Kybernetika

Shanti S. Gupta; Friedrich Liese

Asymptotic distribution of the conditional regret risk for selecting good exponential populations

Kybernetika, Vol. 36 (2000), No. 5, [571]--588

Persistent URL: http://dml.cz/dmlcz/135372

KYBERNETIKA — VOLUME 36 (2000), NUMBER 5, PAGES 571–588

ASYMPTOTIC DISTRIBUTION OF THE CONDITIONAL REGRET RISK FOR SELECTING GOOD EXPONENTIAL POPULATIONS[1]

SHANTI S. GUPTA AND FRIEDRICH LIESE[2]

In this paper empirical Bayes methods are applied to construct selection rules for the selection of all good exponential distributions. We modify the selection rule introduced and studied by Gupta and Liang [10] who proved that the regret risk converges to zero with rate $O(n^{-\lambda/2}), 0 < \lambda \leq 2$. The aim of this paper is to study the asymptotic behavior of the conditional regret risk $\mathcal{R}_n$. It is shown that $n\mathcal{R}_n$ tends in distribution to a linear combination of independent χ^2-distributed random variables. As an application we give a large sample approximation for the probability that the conditional regret risk exceeds the Bayes risk by a given $\varepsilon > 0$. This probability characterizes the information contained in the historical data.

1. INTRODUCTION

The family of exponential 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 [12] and Balakrishnan and Basu [1]. We consider k independent exponential populations $\pi_1, \dots, \pi_k$ with expectations $\theta_1, \dots, \theta_k$ which are unknown. But a control value θ_0 is given. Each population π_i is called good if $\theta_i \geq \theta_0$ and bad otherwise. We study the problem of finding all good populations. This is a typical subset selection problem, see Gupta and Panchapakesan [4]. We use the Bayes approach and assume that the θ_i are realizations of the random variables Θ_i with distributions G_i. Then for a given loss function the best selection rule, being the Bayes selection rule, depends on the distributions G_i. We allow that the distributions G_i are not known. But we suppose that historical data are available and can be included in the decision rule. This is the empirical Bayes approach due to Robbins [17]. Empirical Bayes methods have been applied in different areas of statistics. Deely [2] constructed empirical Bayes subset selection procedures. In a series of papers Gupta and Liang [5, 6, 10]

[1]This research was supported in part by US Army Research Office, Grant DAAH04-95-1-0165 Purdue University.

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

and Gupta, Liang and Rau [7], [8] have studied different selection procedures using the empirical Bayes approach.

Assume $\underline{Y} = (Y_1, \dots, Y_k)$ are the actual data from the populations $\pi_1, \dots, \pi_k$ which are used to make a decision. If L is a given loss function then the risk of the selection rule d is $\mathbb{R}(d) = \mathbb{E}L(d(\underline{Y}), \underline{\Theta})$ where $\underline{\Theta} = (\Theta_1, \dots, \Theta_k)$. The optimal selection rule d_G^0 is easily seen to depend on the unknown joint distribution $G = \prod_{i=1}^k G_i$ of $\underline{\Theta}$. The central idea of the empirical Bayes approach is the construction of a good decision rule d_n^* on the basis of historical data $\underline{X}_n$ being independent of $(\underline{Y}, \underline{\Theta})$. The quality of d_n^* is characterized by the *overall Bayes risk* $\mathbf{R}(d_n^*) = \mathbb{E}L(d_n^*(\underline{X}_n, \underline{Y}), \underline{\Theta})$ and the nonnegative *regret risk* $\mathbf{R}(d_n^*) - \mathbb{R}(d_G^0)$. The aim of the above mentioned papers dealing with empirical Bayes methods was to construct suitable decision rules d_n^* and to evaluate the regret risk. The main goal was to prove the convergence to zero of $\mathbf{R}(d_n^*) - \mathbb{R}(d_G^0)$ at a certain rate. Gupta and Liang [10] constructed for the problem of selecting good exponential populations a selection rule d_n^* and proved $\mathbf{R}(d_n^*) - \mathbb{R}(d_G^0) = O(n^{-\lambda/2})$ where the value of the parameter $0 < \lambda \leq 2$ depends on additional assumptions.

