theta_2 < \infty, \ \alpha > 0,$$

$$= \frac{\alpha\{e^{\lambda\theta_0}\}^\alpha}{\{e^{\lambda\theta_2}\}^{\alpha+1}} (\lambda\ e^{\lambda\theta_2}), \qquad -\infty < \theta_0 \le \theta_2 < \infty, \ \alpha > 0,$$

$$= \frac{\lambda\alpha\ e^{\lambda\alpha\theta_0}}{e^{\lambda\alpha\theta_2}} = \frac{\mu\ e^{-\mu\theta_2}}{e^{-\mu\theta_0}}, \qquad -\infty < \theta_0 \le \theta_2 < \infty, \mu > 0,$$

which is a left-truncated exponential distribution. In this case, the posterior distribution is

$$\begin{aligned} p_2(\theta_2|x_{21},\ldots,x_{2n}) &= \frac{(\alpha+n)\{k_2(\theta_0^*)\}^{\alpha+n}}{\{k_2(\theta_2)\}^{\alpha+n+1}}\, h_2(\theta_2) \\ &= \frac{(\alpha+n)\{e^{\lambda\theta_0^*}\}^{\alpha+n}}{\{e^{\lambda\theta_2}\}^{\alpha+n+1}}\, (\lambda\, e^{\lambda\theta_2}) \\ &= \frac{(\alpha+n)\lambda\, e^{-\lambda(\alpha+n)\theta_2}}{e^{-\lambda(\alpha+n)\theta_0^*}}\,, \qquad -\infty < \theta_0^* \leq \theta_2 < \infty, \end{aligned}$$

which is still a left-truncated exponential distribution.

# Chapter 3

# Estimation of the Mixing Distributions

## 3.1 Introduction

Let $(x_1, \theta_1), \ldots, (x_n, \theta_n)$ be independent and identically distributed pairs of random variables in which the $x_1, \ldots, x_n$ are observable but the $\theta_1, \ldots, \theta_n$ are not, with the conditional distribution of $x_i$ given $\theta_i$ being $f(x_i|\theta_i)$ and $\theta_1, \ldots, \theta_n$ being regarded as a random sample from an unknown mixing distribution $G$. The problem of estimating the mixing distribution is to determine the $G$ from the observations $x_1, \ldots, x_n$. This problem arises in a variety of statistical problems. A useful application of such a problem was introduced in the empirical Bayes decision theory (see Robbins, 1964 and Susarla, 1982) where the mixing distribution is called the prior distribution. Various methods have been proposed for estimating mixing distributions in the literature. In addition to Robbins (1964) and Susarla (1982), see Berger (1985) and Maritz and Lwin (1989) for the general introduction. In particular, Susarla (1982) mentioned that it is very useful in empirical Bayes theory if it is possible to express the mixing distribution in terms of the mixture distribution and then propose the explicit estimator by using the observations $x_1, \ldots, x_n$. Early works on this topic include Fox (1970), Susarla and O'Bryan (1979), Blum and Susarla (1981) and Prasad and Singh (1990). They investigated the relations between the mixing distributions and the mixture distributions for uniform and truncated exponential distributions and constructed estimators for the mixing distributions through the obtained relations. Recently, for two general types of truncation parameter distributions, which include uniform and truncated exponential distributions, with continuous mixing distributions, Ma and Balakrishnan (1997) demonstrated the explicit relations between the mixing distributions and the mixture distributions and studied the asymptotic property of the proposed estimators for mixing distributions based on the relations.

In this work, the previous results of Ma and Balakrishnan (1997) are extended to the case where the mixing distribution $G$ is a general distribution which may or may not be continuous. We first demonstrate the explicit expressions for the mixing distributions in terms of the mixture distributions and establish the identifiability for these mixture distributions. Then, we investigate the mean integrated squared error (MISE) convergence rates for the proposed estimators for the mixing distributions and show that these estimators are asymptotically normal in distribution.

## 3.2 Relations

Consider two different kinds of truncation parameter distribution families with density functions over any interval $(a, b)$, finite or infinite, as follows:

Type I truncation parameter density

$$f_1(x_1|\theta_1) = h_1(x_1)/k_1(\theta_1), \qquad a < \theta_1 \le x_1 < b; \tag{3.2.1}$$

Type II truncation parameter density

$$f_2(x_2|\theta_2) = h_2(x_2)/k_2(\theta_2), \qquad a < x_2 \le \theta_2 < b. \tag{3.2.2}$$

where $h_1(x_1)$ and $h_2(x_2)$ are both positive, continuous and integrable over $(\theta_1, b)$ and $(a, \theta_2)$, respectively, for $\theta_i$, $i = 1, 2$, in the interval $(a, b)$.

Let $G_i(\theta_i)$ be the mixing distributions of truncation parameters $\theta_i$, $i = 1, 2$, respectively, defined in the interval $(a, b)$ with $G_i(a) = 0$ and $G_i(b) = 1$, where $G_i(\theta_i)$ might not be continuous distributions; then the mixture distributions of $x_i$, $i = 1, 2$, are given by

$$f_{G_1}(x_1) = f_1(x_1) = \int_a^{x_1} \{h_1(x_1)/k_1(\theta_1)\} dG_1(\theta_1). \tag{3.2.3}$$

$$f_{G_2}(x_2) = f_2(x_2) = \int_{x_2}^{b} \{h_2(x_2)/k_2(\theta_2)\} dG_2(\theta_2). \tag{3.2.4}$$

In this chapter, we will pay some attention to continuous mixture and finite mixture models. A mixture is called a continuous mixture if

$$f_{G_i}(x_i) = \int f_i(x_i|\theta_i) g_i(\theta_i) d\theta_i \tag{3.2.5}$$

i.e., the mixing distributions $G_i$ $i = 1, 2$, are continuous distributions with density functions $g_i(\theta_i)$, $i = 1, 2$, respectively. A mixture is called a finite mixture if

$$f_{G_i}(x_i) = \sum_{j=1}^{k} p_{ij} f_i(x_i|\theta_{ij}) \tag{3.2.6}$$

i.e., the mixing distributions $G_i$, $i = 1, 2$, are finite discrete distributions with probability mass functions $p(G_i = \theta_{ij}) = p_{ij}$, $j = 1, \ldots, k$, $i = 1, 2$, respectively.

From the above density functions (3.2.1)-(3.2.4), we will demonstrate the explicit expressions for the mixing distributions in terms of the mixture distributions of $x_i$, $i = 1, 2$, respectively, for the two general truncation parameter distribution models in the following Theorems 3.2.1 and 3.2.2.

**Theorem 3.2.1** *For the type I truncation parameter distribution family (3.2.1), we have*

$$G_1(t_1) = F_1(t_1) + \frac{k_1(t_1)}{h_1(t_1)} f_1(t_1) \tag{3.2.7}$$

*no matter what kind of distribution the $G_1$ might be, where $f_1$ is the mixture distribution given by (3.2.3) and $F_1$ is the corresponding cumulative distribution function.*

**Proof.** From the conditional density function $f_1(x_1|\theta_1)$ of (3.2.1), we have

$$\begin{aligned} F_1(t_1|\theta_1) &= \int_a^{t_1} f_1(x_1|\theta_1)dx_1 = \int_a^{t_1} \frac{h_1(x_1)}{k_1(\theta_1)}\, dx_1 \\ &= \frac{1}{k_1(\theta_1)}\,(k_1(\theta_1) - k_1(t_1))I(\theta_1 \le t_1) \\ &= I(\theta_1 \le t_1) - \frac{k_1(t_1)}{k_1(\theta_1)}\, I(\theta_1 \le t_1), \end{aligned}$$

where $k_1(\theta_1) = \int_{\theta_1}^b h_1(x_1)dx_1$ and $I(\cdot)$ is the indicator function. Then we are able to get

$$\begin{aligned} F_1(t_1) &= \int_a^b F_1(t_1|\theta_1)dG_1(\theta_1) \\ &= \int_a^b [I(\theta_1 \le t_1) - \frac{k_1(t_1)}{k_1(\theta_1)}\, I(\theta_1 \le t_1)]dG_1(\theta_1) \\ &= G_1(t_1) - \frac{k_1(t_1)}{h_1(t_1)}\, f_1(t_1). \end{aligned}$$

That is,

$$G_1(t_1) = F_1(t_1) + \frac{k_1(t_1)}{h_1(t_1)}\, f_1(t_1).$$

**Theorem 3.2.2** *For the type II truncation parameter distribution family (3.2.2), we have*

$$G_2(t_2) = F_2(t_2) - \frac{k_2(t_2)}{h_2(t_2)}\, f_2(t_2) \qquad (3.2.8)$$

*no matter what kind of distribution the $G_2$ might be, where $f_2$ is the mixture distribution given by (3.2.4) and $F_2$ is the corresponding cumulative distribution function.*

**Proof.** From the conditional density function $f_2(x_2 \mid \theta_2)$ of (3.2.2), we have

$$F_2(t_2|\theta_2) = \int_a^{t_2} f_2(x_2|\theta_2)dx_2 = \int_a^{t_2} \frac{h_2(x_2)}{k_2(\theta_2)}\, dx_2$$

$$
\begin{aligned}
&= \frac{1}{k_2(\theta_2)}\,[k_2(\theta_2)I(\theta_2 \le t_2) + k_2(t_2)I(t_2 < \theta_2)] \\
&= I(\theta_2 \le t_2) + \frac{k_2(t_2)}{k_2(\theta_2)}\, I(t_2 < \theta_2)
\end{aligned}
$$

by using $k_2(\theta_2) = \int_a^{\theta_2} h_2(x_2)dx_2$; then we are able to get

$$
\begin{aligned}
F_2(t_2) &= \int_a^b F_2(t_2|\theta_2)dG_2(\theta_2) \\
&= \int_a^b \left[ I(\theta_2 \le t_2) + \frac{k_2(t_2)}{k_2(\theta_2)}\, I(t_2 < \theta_2) \right] dG_2(\theta_2) \\
&= G_2(t_2) + \frac{k_2(t_2)}{h_2(t_2)}\, f_2(t_2).
\end{aligned}
$$

That is,

$$
G_2(t_2) = F_2(t_2) - \frac{k_2(t_2)}{h_2(t_2)}\, f_2(t_2).
$$

**Example 3.2.1** Consider the left-truncated exponential distribution

$$
f_1(x_1|\theta_1) = \lambda \exp\{-\lambda(x_1 - \theta_1)\}, \quad 0 < \theta_1 \le x_1 < \infty, \ \lambda > 0,
$$

as a type I truncation parameter distribution with $h_1(x_1) = \lambda e^{-\lambda x_1}$ and $k_1(\theta_1) = e^{-\lambda\theta_1}$. Then by Theorem 3.2.1, we have

$$
\begin{aligned}
G_1(t_1) &= F_1(t_1) + \frac{k_1(t_1)}{h_1(t_1)}\, f_1(t_1) \\
&= F_1(t_1) + \frac{1}{\lambda}\, f_1(t_1).
\end{aligned}
$$

This relation was also obtained by Blum and Susarla (1981) and by Prasad and Singh (1990).

**Example 3.2.2** Consider the Pareto distribution

$$
f_1(x_1|\theta_1) = \frac{\alpha\theta_1^\alpha}{x_1^{\alpha+1}}, \quad 0 < \theta_1 \le x < \infty, \ \alpha > 0,
$$

as a type I truncation parameter distribution with $h_1(x_1) = \alpha/x_1^{\alpha+1}$ and $k_1(\theta_1) = 1/\theta^\alpha$. Then by Theorem 3.2.1, we have

$$
G_1(t_1) = F_1(t_1) + \frac{t_1}{\alpha}\, f_1(t_1).
$$

This relation was also obtained by Tiwari and Zalkikar (1990).

**Example 3.2.3** Consider the right-truncated exponential distribution

$$f_2(x_2|\theta_2) = \lambda \exp\{-\lambda(\theta_2 - x_2)\}, \quad -\infty < x_2 \leq \theta_2 < \infty, \ \lambda > 0,$$

as a type II truncation parameter distribution with $h_2(x_2) = \lambda e^{\lambda x_2}$ and $k_2(\theta_2) = e^{\lambda\theta_2}$. Then by Theorem 3.2.2, we have

$$\begin{aligned} G_2(t_2) &= F_2(t_2) - \frac{k_2(t_2)}{h_2(t_2)} f_2(t_2) \\ &= F_2(t_2) - \frac{1}{\lambda} f_2(t_2). \end{aligned}$$

**Example 3.2.4** Consider the power function distribution

$$f_2(x_2|\theta_2) = \frac{\alpha x_2^{\alpha-1}}{\theta_2^{\alpha}}, \quad 0 < x_2 \leq \theta_2 < \infty, \ \alpha > 0,$$

as a type II truncation parameter distribution with $h_2(x_2) = \alpha x_2^{\alpha-1}$ and $k_2(\theta_2) = \theta_2^{\alpha}$. Then by Theorem 3.2.2, we have

$$G_2(t_2) = F_2(t_2) - \frac{t_2}{\alpha} f_2(t_2).$$

This relation was also obtained by Fox (1970) and by Susarla and O'Bryan (1979).

Note that Ma and Balakrishnan (1997) obtained the results in (3.2.7) and (3.2.8) by using different methods for the two truncation parameter distribution families with continuous mixing distributions. Certainly, the results of (3.2.7) and (3.2.8) are more general than former ones by Ma and Balakrishnan (1997). In particular, if $G_i$, $i = 1, 2$, are finite discrete distributions, we will be able to estimate the mixing proportions in the finite mixture for these truncation parameter distributions by the relations (3.2.7) and (3.2.8). We will discuss finite mixtures later in this chapter.

## 3.3 Identifiability

Before we estimate the unknown mixing distributions $G_i$, $i = 1, 2$, we must discuss the problem of identifiability. Identifiability means that there is a one-to-one correspondence between the mixing distribution $G_i$ and the mixture distribution $f_i$. It is obvious that the estimation of mixing distributions will become feasible only if the identifiability condition is satisfied. For the general introduction about identifiability, see Teicher (1961, 1963).

**Definition 3.3.1** A class of mixtures is called identifiable if, and only if, the condition

$$f_G(x) = f_H(x) \Leftrightarrow G = H \tag{3.3.1}$$

is satisfied for every pair $G$ and $H$.

From the obtained explicit relations of (3.2.7) and (3.2.8) in Theorems 3.2.1 and 3.2.2, we can easily establish the identifiability for the two truncation parameter distribution families with arbitrary mixing distributions, respectively.

**Theorem 3.3.1** *(a) The class of mixtures of the type I truncation parameter distribution (3.2.1) with arbitrary mixing distributions is identifiable.*

*(b) The class of mixtures of the type II truncation parameter distribution (3.2.2) with arbitrary mixing distributions is identifiable.*

**Proof.** (a) and (b) are obvious from the relations (3.2.7) and (3.2.8) in Theorems 3.2.1 and 3.2.2, respectively.

As a special case of the general mixing distribution, when $G_i$, $i = 1, 2$, are finite discrete distributions, we can get the identifiability for the finite mixtures of the two truncation parameter distribution families, respectively, as stated in the following theorem.

**Theorem 3.3.2** *(a) The class of all finite mixtures of the type I truncation parameter distribution is identifiable.*

*(b) The class of all finite mixtures of the type II truncation parameter distribution is identifiable.*

## 3.4 MISE Convergence Rates

Since the mixing distributions $G_i$ can be represented in terms of mixture distributions $f_i$ and $F_i$, $i = 1,2$, respectively, as given by (3.2.7) and (3.2.8), it suggests us to estimate the mixing distributions by simply estimating $f_i$ and $F_i$, $i = 1,2$, respectively.

Let $x_{i1}, \ldots, x_{in}$ denote the observations from mixture distributions $f_i(x_i)$, $i = 1,2$, respectively. In order to estimate $f_i(x_i)$ and $F_i(x_i)$, $i = 1,2$, we denote the empirical distribution functions $F_{in}$, $i = 1,2$, as follows:

$$F_{in}(t_i) = \frac{1}{n}\sum_{j=1}^{n} I(x_{ij} \le t_i). \tag{3.4.1}$$

Then we employ kernel estimators for $f_i$, $i = 1,2$,

$$f_{in}(t_i) = \frac{1}{nh_n}\sum_{j=1}^{n} k\left(\frac{t_i - x_{ij}}{h_n}\right). \tag{3.4.2}$$

where $k(\cdot)$ is the kernel function and $h_n$ is a sequence of positive numbers such that $h_n \to 0$ and $nh_n \to \infty$ as $n \to \infty$. We will propose different conditions on $k(\cdot)$ according to the different estimation situations in the following study.

Utilizing $f_{in}$ and $F_{in}$, we propose the estimators for the mixing distributions $G_i$, $i = 1,2$, respectively, based on the relations (3.2.7) and (3.2.8) in Theorems 3.2.1 and 3.2.2, as follows:

$$G_{1n}(t_1) = F_{1n}(t_1) + \frac{k_1(t_1)}{h_1(t_1)} f_{1n}(t_1). \tag{3.4.3}$$

$$G_{2n}(t_2) = F_{2n}(t_2) - \frac{k_2(t_2)}{h_2(t_2)} f_{2n}(t_2). \tag{3.4.4}$$

Using results available in the literature about the estimators $f_{in}$ and $F_{in}$, $i = 1,2$, we will study the asymptotic optimality properties of mean integrated squared error (MISE) for the proposed estimators $G_{in}(t_i)$, $i = 1,2$, for the continuous mixture and finite mixture cases.

