ASYMPTOTIC DISTRIBUTION OF THE RANDOM REGRET RISK OF THE GUPTA-LIANG RULE FOR SELECTING GOOD GAMMA DISTRIBUTIONS

by

Friedrich Liese
University of Rostock and
Purdue University

Technical Report #98-16

Department of Statistics
Purdue University

June 1998

△

ASYMPTOTIC DISTRIBUTION OF THE RANDOM REGRET RISK OF THE GUPTA-LIANG RULE FOR SELECTING GOOD GAMMA DISTRIBUTIONS

by

Friedrich Liese[1]
University of Rostock and
Purdue University

Abstract

Gupta and Liang (1996) introduced an empirical Bayes selection procedure for selecting a good exponential population and proved that the regret risk $\mathbb{E}\mathcal{R}_n$ is of order $0(n^{-\lambda/2})$ for some $0 < \lambda \leq 2$. In this paper we study the random part $\mathcal{R}_n$ of the regret risk and show that $n\mathcal{R}_n$ tends in distribution to a linear combination of χ^2-distributed random variables.

Short Title: Asymptotic Distribution of Random Regret Risk

Keywords and Phrases: Good gamma populations, random regret risks, empirical Bayes method.

AMS 1991 Subject Classification: 62F07, 62C12.

[1]This research was done while this author was a Visiting Professor at the Department of Statistics, Purdue University.

1 Introduction

The family of gamma distributions has fundamental meaning in reliability theory, survival analysis and general in the area of life time distributions. For an overview and more details we refer to Johnson, Kotz and Balakrishnan (1994) and Basu (1995). We consider k independent gamma populations $\pi_1, \ldots, \pi_k$ with the same shape parameter m and different scale parameters. The stochastically largest population has the largest ϑ_i-value. Let there be a control value ϑ_0. Each population is called good if $\vartheta_i \geq \vartheta_0$ and bad otherwise. We study the problem of finding all good populations. This is a typical subset selection problem, see Gupta and Panchapakesan (1985). We assume that the Θ_i are random and independent distributed according to the unknown distribution G_i. Then for a given loss function the best selection rule, being the Bayes selection rule, depends on the unknown joint distribution $G = \prod_{i=1}^{k} G_i$ of $\underline{\Theta} = (\Theta_1, \ldots, \Theta_k)$. We suppose that historical data are available and can be included in the decision rule. This is the empirical Bayes approach due to Robbins (1956). Empirical Bayes methods have been applied in different areas of statistics. Deely (1965) constructed empirical Bayes subset selection procedures. In a series of papers Gupta and Liang (1988, 1994, 1996) and Gupta, Liang and Rau (1994a, 1994b) have studied different selection procedures using empirical Bayes approach. Assume $\underline{Y}$ are the actual data based on which we wish to make a decision. Denote the risk of the decision d by $\mathcal{R}(d)$. Then the optimal decision $d(G)$ depends on the unknown prior distributions G of $(\Theta_1, \ldots, \Theta_k)$. The central idea of the empirical Bayes approach is the construction of a good decision rule d_n^* on the basis of historical data $\underline{Y}_n$. The quality of d_n^* is then characterized by the non-negative random regret risk $\mathcal{R}_n = \mathcal{R}(d_n^*) - \mathcal{R}(d(G))$. The aim of the above mentioned papers dealing with empirical Bayes methods was to construct suitable decision rules d_n^* and to evaluate the non-random regret risk $\mathbb{E}\mathcal{R}_n$. The main goal of these papers was to prove the convergence of $\mathbb{E}\mathcal{R}_n$ to zero with a certain rate. It turns out that the main part of the Bayes decision rule are some functions $H_i(y)$ which are linear transforms of the unknown distribution G_i. But one can find a family of functions f_y with $H_i(y) = \mathbb{E} f_y(Y_{ij})$, $j = 1, \ldots, n$, so that

$$\widehat{H}_{in}(y) = \frac{1}{n} \sum_{j=1}^{n} f_y(Y_{ij})$$

is an unbiased estimator for $H_i(y)$. Gupta and Liang (1996) used $\widehat{H}_{in}$ instead of H_i to construct an empirical Bayes selection rule d_n^* for the selection of a good exponential population and proved

$$\mathbb{E}[\mathcal{R}(d_n^*) - \mathcal{R}(d(G))] = O(n^{-\lambda/2})$$

with some $0 < \lambda \leq 2$. The information one needs from H_i is only the zero η_i of H_i. Gupta and Liese (1998) constructed an $\sqrt{n}$-consistent M-estimator for η_i, introduced a new empirical Bayes selection rule $\widehat{d}_n$ and proved a limit theorem for $n[\mathcal{R}(\widehat{d}_n) - \mathcal{R}(d(G))]$. In this paper we apply to $\widehat{H}_{in}(y)$ techniques from empirical processes to get a Glivenko-Cantelli Theorem and the convergence of the distributions of the stochastic processes $\sqrt{n}(H_{in}(y) - H_i(y))$, $\varkappa_1 \leq y \leq \varkappa_2$. It turns out that the Gupta-Liang selection rule is asymptotically a quadratic functional of the processes $\widehat{H}_{in}$. Using this fact we will show that the distribution of $n[\mathcal{R}(d_n^*) - \mathcal{R}(d(G))]$ converges weakly to the distribution of a linear combination of independent χ^2-distributed random variables each with one degree of freedom. The coefficients in the linear combination are explicitly evaluated and expressed in terms of the prior distributions G_i.