Denote by $\mathbb{E}_{\underline{X}_n}$ and $\mathbb{E}_{(\underline{Y}, \underline{\Theta})}$, the expectation with respect to $\underline{X}_n$ and $(\underline{Y}, \underline{\Theta})$, respectively. By the independence of $\underline{X}_n$ and $(\underline{Y}, \underline{\Theta})$ the risk $\mathbb{R}(d_n^*)$ of the empirical Bayes selection rule d_n^* is a random variable which may be written in the form

$$\mathbb{R}(d_n^*) = \mathbb{E}_{(\underline{Y}, \underline{\Theta})} L(d_n^*(\underline{X}_n, \underline{Y}), \underline{\Theta}). \tag{1}$$

The difference

$$\begin{aligned} \mathcal{R}_n &= \mathbb{R}(d_n^*) - \mathbb{R}(d_G^0) \\ &= \mathbb{E}_{(\underline{Y}, \underline{\Theta})}(L(d_n^*(\underline{X}_n, \underline{Y}), \underline{\Theta}) - L(d_G^0, \underline{Y})) \end{aligned} \tag{2}$$

is called the *conditional regret risk*, so that the regret risk is the expectation of the conditional regret risk with respect of the previous data

$$\begin{aligned} \mathbf{R}(d_n^*) - \mathbb{R}(d_G^0) &= \mathbb{E}(L(d_n^*(\underline{X}_n, \underline{Y}), \underline{\Theta}) - L(d_G^0, \underline{Y})) \\ &= \mathbb{E}_{\underline{X}_n}\left[\mathbb{E}_{(\underline{Y}, \underline{\Theta})}(L(d_n^*(\underline{X}_n, \underline{Y}), \underline{\Theta}) - L(d_G^0, \underline{Y}))\right] \\ &= \mathbb{E}_{\underline{X}_n} \mathcal{R}_n. \end{aligned}$$

When we study the asymptotic behavior of $\mathcal{R}_n$ and ask for the asymptotic distribution, the situation is comparable with the asymptotic theory of parameter estimation where different types of estimators are compared by the limit distribution of $\sqrt{n}(\widehat{\theta}_n - \theta_0)$ in general and not by a direct evaluation of the variance of $\widehat{\theta}_n$. So, in this paper we study distributions instead of expectations. A new selection rule $\widehat{d}_n$ for the problem of selecting a good exponential distribution is introduced by a modification of the Gupta and Liang rule [10]. We show that $n\mathcal{R}_n$ converges in distributions to a linear combination of independent χ^2-distributed random variables each with one degree of freedom. The coefficients in the linear combination are explicitly calculated. The main idea for the new selection rule is to use the fact that the construction of the optimal selection rule d_G needs only the value of the unique zero η_{i0} of some function m_i which depends on G_i and is consequently unknown. The main part of this paper is the construction of a suitable estimator $\widehat{\eta}_{in}$ for η_{i0}

and the proof of a limit theorem for $\sqrt{n}(\widehat{\eta}_{in} - \eta_{i0})$. The problem of estimating the zero of an unknown function is studied in part 3 of the paper in a general setting and is then applied to the selection problem. We transform the problem of estimation of the zero of the function m_i into the problem of finding the point at which the integrated function has a local maximum. To this end we apply general ideas and techniques from the theory of empirical processes.

2. FORMULATION OF THE SELECTION PROBLEM

Consider k independent exponential populations $\pi_1, \dots, \pi_k$ which we assume to have the density functions $g(y_i|\theta_i) = I_{[0,\infty)}(y_i)\frac{1}{\theta_i}\exp\{-\frac{y_i}{\theta_i}\}$, $i = 1, \dots, k$. Here I_A is the indicator function of the set A. Set $\underline{\theta} = (\theta_1, \dots, \theta_k) \in \Omega = (0,\infty)^k$. Given a standard value $\theta_0 > 0$ we call a population π_j good if $\theta_j \geq \theta_0$. Our aim is to select all good populations. Therefore the decision space is $D = \{0,1\}^k = \{(a_1, \dots, a_k) : a_i \epsilon \{0,1\}\}$ and π_i is selected if and only if $a_i = 1$. Similar as in Gupta and Liang [6, 10] we use the loss function

$$L(\underline{\theta}, \underline{a}) = \sum_{i=1}^{k} \ell(\theta_i, a_i) \tag{3}$$

where

$$\ell(\theta_i, a_i) = a_i\theta_i(\theta_0 - \theta_i)I_{(0,\theta_0)}(\theta_i) + (1 - a_i)\theta_i(\theta_i - \theta_0)I_{(\theta_0,\infty)}(\theta_i). \tag{4}$$