### 3.4.1 Continuous mixture

Under the statistical framework of a continuous mixture model, in order to estimate $f_i(x_i)$, $i = 1, 2$, we employ kernel functions suggested by Müller and Gasser (1979). Let $k(\cdot)$ be any kernel function satisfying the following conditions:

$$\text{(i)} \quad \text{support of } k(\cdot) = [-\tau, \tau], \ \tau > 0; \tag{3.4.5}$$

$$\text{(ii)} \quad \int_{-\tau}^{\tau} k(y) y^{\ell} dy = \begin{cases} 0, & \ell = 1, \ldots, r-1, \\ 1, & \ell = 0; \end{cases} \tag{3.4.6}$$

$$\text{(iii)} \quad \int_{-\tau}^{\tau} k(y) y^{r} dy \neq 0. \tag{3.4.7}$$

where $r$ is a positive integer.

**Theorem 3.4.1** *Let $G_{in}(t_i)$, $i = 1, 2$, be defined by (3.4.3) and (3.4.4), respectively, with kernel function $k(\cdot)$ satisfying (3.4.5)-(3.4.7). Suppose the regularity conditions in Theorem 1 of Müller and Gasser (1979) are satisfied. Also, assume $k_i(t_i)/h_i(t_i)$ is bounded and $\int |t_i| f(t_i) dt_i < \infty$. Then, with the choice of $h_n = n^{-\frac{1}{2r+1}}$, we have*

$$E \int \{G_{in}(t_i) - G_i(t_i)\} dt_i = O\left(n^{-\frac{2r}{2r+1}}\right). \tag{3.4.8}$$

**Proof.** From the definitions of $G_{in}(t_i)$ $i = 1, 2$, we can get

$$\begin{aligned} E \int \{G_{in}(t_i) - G_i(t_i)\}^2 dt_i \ &\leq \ 2E \int \{F_{in}(t_i) - F_i(t_i)\}^2 dt_i \\ &\quad + 2E \int \frac{k_i^2(t_i)}{h_i^2(t_i)} \{f_{in}(t_i) - f_i(t_i)\}^2 dt_i. \end{aligned}$$

Now,

$$\begin{aligned} E \int \{F_{in}(t_i) - F_i(t_i)\}^2 dt_i \ &= \ \int \frac{1}{n} F_i(t_i)(1 - F_i(t_i)) dt_i \\ &\leq \ O(n^{-1}) \int |t_i| f(t_i) dt_i, \end{aligned}$$

and from Theorem 1 in Müller and Gasser (1979),

$$E\int\{f_{in}(t_i)-f_i(t_i)\}^2dt_i = O\left(n^{-\frac{2r}{2r+1}}\right).$$

Then the theorem is proved.

### 3.4.2 Finite mixture

Under the statistical framework of a finite mixture model, since $G_i$, $i = 1, 2$, are finite discrete distributions, $f_i(x_i)$, $i = 1, 2$, are obviously discontinuous at the support points of $G_i$, $\theta_{i1}, \ldots, \theta_{ik}$, $i = 1, 2$, respectively. In this situation, in order to estimate $f_i(x_i)$, $i = 1, 2$, we use kernel functions suggested by Cline and Hart (1991). Let $k(\cdot)$ be any kernel functions satisfying the following conditions:

$$(1) \quad \int k(y)dy = 1, \tag{3.4.9}$$

$$(2) \quad \int k^2(y)dy < \infty, \tag{3.4.10}$$

$$(3) \quad \int |y|k(y)dy < \infty. \tag{3.4.11}$$

**Theorem 3.4.2** *Let $G_{in}(t_i)$ be defined by (3.4.3) and (3.4.4), respectively, with kernel function $k(\cdot)$ satisfying (3.4.9)-(3.4.11). Suppose the regularity conditons in Theorem 1 of Cline and Hart (1991) are satisfied and also assume $k_i(t_i)/h_i(t_i)$ is bounded and $\int |t_i| f(t_i)dt_i < \infty$. Then with the choice of $h_n = n^{-\frac{1}{2}}$, we have*

$$E\int\{G_{in}(t_i)-G_i(t_i)\}^2dt_i = O\left(n^{-\frac{1}{2}}\right). \tag{3.4.12}$$

**Proof.** From Theorem 1 in Cline and Hart (1991),

$$E\int\{f_{in}(t_i)-f_i(t_i)\}^2dt_i = O\left(n^{-\frac{1}{2}}\right).$$

Then, with a discussion similar to that in the proof of Theorem 3.4.1, the theorem is proved.

## 3.5 Asymptotic Normality

In this section, based on the definitions of $G_{in}$ in (3.4.3) and (3.4.4), we examine the asymptotic distributions for these estimators $G_{in}$, respectively, under continuous mixture and finite mixture situations.

### 3.5.1 Continuous mixture

**Theorem 3.5.1** *Let $G_{in}(t_i)$ be defined by (3.4.3) and (3.4.4), respectively. Then $G_{in}$ converges in distribution to a normal distribution, $i = 1, 2$, respectively, as follows:*

$$\frac{G_{in}(t_i) - E\{G_{in}(t_i)\}}{\sqrt{var\{G_{in}(t_i)\}}} \xrightarrow{d} N(0, 1) \text{ as } n \to \infty. \tag{3.5.1}$$

*In addition, if*

$$\frac{G_i(t_i) - E\{G_{in}(t_i)\}}{\sqrt{var\{G_{in}(t_i)\}}} \to 0 \text{ as } n \to \infty,$$

*then*

$$\frac{G_{in}(t_i) - G_i(t_i)}{\sqrt{var\{G_{in}(t_i)\}}} \xrightarrow{d} N(0, 1) \text{ as } n \to \infty. \tag{3.5.2}$$

**Proof.** By Theorem 3.3 in Ma and Balakrishnan (1997) and Section 2 in Parzen (1962), we can get the result of (3.5.1), then the result of (3.5.2) follow by using Slutsky's Theorem.

### 3.5.2 Finite mixture

Suppose that $\theta_{i1}, \ldots, \theta_{ik}$, $i = 1, 2$, are support points of $G_i$, respectively, such that $\theta_{i1} < \cdots < \theta_{ik}$. Then we have

$$\begin{aligned} f_1(x_1) &= p_{11} \frac{h_1(x_1)}{k_1(\theta_{11})} I(\theta_{11} \le x_1) + p_{12} \frac{h_1(x_1)}{k_1(\theta_{12})} I(\theta_{12} \le x_1) \\ &+ \cdots + p_{1k} \frac{h_1(x_1)}{k_1(\theta_{1k})} I(\theta_{1k} \le x_1) \end{aligned} \tag{3.5.3}$$

and

$$
\begin{aligned}
f_2(x_2) &= p_{21}\,\frac{h_2(x_2)}{k_2(\theta_{21})}\,I(x_2<\theta_{21}) + p_{22}\,\frac{h_2(x_2)}{k_2(\theta_{22})}\,I(x_2<\theta_{22}) \\
&\quad + \cdots + p_{2k}\,\frac{h_2(x_2)}{k_2(\theta_{2k})}\,I(x_2<\theta_{2k}).
\end{aligned}
\tag{3.5.4}
$$

It is obvious that $f_i(x_i)$, $i = 1,2$, have jumps at $\theta_{i1},\ldots,\theta_{ik}$, respectively.

In this finite mixture situation, one may be interested in estimating the mixing proportions instead of the mixing distributions $G_i$, $i = 1,2$,. From the mixture distributions $f_i(x_i)$, $i = 1,2$, we can get expressions for the mixing proportions in terms of $f_i(x_i)$, $i = 1,2$, respectively, as follows:

$$
\begin{aligned}
p_{1j} &= G_1(\theta_{1j}) - G_1(\theta_{1j}^-) \\
&= [F_1(\theta_{1j}) - F_1(\theta_{1j}^-)] + \frac{k_1(\theta_{1j})}{h_1(\theta_{1j})}\,f_1(\theta_{1j}) - \frac{k_1(\theta_{1j}^-)}{h_1(\theta_{1j}^-)}\,f_1(\theta_{1j}^-) \\
&= \frac{k_1(\theta_{1j})}{h_1(\theta_{1j})}\,[f(\theta_{1j}) - f(\theta_{1j}^-)], \quad j = 1,\ldots,k.
\end{aligned}
\tag{3.5.5}
$$

and

$$
\begin{aligned}
p_{2j} &= G_2(\theta_{2j}) - G_2(\theta_{2j}^-) \\
&= [F_2(\theta_{2j}) - F_2(\theta_{2j}^-)] - \frac{k_2(\theta_{2j})}{h_2(\theta_{2j})}\,f_2(\theta_{2j}) + \frac{k_2(\theta_{2j}^-)}{h_2(\theta_{2j}^-)}\,f_2(\theta_{2j}^-) \\
&= -\,\frac{k_2(\theta_{2j})}{h_2(\theta_{2j})}\,[f(\theta_{2j}) - f(\theta_{2j}^-)], \quad j = 1,\ldots,k.
\end{aligned}
\tag{3.5.6}
$$

Then we propose the following estimators for the mixing proportions $p_{ij}$, $i = 1,2$, respectively,

$$
p_{1jn} = \frac{k_1(\theta_{1j})}{h_1(\theta_{1j})}\,[f_{1n}(\theta_{1j}) - f_{1n}(\theta_{1j}^-] \tag{3.5.7}
$$

$$
p_{2jn} = -\,\frac{k_2(\theta_{2j})}{h_2(\theta_{2j})}\,[f_{2n}(\theta_{2j}) - f_{2n}(\theta_{2j}^-]. \tag{3.5.8}
$$

**Theorem 3.5.2** *Let $p_{ijn}$, $i = 1,2$, be defined by (3.5.7) and (3.5.8), respectively, with kernel function $k(\cdot)$ satisfying the conditions in Theorem 4 of Cline and Hart*

*(1991). Then $p_{ijn}$ converges in distribution to a normal distribution, $i = 1, 2$, respectively, as follows:*

$$\frac{p_{ijn} - E(p_{ijn})}{\sqrt{var(p_{ijn})}} \xrightarrow{d} \mathcal{N}(0, 1) \text{ as } n \to \infty. \tag{3.5.9}$$

*In addition, if*

$$\frac{\frac{k_i(\theta_{ij})}{h_i(\theta_{ij})}\left[f_i(\theta_{ij}) - f_i(\theta_{ij}^-)\right] - E(p_{ijn})}{\sqrt{var(p_{ijn})}} \to 0 \text{ as } n \to \infty,$$

*then*

$$\frac{p_{ijn} - p_{ij}}{\sqrt{var(p_{ijn})}} \xrightarrow{d} \mathcal{N}(0, 1) \text{ as } n \to \infty. \tag{3.5.10}$$

**Proof.** See Theorem 4 in Cline and Hart (1991).

# Chapter 4

# Posterior Moments

## 4.1 Introduction

Let $\ell(y|\theta)$ be the likelihood function of an independent and identically distributed sample $y = (x_1, \ldots, x_n)$ from distribution $f(x|\theta)$. Then the Bayes estimator and posterior risk, under squared error loss, are the posterior mean $E(\theta|y)$ and posterior variance $\mathrm{var}(\theta|y)$, respectively. For the normal likelihood function with known variance and an arbitrary prior distribution, the explicit expressions for the posterior mean and variance are derived by Pericchi and Smith (1992). For an arbitrary location parameter likelihood function and the normal prior distribution, the exact form of the posterior mean is given by Polson (1991). Pericchi, Sansó and Smith (1993) also discussed the posterior cumulant relations in Bayesian inference assuming the exponential family form either for the likelihood or for the prior. They all have mentioned that analytical Bayesian computations, without the assumption of normality either on the likelihood function or on the prior distribution, were very difficult and needed investigation.

In this chapter, we consider the likelihood functions of random samples from the following two different types of truncation parameter distributions:

Type I truncation parameter density

$$f_1(x_1|\theta_1) = h_1(x_1)/k_1(\theta_1), \quad a < \theta_1 \le x_1 < b; \tag{4.1.1}$$

Type II truncation parameter density

$$f_2(x_2|\theta_2) = h_2(x_2)/k_2(\theta_2), \quad a < x_2 \le \theta_2 < b, \tag{4.1.2}$$

where $h_1(x_1)$ and $h_2(x_2)$ are positive, continuous and integrable over $(\theta_1, b)$ and $(a, \theta_2)$, respectively, for $\theta_i$, $i = 1, 2$, in the interval $(a, b)$, $-\infty \le a < b \le \infty$.

The object of this chapter is to find the exact analytical expressions of posterior moments for the two different truncation parameter likelihood functions with arbitrary prior distributions. In Section 4.2, the explicit forms for posterior moments are derived by using the sufficient statistics for these truncation parameter distributions. In particular, the explicit expressions of the posterior mean and variance are given, respectively, for these two truncation parameter distribution models. In Section 4.3, two examples representing the two different truncation parameter distribution models are discussed to illustrate the obtained results.

## 4.2 Posterior Moments

Let $y_i = (x_{i1}, \ldots, x_{in})$ denote the independent and identically distributed samples of size $n$ from truncation parameter distributions $f_i(x_i|\theta_i)$, $i = 1, 2$, given by (4.1.1) and (4.1.2), respectively. Then $T_1 = x_{1(1)}$ and $T_2 = x_{2(n)}$ are sufficient statistics for $\theta_i$, $i = 1, 2$, respectively, where $x_{1(1)}$ is the smallest order statistic from $y_1 = (x_{11}, \ldots, x_{1n})$ and $x_{2(n)}$ the largest order statistic from $y_2 = (x_{21}, \ldots, x_{2n})$. The conditional density functions of $T_1$ and $T_2$ can be easily derived as follows [Arnold, Balakrishnan and Nagaraja (1992, p. 12)]:

$$\begin{aligned} f_{T_1}(t_1|\theta_1) &= n\{k_1(t_1)\}^{n-1}h_1(t_1)/\{k_1(\theta_1)\}^n \\ &= H_1(t_1)/K_1(\theta_1), \quad a < \theta_1 \le t_1 < b, \end{aligned} \tag{4.2.1}$$

$$
\begin{aligned}
f_{T_2}(t_2|\theta_2) &= n\{k_2(t_2)\}^{n-1}h_2(t_2)/\{k_2(\theta_2)\}^n \\
&= H_2(t_2)/K_2(\theta_2), \quad a < t_2 \le \theta_2 < b, \qquad (4.2.2)
\end{aligned}
$$

where $H_i(t_i) = n\{k_i(t_i)\}^{n-1}h_i(t_i)$ and $K_i(\theta_i) = \{k_i(\theta_i)\}^n$, $i = 1, 2$, and the second equalities in (4.2.1) and (4.2.2) are obtained by using relations

$$
k_1(\theta_1) = \int_{\theta_1}^{b} h_1(x_1)dx_1,
$$

and

$$
k_2(\theta_2) = \int_{a}^{\theta_2} h_2(x_2)dx_2,
$$

respectively. Note that the conditional density functions of $T_1$ and $T_2$ are still type I and type II truncation parameter densities respectively.

It is assumed in this chapter the truncation parameters $\theta_i$, $i = 1, 2$, have arbitrary prior distributions $G_i(\theta_i)$ on $(a, b)$ with $G_i(a) = 0$ and $G_i(b) = 1$, respectively; then the marginal distributions of $T_i$, $i = 1, 2$, are given by

$$
f_{G_1}(t_1) = f_1(t_1) = \int_a^{t_1}\{H_1(t_1)/K_1(\theta_1)\}dG_1(\theta_1). \qquad (4.2.3)
$$

$$
f_{G_2}(t_2) = f_2(t_2) = \int_{t_2}^{b}\{H_2(t_2)/K_2(\theta_2)\}dG_2(\theta_2). \qquad (4.2.4)
$$

From these density functions (4.2.1)-(4.2.4), we are able to demonstrate the explicit expressions of posterior moments in terms of the marginal distributions of $T_i$, $i = 1, 2$, respectively, for the two different truncation parameter likelihood functions with arbitrary prior distributions.

We first demonstrate the exact relations between the posterior means of general functions $g_i(\theta_i)$ and the marginal distributions $f_i(t_i)$, $i = 1, 2$, respectively, in Theorems 4.2.1 and 4.2.2. Then we can easily give the explicit expressions for posterior moments and posterior means and variances in Corollaries 4.2.1-4.2.4.