2 Empirical Bayes Selection Procedures

We consider k independent gamma distributed populations $\pi_1, \ldots, \pi_k$ with densities $f(y|\vartheta_i) = c(\vartheta)u(y)e^{-y/\vartheta_i}I(y > 0)$, where $c(\vartheta) = \frac{1}{\Gamma(m)\vartheta^m}, u(y) = y^{m-1}, m > 0, \vartheta > 0$. Here and in the following $I(A)$ denotes the indicator function of the set A. Given a standard value ϑ_0 we call a population π_i good if $\vartheta_i \geq \vartheta_0$. Our aim is to select all good populations. A selection rule $d = (a_1, \ldots, a_k)$ is a measurable mapping $d : (0, \infty)^k \to \mathbb{D}$ where $\mathbb{D} = \{0, 1\}^k$ is the decision space. π_i is selected iff $a_i = 1$. Similar as in Gupta, Liang (1994, 1996) we use the loss function

$$L(\underline{\vartheta}, \underline{a}) = \sum_{i=1}^{k} l(\vartheta_i, a_i)$$

where $\underline{\vartheta} = (\vartheta_1, \ldots, \vartheta_k), \underline{a} = (a_1, \ldots, a_k)$ and

$$l(\vartheta_i, a_i) = a_i\vartheta_i(\vartheta_0 - \vartheta_i)I(0 < \vartheta_i < \vartheta_0) + (1 - a_i)\vartheta_i(\vartheta_i - \vartheta_0)I(\vartheta_0 \leq \vartheta_i) \ . \tag{2.1}$$

If we have a measurement Y_i from each π_i the risk of the selection rule d is given by

$$R(\underline{\vartheta}, d) = \mathbb{E}L(\underline{\vartheta}, d(\underline{Y})) = \sum_{i=1}^{k} \mathbb{E}l(\vartheta_i, q_i(Y_i)) \tag{2.2}$$

where $\underline{\vartheta} = (\vartheta_1, \ldots, \vartheta_k)$, $\underline{Y} = (Y_1, \ldots, Y_k)$, $q_i(y_i) = \mathbb{E}a_i(Y_1, \ldots, Y_{i-1}, y_i, Y_{i+1}, \ldots, Y_k)$. Furthermore

$$\mathbb{E}l(\vartheta_i, q_i(Y_i)) = \int_0^\infty q_i(y_i)(\vartheta_0 - \vartheta_i)c(\vartheta_i)u(y_i)e^{-\frac{y_i}{\vartheta_i}}\, dy_i + C(\vartheta_i) \tag{2.3}$$

with

$$C(\vartheta_i) = \vartheta_i(\vartheta_i - \vartheta_0)I(\vartheta_i \geq \vartheta_0)\ .$$

We will apply the Bayes approach to the selection problem and assume that the ϑ_i are realizations of independent r. v. Θ_i with distribution G_i. We suppose

$$\int_0^\infty \vartheta_i^2\, dG_i(\vartheta_i) < \infty\ . \tag{2.4}$$

Then the Bayes risk $\mathcal{R}(d)$ is finite and given by

$$\mathcal{R}(d) = \mathbb{E}L(\underline{\Theta}, d(\underline{Y})) = \sum_{i=1}^{k} \int_0^\infty \int_0^\infty q_i(y_i)(\vartheta_0 - \vartheta_i)c(\vartheta_i)u(y_i)e^{-\frac{y_i}{\vartheta_i}}\, dG_i(\vartheta_i)dy_i + \gamma_i$$

where

$$\gamma_i = \int_{\vartheta_0}^\infty \vartheta_i(\vartheta_i - \vartheta_0)\, dG_i(\vartheta_i)$$

Set

$$\psi_{ia}(y) = \int_0^\infty \vartheta^a c(\vartheta)e^{-\frac{y}{\vartheta}}\, dG_i(\vartheta) \tag{2.5}$$

where $c(\vartheta) = (\Gamma(m)\vartheta^m)^{-1}$. As in Gupta, Liang (1996) one obtains by integration by parts

$$\begin{aligned}(\vartheta_0 - \vartheta_i)e^{-\frac{y_i}{\vartheta_i}} &= \int_{y_i}^\infty (\vartheta_0 + y_i - t_i)\frac{1}{\vartheta_i}e^{-\frac{t_i}{\vartheta_i}}\, dt_i \\ \int_0^\infty (\vartheta_0 - \vartheta_i)e^{-\frac{y_i}{\vartheta_i}}\, dG_i(\vartheta_i) &= \vartheta_0\psi_{1i}(y_i) - \psi_{2i}(y_i)\ .\end{aligned}$$

Using these relations we obtain

$$\mathbb{E}L(\underline{\Theta}, d(\underline{Y})) = \sum_{i=1}^{k} \int_0^\infty u(y_i)\left(\vartheta_0\psi_{i1}(y_i) - \psi_{i2}(y_i)\right) q_i(y_i)\, dy_i + \gamma_i \ .$$

This shows that $\inf_d \mathbb{E}L(\underline{\Theta}, d(\underline{Y}))$ is attained by the selection rule $d^0 = (d_1^0, \dots, d_k^0)$ where

$$d_i^0(y_i) = \begin{cases} 1: & \vartheta_0\psi_{i1}(y_i) \le \psi_{i2}(y_i) \\ 0: & \text{otherwise} \ . \end{cases}$$

If G_i is non-degenerate then ψ_{i2}/ψ_{i1} is strictly increasing. This means that the zero η_{i0} of $H_i(y_i) = \vartheta_0\psi_{i1}(y_i) - \psi_{i2}(y_i)$, if there is any, is uniquely determined. To apply the selection rule d^0 we have to know η_{i0}. But the ψ_{i1}, ψ_{i2} as well as η_{i_0} include the unknown prior distribution G_i. Let $Y_{i1}, \dots, Y_{in}, i = 1, \dots, k$ be data from the past where $Y_{i1}, \dots, Y_{ik}$ are i. i. d. with common density

$$f_i(y_i) = \int_0^\infty c(\vartheta_i) u(y_i) e^{-\frac{y_i}{\vartheta_i}}\, dG_i(\vartheta_i) \ . \tag{2.6}$$