If $\theta_i \geq \theta_0$ so that the population π_i is good and we make a false decision, i. e. $a_i = 0$, then the penalty increases with the distance of θ_i from θ_0. For a bad population π_i and a false decision we get the loss $\theta_i(\theta_0 - \theta_i)$ which is an increasing function on $(0, \frac{1}{2}\theta_0)$ and decreasing for $(\frac{1}{2}\theta_0, \theta_0)$. The behavior on $(0, \frac{1}{2}\theta_0)$ is an unpleasant side effect of the loss function $l(\theta_i, a_i)$ which was mainly motivated by the fact that it allows the construction of an unbiased estimator for an unknown function which provides the optimal decision rule d_G^0, see (18). If the r.v. Θ_i take values in $(0, \frac{1}{2}\theta_0)$ only with small probabilities then the behavior of $l(\theta_i, 1)$ has small influence on the risk. Hence the use of the loss function $l(\theta_i, a_i)$ is reasonable if the populations are not too bad.

By a selection rule $d = (a_1, \dots, a_k)$ we shall mean a measurable mapping of the sample space $\mathcal{Y} = (0,\infty)^k$ into the decision space D. As we will apply the Bayes and the empirical Bayes approach to the selection problem we assume that the θ_i are realizations of independent random variables Θ_i. The Θ_i are assumed to take values in $(0,\infty)$ and have distribution G_i. The distribution G of the random vector $\underline{\Theta} = (\Theta_1, \dots, \Theta_k)$ is then the product of the G_i.

The random variables $\Theta_1, \dots, \Theta_k$ are not observable, but we can observe $\underline{Y} = (Y_1, \dots, Y_k)$ where the $Y_1, \dots, Y_k$ are independent and the conditional density of Y_i given $\Theta_i = \theta_i$ is $g(y_i|\theta_i) = I_{[0,\infty)}(y_i)\frac{1}{\theta_i}\exp\{-\frac{y_i}{\theta_i}\}$

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

$$\mathbb{R}(d) = \mathbb{E}L(d(\underline{Y}), \underline{\Theta}). \tag{5}$$

In terms of densities the risk can be written in the form

$$\mathbb{R}(d) = \int \left(\int L(\underline{\theta}, \mathrm{d}(\underline{y})) g(\underline{y}|\underline{\theta}) \, \mathrm{d}\underline{y} \right) \mathrm{d}G(\underline{\theta})$$

where $g(\underline{y}|\underline{\theta}) = \prod_{i=1}^{k} g(y_i|\theta_i)$ and $\mathrm{d}\underline{y} = \mathrm{d}y_1, \ldots \mathrm{d}y_k$. Using the loss function $L(\underline{\theta}, \underline{a})$ (3) we get

$$\mathbb{R}(d) = \sum_{i=1}^{k} \mathbb{E}\ell(\Theta_i, q_i(Y_i)) \tag{6}$$

where

$$q_i(y_i) = \mathbb{E}a_i(Y_1, \ldots, Y_{i-1}, y_i, Y_{i+1}, \ldots, Y_k).$$

The formula (6) shows that due to the special structure of the loss function we may restrict ourselves to randomized decisions q_i which depend on the data of π_i only. We assume

$$\int_0^\infty \theta_i^2 \, \mathrm{d}G_i(\theta_i) < \infty, \quad i = 1, \ldots, k. \tag{7}$$