**Theorem 4.2.1** *For the type I truncation parameter distribution model, if $g_1(\cdot)$ is differentiable and*

$$
\int_a^{t_1} |g_1'(s_1)|\{H_1(t_1)/H_1(s_1)\}dF_1(s_1) \; < \; \infty,
$$

*then we have*

$$
\begin{aligned}
&E(g_1(\theta_1) \mid y_1) = E(g_1(\theta_1) \mid t_1) \\
&= g_1(t_1) - \frac{\int_a^{t_1} g_1'(s_1)\{H_1(t_1)/H_1(s_1)\}dF_1(s_1)}{f_1(t_1)} \\
&= g_1(t_1) - \frac{\int_a^{t_1} g_1'(s_1)[\{k_1(t_1)\}^{n-1}h_1(t_1)/\{k_1(s_1)\}^{n-1}h_1(s_1)]dF_1(s_1)}{f_1(t_1)},
\end{aligned}
\tag{4.2.5}
$$

*where $f_1(\cdot)$ is given by (4.2.3) and $F_1(\cdot)$ is the corresponding cumulative distribution function.*

**Proof.** By Fubini's theorem

$$
\begin{aligned}
\int_a^{t_1} g_1'(s_1)\,\frac{H_1(t_1)}{H_1(s_1)}\,dF_1(s_1) &= \int_a^{t_1} g_1'(s_1)\,\frac{H_1(t_1)}{H_1(s_1)}\,f_1(s_1)ds_1 \\
&= \int_a^{t_1} g_1'(s_1)\,\frac{H_1(t_1)}{H_1(s_1)}\left\{\int_a^{s_1}\,\frac{H_1(s_1)}{K_1(\theta_1)}\,d\pi_1(\theta_1)\right\}ds_1 \\
&= \int_a^{t_1}\int_{\theta_1}^{t_1} g_1'(s_1)\,\frac{H_1(t_1)}{K_1(\theta_1)}\,ds_1 d\pi_1(\theta_1) \\
&= g_1(t_1)f_1(t_1) - \int_a^{t_1} g_1(\theta_1)f_1(t_1 \mid \theta_1)d\pi_1(\theta_1).
\end{aligned}
$$

Then we obtain

$$
\begin{aligned}
E(g_1(\theta_1) \mid t_1) &= \frac{\int_a^{t_1} g_1(\theta_1)f_1(t_1 \mid \theta_1)d\pi_1(\theta_1)}{f_1(t_1)} \\
&= g_1(t_1) - \frac{\int_a^{t_1} g_1'(s_1)\{H_1(t_1)/H_1(s_1)\}dF_1(s_1)}{f_1(t_1)}.
\end{aligned}
$$

**Theorem 4.2.2** *For the type II truncation parameter distribution model, if $g_2(\cdot)$ is differentiable and*

$$
\int_{t_2}^{b} |g_2'(s_2)|\{H_2(t_2)/H_2(s_2)\}dF_2(s_2) < \infty,
$$

*then we have*

$$
E(g_2(\theta_2) \mid y_2) = E(g_2(\theta_2) \mid t_2)
$$

$$
\begin{aligned}
&= g_2(t_2) + \frac{\int_{t_2}^{b} g_2'(s_2)\{H_2(t_2)/H_2(s_2)\}dF_2(s_2)}{f_2(t_2)} \\
&= g_2(t_2) + \frac{\int_{t_2}^{b} g_2'(s_2)[\{k_2(t_2)\}^{n-1}h_2(t_2)/\{k_2(s_2)\}^{n-1}h_2(s_2)]dF_2(s_2)}{f_2(t_2)} .
\end{aligned} \tag{4.2.6}
$$

*where $f_2(\cdot)$ is given by (4.2.4) and $F_2(\cdot)$ is the corresponding cumulative distribution function.*

**Proof.** It is similar to the proof of Theorem 4.2.1 and is omitted.

**Corollary 4.2.1** *For the type I truncation parameter distribution model, we have*

$$
\begin{aligned}
E(\theta_1^r|y_1) &= E(\theta^r|t_1) = t_1^r - \frac{\int_a^{t_1} rs_1^{r-1}\{H_1(t_1)/H_1(s_1)\}dF_1(s_1)}{f_1(t_1)} \\
&= t_1^r - \frac{\int_a^{t_1} rs_1^{r-1}[\{k_1(t_1)\}^{n-1}h_1(t_1)/\{k_1(s_1)\}^{n-1}h_1(s_1)]dF_1(s_1)}{f_1(t_1)} .
\end{aligned} \tag{4.2.7}
$$

*where $f_1(\cdot)$ is given by (4.2.3) and $F_1(\cdot)$ is the corresponding cumulative distribution function.*

**Corollary 4.2.2** *For the type II truncation parameter distribution model, we have*

$$
\begin{aligned}
E(\theta_2^r|y_2) &= E(\theta_2^r|t_2) = t_2^r + \frac{\int_{t_2}^{b} rs_2^{r-1}\{H_2(t_2)/H_2(s_2)\}dF_2(s_2)}{f_2(t_2)} \\
&= t_2^r + \frac{\int_{t_2}^{b} rs_2^{r-1}[\{k_2(t_2)\}^{n-1}h_2(t_2)/\{k_2(s_2)\}^{n-1}h_2(s_2)]dF_2(s_2)}{f_2(t_2)} .
\end{aligned} \tag{4.2.8}
$$

*where $f_2(\cdot)$ is given by (4.2.4) and $F_2(\cdot)$ is the corresponding cumulative distribution function.*

**Corollary 4.2.3** *For the type I truncation parameter distribution model, the posterior mean and variance are given by, respectively,*

$$
E(\theta_1|y_1) = E(\theta_1|t_1) = t_1 - \frac{u_1(t_1)}{f_1(t_1)} , \tag{4.2.9}
$$

$$
var(\theta_1|y_1) = var(\theta_1|t_1) = \frac{2t_1u_1(t_1) - v_1(t_1)}{f_1(t_1)} - \frac{u_1^2(t_1)}{f_1^2(t_1)} , \tag{4.2.10}
$$

*where*

$$\begin{aligned} u_1(t_1) &= \int_a^{t_1} \{H_1(t_1)/H_1(s_1)\} dF_1(s_1), \\ v_1(t_1) &= \int_a^{t_1} 2s_1 \{H_1(t_1)/H_1(s_1)\} dF_1(s_1). \end{aligned}$$

**Corollary 4.2.4** *For the type II truncation parameter distribution model, the posterior mean and variance are given by, respectively,*

$$E(\theta_2|y_2) = E(\theta_2|t_2) = t_2 + \frac{u_2(t_2)}{f_2(t_2)}. \tag{4.2.11}$$

$$var(\theta_2|y_2) = var(\theta_2|t_2) = \frac{v_2(t_2) - 2t_2u_2(t_2)}{f_2(t_2)} - \frac{u_2^2(t_2)}{f_2^2(t_2)}. \tag{4.2.12}$$

*where*

$$\begin{aligned} u_2(t_2) &= \int_{t_2}^{b} \{H_2(t_2)/H_2(s_2)\} dF_2(s_2), \\ v_2(t_2) &= \int_{t_2}^{b} 2s_2 \{H_2(t_2)/H_2(s_2)\} dF_2(s_2). \end{aligned}$$

## 4.3 Examples

**Example 4.3.1** Let $x_{11}, \ldots, x_{1n}$ be independent and identically distributed according to the truncated exponential distribution as follows:

$$p_1(x_1|\theta_1) = \lambda \exp\{-\lambda(x_1 - \theta_1)\}, \quad -\infty < \theta_1 \le x_1 < \infty, \; \lambda > 0.$$

This is a type I truncation parameter density with $h_1(x_1) = \lambda e^{-\lambda x_1}$, $k_1(\theta_1) = e^{-\lambda\theta_1}$ and $T_1 = \min(x_{11}, \ldots, x_{1n})$ is the sufficient statistic for $\theta_1$. Then we have by Corollary 4.2.1,

$$E(\theta_1^r|t_1) = t_1^r - \frac{\int_{-\infty}^{t_1} r s_1^{r-1} \{e^{-n\lambda(t_1 - s_1)}\} dF_1(s_1)}{f_1(t_1)} \tag{4.3.1}$$

and when $r = 1$, the posterior mean is given by

$$E(\theta_1|t_1) = t_1 - \frac{\int_{-\infty}^{t_1} e^{-n\lambda(t_1 - s_1)} dF_1(s_1)}{f_1(t_1)}; \tag{4.3.2}$$

for the special case $n = 1$, $\lambda = 1$, this result of (4.3.2) is exactly the same as that given by Fox (1978).

**Example 4.3.2** Let $x_{21}, \ldots, x_{2n}$ be independent and identically distributed according to the power function distribution as follows:

$$p_2(x_2|\theta_2) = \alpha x_2^{\alpha-1}/\theta_2^{\alpha}, \quad 0 < x_2 \le \theta_2 < \infty, \ \alpha > 0.$$

This is a type II truncation parameter density with $h_2(x_2) = \alpha x_2^{\alpha-1}$, $k_2(\theta_2) = \theta_2^{\alpha}$ and $T_2 = \max(x_{21}, \ldots, x_{2n})$ is the sufficient statistic for $\theta_2$. Then we have by Corollary 4.2.2,

$$E(\theta_2^r|t_2) = t_2^r + \frac{\int_{t_2}^{\infty} r s_2^{r-1}(t_2^{n\alpha-1}/s_2^{n\alpha-1})dF_2(s_2)}{f_2(t_2)} \tag{4.3.3}$$

and when $r = 1$, the posterior mean is given by

$$E(\theta_2|t_2) = t_2 + \frac{\int_{t_2}^{\infty}(t_2^{n\alpha-1}/s_2^{n\alpha-1})dF_2(s_2)}{f_2(t_2)}\ ; \tag{4.3.4}$$

for the special case $n = 1$, $\alpha = 1$, this result of (4.3.4) is exactly the same as that given by Fox (1978) for the uniform distribution.

# Chapter 5

# Empirical Bayes Estimation

## 5.1 Introduction

Consider the two types of truncation parameter density functions with different truncation parameters $\theta_i$, $i = 1, 2$, as follows:

Type I truncation parameter density

$$f_1(x_1|\theta_1) = h_1(x_1)/k_1(\theta_1), \quad a < \theta_1 \le x_1 < b; \tag{5.1.1}$$

Type II truncation parameter density

$$f_2(x_2|\theta_2) = h_2(x_2)/k_2(\theta_2), \quad a < x_2 \le \theta_2 < b; \tag{5.1.2}$$

where $-\infty \le a < b \le \infty$ in (5.1.1) and (5.1.2).

In this chapter, we consider the problems of estimating the truncation parameters $\theta_i$, $i = 1, 2$, under the squared error loss. We know the Bayes estimator of $\theta_i$ relative to the prior $G_i(\theta_i)$ is given by

$$d_{G_i}(x_i) = E(\theta_i|x_i) = \frac{\int \theta_i f_i(x_i|\theta_i)dG_i(\theta_i)}{\int f_i(x_i|\theta_i)dG_i(\theta_i)} \tag{5.1.3}$$

and the Bayes risk associated with $d_{G_i}(x_i)$ is

$$r_i(d_{G_i}, G_i) = \inf_{d_i} r_i(d_i, G_i) = r_i(G_i). \tag{5.1.4}$$

However, the prior distributions $G_i(\theta_i)$, $i = 1, 2$, are usually unknown in practice. We adopt the empirical Bayes approach in this chapter. The empirical Bayes procedure was first formulated by Robbins (1955) and used rather extensively for various statistical problems by many authors, including Robbins (1963, 1964), Johns and Van Ryzin (1971, 1972), Lin (1975), Singh (1979), and Singh and Wei (1992). Also see Maritz and Lwin (1989) for a general description. For certain nonexponential distribution models, the empirical Bayes method has been applied by Fox (1978), Van Houwelingen (1987), Nogami (1988), Prasad and Singh (1990), Datta (1991), Liang (1993), Huang (1995), and Singh (1995) among others.

In this chapter, we consider the empirical Bayes estimation for the two general types of truncation parameter distribution families (5.1.1) and (5.1.2). In Section 5.2, we demonstrate the relation between the Bayes estimator and the marginal distribution of $x_i$, $i = 1, 2$, respectively, for distribution models (5.1.1) and (5.1.2) and propose the corresponding empirical Bayes estimator. In Section 5.3, we investigate the asymptotic optimality properties of the proposed empirical Bayes estimators. Finally, we discuss some important examples in Section 5.4.

## 5.2 Bayes and Empirical Bayes Estimators

In this section, we first demonstrate the explicit expression for the Bayes estimator $d_{G_i}(x_i)$ in terms of the marginal distribution of $x_i$, $i = 1, 2$, respectively, as follows

**Lemma 5.2.1** *For the type I truncation parameter density (5.1.1), the Bayes estimator under squared error loss is given by*

$$d_{G_1}(x_1) = x_1 - \psi_1(x_1), \tag{5.2.1}$$

*where*

$$\psi_1(x_1) = \int_a^{x_1} \frac{h_1(x_1)}{h_1(t_1)} \, dF_1(t_1)/f_1(x_1) \stackrel{d}{=} w_1(x_1)/f_1(x_1)$$

*with $f_1(x_1)$ given by (4.2.3) and $F_1(x_1)$ the corresponding cumulative distribution.*

**Proof.** See Corollary 4.2.1 in Chapter 4.

**Lemma 5.2.2** *For the type II truncation parameter density (5.1.2), the Bayes estimator under squared error loss is given by*

$$d_{G_2}(x_2) = x_2 + \psi_2(x_2), \tag{5.2.2}$$

*where*

$$\psi_2(x_2) = \int_{x_2}^{b} \frac{h_2(x_2)}{h_2(t_2)}\, dF_2(t_2)/f_2(x_2) \stackrel{d}{=} w_2(x_2)/f_2(x_2)$$

*with $f_2(x_2)$ given by (4.2.4) and $F_2(x_2)$ the corresponding cumulative distribution.*

**Proof.** See Corollary 4.2.2 in Chapter 4.

Now we consider some examples.

**Example 5.2.1** Consider the truncated exponential distribution

$$f_1(x_1|\theta_1) = \frac{1}{\sigma}\, \exp\{-\frac{1}{\sigma}(x_1 - \theta_1)\}, \qquad -\infty < \theta_1 \le x_1 < \infty.$$

This is a type I truncation parameter distribution with $h_1(x_1) = \frac{1}{\sigma}\, e^{-x_1/\sigma}$, $k_1(\theta_1) = e^{-\theta_1/\sigma}$. Then we have, by Lemma 5.2.1,

$$d_{G_1}(x_1) = x_1 - \int_{-\infty}^{x_1} e^{-\frac{1}{\sigma}(x_1 - t_1)} dF_1(t_1)/f_1(x_1).$$

This relation is also obtained by Fox (1978).

**Example 5.2.2** Consider the power function distribution

$$f_2(x_2|\theta_2) = \alpha x_2^{\alpha-1}/\theta_2^{\alpha}, \qquad 0 < x_2 \le \theta_2 < \infty.$$

This is a type II truncation parameter distribution with $h_2(x_2) = \alpha x_2^{\alpha-1}$, $k_2(\theta_2) = \theta_2^{\alpha}$. Then we have, by Lemma 5.2.2,

$$d_{G_2}(x_2) = x_2 + \int_{x_2}^{\infty} (x_2^{\alpha-1}/t_2^{\alpha-1}) dF_2(t_2)/f_2(x_2).$$

When $\alpha = 1$, the Bayes estimator for the uniform distribution $U(0, \theta)$ is given by

$$\begin{aligned} d_{G_2}(x_2) &= x_2 + \int_{x_2}^{\infty} dF_2(t_2)/f_2(x_2) \\ &= x_2 + (1 - F_2(x_2))/f_2(x_2). \end{aligned}$$

This result is also obtained by Fox (1978).

In the empirical Bayes framework, we assume there are sequences $(x_{i1}, \theta_{i1}), \ldots, (x_{in}, \theta_{in})$ (past data) and $(x_{in+1}, \theta_{in+1}) = (x_i, \theta_i)$ (present data) of independent pairs of random variables, where $\theta_{i1}, \ldots, \theta_{in}, \theta_i$ are not observable with unknown prior distribution $G_i$, $i = 1, 2$, and $x_{i1}, \ldots, x_{in}, x_i$ are observable with marginal distribution $f_i(x_i)$, $i = 1, 2$, respectively.

Based on the past data sets $(x_{i1}, \ldots, x_{in})$, $i = 1, 2$, we define the unbiased estimators for the functions $w_i(x_i)$, $i = 1, 2$, as follows

$$w_{1n}(x_1) = \frac{1}{n} \sum_{j=1}^{n} \frac{h_1(x_1)}{h_1(x_{1j})} I_{(a, x_1)}(x_{1j}), \tag{5.2.3}$$

$$w_{2n}(x_2) = \frac{1}{n} \sum_{j=1}^{n} \frac{h_2(x_2)}{h_2(x_{2j})} I_{(x_2, b)}(x_{2j}). \tag{5.2.4}$$

In order to estimate $f_i(x_i)$, $i = 1, 2$, we employ kernel functions suggested by Müller and Gasser (1979). Let $k$ be any kernel function satisfying the following conditions:

(i) support of $k = [-\tau, \tau]$, $\tau > 0$;

(ii) $\int_{-\tau}^{\tau} k(y) y^{\ell} dy = \begin{cases} 0, & \ell = 1, \ldots, r-1, \\ 1, & \ell = 0; \end{cases}$

(iii) $\int_{-\tau}^{\tau} k(y) y^{r} dy \neq 0$;

where $r$ is a positive integer. Then we define

$$f_{in}(x_i) = \frac{1}{nh_n} \sum_{j=1}^{n} k\left(\frac{x_i - x_{ij}}{h_n}\right) \tag{5.2.5}$$

where $h_n$ is a sequence of positive numbers such that $h_n \to 0$ as $n \to \infty$. Recently, these kernel functions have been used by Karunamuni (1996) to consider empirical Bayes testing problems.

Utilizing $w_{in}(x_i)$ and $f_{in}(x_i)$, $i = 1, 2$, we propose the empirical Bayes estimators $d_{in}(x_i)$ for truncation parameters $\theta_i$, $i = 1, 2$, respectively, as follows:

$$d_{1n}(x_1) = [x_1 - w_{1n}(x_1)/f_{1n}(x_1)]_{h_n^{-1}} \tag{5.2.6}$$

and

$$d_{2n}(x_2) = [x_2 + w_{2n}(x_2)/f_{2n}(x_2)]_{h_n^{-1}} \tag{5.2.7}$$

where $[u]_c$ is defined by

$$[u]_c = \begin{cases} -c, & \text{if} \quad u < -c \\ u, & \text{if} \quad |u| \leq c \\ c, & \text{if} \quad u > c. \end{cases} \tag{5.2.8}$$

## 5.3 Asymptotic Optimality of Empirical Bayes Estimators

In this section, we study the asymptotic optimality property of the proposed empirical Bayes estimators in the last section. Under the squared error loss, for each $i = 1, 2$, the Bayes risk of the proposed empirical Bayes estimators $d_{in}(x_i)$ and the Bayes estimators $d_{G_i}(x_i)$ are, respectively,

$$r_i(d_{in}, G_i) = E(d_{in}(x_i) - \theta_i)^2 \tag{5.3.1}$$

and

$$r_i(d_{G_i}, G_i) = E(d_{G_i}(x_i) - \theta_i)^2 = r_i(G_i). \tag{5.3.2}$$

It is well known that

$$\begin{aligned} E_n\{r_i(d_{in}, G_i)\} - r_i(G_i) &= \int f_i(x_i) E_n(d_{in}(x_i) - d_{G_i}(x_i))^2 dx_i \\ &= E\, E_n(d_{in}(x_i) - d_{G_i}(x_i))^2. \end{aligned} \tag{5.3.3}$$

where $E_{in}$ indicates the expectation with respect to $(x_{i1}, \ldots, x_{in})$, $i = 1, 2$. As in many papers on empirical Bayes inference, we use the nonnegative difference $E_n\{r_i(d_{in}, G_i)\} - r_i(G_i)$ as the measure to evaluate the performance of the proposed empirical Bayes estimator $d_{in}(x_i)$.