The idea of the empirical Bayes approach is to estimate the unknown optimal selection rule d_i^0 with the help of the historical data $Y_{i1}, \dots, Y_{in}$, i. e. an empirical Bayes selection rule $\widehat{d}_{i,n}$ is a measurable mapping from $(0, \infty)^{n+1}$ into $[0, 1]$.
The conditional Bayes risk of the empirical Bayes selection rule $\widehat{d}_n = (\widehat{d}_{1n}, \dots, \widehat{d}_{kn})$ is

$$\sum_{i=1}^{k} \int_0^\infty \widehat{d}_{in}(Y_{i1}, \dots, Y_{in}, y_i) u(y_i) H_i(y_i)\, dy_i + \gamma_i$$

where $H_i = \vartheta_0\psi_{i1} - \psi_{i2}$. We call the nonnegative random variable

$$\mathcal{R}(\widehat{d}_n) - \mathcal{R}(d^0) = \sum_{i=1}^{k} \int_0^\infty \left[\widehat{d}_{in}(Y_{i1}, \dots, Y_{in}, y_i) - d_i^0(y_i)\right] u(y_i) H_i(y_i)\, dy_i \tag{2.7}$$

the random regret risk. The regret risk studied in Gupta and Liang (1996) is then $\mathbb{E}\mathcal{R}(\widehat{d}_n) - \mathcal{R}(d^0)$. The definition of ψ_{ia} in (2.5) and partial integration show that

$$\begin{aligned} \psi_{1i}(y_i) &= \int_0^\infty \int_{y_i}^\infty c(\vartheta_i) e^{-\frac{t_i}{\vartheta_i}}\, dt_i dG_i(\vartheta_i) = \mathbb{E}\frac{1}{u(Y_i)} I(Y_i \ge y_i) \\ \psi_{2i}(y_i) &= \int_0^\infty \int_{y_i}^\infty (t_i - y_i) c(\vartheta_i) e^{-\frac{t_i}{\vartheta_i}}\, dt_i dG_i(\vartheta_i) = \mathbb{E}\frac{1}{u(Y_i)})\,(Y_i - y_i)\, I(Y_i > y_i) \ . \end{aligned}$$

Consequently,

$$\widehat{H}_{in}(y) = \frac{1}{n}\sum_{l=1}^{n}(\vartheta_0 + y - Y_{il})I(Y_{il} > y)\frac{1}{u(Y_i)}$$

is an unbiased and consistent estimator for

$$H_i(y) = \vartheta_0\psi_{i1}(y) - \psi_{i2}(y)$$

for every fixed y. Using this fact Gupta und Liang (1996) introduced an empirical Bayes selection procedure d_n^* by setting

$$d_{in}^*(y_i) = \begin{cases} 1: & \widehat{H}_{in}(y_i) \leq 0 \\ 0: & \text{otherwise} . \end{cases}$$

The random regret risk of d_n^* is in view of (2.7) given by

$$\mathcal{R}(d_n^*) - \mathcal{R}(d^0) = \sum_{i=1}^{k}\int_0^\infty \left[I(\widehat{H}_{in}(y_i) \leq 0) - I(H_i(y_i)) \leq 0)\right] u(y_i)H_i(y_i)\,dy_i \tag{2.8}$$

and the regret risk is

$$\mathbb{E}\mathcal{R}(d_n^*) - \mathcal{R}(d^0) = \mathbb{E}\sum_{i=1}^{k}\int_0^\infty \left[I(\widehat{H}_{in}(y_i) \leq 0) - I(H_i(y_i) \leq 0)\right] u(y_i)H_i(y_i)\,dy_i \ . \tag{2.9}$$

Gupta und Liang (1996) studied the rate of convergence of the regret risk $\mathbb{E}\mathcal{R}(d_n^*) - \mathcal{R}(d^0)$ to zero. In this paper we investigate the asymptotic distribution of $n\,[\mathcal{R}(d_n^*) - \mathcal{R}(d^0)]$.

A limit theorem for the random regret risk of a modified Gupta-Liang selection rule was proved in Gupta and Liese (1998). In this paper we apply techniques from the theory of empirical processes to prove the uniform consistency of $\widehat{H}_{in}$ and to establish a limit theorem for $n\,[\mathcal{R}(d_n^*) - \mathcal{R}(d^0)]$.

3 Uniform consistency of $\widehat{H}_{in}$

To simplify the notations we concentrate ourselves to one population in this chapter. Assume $Y_1, \ldots, Y_n$ are i. i. d. with density function

$$f(y) = \int_0^\infty c(\vartheta)y^{m-1}e^{-\frac{y}{\vartheta}}I(y > \vartheta)\,dG(\vartheta) \ .$$

Integration by parts shows that

$$\int_0^\infty \vartheta^k \, dG(\vartheta) < \infty \tag{3.1}$$

implies

$$\mathbb{E}Y_i^k < \infty \ . \tag{3.2}$$

Set $c(\vartheta) = (\Gamma(m)\vartheta^{m-1})^{-1}$ and

$$\psi_1(y) = \int_0^\infty c(\vartheta)\vartheta e^{-\frac{y}{\vartheta}} \, dG(y), \quad \psi_2(y) = \int_0^\infty c(\vartheta)\vartheta^2 e^{-\frac{y}{\vartheta}} \, dG(\vartheta) \ . \tag{3.3}$$

Then

$$\begin{aligned} \psi_1(y) &= \mathbb{E}I(Y_1 > y)\frac{1}{u(Y_1)} \\ \psi_2(y) &= \mathbb{E}(Y_1 - y)I(Y_1 > y)\frac{1}{u(Y_1)} \end{aligned} \tag{3.4}$$

put

$$H(y) = \vartheta_0\psi_1(y) - \psi_2(y) \ . \tag{3.5}$$

We study the structure of the function H in more details. For this aim we put for any measure μ on the Borel sets of $[0, \infty)$

$$L_\mu(y) = \int_0^\infty e^{-\frac{y}{\vartheta}} \, \mu(d\vartheta) \ .$$

Then for $0 < \alpha < 1$ by Hölders inequality

$$L_\mu(\alpha y_1 + (1-\alpha)y_2) \leq (L_\mu(y_1))^\alpha (L_\mu(y_2))^{1-\alpha} \ .$$