Then the Bayes risk $\mathbb{R}(d) = \mathbb{E}L(d(\underline{Y}), \underline{\Theta})$ is finite and the following holds

$$\begin{aligned} \mathbb{R}(d) &= \sum_{i=1}^{k} \int_0^\infty \int_0^\infty q_i(y_i)(\theta_0 - \theta_i) e^{-\frac{y_i}{\theta_i}} \, \mathrm{d}G_i(\theta_i) \, \mathrm{d}y_i + \gamma_i \\ &= \sum_{i=1}^{k} \int_0^\infty q_i(y_i) m_i(y_i) \, \mathrm{d}y_i + \gamma_i \end{aligned} \tag{8}$$

where

$$\gamma_i = \int_0^\infty \theta_i(\theta_i - \theta_0) I_{(\theta_0, \infty)}(\theta_i) \, \mathrm{d}G_i(\theta_i)$$

$$m_i(y_i) = \int_0^\infty (\theta_0 - \theta_i) e^{-\frac{y_i}{\theta_i}} \, \mathrm{d}G_i(\theta_i).$$

Using the relation (8) we see that $\inf_d \mathbb{R}(d)$ is attained by the selection rule $d_G^0 = (d_1^0, \ldots, d_k^0)$, where

$$d_i^0(y_i) = \begin{cases} 1 & \text{if } m_i(y_i) < 0 \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

As in Gupta and Liang [10] one obtains by integration by parts

$$m_i(y_i) = \theta_0 \psi_{i1}(y_i) - \psi_{i2}(y_i) \tag{10}$$

where

$$\begin{aligned}
\psi_{i1}(y_i) &= \int_0^\infty e^{-\frac{y_i}{\theta_i}}\,\mathrm{d}G_i(\theta_i) \\
&= \int_0^\infty \int_{y_i}^\infty \frac{1}{\theta_i} e^{-\frac{t_i}{\theta_i}}\,\mathrm{d}t_i\,\mathrm{d}G_i(\theta_i) = \mathbb{E}I_{[0,\infty)}(Y_i - y_i) && (11)\\
\psi_{i2}(y_i) &= \int_0^\infty \theta_i e^{-\frac{y_i}{\theta_i}}\,\mathrm{d}G_i(\theta_i) \\
&= \mathbb{E}(Y_i - y_i) I_{[0,\infty)}(Y_i - y_i). && (12)
\end{aligned}$$

We see from (9) that the Bayes selection rule d_i^0 is trivial, i.e. takes on the value 0 or 1 for every y_i, unless m_i has a zero on $(0,\infty)$. To give necessary and sufficient conditions for the existence of a zero we need some auxiliary results. For any not necessarily finite measure μ on the Borel sets of $(0,\infty)$ we set

$$\kappa_\mu(y) = \int_{(0,\infty)} \exp\left\{-\frac{y}{\theta}\right\} \mu(\mathrm{d}\theta)$$

and assume that $\kappa_\mu(y) < \infty$ for every $y > 0$. Then by Hölder's inequality for $0 < \alpha < 1$

$$\kappa_\mu(\alpha y_1 + (1-\alpha) y_2) \le [\kappa_\mu(y_1)]^\alpha\, [\kappa_\mu(y_2)]^{1-\alpha}.$$

Consequently, $\ln \kappa_\mu$ is convex and $\frac{\kappa'_\mu}{\kappa_\mu}$ is nondecreasing. Moreover, if μ is nondegenerate, i.e. μ is not concentrated at one point, then $\ln \kappa_\mu$ is strictly convex and $\frac{\kappa'_\mu}{\kappa_\mu}$ strictly increasing. Put for any $a > 0$

$$m_\mu(y) = -a\kappa'_\mu(y) - \kappa_\mu(y).$$

As $\kappa_\mu(y) > 0$ and $\frac{\kappa'_\mu}{\kappa_\mu}$ is strictly increasing we see that m_μ has at most one zero on $(0,\infty)$. The derivative m'_μ of m_μ can be written as