**Definition 5.3.1** For $i = 1, 2$, a sequence of empirical Bayes estimators $\{\varphi_{in}(x_i)\}$ is said to be asymptotically optimal at least of order $\alpha_{in}$ relative to the prior $G_i$ if $E_n\{r_i(\varphi_{in}, G_i)\} - r_i(G_i) \leq O(\alpha_{in})$ as $n \to \infty$, where $\{\alpha_{in}\}$ is a sequence of positive numbers such that $\lim_{n\to\infty} \alpha_{in} = 0$.

Before giving the main results on asymptotic optimality of the proposed empirical Bayes estimators, we first present some useful lemmas. In the following, $c_1$, $c_2$ and $c$ always stand for some positive constants and they may be different even with the same notations.

**Lemma 5.3.1** *Let $y, z \neq 0$ and $L \geq 0$ be real numbers, and $Y, Z$ be two real valued random variables. Then for any $0 < \tau \leq 2$,*

$$E\left\{\left(\left[\frac{y}{z} - \frac{Y}{Z}\right]_L\right)^\tau\right\} \leq 2|z|^{-\tau}\left\{E(|y - Y|^\tau) + \left(\left|\frac{y}{z}\right| + L\right)^\tau E(|z - Z|^\tau)\right\}. \quad (5.3.4)$$

**Proof.** This is Lemma 3.1 in Singh and Wei (1992).

**Lemma 5.3.2** *Let $w_{in}(x_i)$, $i = 1, 2$, be defined by (5.2.3) and (5.2.4), respectively. Then for any $0 < \delta \leq 2$,*

$$E_{1n}(|w_{1n}(x_1) - w_1(x_1)|^\delta) \leq (n^{-\delta/2})\left\{\int_a^{x_1} \left(\frac{h_1(x_1)}{h_1(t_1)}\right)^2 dF_1(t_1)\right\}^{\delta/2} \quad (5.3.5)$$

*and*

$$E_{2n}(|w_{2n}(x_2) - w_2(x_2)|^\delta) \leq (n^{-\delta/2})\left\{\int_{x_2}^b \left(\frac{h_2(x_2)}{h_2(t_2)}\right)^2 dF_2(t_2)\right\}^{\delta/2}. \quad (5.3.6)$$

**Proof.** From the definition of $w_{1n}(x_1)$, we know that $w_{1n}(x_1)$ is an unbiased estimator for $w_1(x_1)$. Then

$$\begin{aligned} E_{1n}(|w_{1n}(x_1) - w_1(x_1)|^2) &= \text{var}(w_{1n}(x_1)) \\ &= \frac{1}{n}\text{Var}\left\{\frac{h_1(x_1)}{h_1(t_1)} I_{(a,x_1)}(t_1)\right\} \\ &\leq \frac{1}{n}\int_a^{x_1}\left(\frac{h_1(x_1)}{h_1(t_1)}\right)^2 dF_1(t_1) \end{aligned}$$

and by Hölder's inequality, for any $0 < \delta \leq 2$,

$$\begin{aligned} E_{1n}(|w_{1n}(x_1) - w_1(x_1)|^\delta) &\leq \{E_{1n}(|w_{1n}(x_1) - w_1(x_1)|^2)\}^{\delta/2} \\ &\leq (n^{-\delta/2})\left\{\int_a^{x_1}\left(\frac{h_1(x_1)}{h_1(t_1)}\right)^2 dF_1(t_1)\right\}^{\delta/2}. \end{aligned}$$

Thus (5.3.5) is true. Similarly, we can show (5.3.6) is also true.

**Lemma 5.3.3** *For $i = 1, 2$, let $f_{in}(x_i)$ be defined by (5.2.5) with $h_n = n^{-1/(2r+1)}$. If the r-th derivative of $f_i(x_i)$ is finite, then for any $0 < \delta \leq 2$,*

$$E_{in}(|f_{in}(x_i) - f_i(x_i)|^\delta) \leq O(n^{-\frac{\delta r}{2r+1}})\left[\{f_i(x_i)\}^{\delta/2} + \{|f_i^{(r)}(x_i)|\}^\delta\right]. \quad (5.3.7)$$

**Proof.** When $h_n = n^{-1/(2r+1)}$, from Lemma 3.1 in Karunamuni (1996), we have

$$E_{in}(|f_{in}(x_i) - f_i(x_i)|^2) \leq O(n^{\frac{2r}{2r+1}})[f_i(x_i) + \{f_i^{(r)}(x_i)\}^2].$$

Then for any $0 < \delta \leq 2$, by Hölder's inequality

$$\begin{aligned} E_{in}(|f_{in}(x_i) - f_i(x_i)|^\delta) &\leq \{E_{in}(|f_{in}(x_i) - f_i(x_i)|^2\}^{\delta/2} \\ &\leq O(n^{-\frac{\delta r}{2r+1}})[\{f_i(x_i)\}^{\delta/2} + \{|f_i^{(r)}(x_i)|\}^\delta]. \end{aligned}$$

**Theorem 5.3.1** *Let $\{d_{1n}(x_1)\}$ be the sequence of empirical Bayes estimators for the truncation parameter $\theta_1$ defined by (5.2.6). If the r-th derivative of $f_1(x_1)$ is finite and for any $0 < \delta < 2$,*

*(a)* $\int |\theta_1|^{r\delta} dG_1(\theta_1) < \infty$;

(b) $\int_a^b \{f_1(x_1)\}^{1-\delta}\left\{\int_a^{x_1}\left(\frac{h_1(x_1)}{h_1(t_1)}\right)^2 dF_1(t_1)\right\}^{\delta/2} dx_1 < \infty$;

(c) $\int_a^b \{f_1(x_1)\}^{1-\delta}[\{f_1(x_1)\}^{\delta/2} + \{|f_1^{(r)}(x_1)|\}^{\delta}]dx_1 < \infty$:

*then with the choice of* $h_n = n^{-1/(2r+1)}$, *we have*

$$E_{1n}\{r_1(d_{1n}, G_1)\} - r_1(G_1) \leq O(n^{-\frac{\delta r-2}{2r+1}}). \tag{5.3.8}$$

**Proof.** From the definition of $d_{1n}(x_1)$, we have

$$\begin{aligned}
(d_{1n}(x_1) - d_{G_1}(x_1))^2 &\leq 2(d_{1n} - [d_{G_1}]_{h_n^{-1}})^2 + 2([d_{G_1}]_{h_n^{-1}} - d_{G_1})^2 \\
&\leq 2\{4h_n^{-2}I(|d_{G_1}| \geq h_n^{-1}) + 2(d_{1n} - [d_{G_1}]_{h_n^{-1}})^2 I(|d_{G_1}| < h_n^{-1})\} + 2([d_{G_1}]_{h_n^{-1}} - d_{G_1})^2 \\
&\leq c_1|d_{G_1}|^2 I(|d_{G_1}| \geq h_n^{-1}) + c_2(d_{1n} - [d_{G_1}]_{h_n^{-1}})^2 I(|d_{G_1}| < h_n^{-1}) \\
&\stackrel{d}{=} c_1\, A_{1n} + c_2\, B_{1n}.
\end{aligned}$$

By Hölder's inequality and Markov's inequality,

$$\begin{aligned}
E\,E_{1n}(A_{1n}) &= E\,E_{1n}\{|d_{G_1}|^2 I(|d_{G_1}| \geq h_n^{-1})\} \\
&\leq \{E\,E_{1n}(|d_{G_1}|^{2\beta})\}^{1/\beta}\{E\,E_{1n}(I|d_{G_1}| \geq h_n^{-1}))^{\alpha}\}^{1-1/\beta} (\beta > 1) \\
&\leq \{E\,E_{1n}(|d_{G_1}|^{2\beta})\}^{1/\beta}\{E\,E_{1n}(|d_{G_1}|^{\alpha})h_n^{\alpha}\}^{1-1/\beta}.
\end{aligned}$$

Let $\alpha = 2\beta = r\delta$: then by Jensen's inequality

$$\begin{aligned}
E\,E_{1n}(A_{1n}) &\leq E\,E_{1n}(|d_{G_1}|^{2\beta})h_n^{2(\beta-1)} \\
&\leq E(|\theta_1|^{2\beta})h_n^{2(\beta-1)} \\
&\leq O(n^{-\frac{\delta r-2}{2r+1}})
\end{aligned}$$

by the assumption (a) and the choice of $h_n = n^{-1/(2r+1)}$. Next, by Lemma 5.3.1, Lemma 5.3.2 and Lemma 5.3.3, we have

$$\begin{aligned}
E\,E_{1n}(B_{1n}) &= E\,E_{1n}\{(d_{1n} - [d_{G_1}]_{h_n^{-1}})^2 I(|d_{G_1}| < h_n^{-1})\} \\
&\leq E\,E_{1n}\left\{\left(\left[\frac{w_{1n}}{f_{1n}} - \frac{w_1}{f_1}\right]_{2h_n^{-1}}\right)^2 I(|d_{G_1}| < h_n^{-1})\right\}
\end{aligned}$$

$$
\begin{aligned}
&\le E\,E_{1n}\left\{(2h_n^{-1})^{2-\delta}\left(\left[\frac{w_{1n}}{f_{1n}}-\frac{w_1}{f_1}\right]_{2h_n^{-1}}\right)^{\delta} I(|d_{G_1}|<h_n^{-1})\right\}\\
&\le c_1 h_n^{\delta-2}E[|f_1|^{-\delta}\{E_{1n}(|w_{1n}-w_1|^{\delta})+c_2h_n^{-\delta}E_{1n}(|f_{1n}-f_1|^{\delta})\}]\\
&\le c_1 h_n^{\delta-2}n^{-\delta/2}E\left[f_1^{-\delta}\left\{\int_a^{x_1}\left(\frac{h_1(x_1)}{h_1(t_1)}\right)^2 dF_1(t_1)\right\}^{\delta/2}\right]\\
&\quad +c_2h_n^{-2}n^{-\delta r/(2r+1)}E[f_1^{-\delta}\{(f_1)^{\delta/2}+(|f_1^{(r)}|)^{\delta}\}]\\
&\le O(n^{-\frac{\delta r-2}{2r+1}})
\end{aligned}
$$

by the assumptions (b) and (c) and the choice of $h_n = n^{-1/(2r+1)}$. Therefore.

$$
\begin{aligned}
E_{1n}\{r_1(d_{1n}.G_1)\}-r_1(G_1) &= E_1E_n(d_{1n}(x_1)-d_{G_1}(x_1))^2\\
&\le O(n^{-\frac{\delta r-2}{2r+1}}).
\end{aligned}
$$

**Theorem 5.3.2** *Let $\{d_{2n}(x_2)\}$ be the sequence of empirical Bayes estimators for truncation parameter $\theta_2$ defined by (5.2.7). If the r-th derivatives of $f_2(x_2)$ is finite and for any $0<\delta<2$.*

*(a)* $\int |\theta_2|^{r\delta}dG_2(\theta_2)<\infty$;

*(b)* $\int_a^b\{f_2(x_2)\}^{1-\delta}\left\{\int_{x_2}^b\left(\frac{h_2(x_2)}{h_2(t_2)}\right)^2 dF_2(t_2)\right\}^{\delta/2}dx_2<\infty$;

*(c)* $\int_a^b\{f_2(x_2)\}^{1-\delta}\left[\{f_2(x_2)\}^{\delta/2}+\{|f_2(x_2)|\}^{\delta}\right]dx_2<\infty$;

*then with the choice of $h_n = n^{-1/(2r+1)}$, we have*

$$
E_{2n}\{r_2(d_{2n}.G_2)\}-r_2(G_2)\le O(n^{-\frac{\delta r-2}{2r+1}}). \tag{5.3.9}
$$

**Proof.** It is similar to the proof of Theorem 5.3.1.

## 5.4 Some Examples

Finally, we discuss some important examples to illustrate the results of Theorem 5.3.1 and Theorem 5.3.2 in Section 5.3.

**Example 5.4.1** Consider the truncated exponential distribution

$$f_1(x_1|\theta_1) = \frac{1}{\sigma}\exp\left\{-\frac{1}{\sigma}(x_1-\theta_1)\right\}, \qquad 0 < \theta_1 \le x_1 < \infty. \tag{5.4.1}$$

Let $G_1(\theta_1)$ be a prior distribution with density function

$$g_1(\theta_1) = \left(1-\frac{\lambda}{\sigma}\right)\lambda^2\theta_1\epsilon^{-\lambda\theta_1} + \left(\frac{\lambda}{\sigma}\right)\lambda\epsilon^{-\lambda\theta_1}, \qquad \theta_1 > 0. \tag{5.4.2}$$

Then, we obtain

$$\begin{aligned} f_1(x_1) &= \int_0^{x_1} f_1(x_1|\theta_1)dG_1(\theta_1) \\ &= \lambda^2 x_1\epsilon^{-\lambda x_1}, \qquad x_1 > 0. \end{aligned}$$

$$|f_1^{(r)}(x_1)| \le c_1\,\lambda\epsilon^{-\lambda x_1} + c_2\,\lambda^2 x_1\epsilon^{-\lambda x_1}$$

and

$$\begin{aligned} \int_0^{x_1}\epsilon^{-\frac{2}{\sigma}(x_1-t_1)}f_1(t_1)dt_1 &= \int_0^{x_1}\epsilon^{-\frac{2}{\sigma}x_1}\lambda^2 t_1 e^{(\frac{2}{\sigma}-\lambda)t_1}dt_1 \\ &\le c\,f_1(x_1) \end{aligned}$$

where $c_1$, $c_2$ and $c$ may depend on $\lambda$ and $r$. Thus, for any $0 < \delta < 2$,

$$\begin{aligned} &\int_0^\infty \{f_1(x_1)\}^{1-\delta}\left\{\int_0^{x_1}\left(\frac{h_1(x_1)}{h_1(t_1)}\right)^2 dF_1(t_1)\right\}^{\delta/2}dx_1 \\ &\quad\le c\int_0^\infty \{f_1(x_1)\}^{1-\delta/2}dx_1 < \infty \\ &\int_0^\infty \{f_1(x_1)\}^{1-\delta}[\{f_1(x_1)\}^{\delta/2} + \{|f_1^{(r)}(x_1)\}^{\delta}]dx_1 \\ &\quad\le \int_0^\infty [c\{f_1(x_1)\}^{1-\delta/2} + c_1\{f_1(x_1)\}^{1-\delta}\{\lambda\,\epsilon^{-\lambda x_1}\}^{\delta} + c_2\{f_1(x_1)\}]dx_1 < \infty \end{aligned}$$

and

$$\int_0^\infty \theta_1^{r\delta}g_1(\theta_1)d\theta_1 < \infty.$$

Therefore, conditions (a), (b) and (c) in Theorem 5.3.1 are satisfied for the prior distribution (5.4.2) for any $\delta$ arbitrarily close to 2 and any arbitrarily large $r$. Then the convergence rate can be arbitrarily close to $O(n^{-1})$ in this case.

**Example 5.4.2** Consider the power function distribution

$$f_2(x_2|\theta_2) = \alpha x_2^{\alpha-1}/\theta_2^{\alpha}, \qquad 0 < x_2 \leq \theta_2 < \infty, \alpha \geq 1. \tag{5.4.3}$$

Let $G_2(\theta_2)$ be a prior distribution with density function

$$g_2(\theta_2) = \left(1 - \frac{1}{\alpha}\right) \lambda e^{-\lambda\theta_2} + \left(\frac{1}{\alpha}\right) \lambda^2 \theta_2 e^{-\lambda\theta_2}, \quad \theta_2 > 0, \ \alpha \geq 1. \tag{5.4.4}$$

Then we obtain

$$\begin{aligned} f_2(x_2) &= \int_{x_2}^{\infty} f_2(x_2|\theta_2) dG_2(\theta_2) \\ &= \lambda e^{-\lambda x_2}, \end{aligned}$$

$$|f_2^{(r)}(x_2)| \leq \lambda^r e^{-\lambda x_2} = c_1 f_2(x_2)$$

and

$$\begin{aligned} \int_{x_2}^{\infty} \left(\frac{x_2^{\alpha-1}}{t_2^{\alpha-1}}\right)^2 f_2(t_2) dt_2 &= \int_{x_2}^{\infty} x_2^{2(\alpha-1)} t_2^{-2(\alpha-1)} \lambda e^{-\lambda t_2} dt_2 \\ &\leq c_2 f_2(x_2) \end{aligned}$$

where $c_1$, $c_2$ depend on $\lambda$. Thus, for any $0 < \delta < 2$,

$$\int_0^{\infty} \{f_2(x_2)\}^{1-\delta} \left\{\int_{x_2}^{\infty} \left(\frac{h_2(x_2)}{h_2(t_2)}\right)^2 dF_2(t_2)\right\}^{\delta/2} dx_2$$
$$\leq c_1 \int_0^{\infty} \{f_2(x_2)\}^{1-\delta/2} dx_2 < \infty$$

$$\int_0^{\infty} \{f_2(x_2)\}^{1-\delta} [\{f_2(x_2)\}^{\delta/2} + \{|f_2^{(r)}(x_2)|\}^{\delta}] dx_2$$
$$\leq \int_0^{\infty} [\{f_2(x_2)\}^{1-\delta/2} + c_2\{f_2(x_2)\}] dx_2 < \infty$$

and

$$\int_0^{\infty} \theta_2^{r\delta} g_2(\theta_2) d\theta_2 < \infty.$$

Therefore, conditions (a), (b) and (c) in Theorem 5.3.2 are satisfied for the prior distribution (5.4.4) for any $\delta$ arbitrarily close to 2 and any arbitrarily large $r$. Then the convergence rate in this case can also be arbitrarily close to $O(n^{-1})$.