Hence $\ln L_\mu(y)$ is convex. The set $\{y : L_\mu(y) < \infty\}$ is some interval and L_μ is known to be infinitely often differentiable in the interior, say (a, b), of $\{y : L_\mu(y) < \infty\}$. As $\ln L_\mu(y)$ is convex we see that $(\ln L_\mu)'$ is increasing. Moreover, if μ is nondegenerated, then L_μ is strictly convex and $(\ln L_\mu)'$ is strictly increasing. Hence both L'_μ/L_μ and $\vartheta_0 L'_\mu - L_\mu$ have at most one zero. As $L_\mu^{(k)} = (-1)^k L_{\mu_k}$, where $\mu_k(d\vartheta) = \vartheta^k \mu(d\vartheta)$ we see that also the function $\vartheta_0 L''_\mu - L'_\mu$ has at most one zero. This means that with $\mu(d\vartheta) = \vartheta^2\, dG(\vartheta)$ under the assumption

$$\lim_{y\downarrow 0} \frac{\psi_2(y)}{\psi_1(y)} < \vartheta_0 < \lim_{y\uparrow\infty} \frac{\psi_2(y)}{\psi_1(y)} \tag{3.6}$$

the function H from (3.5) has a unique zero η and it holds

$$\begin{aligned} H(y) &> 0, \quad 0 < y < \eta \\ H(y) &< 0, \quad \eta < y < \infty\,. \end{aligned} \tag{3.7}$$

As the function $H'(y)$ has again at least one zero, say λ, and $\lim_{y\to\infty} H(y) = 0$ we see that in view of $H(y) < 0, \eta < y < \infty$, there exists a unique local minimum at λ being the zero of H' and H strictly decreasing in $(0,\lambda)$ and strictly increasing in (λ,∞).
In view of the relations (3.4)

$$\widehat{H}_n(y) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{u(Y_i)} \left(\vartheta_0 I(Y_i \ge y) - (y_i - y) I(Y_i \ge y)\right) \tag{3.8}$$

is an unbiased estimator for $H(y)$.
To prove an uniform law of large numbers we apply methods from empirical process theory for which we refer to van der Vaart and Wellner (1996). We introduce the classes of functions

$$t \mapsto \vartheta_0 I(t-y>0), \quad t \mapsto (t-y)I(t-y>0), \quad t \ge 0, y > 0\,.$$

These classes are translates of the monotone function $\vartheta_0 I(x > 0)$, $xI(x > 0)$ and are consequently VC-subgraph classes in view of Lemma 2.6.16 in van der Vaart and Wellner (1996). Set $g(x) = \frac{1}{u(x)} I(x>0) = \frac{1}{x^{m-1}} I(x>0)$. Then by Lemma 2.6.18 in van der Vaart and Wellner the classes of functions $\mathcal{F}_i = \{f_{i,y}(\cdot), y > 0\}$

$$\begin{aligned} f_{1,y}(\cdot) &= g(\cdot)\vartheta_0 I(\cdot - y > 0) \\ f_{2,y}(\cdot) &= g(\cdot)(\cdot - y) I(\cdot - y > 0) \end{aligned}$$

are again VC-subgraph classes. Note that for $t > 0$

$$\begin{aligned} |f_{1,y}(t)| &\le \vartheta_0 g(t) \\ |f_{2,y}(t)| &\le t g(t)\,. \end{aligned}$$

If $\mathbb{E}|Y_1| < \infty$ then both $\mathcal{F}_1$ and $\mathcal{F}_2$ have integrable envelope. Hence by the Glivenko-Cantelli theorem (Theorem 2.4.3 in van der Vaart and Wellner (1996))

$$\sup_{0<y<\infty} \left| \frac{1}{n} \sum_{j=1}^{n} f_{i,y}(Y_j) - \mathbb{E} f_{i,y}(Y_1) \right| \xrightarrow[n\to\infty]{} 0 \quad \text{a. s.}\ .$$

As $H(y) = \mathbb{E}[f_{1,y}(Y_1) - f_{2,y}(Y_1)]$ we have obtained the following statement.

Proposition 1 *If* $\int_0^\infty \vartheta \, dG(\vartheta) < \infty$ *then* $\widehat{H}_n$ *from* (3.8) *fulfils a. s.*

$$\sup_{0<y<\infty} \left| \widehat{H}_n(y) - H(y) \right| \xrightarrow[n\to\infty]{} 0\ .$$

Denote by L_2 the set of all measurable f such that $\mathbb{E} f^2(Y_1) < \infty$ and set for $f_i \in L_2, i = 1, 2,$

$$\varrho(f_1, f_2) = \left(\mathbb{E} \left(f_1(Y_1) - f_2(Y_1) \right)^2 \right)^{\frac{1}{2}}\ .$$

The definition of $f_{i,y}(\cdot)$ yields

$$\begin{aligned} f_{1,y}^2(t) &= \left[\frac{1}{t^{m-1}} \vartheta_0 I(t \geq y) \right]^2 \leq \vartheta_0^2 t^{2(1-m)} I(t \geq y) \\ f_{2,y}^2(t) &= \left[\frac{1}{t^{m-1}} (t-y) I(t \geq y) \right]^2 \leq t^{4-2m} I(t \geq y)\ . \end{aligned}$$

If $\mathbb{E} Y_1^4 < \infty$ then for every $0 < \varkappa_1 < \varkappa_2 < \infty$ the classes

$$\mathcal{F}_{i,\varkappa_1,\varkappa_2} = \{ f_{i,y}(\cdot), \varkappa_1 \leq y \leq \varkappa_2 \}$$

have square integrable envelops. Furthermore, as Y_1 has a Lebesgue density

(3.9) $$\lim_{y_1 \to y_2} \varrho\left(f_{i,y_1}, f_{i,y_2} \right) = 0\ .$$

As the $\mathcal{F}_i$ are VC-subgraph classes the $\mathcal{F}_{i,\varkappa_1,\varkappa_2}$ have the same property and square integrable envelops. Consequently they are Donsker classes in the sense of van der Vaart and Wellner (1996) p. 81. This implies (see p. 89) that the processes

$$\Gamma_{i,n}(y) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left[f_{i,y}(Y_j) - \mathbb{E} f_{i,y}(Y_j) \right]$$