# Chapter 6

# Empirical Bayes Estimation and Selection for Exponential Populations With Location Parameters

## 6.1 Introduction

The empirical Bayes approach, formulated by Robbins (1955), is appropriate when one is confronted repeatedly and independently with the same decision problem. This approach has been used extensively for various statistical problems by many authors including Robbins (1963, 1964) Johns and Van Ryzin (1971, 1972), Singh (1979), Gupta and Liang (1986, 1988), and Gupta, Liang and Rau (1994).

The exponential distribution with location (or threshold) parameter arises in many areas of applications including reliability and life-testing, survival analysis, and engineering problems; for example, see Balakrishnan and Basu (1995). In the literatures, the location or the threshold parameter is interpreted to be the guaranteed life-time. For this distributional model, Singh and Prasad (1989) and

Prasad and Singh (1990) have discussed the empirical Bayes estimation under squared error loss. They have also discussed some asymptotic properties of their empirical Bayes estimators under the assumption of the class of all prior distributions having support in a compact interval of the real line. Lin and Leu (1994) have studied the problem of selecting the best population from $k$ independent exponential populations with different location parameters through the empirical Bayes approach; however, they have not examined the convergence rates of their empirical Bayes selection rule.

In this chapter, we consider the problems of the empirical Bayes estimation for the location parameters without the assumption of the compact support for the prior distributions and the empirical Bayes selection of the best of $k$ exponential populations with different location parameters. In Section 6.2, we formulate the selection problem and derive the Bayes estimators for the location parameters and the Bayes selection rule. In Section 6.3, we construct the empirical Bayes estimators for the location parameters and the empirical Bayes selection rule. Finally, we discuss the asymptotic optimality properties of these empirical Bayes estimators and the empirical Bayes selection rule in Sections 6.4 and 6.5, respectively, and give an example to illustrate the obtained results in Section 6.6.

## 6.2 Bayes Estimators and Bayes Selection Rule

Consider $k$ independent populations $\pi_1, \ldots, \pi_k$, where an observation $x_i$ from $\pi_i$ has an exponential distribution with location parameter $\theta_i$ and scale parameter $\sigma$ as follows:

$$f_i(x_i|\theta_i) = \frac{1}{\sigma} \exp\left\{-\frac{x_i - \theta_i}{\sigma}\right\}, \qquad x_i > \theta_i,\ \theta_i, \sigma > 0 . \qquad (6.2.1)$$

The density function (6.2.1) provides a model for life length data when we assume a minimum guaranteed life-time $\theta_i$, which is here a location (or threshold) parameter. It is assumed that all the $k$ populations have a common known scale parameter $\sigma$. The $\theta_i$'s are unknown and let $\theta_{[1]} \leq \cdots \leq \theta_{[k]}$ be the ordered parameters of $\theta_1, \ldots, \theta_k$. A population $\pi_i$ with $\theta_i = \theta_{[k]}$ is considered as the best population. Our interest is to select the population $\pi_i$ associated with the largest guaranteed life-time $\theta_i = \theta_{[k]}$. Note that the best population also has the largest mean life-time $\theta_{[k]} + \sigma$.

Let $\Omega = \{\boldsymbol{\theta} \mid \boldsymbol{\theta} = (\theta_1, \ldots, \theta_k),\ \theta_i > 0,\ i = 1, \ldots, k\}$ be the parameter space. It is assumed that the parameter $\boldsymbol{\theta}$ has a prior distribution $G(\boldsymbol{\theta})$ with a joint distribution function $G(\boldsymbol{\theta}) = \prod_{i=1}^{k} G_i(\theta_i)$. Note that $\theta_i$'s are assumed to be independently distributed.

Let $A = \{i \mid i = 1, \ldots, k\}$ be the action space; when action $i$ is taken, it means that population $\pi_i$ is selected as the best population. For the parameter $\boldsymbol{\theta}$ and action $i$, the loss function is defined by

$$L(\boldsymbol{\theta}, i) = \theta_{[k]} - \theta_i, \tag{6.2.2}$$

the difference between the life-times of the best population and the selected population. This loss function is very common in Bayes and empirical Bayes selection problems.

Let $\mathcal{X}$ be the sample space generated by $\boldsymbol{x} = (x_1, \ldots, x_k)$. A selection rule $\boldsymbol{\delta} = (d_1, \ldots, d_k)$ is a mapping from the sample space $\mathcal{X}$ to $[0, 1]^k$ such that for each $\boldsymbol{x} \in \mathcal{X}$, the function $\boldsymbol{\delta}(\boldsymbol{x}) = (d_1(\boldsymbol{x}), \ldots, d_k(\boldsymbol{x}))$ satisfies $0 \leq d_i(\boldsymbol{x}) \leq 1$, $i = 1, \ldots, k$, and $\sum_{i=1}^{k} d_i(\boldsymbol{x}) = 1$. Note that $d_i(\boldsymbol{x})$, $i = 1, \ldots, k$, is the probability of selecting the population $\pi_i$ as the best population when $\boldsymbol{x}$ is observed.

Let $D$ be the set of all selection rules. For each $\boldsymbol{\delta} \in D$, let $r(G, \boldsymbol{\delta})$ denote the associated Bayes risk. Then from the loss function (6.2.2), the Bayes risk associated with $\boldsymbol{\delta}$ can be written as

$$\begin{aligned} r(G, \boldsymbol{\delta}) &= \int \int \sum_{i=1}^{k} L(\boldsymbol{\theta}, i) d_i(\boldsymbol{x}) f(\boldsymbol{x} \mid \boldsymbol{\theta}) dG(\boldsymbol{\theta}) d\boldsymbol{x} \\ &= C - \int \left\{ \sum_{i=1}^{k} d_i(\boldsymbol{x}) \varphi_i(x_i) \right\} f(\boldsymbol{x}) d\boldsymbol{x}, \end{aligned} \tag{6.2.3}$$

where

$$f(\boldsymbol{x}) = \prod_{i=1}^{k} f_i(x_i), \quad f_i(x_i) = \int_0^{x_i} f(x_i \mid \theta_i) dG_i(\theta_i), \tag{6.2.4}$$

$$C = \int\int \theta_{[k]} f(\boldsymbol{x} \mid \boldsymbol{\theta}) dG(\boldsymbol{\theta}) d\boldsymbol{x} = \int \theta_{[k]} dG(\boldsymbol{\theta}), \tag{6.2.5}$$

and

$$\varphi_i(x_i) = E(\theta_i \mid x_i) = \frac{\int_0^{x_i} \theta_i f_i(x_i \mid \theta_i) dG_i(\theta_i)}{f_i(x_i)}. \tag{6.2.6}$$

Note that $\varphi_i(x_i)$, $i = 1, \ldots, k$, is the Bayes estimator of location parameter $\theta_i$ under the squared error loss.

The minimum Bayes risk among the class $D$ is defined by $r(G) \equiv \inf_{\boldsymbol{\delta} \in D} r(G, \boldsymbol{\delta})$; any selection rule $\boldsymbol{\delta}$ such that $r(G, \boldsymbol{\delta}) = r(G)$ is called a Bayes selection rule. For each $\boldsymbol{x} \in \mathcal{X}$, let

$$A(\boldsymbol{x}) = \{i \mid \varphi_i(x_i) = \max_{1 \le j \le k} \varphi_j(x_j)\} \tag{6.2.7}$$

and

$$i^* \equiv i^*(\boldsymbol{x}) = \min\{i \mid i \in A(\boldsymbol{x})\}; \tag{6.2.8}$$

then from (6.2.3), a Bayes selection rule $\boldsymbol{\delta}_G = (d_{1G}, \ldots, d_{kG})$ is given by

$$d_{iG}(\boldsymbol{x}) = \begin{cases} 1 & \text{if } i = i^* \\ 0 & \text{otherwise}. \end{cases} \tag{6.2.9}$$

## 6.3 Empirical Bayes Estimators and Empirical Bayes Selection Rule

Since the Bayes estimators $\varphi_i(x_i)$, $i = 1, \ldots, k$, and Bayes selection rule $\boldsymbol{\delta}_G$ are both dependent on the prior distributions $G_i(\theta_i)$, $i = 1, \ldots, k$, which may not be known, it is impossible to apply the Bayes estimators and Bayes selection rule in practice. Hence, we adopt now the empirical Bayes approach.

Let $x_{ij}$ denote the observations from population $\pi_i$ at stage $j$, $j = 1, \ldots, n$. It is assumed that conditional on $\theta_{ij}$, $x_{ij}$ follows an exponential distribution with location parameter $\theta_{ij}$, as follows

$$x_{ij} \mid \theta_{ij} \sim \frac{1}{\sigma} \exp\left\{ -\frac{x_{ij} - \theta_{ij}}{\sigma} \right\}, \qquad x_{ij} > \theta_{ij},\ \theta_{ij}, \sigma > 0 \ . \tag{6.3.1}$$

Denote $\boldsymbol{\theta}_j = (\theta_{1j}, \ldots, \theta_{kj})$ and assume that $\boldsymbol{\theta}_j$, $j = 1, \ldots, n$, are i.i.d. with the prior distribution $G(\boldsymbol{\theta}) = \prod_{i=1}^{k} G_i(\theta_i)$. Let $\boldsymbol{x}_j = (x_{1j}, \ldots, x_{kj})$ denote the observations at the $j$th stage, $j = 1, \ldots, n$. We also let $\boldsymbol{x}_{n+1} = \boldsymbol{x} = (x_1, \ldots, x_k)$ denote the observations at the present stage.

Under this statistical framework, for each $i = 1, \ldots, k$, by a method similar to the one used in Lemma 4.1 of Fox (1978) we can derive

$$\varphi_i(x_i) = E(\theta_i | x_i) = x_i - \psi_i(x_i), \tag{6.3.2}$$

where

$$\psi_i(x_i) = \frac{\int_0^{x_i} e^{-\frac{1}{\sigma}(x_i - t_i)} dF_i(t_i)}{f_i(x_i)} \stackrel{d}{=} \frac{w_i(x_i)}{f_i(x_i)} \tag{6.3.3}$$

with $f_i(x_i)$ being the density function of $x_i$ and $F_i(x_i)$ the corresponding cumulative distribution function. Note that under the statistical model (6.2.1), $0 \leq \varphi_i(x_i) \leq x_i$, then we have $0 \leq \psi_i(x_i) \leq x_i$.

For each $i = 1, \ldots, k$, based on the past data $x_{i1}, \ldots, x_{in}$, we define

$$w_{in}(x_i) = \frac{1}{n} \sum_{j=1}^{n} e^{-\frac{1}{\sigma}(x_i - x_{ij})} I_{(0, x_i)}(x_{ij}) \tag{6.3.4}$$

as the estimator of $w_i(x_i)$. In order to estimate $f_i(x_i)$, we employ kernel functions used by Johns and Van Ryzin (1972) and Singh (1977, 1979). Let $k_r$ be the class of all Borel-measurable real-valued bounded functions $k(y)$ vanishing off $(0, 1)$ such that

$$\int k(y)dy = 1, \ \int y^{\ell} k(y)dy = 0 \text{ for } \ell = 1, \ldots, r-1, \tag{6.3.5}$$

where $r$ is an arbitrary but fixed positive integer. Define

$$f_{in}(x_i) = \frac{1}{nh_n}\sum_{j=1}^{n} k\left(\frac{x_{ij} - x_i}{h_n}\right), \tag{6.3.6}$$

where $h_n$ is a positive function of $n$ such that $h_n \to 0$ and $nh_n \to \infty$ as $n \to \infty$.

Utilizing $w_{in}(x_i)$ and $f_{in}(x_i)$, we propose the empirical Bayes estimators $\varphi_{in}(x_i)$ for location parameters $\theta_i$, $i = 1, 2, \ldots, k$, defined by

$$\begin{aligned} \varphi_{in}(x_i) &= x_i - 0 \vee \left(\frac{w_{in}(x_i)}{f_{in}(x_i)}\right) \wedge x_i \\ &\stackrel{d}{=} x_i - \psi_{in}(x_i), \end{aligned} \tag{6.3.7}$$

where $a \vee b = \max(a, b)$ and $a \wedge b = \min(a, b)$.

Next, we propose an empirical Bayes selection rule $\boldsymbol{\delta}_n^* = (d_{1n}^*, \ldots, d_{kn}^*)$ for the selection problem under study as follows:

For each $\boldsymbol{x} \in \mathcal{X}$, let

$$A_n^*(\boldsymbol{x}) = \{i \mid \varphi_{in}(x_i) = \max_{1\le j\le k} \varphi_{jn}(x_j)\}\ , \tag{6.3.8}$$

where $\varphi_{in}(x_i)$ is given by (6.3.7), and

$$i_n^* \equiv i_n^*(\boldsymbol{x}) = \min\{i \mid i \in A_n^*(\boldsymbol{x})\}; \tag{6.3.9}$$

then define the empirical Bayes rule as

$$d_{in}^*(\boldsymbol{x}) = \begin{cases} 1 & \text{if } i = i_n^* \\ 0 & \text{otherwise.} \end{cases} \tag{6.3.10}$$

## 6.4 Asymptotic Optimality of the Empirical Bayes Estimators

Under the squared error loss function, for each $i = 1, \ldots, k$, the Bayes risk of the proposed empirical Bayes estimators $\varphi_{in}(x_i)$ and the Bayes estimators $\varphi_i(x_i)$ are, respectively,

$$R_i(G_i, \varphi_{in}(x_i)) = E(\theta_i - \varphi_{in}(x_i))^2, \tag{6.4.1}$$

and

$$R_i(G_i) = R_i(G_i, \varphi(x_i)) = E(\theta_i - \varphi_i(x_i))^2 . \tag{6.4.2}$$

Since $\varphi_i(x_i)$ is the Bayes estimator of $\theta_i$, obviously,

$$E_{in}\{R_i(G_i, \varphi_{in}(x_i))\} - R_i(G_i) \geq 0 , \tag{6.4.3}$$

where $E_{in}$ denotes the expectation with respect to $(x_{i1}, \ldots, x_{in})$. It can be shown that

$$\begin{aligned} E_{in}\{R_i(G_i, \varphi_{in}(x_i))\} - R_i(G_i) &= \int_{x_i} f_i(x_i) E_n(\varphi_{in}(x_i) - \varphi_i(x_i))^2 dx_i \\ &= \tilde{E}_{in}(\varphi_{in}(x_i) - \varphi_i(x_i))^2 \end{aligned} \tag{6.4.4}$$

where $\tilde{E}_{in}$ denotes the expectation with respect to $(x_i, x_{i1}, \ldots, x_{in})$.

**Definition 6.4.1** A sequence of empirical Bayes estimators $\{\varphi_{in}\}$ is said to be asymptotically optimal at least of order $\alpha_n$ relative to the prior $G_i$ if

$$E_{in}\{R_i(G_i, \varphi_{in}(x_i))\} - R_i(G_i) \leq O(\alpha_n) \qquad \text{as } n \to \infty ,$$

where $\{\alpha_n\}$ is a sequence of positive numbers satisfying $\lim_{n\to\infty} \alpha_n = 0$.

In order to investigate the asymptotic optimality of the proposed empirical Bayes estimators $\varphi_{in}(x_i)$, we present some useful lemmas.

**Lemma 6.4.1** *Let $y$, $z \neq 0$ and $L \geq 0$ be real numbers, and $Y$, $Z$ be two real valued random variables; then for any $0 < \tau \leq 2$,*

$$E\left(\left|\frac{y}{z} - \frac{Y}{Z}\right| \wedge L\right)^{\tau} \leq 2|z|^{-\tau}\left\{E|y - Y|^{\tau} + \left(\left|\frac{y}{z}\right| + L\right)^{\tau} E|z - Z|^{\tau}\right\} . \tag{6.4.5}$$

**Proof.** This is Lemma 3.1 in Singh and Wei (1992).

**Lemma 6.4.2** *For $i = 1, 2, \ldots, k$, let $f_{in}(x_i)$ be defined by (6.3.6); if for $r \geq 1$, the $r$-th derivative of $f_i(x_i)$ exists, then for any $\varepsilon > 0$,*

*(a)*

$$E_{in}(|f_{in}(x_i) - f_i(x_i)|^2) \le O(h_n^{2r})[f_{i\varepsilon}^{(r)}(x_i)]^2 + O(n^{-1}h_n^{-1})f_{i\varepsilon}(x_i) \quad (6.4.6)$$

*where* $f_{i\varepsilon}(x_i) = \sup_{0\le u_i \le \varepsilon} f(x_i + u_i)$, $f_{i\varepsilon}^{(r)}(x_i) = \sup_{0 \le u_i \le \varepsilon} |f^{(r)}(x_i + u_i)|$;

*(b) for any* $0 < \delta \le 2$, *when* $h_n = n^{-1/(2r+1)}$,

$$E_{in}(|f_{in}(x_i) - f_i(x_i)|^\delta) \le O\left(n^{-\frac{\delta r}{2r+1}}\right)\{[f_{i\varepsilon}(x_i)]^{\delta/2} + [f_{i\varepsilon}^{(r)}(x_i)]^\delta\}. \quad (6.4.7)$$

**Proof.** (a) is easily proved by Theorem 3.3 in Singh (1977). If $h_n = n^{-1/(2r+1)}$, from (a), we have

$$E_{in}(|f_{in}(x_i) - f_i(x_i)|^2) \le O\left(n^{-\frac{2r}{2r+1}}\right)\{f_{i\varepsilon}(x_i) + [f_{i\varepsilon}^{(r)}(x_i)]^2\};$$

then for any $0 < \delta \le 2$, by Hölder's inequality

$$\begin{aligned} E_{in}(|f_{in}(x_i) - f_i(x_i)|^\delta) &\le \left[E_{in}(|f_{in}(x_i) - f_i(x_i)|^2)\right]^{\delta/2} \\ &\le cn^{-\frac{\delta r}{2r+1}}\left\{[f_{i\varepsilon}(x_i)]^{\delta/2} + [f_{i\varepsilon}^{(r)}(x_i)]^\delta\right\}. \end{aligned}$$

This proves part (b).