$\varkappa_1 \leq y \leq \varkappa_2$ are asymptotically equicontinuous in the following sense

$$\lim_{h \downarrow 0} \limsup_{n\to\infty} P \left(\sup_{\substack{y_1, y_2 \in [\varkappa_1, \varkappa_2] \\ \varrho(f_{i,y_1}, f_{i,y_2}) < h}} \left| \Gamma_{i,n}(y_1) - \Gamma_{i,n}(y_2) \right| > \varepsilon \right) = 0\ .$$

for every $\varepsilon > 0$. As $y \mapsto f_{i,y}$ is continuous in L_2-sense (see (3.9)) we obtain

$$\lim_{h\downarrow 0} \limsup_{n\to\infty} P\left(\sup_{\substack{y_1,y_2\in[\varkappa_1,\varkappa_2] \\ |y_1-y_2|<h}} |\Gamma_{i,n}(y_1) - \Gamma_{i,n}(y_2)| > \varepsilon \right) = 0$$

for every $\varepsilon > 0$. Set $S_n(y) = \Gamma_{1,n}(y) + \Gamma_{2,n}(y) = \sqrt{n}\left(\widehat{H}_n(y) - H(y)\right)$ then

$$\lim_{h\downarrow 0} \limsup_{n\to\infty} P\left(\sup_{\substack{y_1,y_2\in[\varkappa_1,\varkappa_2] \\ |y_1-y_2|<h}} |S_n(y_1) - S_n(y_2)| > \varepsilon \right) = 0 \tag{3.10}$$

for every $\varepsilon > 0$. Put

$$f_y(t) = f_{1,y}(t) + f_{2,y}(t) = \frac{1}{u(t)} \left[\vartheta_0 I(t > y) - (t-y) I(t > y)\right]$$

Then

$$S_n(y) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[f_y(Y_i) - \mathbb{E} f_y(Y_i)\right] \tag{3.11}$$

and by the central limit theorem the finite dimensional distributions of S_n converge to the corresponding finite dimensional distribution of a Gaussian process $S(y)$ with expectation zero and covariance function

$$C(y_1, y_2) = \operatorname{cov}\left(f_{y_1}(Y_1), f_{y_2}(Y_1)\right) . \tag{3.12}$$

Let $D[\varkappa_1, \varkappa_2]$ be the Skorokhod space of all functions on $[\varkappa_1, \varkappa_2]$ which are right continuous and have limits from the left. For distributions $P, P_1, P_2, \ldots$ defined on the Borel sets of $D[\varkappa_1, \varkappa_2]$ the weak convergence is denoted by $P_n \Rightarrow P$. The convergence of the finite dimensional distributions together with the asymptotic equicontinuity (3.10) imply the weak convergence of the distributions $\mathcal{L}(S_n)$ to the distribution $\mathcal{L}(S)$ of S. Note that (3.10) implies the continuity of S (see Billingsley (1968), Theorem 15.5). Thus we have obtained the following statement.

Theorem 1 *If* $\int_0^\infty \vartheta^4 \, dG(\vartheta) < \infty$ *then for the processes* S_n *defined in* (3.11) *it holds*

$$\mathcal{L}(S_n) \Rightarrow \mathcal{L}(S)$$

where S *is a continuous zero mean Gaussian process with covariance function* (3.12).

In order to study the asymptotic distribution of $n\left[\mathcal{R}(d_n^*) - \mathcal{R}(d^0)\right]$ we have to deal with the following integral

$$n\int_0^\infty \left[I\left(\widehat{H}_n(y) \leq 0\right) - I(H(y) \leq 0)\right] u(y)H(y)\,dy\,.$$

In a first step we will show that for intervals $[\alpha, \beta]$ with $\inf_{\alpha\leq y\leq\beta} |H(y)| > 0$ the corresponding integrals are $o\left(\frac{1}{n}\right)$ and can therefore neglected as $n \to \infty$. We set

$$W_i(y) = Y_i = \frac{1}{u(Y_i)}\left[\vartheta_0 I(Y_i > y) - (Y_i - y)I(Y_i > y)\right]\ .$$

Note $\mathbb{E}W_i(y) = H(y)$ and set

$$\begin{aligned}\sigma^2(y) &= V(W_i(y)) = \mathbb{E}\left(W_i(y) - H(y)\right)^2\\ \mu_4(y) &= \mathbb{E}\left(W_i(y) - H(y)\right)^4\ .\end{aligned}$$

Gupta and Liang (1996) estimated $\sigma^2(y)$ under the assumption $\int_0^\infty \vartheta^2 dG(\vartheta)$ and obtained

$$\sigma^2(y) \leq \frac{1}{u(y)}\left(\vartheta_o^2\psi_1(y) + 2\psi_3(y)\right)\ . \tag{3.13}$$

This yields

$$V\left(\widehat{H}_n(y)\right) \leq \frac{1}{nu(y)}\left(\vartheta_0^2\psi_1(y) + 2\psi_3(y)\ .\right) \tag{3.14}$$

If $X_1, \ldots, X_n$ are i. i. d. with $\mathbb{E}X_i^4 < \infty$ and $\mathbb{E}X_i = 0$ then

$$\mathbb{E}\left(\sum_{i=1}^n X_i\right)^4 = n\mathbb{E}X_1^4 + 3(n-1)n\left(V(X_1)\right)^2\ .$$

This yields

$$\mathbb{E}\left(\widehat{H}_n(y) - H(y)\right)^4 = \frac{1}{n^3}\mathbb{E}\left(W_i(y) - H(y)\right)^4 + \frac{3(n-1)}{n^3}\left(\sigma^2(y)\right)^2\ . \tag{3.15}$$

Lemma 1 *If* $\int_0^\infty \vartheta^4\, dG(\vartheta) < \infty$ *then for every* $a < \eta < b$ *and every* $0 < \alpha < \beta < \infty$

with $d = \inf_{\alpha \le y \le \beta} |H(y)| > 0$

$$\mathbb{E} \int_0^a n \left[I\left(\widehat{H}_n(y) \le 0\right) - I(H(y) \le 0) \right] u(y) H(y)\, dy \tag{3.16}$$

$$\le \frac{1}{H(a)} \int_0^a \left(\vartheta_0^2 \psi_1(y) + 2\psi_3(y)\right) H(y)\, dy$$

$$\mathbb{E} \int_b^\infty n \left[I\left(\widehat{H}_n(y) \le 0\right) - I(H(y) \le 0) \right] u(y) H(y)\, dy \tag{3.17}$$

$$\le \int_b^\infty \frac{\sigma^2(y)}{|H(y)|}\, dy$$

$$\mathbb{E} \int_\alpha^\beta n \left[I\left(\widehat{H}_n(y) \le 0\right) - I(H(y) \le 0) \right] u(y) H(y)\, dy \tag{3.18}$$

$$\le \frac{1}{d^2 n} \int_\alpha^\beta \mathbb{E}\, (W_1(y) - H(y))^4\, u(y)\, dy$$

$$+ \frac{3}{d^2 n} \int_\alpha^\beta \left(\sigma^2(y)\right)^2 u(y)\, dy =: \frac{1}{d^2 n} C(\alpha, \beta)\ .$$

Proof :
We have $\left[I\left(\widehat{H}_n(y) \le 0\right) - I(H(y) \le 0) \right] H(y) \ge 0$. By Tschebyshev's inequality

$$\begin{aligned}
&\mathbb{E}\left[I\left(\widehat{H}_n(y) \le 0\right) - I(H(y) \le 0) \right] H(y) \\
&\quad = \mathbb{E}\left[I\left(\widehat{H}_n(y) - H(y) \le -H(y)\right) - I(H(y) \le 0) \right] H(y) \\
&\quad \le \mathbb{E}\left[I\left(\left|\widehat{H}_n(y) - H(y)\right| \ge |H(y)|\right) \right] |H(y)| \\
&\quad \le \frac{\mathbb{E}\left|\widehat{H}_n(y) - H(y)\right|^k}{|H(y)|^{k-1}}\ . \qquad (3.19)
\end{aligned}$$

This inequality for $k = 2$ implies (3.17). The inequality (3.16) follows now from (3.14) and $H(y) \ge H(a)$ as H is positive and decreasing for $0 < y < \eta$. To prove (3.18) we apply (3.19) for $k = 4$ and use (3.15). □

Now we study the integrand in an interval which contains the zero η of H. Denote λ the unique minimum point of H. Assume $0 < \varkappa_1 < \eta < \varkappa_2 \le \lambda$. Then H is strictly decreasing in $[\varkappa_1, \varkappa_2]$. Denote by $\varkappa$ the inverse function and note that both H and $\varkappa$ are continuously differentiable in $(\varkappa_1, \varkappa_2)$. For $S_n(y)$ from (3.11). We set

$$\|S_n\| = \sup_{\varkappa_1 \le y \le \varkappa_2} |S_n(y)| \ .$$

Note that by Theorem 1 the sequence $\|S_n\|$ is stochastically bounded, i. e.

$$\lim_{c\to\infty} \limsup_{n\to\infty} P\left(\|S_n\| > c\right) = 0 \ . \tag{3.20}$$

We have

$$\int_{\varkappa_1}^{\varkappa_2} \left[I\left(\widehat{H}_n(y) < 0\right) - I(H(y) < 0) \right] u(y)H(y)\, dy$$

$$= \int_{\gamma_1}^{\gamma_2} \left[I\left(s + \frac{1}{\sqrt{n}} S_n(\varkappa(s)) < 0\right) - I(s < 0)\right] su(\varkappa(s))\, |\varkappa'(s)|\, ds$$

where $\gamma_2 = H(\varkappa_1), \gamma_1 = H(\varkappa_2)$ and $\gamma_1 < 0 < \gamma_2$. It holds

$$\left[I\left(s + \frac{1}{\sqrt{n}} S_n(\varkappa(s)) < 0\right) - I(s < 0)\right] s \ge 0$$

and

$$\left[I\left(s + \frac{1}{\sqrt{n}} S_n(\varkappa(s)) < 0\right) - I(s < 0)\right] s = 0$$

if $|s| > \frac{1}{\sqrt{n}}\|S_n\|$. We have for any real numbers $x, y \in S$ with $|x - y| \le w$

$$|I(s + x < 0) - I(s + y < 0)| \le I(-x - w \le s \le -x + w) \ .$$

Set

$$\omega_n = \sup_{|s| \le \frac{1}{\sqrt{n}}\|S_n\|} |S_n(\varkappa(s)) - S_n(\varkappa(0))| \, . \tag{3.21}$$

Then with $w = \frac{1}{\sqrt{n}}\omega_n$

$$n\left|\int_{\gamma_1}^{\gamma_2}\left\{\left[I\left(s+\frac{1}{\sqrt{n}}S_n(\varkappa(s))<0\right)-I(s<0)\right]\right.\right.$$
$$\left.\left.-\left[I\left(s+\frac{1}{\sqrt{n}}S_n(\varkappa(0))<0\right)-I(s<0)\right]\right\}su(\varkappa(s))|\varkappa'(s)|\,ds\right|$$
$$\le n\int I\Bigg(-\frac{1}{\sqrt{n}}S_n(\varkappa(0))-\frac{1}{\sqrt{n}}\omega_n\le s$$
$$\le -\frac{1}{\sqrt{n}}S_n(\varkappa(0))+\frac{1}{\sqrt{n}}\omega_n\Bigg)\ |su(\varkappa(s))\varkappa'(s)|\ ds$$
$$\le \frac{1}{\sqrt{n}}w_n\frac{1}{\sqrt{n}}\left(|S_n(\varkappa(0))|+w_n\right)C$$
$$\le C\omega_n\left(|S_n(\varkappa(0))|+\omega_n\right) \tag{3.22}$$

where

$$C=\sup_{\gamma_1\le s\le\gamma_2}|u(\varkappa(s))\varkappa'(s)|\ . \tag{3.23}$$