The following theorem is one of the main results in this chapter concerning the convergence rate of the empirical Bayes estimators. In the rest of this chapter, $c_1$, $c_2$, $c_3$ and $c$ always stand for some positive constants, and they may be different even with the same notation.

**Theorem 6.4.1** *For* $i = 1, \dots, k$, *let* $\{\varphi_{in}(x_i)\}$ *be the sequence of empirical Bayes estimators defined by (6.3.7); if for* $r \ge 1$, *the r-th derivative of* $f_i(x_i)$ *exists and if for* $\varepsilon > 0$, $0 < \delta < 2$,

(a) $$\int_0^\infty x_i^{2-\delta}[f_i(x_i)]^{1-\delta}\left[\int_0^{x_i} e^{-\frac{2}{\sigma}(x_i - t_i)}dF_i(t_i)\right]^{\delta/2} dx_i < \infty;$$

(b) $$\int_0^\infty x_i^2[f_i(x_i)]^{1-\delta}\left\{[f_{i\varepsilon}(x_i)]^{\delta/2} + [f_{i\varepsilon}^{(r)}(x_i)]^\delta\right\} dx_i < \infty;$$

*then with the choice of* $h_n = n^{-1/(2r+1)}$, *we have*

$$E_{in}\{R_i(G_i, \varphi_{in}(X_i))\} - R_i(G_i) \le O\left(n^{-\frac{\delta r}{2r+1}}\right). \quad (6.4.8)$$

**Proof.** For each $i = 1, \ldots, k$, from the definition of $w_{in}(x_i)$, we know that $w_{in}(x_i)$ is an unbiased estimator of $w_i(x_i)$; then we have

$$\begin{aligned} E_{in}(|w_{in}(x_i) - w_i(x_i)|^2) &= \mathrm{Var}(x_{in}) = \frac{1}{n}\,\mathrm{Var}\left\{e^{-\frac{1}{\sigma}(x_i-t_i)} I_{(0,x_i)}(t_i)\right\} \\ &\leq \frac{1}{n}\,E\left(e^{-\frac{1}{\sigma}(x_i-t_i)} I_{(0,x_i)}(t_i)\right)^2 \\ &= \frac{1}{n}\int_0^{x_i} e^{-\frac{2}{\sigma}(x_i-t_i)} dF_i(t_i) \end{aligned}$$

and by Hölder's inequality,

$$\begin{aligned} E_{in}(|w_{in}(x_i) - w_i(x_i)|^{\delta}) &\leq [E_{in}(|w_{in}(x_i) - w_i(x_i)|^2)]^{\delta/2} \\ &= (n^{-\delta/2})\left[\int_0^{x_i} e^{-\frac{2}{\sigma}(x_i-t_i)} dF_i(t_i)\right]^{\delta/2}. \end{aligned}$$

Now, by Lemma 6.4.1 and Lemma 6.4.2,

$$\begin{aligned} & E_{in}\{R_i(G_i, \varphi_{in}(x_i))\} - R_i(G_i) = \tilde{E}_{in}(\varphi_{in}(x_i) - \varphi_i(x_i))^2 \\ &= \tilde{E}_{in}(\psi_{in}(x_i) - \psi_i(x_i))^2 \\ &\leq \tilde{E}_{in}\left(\left|\frac{w_{in}(x_i)}{f_{in}(x_i)} - \frac{w_i(x_i)}{f_i(x_i)}\right| \wedge x_i\right)^2 \\ &\leq \tilde{E}_{in}\left[x_i^{2-\delta}\left(\left|\frac{w_{in}(x_i)}{f_{in}(x_i)} - \frac{w_i(x_i)}{f_i(x_i)}\right| \wedge x_i\right)^{\delta}\right] \\ &\leq E\left\{2x_i^{2-\delta}|f_i(x_i)|^{-\delta}[E_{in}(|w_{in} - w_i|^{\delta}) + cx_i^{\delta}E_{in}(|f_{in} - f_i|^{\delta})]\right\} \\ &\leq c_1 E\left[x_i^{2-\delta} f_i^{-\delta}\left(\int_0^{x_i} e^{-\frac{2}{\sigma}(x_i-t_i)} dF_i(t_i)\right)^{\delta/2}\right]\left(n^{-\frac{\delta}{2}}\right) \\ &\quad + c_2 E\left\{x_i^2 f_i^{-\delta}[(f_{ie}(x_i))^{\delta/2} + (f_{ie}^{(r)}(x_i))^{\delta}]\right\}\left(n^{-\frac{\delta r}{2r+1}}\right) \\ &\leq O\left(n^{-\frac{\delta r}{2r+1}}\right) \end{aligned}$$

by assumptions (a) and (b). This proves the theorem.

## 6.5 Asymptotic Optimality of the Empirical Bayes Selection Rule

Now we investigate the convergence rate of the proposed empirical Bayes selection rule $\{\boldsymbol{\delta}_n^*\}$. By the formula (6.2.3), the Bayes risk associated with the selection rule $\{\boldsymbol{\delta}_n^*\}$ and the Bayes selection rule $\boldsymbol{\delta}_G$ are given by

$$r(G, \boldsymbol{\delta}_n^*) = c - \int \left\{ \sum_{i=1}^{k} d_{in}^*(\boldsymbol{x}) \varphi_i(x_i) \right\} f(\boldsymbol{x}) d\boldsymbol{x} \tag{6.5.1}$$

and

$$r(G) = r(G, \boldsymbol{\delta}_G) = c - \int \left\{ \sum_{i=1}^{k} d_{iG}(\boldsymbol{x}) \varphi_i(x_i) \right\} f(\boldsymbol{x}) d\boldsymbol{x} \ , \tag{6.5.2}$$

respectively. It is obvious that

$$E_n\{r(G, \boldsymbol{\delta}_n^*)\} - r(G) \geq 0 \tag{6.5.3}$$

because the Bayes rule $\boldsymbol{\delta}_G$ achieves the minimum Bayes risk $r(G)$ and the expectation $E_n$ is taken with respect to $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$.

From the following straightforward computation, we have

$$\begin{aligned}
0 &\leq E_n\{r(G, \boldsymbol{\delta}_n^*)\} - r(G) \\
&= \int \left\{ \sum_{i=1}^{k} (d_{iG}(\boldsymbol{x}) - E_n(d_{in}^*(\boldsymbol{x}))\varphi_i(x_i) \right\} f(\boldsymbol{x}) d\boldsymbol{x} \\
&= \int E_n\{d_{i^*G}(\boldsymbol{x})\varphi_{i^*}(x_{i^*}) - d_{i_n^* n}(\boldsymbol{x})\varphi_{i_n^* n}(x_{i_n^*})\} f(\boldsymbol{x}) d\boldsymbol{x} \\
&= \int E_n \left\{ \sum_{i=1}^{k} \sum_{j=1}^{k} \mathrm{I}_{(i^*=i,\, i_n^*=j)}(\varphi_i(x_i) - \varphi_j(x_j)) \right\} f(\boldsymbol{x}) d\boldsymbol{x} \\
&= \int \left\{ \sum_{i=1}^{k} \sum_{j=1}^{k} P\{i^* = i,\ i_n^* = j\}(\varphi_i(x_i) - \varphi_j(x_j)) \right\} f(\boldsymbol{x}) d\boldsymbol{x} \\
&\leq \sum_{i=1}^{k} \sum_{j=1}^{k} \int_{x_i} \int_{x_j} \left\{ \Pr\Big( |\varphi_{in}(x_i) - \varphi_i(x_i)| > \frac{1}{2}|\varphi_i(x_i) - \varphi_j(x_j)| \Big) \right. \\
&\quad \left. + \Pr\Big( |\varphi_{jn}(x_j) - \varphi_j(x_j)| > \frac{1}{2}|\varphi_i(x_i) - \varphi_j(x_j)| \Big) \right\} \\
&\quad \times |\varphi_i(x_i) - \varphi_j(x_j)| f(x_i) f(x_j)\, dx_i dx_j \ .
\end{aligned} \tag{6.5.4}$$

**Definition 6.5.1** A sequence of empirical Bayes selection rules $\{\boldsymbol{\delta}_n\}$ is said to be asymptotically optimal at least of order $\beta_n$ relative to the unknown prior distribution $G$ if

$$E_n\{r(G,\boldsymbol{\delta}_n)\} - r(G) \leq O(\beta_n) \text{ as } n \to \infty$$

where $\beta_n$ is a sequence of positive numbers such that $\lim_{n\to\infty} \beta_n = 0$.

The following theorem is another main result in this chapter; it establishes the convergence rate of the empirical Bayes selection rule $\{\boldsymbol{\delta}_n^*\}$.

**Theorem 6.5.1** *Let $\{\boldsymbol{\delta}_n^*\}$ be the sequence of empirical Bayes selection rules defined by (6.3.10); if for $r \geq 1$, the r-th derivative of $f_i(x_i)$ exists and, if for $\varepsilon > 0$ and $0 < \delta < 2$,*

$$(a) \qquad \int_0^\infty [f_i(x_i)]^{1-\delta} x_i^{2-\delta} \left[\int_0^{x_i} e^{-\frac{2}{\sigma}(x_i - t_i)} dF_i(t_i)\right]^{\delta/2} dx_i < \infty;$$

$$(b) \qquad \int_0^\infty [f_i(x_i)]^{1-\delta} x_i^2 \left\{[f_{i\varepsilon}(x_i)]^{\delta/2} + [f_{i\varepsilon}^{(r)}(x_i)]^\delta\right\} dx_i < \infty;$$

*then with the choice of $h_n = n^{-1/(2r+1)}$, we have*

$$E_n\{r(G,\boldsymbol{\delta}_n^*)\} - r(G) \leq O\left(n^{-\frac{\delta r}{2(2r+1)}}\right). \qquad (6.5.5)$$

**Proof.** Let

$$X_{ijn} = \left\{(x_i, x_j)\big| |\varphi_i(x_i) - \varphi_j(x_j)| \leq \varepsilon_n\right\}$$

where $\varepsilon_n > 0$ and $\varepsilon_n \to 0$ as $n \to \infty$. Then by Markov's inequality, and Theorem 6.4.1, we have

$$\begin{aligned}
&E_n\{r(G,\boldsymbol{\delta}_n^*)\} - r(G) \\
&\leq \sum_{i=1}^k \sum_{j=1}^k 2\varepsilon_n \int\int_{X_{ijn}} f_i(x_i) f_j(x_j) dx_i dx_j \\
&\quad + \sum_{i=1}^k \sum_{j=1}^k \int\int_{X_{ijn}^c} \Big\{\Pr(|\varphi_{in}(x_i) - \varphi_i(x_i)| > \frac{1}{2}|\varphi_i(x_i) - \varphi_j(x_j)|) \\
&\qquad + \Pr(|\varphi_{jn}(x_j) - \varphi_j(x_j)| > \frac{1}{2}|\varphi_i(x_i) - \varphi_j(x_j)|)\Big\}
\end{aligned}$$

$$
\begin{aligned}
&\times|\varphi_i(x_i)-\varphi_j(x_j)|f_i(x_i)f_j(x_j)dx_i\,dx_j\\
\leq\ & c_1\bar{\varepsilon}_n + c_2\sum_{i=1}^{k}\sum_{j=1}^{k}\int\int_{X^c_{ijn}}\Big(\frac{E_{in}|\varphi_{in}(x_i)-\varphi_i(x_i)|^2}{|\varphi_i(x_i)-\varphi_j(x_j)|}\\
&+\frac{E_{jn}|\varphi_{jn}(x_j)-\varphi_j(x_j)|^2}{|\varphi_i(x_i)-\varphi_j(x_j)|}\Big)f_i(x_i)f_j(x_j)dx_i\,dx_j\\
\leq\ & c_1\bar{\varepsilon}_n + c_2\sum_{i=1}^{k}\sum_{j=1}^{k}\int\int_{X^c_{ijn}}\frac{1}{\bar{\varepsilon}_n}\ E_{in}|\varphi_{in}(x_i)-\varphi_i(x_i)|^2 f(x_i)f(x_j)dx_idx_j\\
\leq\ & c_1\bar{\varepsilon}_n + c_2\,\frac{1}{\bar{\varepsilon}_n}\sum_{i=1}^{k}\tilde{E}_{in}(|\varphi_{in}(x_i)-\varphi_i(x_i)|^2)\\
\leq\ & c_1\bar{\varepsilon}_n + c_2\bar{\varepsilon}_n^{-1}\left(n^{-\frac{\delta r}{2r+1}}\right).
\end{aligned}
$$

Thus, letting $\bar{\varepsilon}_n = n^{-\frac{\delta r}{2(2r+1)}}$, we obtain

$$
E_n\{r(G,\boldsymbol{\delta}^*_n)\} - r(G) \leq O\left(n^{-\frac{\delta r}{2(2r+1)}}\right).
$$

## 6.6 An Example

Finally, we discuss an example to illustrate the results obtained in Theorems 6.4.1 and 6.5.1. Without loss of generality, we assume the common known scale parameter $\sigma = 1$ in (6.2.1), i.e., for $i = 1,\ldots,k$,

$$
f_i(x_i|\theta_i) = \exp\{-(x_i-\theta_i)\},\qquad x_i > \theta_i,\ \theta_i > 0. \tag{6.6.1}
$$

Let $G_i(\theta_i)$ be a prior distribution with density function

$$
g_i(\theta_i) = \left(1-\frac{1}{\tau}\right)\frac{\theta_i}{\tau^2}\,e^{-\frac{\theta_i}{\tau}} + \left(\frac{1}{\tau}\right)\frac{1}{\tau}\,e^{-\frac{\theta_i}{\tau}},\qquad \theta_i > 0,\ \tau > 1. \tag{6.6.2}
$$

With such a prior distribution on $\theta_i$, we have

$$
\begin{aligned}
f_i(x_i) &= \int_0^{x_i} f_i(x_i|\theta_i)dG_i(\theta_i)\\
&= e^{-x_i}\left(\int_0^{x_i}\left(1-\frac{1}{\tau}\right)\frac{\theta_i}{\tau^2}\,e^{-\frac{\theta_i}{\tau}}e^{\theta_i}d\theta_i + \int_0^{x_i}\frac{1}{\tau^2}e^{-\frac{\theta_i}{\tau}}e^{\theta_i}d\theta_i\right)\\
&= \frac{x_i}{\tau^2}\,e^{-\frac{x_i}{\tau}},\qquad x_i > 0,\ \tau > 1. \qquad (6.6.3)
\end{aligned}
$$

Then

$$\begin{aligned}\int_0^{x_i} e^{-2(x_i-t_i)} f_i(t_i)dt_i &= \int_0^{x_i} e^{-2x_i} \frac{t_i}{\tau^2} e^{(2-\frac{1}{\tau})t_i}dt_i \\ &\leq c \frac{x_i}{\tau^2} e^{-\frac{x_i}{\tau}} = cf_i(x_i)\end{aligned}$$

and

$$f_i^{(r)}(x_i) = c_1 \left(\frac{1}{\tau} e^{-\frac{x_i}{\tau}}\right) + c_2 \left(\frac{x_i}{\tau^2} e^{-\frac{x_i}{\tau}}\right)$$

where $c_1$, $c_2$ and $c$ are dependent on $r$ and $\tau$. Since $xe^{-\frac{x}{\tau}}$ is increasing when $0 < x < \tau$ and decreasing when $x \geq \tau$, then, for any $\varepsilon > 0$,

$$f_{i\varepsilon}(x_i) = \sup_{0\leq u_i \leq \varepsilon} f_i(x_i + u_i) \leq \begin{cases} f_i(x_i) & \text{if } x_i \geq \tau \\ f_i(\tau) & \text{if } x_i < \tau \end{cases}$$

and

$$f_{i\varepsilon}^{(r)}(x_i) = \sup_{0\leq u_i \leq \varepsilon} f_i^{(r)}(x_i + u_i) \leq \begin{cases} c_1 e^{-\frac{x_i}{\tau}} + c_2 x_i e^{-\frac{x_i}{\tau}} & \text{if } x_i \geq \tau \\ c_1 + c_2 f_i(\tau) & \text{if } x_i < \tau. \end{cases}$$

Thus, for any $0 < \delta < 2$,

$$\begin{aligned}&\int_0^\infty [f_i(x_i)]^{1-\delta} x_i^{2-\delta} \left[\int_0^{x_i} e^{-2(x_i-t_i)} dF_i(t_i)\right]^{\delta/2} dx_i \\ &\leq c\int_0^\infty x_i^{2-\delta}[f_i(x_i)]^{1-\delta/2}dx_i \\ &= c\int_0^\infty x_i^{3(2-\delta)/2} e^{-\frac{(2-\delta)x_i}{2\tau}} dx_i < \infty \qquad (6.6.4)\end{aligned}$$

and

$$\begin{aligned}&\int_0^\infty x_i^2[f_i(x_i)]^{1-\delta}\{[f_{i\varepsilon}(x_i)]^{\delta/2} + [f_{i\varepsilon}^{(r)}(x_i)]^{\delta}\}dx_i \\ \leq\ & c_1 \int_0^\tau x_i^2 \left(x_i^{1-\delta} e^{-\frac{(1-\delta)x_i}{\tau}}\right) dx_i \\ &+ c_2 \int_\tau^\infty x_i^2[f_i(x_i)]^{1-\delta}\{[f_i(x_i)]^{\delta/2} + c_3[f_i(x_i)]^{\delta}\}dx_i \\ =\ & c_1 \int_0^\tau x_i^{3-\delta} e^{-\frac{(1-\delta)x_i}{\tau}} dx_i + c_2 \int_\tau^\infty x_i^2[f_i(x_i)]^{1-\delta/2}dx_i \\ &+ c_3 \int_\tau^\infty x_i^2 f_i(x_i)dx_i < \infty. \qquad (6.6.5)\end{aligned}$$

Therefore, conditions (a) and (b) in Theorems 6.4.1 and 6.5.1 are satisfied for the prior distribution (6.6.2) for any $\delta$ arbitrarily close to 2, so that the convergence rates can be arbitrarily close to $O(n^{-1})$ and $O(n^{-1/2})$, respectively, if integer $r$ is sufficiently large.