Consequently

$$n\int_{\gamma_1}^{\gamma_2}\left[I\left(s+\frac{1}{\sqrt{n}}S_n(\varkappa(s))<0\right)-I(s<0)\right]su(\varkappa(s))|\varkappa'(s)|\,ds$$
$$\le n\int_{\gamma_1}^{\gamma_2}\left[I\left(s+\frac{1}{\sqrt{n}}S_n(\varkappa(0))<0\right)-I(s<0)\right]su(\varkappa(s))|\varkappa'(s)|\,ds+R_{1,n}$$
$$= n\int_{\gamma_1}^{\gamma_2}\left[I\left(s+\frac{1}{\sqrt{n}}S_n(\varkappa(0))<0\right)-I(s<0)\right]su(\varkappa(0))\varkappa'(0)\,ds+R_{1,n}+R_{2,n}$$

if $\gamma_1 \le \frac{1}{\sqrt{n}}S_n(\varkappa(0)) \le \gamma_2$. In view of (3.22)

$$|R_{1.n}|\le Cw_n\left(|S_n(\varkappa(0))|+w_n\right)\ .$$

As

$$|u(\varkappa(s))\varkappa'(s)-u(\varkappa(0))\varkappa'(0)|\le L|s| \tag{3.24}$$

for every s with $\gamma_1 \le s \le \gamma_2$ with some L the remainder term $R_{2,n}$ fulfils

$$|R_{2,n}| \le n \int\limits_{\gamma_1}^{\gamma_2} \left| I\left(s + \frac{1}{\sqrt{n}} S_n(\varkappa(0)) < 0\right) - I(s<0) \right| Ls^2\, ds$$

$$\le n \int\limits_{-\frac{\|S_n\|}{\sqrt{n}}}^{\frac{\|S_n\|}{\sqrt{n}}} 2Ls^2\, ds = L\frac{2}{3}\frac{1}{\sqrt{n}}\|S_n\|^3 .$$

Lemma 2 *If* $0 < \varkappa_1 < \eta < \varkappa_2 \le \lambda$ *then for* $\gamma_1 \le \frac{1}{\sqrt{n}} S_n(\eta) \le \gamma_2$

$$\left| n \int\limits_{\varkappa_1}^{\varkappa_2} \left[I\left(H(y) + \frac{1}{\sqrt{n}} S_n(y) \le 0\right) - I(H(y) \le 0)\right] u(y)H(y)\, dy - \frac{u(\eta)}{2|H'(\eta)|} S_n^2(\eta) \right|$$

$$\le c\omega_n \left(|S_n(\eta)| + \omega_n\right) + \frac{2L}{3}\frac{1}{\sqrt{n}}\|S_n\|^3$$

where w_n, c *and* L *are defined in* (3.21), (3.23), (3.24).

4 Asymptotic distribution of the random regret risk

We suppose that $Y_{i1}, \ldots, Y_{in}$, $i = 1, \ldots, k$ is a sample from $\pi_1, \ldots, \pi_k$ and assume that the Y_{ij}, $j = 1, \ldots, n$ have the density (2.6). The prior distributions are supposed to fulfil

$$\int\limits_0^\infty \vartheta^4\, dG_i(\vartheta) < \infty . \tag{4.1}$$

Furthermore we suppose that the function ψ_{ia} introduced in (2.5) satisfy

$$\lim_{y \downarrow 0} \frac{\psi_{i2}(y)}{\psi_{i1}(y)} < \vartheta_0 < \lim_{y \uparrow \infty} \frac{\psi_{i2}(y)}{\psi_{i1}(y)} \tag{4.2}$$

for every $1 \le i \le k$. We suppose that for

$$\sigma_i^2(y) = V\left(\vartheta_0 I(Y_i > y) - (Y_i - y) I(Y_i > y)\right)$$

and every $1 \le i \le k$

$$\int\limits_{b_0}^\infty \frac{\sigma_i^2(y)}{|H_i(y)|} u(y)\, dy < \infty \tag{4.3}$$

for some $b_0 > \eta_i$, where η_i is the zero of $H_i(y)$. Recall that the random regret risk $\mathcal{R}_n = \mathcal{R}(d_n^*) - \mathcal{R}(d^0)$ of the Gupta-Liang selection rule d_n^* is given by

$$\mathcal{R}(d_n^*) - \mathcal{R}(d^0) = \sum_{i=1}^{k} \int_0^{\infty} \left[I\left(\widehat{H}_{i,n}(y) \le 0\right) - I(H_i(y) \le 0) \right] u(y) H_i(y)\, dy \;.$$

Now we are ready to formulate and to prove the main result of this paper.

Theorem 2 *Assume the conditions* (4.1), (4.2) *and* (4.3) *are fulfilled. Then*

$$\mathcal{L}\left(n(\mathcal{R}(d_n^*) - \mathcal{R}(d^0))\right) \Rightarrow \mathcal{L}\left(\sum_{i=1}^{k} \beta_i \chi_i^2\right)$$

where $\chi_1^2, \ldots, \chi_k^2$ *are independent r. v. where each has a* χ^2*-distribution with one degree of freedom and*

$$\beta_i = \frac{u(\eta_i)}{2H_i'(\eta_i)} \sigma_i^2(\eta_i) \;.$$

The proof is divided into several steps.

Proof :

1) If $U_n = X_{n,\varepsilon} + Y_{n,\varepsilon}$ are random variables with $\mathcal{L}(X_{n,\varepsilon}) \Rightarrow \mu$ for every $\varepsilon > 0$ and $\mathbb{E}|Y_{n,\varepsilon}| < \varepsilon$ then for every Lipschitz continuous function φ

$$\begin{aligned} |\varphi(x) - \varphi(y)| &\le L(|x-y|) \\ |\mathbb{E}\varphi(U_n) - \mathbb{E}\varphi(X_{n,\varepsilon})| &\le \varepsilon. \end{aligned}$$

Hence $\mathbb{E}\varphi(Z_n) \to \int \varphi \, d\mu$ which implies $\mathcal{L}(U_n) \Rightarrow \mu$.