# Chapter 7

# Empirical Bayes Estimation and Testing for a Location Parameter Family of Gamma Distributions

## 7.1 Introduction

The gamma distribution with location parameter is useful in many areas of application including survival analysis, life-testing, and reliability theory (in these cases, the location parameter is often referred to as the 'threshold parameter'); for example, see Balakrishnan and Cohen (1991), and Johnson, Kotz, and Balakrishnan (1994). For this distribution model, Fox (1978) studied empirical Bayes estimation of the location parameter under squared-error loss; however, Fox did not examine the convergence rate of that empirical Bayes estimator.

In this chapter, we consider the empirical Bayes estimation of the location parameter as well as the empirical Bayes two-action testing problem. In Section 7.2, we formulate the two problems and derive the Bayes estimator and the Bayes testing rule. In Section 7.3, we derive the empirical Bayes estimator of the location parameter and the empirical Bayes testing rule for the two-action problem. In Sections 7.4 and 7.5, we examine the asymptotic optimality properties of the proposed empirical Bayes estimator and the empirical Bayes testing rule, respectively.

## 7.2 Bayes Estimator and Bayes Testing Rule

Consider the family of gamma distributions (with location parameter $\theta$ and shape parameter $\alpha$) with conditional density

$$f(x \mid \theta) = \frac{1}{\Gamma(\alpha)} e^{-(x-\theta)} (x-\theta)^{\alpha-1}, \qquad x \geq \theta,\ \theta > 0, \tag{7.2.1}$$

where $\Gamma(\cdot)$ is the complete gamma function and $\alpha \geq 2$ is fixed. In life-testing situations, $\theta$ in (7.2.1) is interpreted as 'minimum guaranteed life-time'.

### 7.2.1 Bayes estimation

Under the squared-error loss, it is known that the Bayes estimator relative to the prior $G(\theta)$ is

$$\varphi_G(x) = \frac{\int \theta f(x \mid \theta)\, dG(\theta)}{\int f(x \mid \theta)\, dG(\theta)} . \tag{7.2.2}$$

For the gamma model in (7.2.1), Fox (1978) obtained the Bayes estimator of $\theta$ as

$$\varphi_G(x) = E(\theta \mid x) = x - v(x) . \tag{7.2.3}$$

where

$$v(x) = \frac{\alpha \int_0^x e^{-(x-t)}\, dF(t)}{f(x)} \stackrel{d}{=} \frac{w(x)}{f(x)} \tag{7.2.4}$$

with

$$f(x) = \int_0^x f(x \mid \theta)\, dG(\theta)$$

being the density function of $X$ and $F(x)$ the corresponding cumulative distribution function.

### 7.2.2 Bayes testing

Consider the problem of testing the hypothesis $H_0 : \theta \leq \theta_0$ vs. $H_1 : \theta > \theta_0$ with the linear loss function

$$\begin{aligned} L(a_i, \theta) &= b \max(0, \theta - \theta_0) \quad \text{if } i = 0 \\ &= b \max(0, \theta_0 - \theta) \quad \text{if } i = 1 \; . \end{aligned} \tag{7.2.5}$$

where $\theta_0$ is a given positive constant and $a_i$ is the action in favour of $H_i$, $i = 0, 1$. $L(a_i, \theta)$ denotes the loss when action $a_i$ is taken ($i = 0, 1$), and $b$ is a positive constant. Let

$$d(x) = \Pr\{\text{accepting } H_0 \mid x\} \tag{7.2.6}$$

be the decision rule for the two-action problem considered. Then the Bayes risk associated with $d(x)$ under prior $G(\theta)$ is given by [Johns and Van Ryzin (1971, 1972)]

$$r(G, d) = b \int \alpha(x) d(x) \, dx + C_G \; . \tag{7.2.7}$$

where

$$\begin{aligned} \alpha(x) &= \int \theta \, f(x \mid \theta) \, dG(\theta) - \theta_0 \, f(x) \\ &= (x - \theta_0) f(x) - \alpha \int_0^x \epsilon^{-(x-t)} \, dF(t) \\ &= (x - \theta_0) f(x) - w(x) \end{aligned} \tag{7.2.8}$$

and

$$C_G = \int L(a_1, \theta) \, dG(\theta) \; . \tag{7.2.9}$$

From (7.2.7), a Bayes testing rule $d_G(x)$ is then given by

$$\begin{aligned} d_G(x) &= 1 \quad \text{if } \alpha(x) \leq 0 \\ &= 0 \quad \text{if } \alpha(x) > 0 \; . \end{aligned} \tag{7.2.10}$$

## 7.3 Empirical Bayes Estimator and Empirical Bayes Testing Rule

Since the Bayes estimator and the Bayes testing rule presented in the last section are both dependent on the prior distribution $G(\theta)$ which may not be known, we adopt the empirical Bayes approach in this section.

Let $x_i$ and $\theta_i$ ($i = 1, 2, \ldots, n$) denote the observation and the location parameter at stage $i$, and assume that (conditional on $\theta_i$) $x_i$ follows a gamma distribution (with location parameter $\theta_i$ and shape parameter $\alpha$) with density

$$f(x_i \mid \theta_i) = \frac{1}{\Gamma(\alpha)} e^{-(x_i-\theta_i)}(x_i - \theta_i)^{\alpha-1}, \qquad x_i \geq \theta_i, \ \theta_i > 0. \tag{7.3.1}$$

We assume that $\theta_1, \theta_2, \ldots, \theta_n$ are i.i.d. with the unknown prior distribution $G(\theta)$ and denote $x_{n+1} = x$ for the observation at the present stage.

Based on the past data, *viz.*, $x_1, x_2, \ldots, x_n$, we define the estimator for the function $w(x)$ in (7.2.4) as

$$w_n(x) = \frac{\alpha}{n} \sum_{i=1}^{n} e^{-(x-x_i)} I_{(0,x)}(x_i). \tag{7.3.2}$$

Further, let $k_r$ be the class of all Borel measurable real-valued bounded functions vanishing off $(0, 1)$ such that

$$\int k(y)dy = 1, \ \int y^{\ell}k(y)dy = 0 \text{ for } \ell = 1, \ldots, r-1, \tag{7.3.3}$$

where $r$ is an arbitrary, but fixed, positive integer. Then, define the kernel estimator for the density function $f(x)$ as

$$f_n(x) = \frac{1}{nh_n} \sum_{i=1}^{n} k\left(\frac{x_i - x}{h_n}\right) \tag{7.3.4}$$

where $h_n$ is a positive function of $n$ such that $h_n \to 0$ and $nh_n \to \infty$ as $n \to \infty$. These kernel estimators have been used by Johns and Van Ryzin (1972) and Singh (1977, 1979).

### 7.3.1 Empirical Bayes estimator

Note that under the statistical model (7.2.1), $0 \le \phi_G(x) = E(\theta|x) \le x$; then, from (7.2.3), $0 \le \psi(x) \le x$. Utilizing $w_n(x)$ and $f_n(x)$ defined in (7.3.2) and (7.3.4), respectively, we propose the empirical Bayes estimator for the location parameter $\theta$ as [see Eqs. (7.2.3) and (7.2.4)]

$$\begin{aligned} \phi_n(x) &= x - 0 \vee \left( \frac{w_n(x)}{f_n(x)} \right) \wedge x \\ &\stackrel{d}{=} x - \psi_n(x), \end{aligned} \tag{7.3.5}$$

where $a \vee b = \max(a, b)$ and $a \wedge b = \min(a, b)$.

### 7.3.2 Empirical Bayes testing rule

Next, we propose an empirical Bayes testing rule [from Eqs. (7.2.8) and (7.2.10)] as follows.

Let

$$\alpha_n(x) = (x - \theta_0) f_n(x) - w_n(x) \ . \tag{7.3.6}$$

Then, the empirical Bayes testing rule is given by

$$\begin{aligned} d_n(x) &= 1 \quad \text{if } \alpha_n(x) \le 0 \\ &= 0 \quad \text{if } \alpha_n(x) > 0 \ . \end{aligned} \tag{7.3.7}$$

## 7.4 Asymptotic Optimality of the Empirical Bayes Estimator

Under the squared-error loss function, the Bayes risk of the empirical Bayes estimator $\phi_n(x)$ in (7.3.5) and the Bayes estimator $\varphi_G(x)$ in (7.2.3) are

$$R(G, \phi_n(x)) = E(\theta - \phi_n(x))^2 \tag{7.4.1}$$

and

$$R(G) = R(G, \phi_G(x)) = E(\theta - \phi_G(x))^2 . \tag{7.4.2}$$

Since $\phi_G(x)$ is the Bayes estimator of $\theta$, we have

$$E_n\{R(G, \phi_n(x))\} - R(G) \geq 0 , \tag{7.4.3}$$

where $E_n$ denotes the expectation with respect to $(x_1, \ldots, x_n)$. It is known that

$$\begin{aligned} E_n\{R(G, \phi_n(x))\} - R(G) &= \int f(x) E_n(\phi_n(x) - \phi_G(x))^2 dx \\ &= E_n^*(\phi_n(x) - \phi_G(x))^2 . \end{aligned} \tag{7.4.4}$$

where $E_n^*$ denotes the expectation with respect to $(x, x_1, \ldots, x_n)$.

**Definition 7.4.1** A sequence of empirical Bayes estimators $\{\psi_n\}$ is said to be asymptotically optimal at least of order $\alpha_n$ relative to the prior $G$ if

$$E_n\{R(G, \psi_n(x))\} - R(G) \leq O(\alpha_n) \qquad \text{as } n \to \infty , \tag{7.4.5}$$

where $\{\alpha_n\}$ is a sequence of positive numbers such that $\lim_{n\to\infty} \alpha_n = 0$.

In order to examine the asymptotic optimality of the empirical Bayes estimator $\phi_n(x)$ proposed in (7.3.5), we need the following lemmas.

**Lemma 7.4.1** *Let $y, z \neq 0$ and $L > 0$ be real numbers, and $Y$ and $Z$ be two real-valued random variables. Then, for any $0 < \tau \leq 2$*

$$E\left\{\left[\left(\frac{y}{z} - \frac{Y}{Z}\right) \wedge L\right]^\tau\right\} \leq 2|z|^{-\tau}\left[E\left(|y - Y|^\tau\right) + \left(\left|\frac{y}{z}\right| + L\right)^\tau E\left(|z - Z|^\tau\right)\right] \tag{7.4.6}$$

**Proof.** This is Lemma 3.1 in Singh and Wei (1992).

**Lemma 7.4.2** *(a) Let $f_n(x)$ be defined by (7.3.4) with kernel function $k \in k_{[\alpha-1]}$, where $\alpha$ is the shape parameter; if $h_n = n^{-1/(2[\alpha-1]+1)}$, then for any $0 < \delta \leq 2$,*

$$E(|f_n(x) - f(x)|^\delta) \leq O(n^{-\delta[\alpha-1]/(2[\alpha-1]+1)}). \tag{7.4.7}$$

*(b) Let $w_n(x)$ be defined by (7.3.2); then for any $0 < \delta \leq 2$,*

$$E(|w_n(x) - w(x)|^\delta) \leq O(n^{-\delta/2}). \tag{7.4.8}$$

**Proof.** (a) Since

$$f(x) = \int_0^x \frac{1}{\Gamma(\alpha)} (x-\theta)^{\alpha-1} e^{-(x-\theta)} dG(\theta)$$

and

$$f^{(r)}(x) = \sum_{i=0}^{r} (-1)^i \binom{r}{i} \int_0^x \frac{1}{\Gamma(\alpha)} (x-\theta)^{\alpha-1+i-r} e^{-(x-\theta)} dG(\theta),$$

it is easy to show that both $f(x)$ and $f^{([\alpha-1])}(x)$ are finite because the function $v(x) = x^a e^{-x}$, $x > 0$, $a > 0$, attains its maximum value at $x = a$. Then with $h_n = n^{-1/(2[\alpha-1]+1)}$, by Corollary 3.3.4 of Singh (1977), we obtain

$$E(|f_n(x) - f(x)|^2) \le O(n^{-2[\alpha-1]/(2[\alpha-1]+1)}).$$

Next, for any $0 < \delta \le 2$, by Hölder's inequality,

$$E(|f_n(x) - f(x)|^\delta) \le [E(|f_n(x) - f(x)|^2)]^{\delta/2} \le O(n^{-\delta[\alpha-1]/(2[\alpha-1]+1)}).$$

(b) Since $w_n(x)$ is an unbiased estimator of $w_n(x)$, then

$$E(|w_n(x) - w(x)|^2) = \text{Var}(w_n(x)) \le \frac{\alpha^2}{n}$$

and

$$E(|w_n(x) - w(x)|^\delta) \le O(n^{-\delta/2}).$$

The following theorem presents the convergence rate of the empirical Bayes estimator $\phi_n(x)$ proposed in (7.3.5). We shall use $c_1$, $c_2$ and $c$ to denote some positive constants which may be different with the same notation.

**Theorem 7.4.1** *Let $\{\phi_n(x)\}$ be the sequence of empirical Bayes estimators of $\theta$ proposed in (7.3.5); if for* $0 < \delta < 1$,

$$E\left(\theta^{3/(1-\delta)}\right) < \infty, \tag{7.4.9}$$

*then with the choice of $h_n = n^{-1/(2[\alpha-1]+1)}$, we have*

$$E_n\{R(G, \phi_n(x))\} - R(G) \le O(n^{-\delta[\alpha-1]/(2[\alpha-1]+1)}). \tag{7.4.10}$$

**Proof.** From the definition of $\psi_n(x)$, we have by Lemma 7.4.1 and Lemma 7.4.2,

$$
\begin{aligned}
E_n^*(|\psi_n(x) - \psi(x)|^2) &\le E_n^*\left[\left(\left|\frac{w_n(x)}{f_n(x)} - \frac{w(x)}{f(x)}\right| \wedge x\right)^2\right] \\
&\le E_n^*\left[x^{2-\delta}\left(\left|\frac{w_n(x)}{f_n(x)} - \frac{w(x)}{f(x)}\right| \wedge x\right)^{\delta}\right] \\
&\le E\{2x^{2-\delta}f^{-\delta}[E(|w_n(x) - w(x)|^{\delta}) \\
&\quad + \left(\frac{w(x)}{f(x)} + x\right)^{\delta} E(|f_n(x) - f(x)|^{\delta})]\} \\
&\le c_1 A_1(n^{-\delta/2}) + c_2 A_2(n^{-\delta[\alpha-1]/(2[\alpha-1]+1)}),
\end{aligned}
$$

where

$$
\begin{aligned}
A_1 &= E(x^{2-\delta}f^{-\delta}) = \int x^{2-\delta}\left(\int f(x|\theta)dG(\theta)\right)^{1-\delta} dx, \\
A_2 &= E(x^2 f^{-\delta}) = \int x^2\left(\int f(x|\theta)dG(\theta)\right)^{1-\delta} dx.
\end{aligned}
$$

To show that $A_2$ is finite, we use Hölder's inequality and observe that

$$
\begin{aligned}
&\int_0^1 (x^2)\left(\int_0^x f(x|\theta)dG(\theta)\right)^{1-\delta} dx \\
&\le \left[\int_0^1\left(\int_0^x f(x|\theta)dG(\theta)\right)dx\right]^{1-\delta}\left[\int_0^1 (x^{2/\delta})dx\right]^{\delta} \\
&\le c\left[\int_0^1\left(\int_\theta^1 f(x|\theta)dx\right)dG(\theta)\right]^{1-\delta} < \infty
\end{aligned}
$$

and

$$
\begin{aligned}
&\int_1^\infty (x^2)\left(\int_0^x f(x|\theta)dG(\theta)\right)^{1-\delta} dx \\
&= \int_1^\infty (x^{-1})\left[x^3\left(\int_0^x f(x|\theta)dG(\theta)\right)^{1-\delta}\right] dx \\
&\le \left[\int_1^\infty (x^{-1})^{1/\delta}dx\right]^{\delta}\left[\int_1^\infty x^{3/(1-\delta)}\left(\int_0^x f(x|\theta)dG(\theta)\right)dx\right]^{1-\delta} \\
&\le c\left(\int_0^\infty \int_\theta^\infty x^{3/(1-\delta)} f(x|\theta)dx\, dG(\theta)\right)^{1-\delta} \\
&= c\left(\int_0^\infty \int_\theta^\infty x^{3/(1-\delta)}\frac{1}{\Gamma(\alpha)}(x-\theta)^{\alpha-1}\epsilon^{-(x-\theta)}dx\, dG(\theta)\right)^{1-\delta}
\end{aligned}
$$

$$\leq c\left\{\int_0^\infty \int_\theta^\infty \left[c_1(x-\theta)^{3/(1-\delta)} + c_1\theta^{3/(1-\delta)}\right] \frac{1}{\Gamma(\alpha)} (x-\theta)^{\alpha-1} e^{-(x-\theta)} dx\, dG(\theta)\right\}^{1-\delta}$$

$$\leq \left[\int_0^\infty \left(c_1 + c_2\theta^{3/(1-\delta)}\right) dG(\theta)\right]^{1-\delta} < \infty$$

by assumption (7.4.9). Similarly, we can show that $A_1$ is also finite under the assumption. Then, with the choice of $h_n = n^{-1/(2[\alpha-1]+1)}$, we have

$$E_n^*(|\psi_n(x) - \psi(x)|^2) \leq cn^{-\delta[\alpha-1]/(2[\alpha-1]+1)}.$$

Therefore, we obtain from the definition of $\phi_n(x)$,

$$\begin{aligned} E_n\{R(G,\phi_n)\} - R(G) &= E_n^*(|\psi_n(x) - \psi(x)|^2) \\ &\leq O(n^{-\delta[\alpha-1]/(2[\alpha-1]+1)}). \end{aligned}$$

Hence, the theorem.

## 7.5 Asymptotic Optimality of the Empirical Bayes Testing Rule

In this section, we examine the convergence rate of the empirical Bayes testing rule $d_n(x)$ proposed in (7.3.7). From (7.2.7), we have the Bayes risk associated with the empirical Bayes testing rule $d_n(x)$ and the Bayes testing rule $d_G(x)$ as

$$r(G,d_n) = b\int \alpha(x)\, d_n(x)\, dx + C_G \qquad (7.5.1)$$

and

$$r(G) = r(G,d_G) = b\int \alpha(x)\, d_G(x)\, dx + C_G\ . \qquad (7.5.2)$$

respectively. Obviously,

$$E_n\{r(G,d_n)\} - r(G) \geq 0 \qquad (7.5.3)$$

since the Bayes testing rule $d_G$ achieves the minimum Bayes risk $r(G)$; the expectation $E_n$ is taken with respect to $(x_1, x_2, \ldots, x_n)$.

From Lemma 1 of Johns and Van Ryzin (1972), we have

$$\begin{aligned} E_n\{r(G, d_n)\} - r(G) &\leq b\int |\alpha(x)| \Pr\{|\alpha_n(x) - \alpha(x)| \geq |\alpha(x)|\}dx \\ &\leq b\int |\alpha(x)|^{1-\gamma}\, E_n(|\alpha_n(x) - \alpha(x)|^{\gamma})dx \qquad (7.5.4) \end{aligned}$$

for $0 < \gamma < 1$.

**Definition 7.5.1** A sequence of empirical Bayes testing rules $\{\delta_n\}$ is said to be asymptotically optimal at least of order $\beta_n$ relative to the unknown prior $G$ if

$$E_n\{r(G, \delta_n)\} - r(G) \leq O(\beta_n) \qquad \text{as } n \to \infty\ . \qquad (7.5.5)$$

where $\beta_n$ is a sequence of positive numbers such that $\lim_{n\to\infty} \beta_n = 0$.

Now we present the convergence rate of the empirical Bayes testing rule $d_n(x)$ in (7.3.7) in the following theorem.

**Theorem 7.5.1** *Let $\{d_n(x)\}$ be the sequence of empirical Bayes testing rules proposed in (7.3.7); if for $0 < \delta < 1$,*

$$E\left(\theta^{2/(1-\delta)}\right) < \infty, \qquad (7.5.6)$$

*then with the choice of $h_n = n^{-1/(2[\alpha-1]+1)}$, we have*

$$E_n\{r(G, d_n)\} - r(G) \leq O(n^{-\delta[\alpha-1]/(2[\alpha-1]+1)}). \qquad (7.5.7)$$

**Proof.** With

$$E_n\{r(G, d_n)\} - r(G) \leq b\int |\alpha(x)|^{1-\delta} E_n(|\alpha_n(x) - \alpha(x)|^{\delta})dx$$

and since

$$\begin{aligned} \alpha(x) &= \int \theta f(x|\theta)dG(\theta) - \theta_0 f(x) \\ &= (x - \theta_0)f(x) - w(x) \end{aligned}$$

and

$$\alpha_n(x) = (x - \theta_0)f_n(x) - w_n(x),$$

we have, by Lemma 7.4.2,

$$\begin{aligned}
&E_n\{r(G, d_n)\} - r(G) \\
&\quad\leq\; b\int \left|\int \theta f(x|\theta)dG(\theta) - \theta_0 f(x)\right|^{1-\delta} \\
&\qquad\times\; E_n(|x(f_n(x) - f(x)) - \theta_0(f_n(x) - f(x)) - (w_n(x) - w(x))|^{\delta})dx \\
&\quad\leq\; O(n^{-\delta[\alpha-1]/(2[\alpha-1]+1)})(A_1 + A_2 + A_3 + A_4),
\end{aligned}$$

where

$$\begin{aligned}
A_1 &= \int x^{\delta}\left(\int \theta f(x|\theta)dG(\theta)\right)^{1-\delta} dx \\
A_2 &= \int \left(\int \theta f(x|\theta)dG(\theta)\right)^{1-\delta} dx \\
A_3 &= \int x^{\delta}\left(\int f(x|\theta)dG(\theta)\right)^{1-\delta} dx \\
A_4 &= \int \left(\int f(x|\theta)dG(\theta)\right)^{1-\delta} dx.
\end{aligned}$$

To show that $A_1$ is finite, we use Hölder's inequality to observe

$$\begin{aligned}
&\int_0^1 (x^{\delta})\left[\left(\int_0^x \theta f(x|\theta)dG(\theta)\right)^{1-\delta}\right] dx \\
&\quad\leq\; \left[\int_0^1 x\,dx\right]^{\delta}\left[\int_0^1 \left(\int_0^x \theta f(x|\theta)dG(\theta)\right) dx\right]^{1-\delta} \\
&\quad\leq\; c\left[\int_0^1 \left(\int_\theta^1 \theta f(x|\theta)dx\right) dG(\theta)\right]^{1-\delta} \\
&\quad\leq\; c\left(\int_0^\infty \theta dG(\theta)\right)^{1-\delta} < \infty
\end{aligned}$$

and

$$\begin{aligned}
&\int_0^\infty (x^{\delta})\left[\left(\int_0^x \theta f(x|\theta)dG(\theta)\right)^{1-\delta}\right] dx \\
&\quad=\; \int_1^\infty (x^{-1})\left[x^{1+\delta}\left(\int_0^x \theta f(x|\theta)dG(\theta)\right)^{1-\delta}\right] dx \\
&\quad\leq\; \left[\int_1^\infty (x^{-1})^{1/\delta}dx\right]^{\delta}\left[\int_1^\infty \left(x^{1+\delta/(1-\delta)}\right)\left(\int_0^x \theta f(x|\theta)dG(\theta)\right) dx\right]^{1-\delta}
\end{aligned}$$

$$\leq c\left[\int_0^\infty \int_\theta^\infty \left(x^{1+\delta/(1-\delta)}\right)(\theta f(x|\theta))dx\, dG(\theta)\right]^{1-\delta}$$

$$= c\left[\int_0^\infty \int_\theta^\infty \theta x^{1+\delta/(1-\delta)} \frac{1}{\Gamma(\alpha)}(x-\theta)^{\alpha-1}e^{-(x-\theta)}dx\, dG(\theta)\right]^{1-\delta}$$

$$\leq c\left\{\int_0^\infty \int_\theta^\infty \theta\left[c_1(x-\theta)^{1+\delta/(1-\delta)} + c_1\theta^{1+\delta/(1-\delta)}\right]\right.$$

$$\left.\cdot \frac{1}{\Gamma(\alpha)}(x-\theta)^{\alpha-1}e^{-(x-\theta)}dx\, dG(\theta)\right\}^{1-\delta}$$

$$\leq \left[\int_0^\infty \left(c_1\theta + c_2\theta^{2/(1-\delta)}\right)dG(\theta)\right]^{1-\delta} < \infty$$

by assumption (7.5.6). We can similarly show that $A_2$, $A_3$ and $A_4$ are also finite. Then, we finally obtain that

$$E_n\{r(G,d_n)\} - r(G) \leq O\left(n^{-\delta[\alpha-1]/(2[\alpha-1]+1)}\right).$$

Hence, the theorem.

# Remarks

1. The location parameter family of gamma distributions in (7.2.1) is not included in the typical truncation parameter density family as considered by Datta (1991) and Huang (1995). The importance of (7.2.3) obtained by Fox (1978) is that it gives us an explicit expression for the Bayes estimator $\varphi_G(x)$ in terms of the marginal distribution of $x$, which enables us to estimate $\varphi_G(x)$ from the past observations $x_1, x_2, \ldots, x_n$.

2. The convergence rates in Theorem 7.4.1 and Theorem 7.5.1 are dependent on $\delta$, $0 < \delta < 1$, and the shape parameter $\alpha$. If $\alpha$ is larger, then the convergence rates are faster. If the conditions of Theorem 7.4.1 and Theorem 7.5.1 are satisfied for $\delta$ arbitrarily close to 1, then the convergence rates can be arbitrarily close to $O(n^{-[\alpha-1]/(2[\alpha-1]+1)})$.

3. For the more general location parameter family of gamma distributions when the location parameter $\theta \in (-\infty, \infty)$ instead of $\theta \in (0, \infty)$ as in (7.2.1), the

relation $0 \leq \phi_G(x) = E(\theta|x) \leq x$ is no longer true; we just get $\phi_G(x) \leq x$. Then we can similarly propose empirical Bayes estimator for $\theta$ and an empirical Bayes testing rule for the two-action problem and obtain convergence rates under some moment conditions on the prior $G$. However, since the distribution family (7.2.1) is more useful in applications, we just present the asymptotic optimality results for this distribution family in Theorem 7.4.1 and Theorem 7.5.1.

# Bibliography


Arnold, B. C., Balakrishnan, N. and Nagaraja, H. N. *A First Course in Order Statistics*. New York: John Wiley & Sons, 1992.

Balakrishnan, N. and Basu, A. P. *The Exponential Distribution: Theory, Methods and Applications*. Langhorne, PA: Gordon and Breach, 1995.

Balakrishnan, N. and Cohen, A. C. *Order Statistics and Inference: Estimation Methods*. San Diego: Academic Press, 1991.

Berger, J. O. *Statistical Decision Theory and Bayesian Analysis*, Second Edition. New York: Springer-Verlag, 1985.

Blum, J. R. and Susarla, V. "Maximal derivation theory of some estimators of prior distribution functions", *Annals of the Institute of Statistical Mathematics*, **33** (1981), 425–436.

Cline, D. B. H. and Hart, J. D. "Kernel estimation of densities with discontinuities or discontinuous derivatives", *Statistics*, **22** (1991), 69–84.

Datta, S. "Nonparametric empirical Bayes estimation with $O(n^{-1/2})$ rate of a truncation parameter", *Statistics & Decisions*, **9** (1991), 45–61.

DeGroot, M. H. *Optimal Statistical Decisions*. New York: McGraw-Hill, 1970.

Fox, R. "Estimating the empirical distribution function of a parameter sequence", *Annals of Mathematical Statistics*, **41** (1970), 1845–1852.

Fox, R. "Solutions to empirical Bayes squared error loss estimation problems", *Annals of Statistics*, **6** (1978), 846–853.

Gupta, S. S. and Liang, T. "Empirical Bayes rules for selecting good binomial population", In *Adaptive Statistical Procedures and Related Topics* (Ed., J. Van Ryzin), IMS Lecture Notes - Monograph Series, **8** (1986), 110–128.

Gupta, S. S. and Liang, T. "Empirical Bayes rules for selecting the best binomial population", In *Statistical Decision Theory and Related Topics-IV* (Eds., S. S. Gupta and J. O. Berger). New York: Springer-Verlag (1988), 213–224.

Gupta, S. S., Liang, T. and Rau, R. "Empirical Bayes rules for selecting the best normal population compared with a control", *Statistics & Decisions*, **12** (1994), 125–147.

Huang, S. Y. "Empirical Bayes testing procedures in some nonexponential families using asymmetric Linex loss function", *Journal of Statistical Planning and Inference*, **46** (1995), 293–309.

Johns, M. V. and Van Ryzin, J. "Convergence rates for empirical Bayes two-action problem I. Discrete case", *Annals of Mathematical Statistics*, **42** (1971), 1521–1539.

Johns, M. V. and Van Ryzin, J. "Convergence rates for empirical Bayes two-action problem II. Continuous case", *Annals of Mathematical Statistics*, **43** (1972), 934–947.

Johnson, N. L., Kotz, S. and Balakrishnan, N. *Continuous Univariate Distributions - Vol. 1*, Second Edition, New York: John Wiley & Sons, 1994.

Karunamuni, R. J. "Optimal rates of convergence of empirical Bayes tests for the continuous one-parameter exponential family", *Annals of Statistics*, **24** (1996), 212–231.

Lehmann, E. L. *Testing Statistical Hypotheses*, Second Edition, New York: John Wiley & Sons, 1986.

Liang, T. "Convergence rates for empirical Bayes estimation of the scale parameter in a Pareto distribution", *Computational Statistics and Data Analysis*, **16** (1993), 35–46.

Lin, I. C. and Leu, L. Y. "Empirical Bayes rules for selecting the best exponential population with location parameter", *Communications in Statistics—Theory and Methods*, **23** (1994), 1797–1809.

Lin, P. E. "Rates of convergence in empirical Bayes estimation problems: continuous case", *Annals of Statistics*, **3** (1975), 155–164.

Ma, Y. and Balakrishnan, N. "On the mixtures of truncation parameter distribution families", *Statistics and Probability Letters*, (1997) (submitted).

Maritz, J. S. and Lwin, T. *Empirical Bayes Methods*, Second Edition, London: Chapman & Hall, 1989.

Müller, H. G. and Gasser, T. "Optimal convergence properties of kernel estimates of derivatives of a density function", In *Lecture Notes in Mathematics*, **757**, Berlin: Springer-Verlag (1979), 144–154.

Nogami, Y. "Convergence rates for empirical Bayes estimation in the uniform $U(0,\theta)$ distribution", *Annals of Statistics*, **16** (1988), 1335–1341.

Parzen, E. "On estimation of a probability density function and mode", *Annals of Mathematical Statistics*, **33** (1962), 1065–1076.

Pericchi, L. R., Sansó, B. and Smith, A. F. M. "Posterior cumulant relationships in Bayesian inference involving the exponential family", *Journal of the American Statistical Association*, **88** (1993), 1419–1426.

Pericchi, L. R. and Smith, A. F. M. "Exact and approximate posterior moments for a normal location parameter", *Journal of the Royal Statistical Society, Series B*, **54** (1992), 793–804.

Polson, N. G. "A representation of the posterior mean for a location model", *Biometrika*, **78** (1991), 426–430.

Prasad, B. and Singh, R. S. "Estimation of prior distribution and empirical Bayes estimation in a non-exponential family", *Journal of Statistical Planning and Inference*, **24** (1990), 81–86.

Raiffa, H. and Schlaifer, R. *Applied Statistical Decision Theory*, Division of Research, Graduate School of Business Administration, Harvard University, 1961.

Robbins, H. "An empirical Bayes approach to statistics", In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, **1**, Berkeley: University of Carlifornia Press (1955), 157–163.

Robbins, H. "The empirical Bayes approach to testing statistical hypothesis", *Reviews of the International Statistical Institute*, **31** (1963), 195–208.

Robbins, H. "The empirical Bayes approach to statistical decision problems", *Annals of Mathematical Statistics*, **35** (1964), 1–10.

Robert, C. P. *The Bayesian Choice: A Decision-Theoretic Motivation*, New York: Springer-Verlag, 1994.

Singh, R. S. "Improvement on some known nonparametric uniformly consistent estimators of derivatives of a density", *Annals of Statistics*, **5** (1977), 394–399.

Singh, R. S. "Empirical Bayes estimation in Lebesgue exponential families with rates near the best possible rate", *Annals of Statistics*, **7** (1979), 890–902.

Singh, R. S. "Empirical Bayes linear loss hypothesis testing in a non-regular exponential family", *Journal of Statistical Planning and Inference*, **43** (1995), 107–120.

Singh, R. S. and Prasad, B. "Uniformly strongly consistent prior-distribution and empirical Bayes estimators with asymptotic optimality and rates in a non-exponential family", *Sankhyā, Series A*, **51** (1989), 334–342.

Singh, R. S. and Wei, L. "Empirical Bayes with rates and best rates of convergence in $u(x)c(\theta)\exp(-x/\theta)$-family: estimation case", *Annals of the Institute of Statistical Mathematics*, **44** (1992), 435–449.

Susarla, V. "Empirical Bayes theory", In *Encyclopedia of Statistical Sciences*, New York: John Wiley & Sons (1982), 490–503.

Susarla, V. and O'Bryan, T. "Empirical Bayes interval estimates involving uniform distributions", *Communications in Statistics*, **8** (1979), 385–397.

Teicher, H. "Identifiability of mixtures", *Annals of Mathematical Statistics*, **32** (1961), 244–248.

Teicher, H. "Identifiability of finite mixtures", *Annals of Mathematical Statistics*, **34** (1963), 1265–1269.

Tiwari, R. C. and Zalkikar, J. N. "Empirical Bayes estimation of the scale parameter of a Pareto distribution", *Computational Statistics & Data Analysis*, **10** (1990), 261–270.

Van Houwelingen, H. C. "Monotone empirical Bayes test for uniform distribution using the maximum likelihood estimator of a decreasing density", *Annals of Statistics*, **15** (1987), 875–879.