2) As the populations π_i are independent, it is enough to show that

$$\mathcal{L}\left(n \int_0^{\infty} \left[I\left(\widehat{H}_{in}(y) \le 0\right) - I\left(H_i(y) \le 0\right) \right] u(y) H_i(y)\, dy \right) \Rightarrow \mathcal{L}\left(\beta_i \chi_i^2\right) \;.$$

To this end we choose for a given $\varepsilon > 0$ $\;0 < \alpha_\varepsilon < \eta_i < b_\varepsilon$ such that by (3.13) and (3.14)

$$\text{(4.4)} \qquad \mathbb{E}n \left| \int_0^{\alpha_\varepsilon} \left[I\left(\widehat{H}_{in}(y) < 0\right) - I(H_i(y) < 0) \right] u(y) H_i(y)\, dy \right| < \frac{\varepsilon}{2}$$

$$\text{(4.5)} \qquad \mathbb{E}n \left| \int_{b_\varepsilon}^{\infty} \left[I\left(\widehat{H}_{in}(y) < 0\right) - I(H_i(y) < 0) \right] u(y) H_i(y)\, dy \right| < \frac{\varepsilon}{2}$$

for every $n = 1, 2, \ldots$. Now we choose $\alpha_\varepsilon < \varkappa_1 < \eta_i < \varkappa_2 < b_\varepsilon$ such that H_i is strictly decreasing in $[\varkappa_1, \varkappa_2]$. Note that by (3.15)

$$\text{(4.6)} \qquad \mathbb{E}n \left| \left(\int_{\alpha_\varepsilon}^{\varkappa_1} + \int_{\varkappa_2}^{b_\varepsilon} \right) \left[I\left(\widehat{H}_{in}(y) < 0\right) - I(H_i(y) < 0) \right] u(y) H_i(y)\, dy \right|$$
$$\leq \frac{1}{nd^2} \left[C(\alpha_\varepsilon . \varkappa_1) + C(\varkappa_2, b_\varepsilon) \right]$$

where

$$d = \inf_{y \in [\alpha_\varepsilon, \varkappa_1] \cup [\varkappa_2, b_\varepsilon]} |H(y)| \,.$$

Theorem 1 implies that $\|S_n\|$ is stochastically bounded. Hence $P\left(\gamma_1 \leq \frac{1}{\sqrt{n}} S_n(\eta) \leq \gamma_2\right) \xrightarrow[n \to \infty]{} 1$. If we combine (3.10) and (3.20) we get that ω_n tends stochastically to zero. Hence the right hand term in the inequality of Lemma 2 tends stochastically to zero. This means that in view of (4.6)

$$X_{n,\varepsilon} = n \int_{\alpha_\varepsilon}^{b_\varepsilon} \left[I\left(\widehat{H}_{in}(y) \leq 0\right) - I(H_i(y) \leq 0) \right] u(y) H_i(y)\, dy$$

and $\frac{u(\eta_i)}{2|H_i'(\eta_i)|}$, $S_n^2(\eta_i)$ have the same limit distribution, if there is any. By the central limit theorem $S_n(\eta_i)$ tends in distribution to a normal distribution with expectation zero and variance

$$\sigma_i^2(\eta_i) = \mathrm{V}\left(\frac{1}{u(Y_{i1})} \left(\vartheta_0 - (Y_{i1} - \eta_i)\right) I(Y_{i1} > \eta) \right) \,.$$

This means that

$$\mathcal{L}\left(X_{n,\varepsilon}\right) \Rightarrow \mathcal{L}(\beta_i \chi_i^2)$$

where χ_i^2 has a χ^2-distribution with one degree of freedom. To complete the proof we set

$$Y_{n,\varepsilon} = n \left(\int_0^{\alpha_\varepsilon} + \int_{b_\varepsilon}^{\infty} \right) \left[I\left(\widehat{H}_n(y) \leq 0\right) - I(H(y) \leq 0) \right] u(y) H(y)\, dy$$

and apply the first part of the proof. □

References

Balakrishnan, N. and Basu, A. P. (eds.) (1995). The Exponential Distribution: Theory, Method and Applications. Gordon and Breach Publishers. Langliorne, Pennsylvania.

Billingsley, P. (1968). Convergence of Probability Measures., J. Wiley & Sons, New York.

Deely, J. J. (1965). Multiple decision procedures from empirical Bayes approach, Ph.D. Thesis (Mimeo. Ser. No. 45). Department of Statistics, Purdue University, West Lafayette, Indiana.

Gupta, S. S. and Panchapakesan, S. (1985). Subset selection procedures: review and assessment. *Amer. J. Management Math. Sci.*, 5, 235–311.

Gupta, S. S. and Liang T. (1986). Empirical Bayes rules for selecting the best binomial population. *Statistical Decision Theory and Related Topics IV* (Eds. S. S. Gupta and J. O. Berger) Vol. 1, 213–224, Springer-Verlag.

Gupta, S. S. and Liang, T. (1994). On empirical Bayes selection rules for sampling inspection. *J. Statist. Plann. Inference*, 38, 43–64.

Gupta, S. S., Liang, T. and Rau, R.-B. (1994a). Empirical Bayes two stage procedures for selecting the best Bernoulli population compared with a control. *Statistical Decision Theory and Related Topics V.* (Eds. S. S. Gupta and J. O. Berger), 277–292, Springer Verlag.

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

Gupta, S. S., Liang, T. (1996). Selecting good exponential populations compared with a control: nonparametric empirical Bayes approach, Technical Report # 96-18C, Center for Statistical Decision Sciences and Department of Statistics.

Gupta, S. S., Liese, F. (1998). Asymptotic distribution of the random regret risk for selecting exponential populations. Technical Report # 98-08C, Center for Statistical Decision Sciences and Department of Statistics.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1994). Continuous Univariate Distributions, Vol. 1, Second Ed., John Wiley & Sons, New York.

Jurečkova, J. and Sen, P. K. (1996). Robust Statistical Procedures, Asymptotics and Interrelations, J. Wiley & Sons, New York.

Liese, F. and Vajda, I. (1994). Consistency of M-estimates in general regression models. *Journ. Multiv. Analysis*, 50, 93–114.

Pfanzagl, J. (1969). On measurability and consistency of minimum contrast estimators, *Metrika*, 14, 249–272.

Pollard, D. (1991). Asymptotics for least absolute deviation regression estimators. *Econometric Theory*, 7, 186–199.

Robbins, H. (1956). An empirical Bayes approach to statistics. Proc. Third Berkeley Symp., *Math. Statist. Probab.* 1, 157–163, Univ. of California Press.

Skorack, G. R. and Wellner, J. A. (1986). Empirical Processes with Applications to Statistics. J. Wiley & Sons, New York.

Wald, A. (1949). Note on the consistency of the maximum likelihood estimate. *Amer. Math. Stat.* 20, 595–601.

Van der Vaart, A. W. and Wellner, J. A. (1996). Weak Convergence and Ernpirical Processes. Springer Series in Statistics, Springer Verlag.