$$m'_\mu = -m'_{\widetilde{\mu}},$$

where the measure $\widetilde{\mu}$ is defined by $\widetilde{\mu}(\mathrm{d}\theta) = \frac{1}{\theta}\mu(\mathrm{d}\theta)$. Consequently, m'_μ has again at most one zero on $(0,\infty)$ and the same statement holds also for every derivative of m_μ of higher order. As $\kappa_\mu(y) > 0$ and $\frac{\kappa'_\mu}{\kappa_\mu}$ is strictly increasing a zero of m_μ exists on $(0,\infty)$ if and only if

$$\lim_{y\downarrow 0} \frac{\kappa'_\mu(y)}{\kappa_\mu(y)} < -\frac{1}{a} < \lim_{y\uparrow\infty} \frac{\kappa'_\mu(y)}{\kappa_\mu(y)}. \tag{13}$$

The limit on the left hand side is

$$\lim_{y\downarrow 0} \frac{\kappa'_\mu(y)}{\kappa_\mu(y)} = -\frac{\int_{(0,\infty)} \frac{1}{\theta}\mu(\mathrm{d}\theta)}{\mu((0,\infty))} \tag{14}$$

provided that both denominator and the numerator are finite. The calculation of the limit on the right hand side of (13) requires more effort. We set for any finite measure μ

$$b_\mu := \inf\{t : \mu((t,\infty)) = 0\},$$

where we have used the convention $\inf \emptyset = \infty$. If $b_\mu < \infty$ then the measure μ is concentrated on $(0, b_\mu]$.

Lemma 1. For any measure μ on the Borel sets of $(0,\infty)$ with $0 < \mu((0,\infty)) < \infty$, it holds

$$\lim_{y\uparrow\infty} \frac{\kappa'_\mu(y)}{\kappa_\mu(y)} = -\frac{1}{b_\mu} \tag{15}$$

where $\frac{1}{b_\mu} = 0$ for $b_\mu = \infty$.

Proof. First, we introduce the abbreviation

$$\begin{aligned} J_b^{(1)}(y) &= \int_{(0,b)} \frac{1}{\theta} \exp\left\{-y\left(\frac{1}{\theta} - \frac{1}{b}\right)\right\} \mu(\mathrm{d}\theta) \\ J_b^{(2)}(y) &= \int_{[b,\infty)} \frac{1}{\theta} \exp\left\{-y\left(\frac{1}{\theta} - \frac{1}{b}\right)\right\} \mu(\mathrm{d}\theta) \\ K_b^{(1)}(y) &= \int_{(0,b)} \exp\left\{-y\left(\frac{1}{\theta} - \frac{1}{b}\right)\right\} \mu(\mathrm{d}\theta) \\ K_b^{(2)}(y) &= \int_{[b,\infty)} \exp\left\{-y\left(\frac{1}{\theta} - \frac{1}{b}\right)\right\} \mu(\mathrm{d}\theta). \end{aligned}$$

Fix $0 < b < b_\mu$ and notice that for $0 < \theta \le b$

$$\begin{aligned} \frac{1}{\theta}\exp\left\{-y(\frac{1}{\theta} - \frac{1}{b})\right\} &\le \frac{1}{\theta}\exp\left\{-\frac{y}{2\theta}\right\} && \text{for } 0 < \theta \le \frac{b}{2} \\ \frac{1}{\theta}\exp\left\{-y(\frac{1}{\theta} - \frac{1}{b})\right\} &\le \frac{2}{b} && \text{for } \frac{b}{2} < \theta. \end{aligned}$$

Consequently, by the Theorem of Lebesgue

$$\lim_{y\to\infty} J_b^{(1)}(y) = 0, \qquad \lim_{y\to\infty} K_b^{(1)}(y) = 0. \tag{16}$$

For every $b < b_\mu$ it holds $K_b^{(2)}(y) > 0$, so that

$$\begin{aligned} 0 &\le \limsup_{y\to\infty}\left[-\frac{\kappa'_\mu(y)}{\kappa_\mu(y)}\right] = \limsup_{y\to\infty} \frac{J_b^{(1)}(y) + J_b^{(2)}(y)}{K_b^{(1)}(y) + K_b^{(2)}(y)} \\ &= \limsup_{y\to\infty} \frac{J_b^{(2)}(y)}{K_b^{(2)}(y)} \\ &\le \limsup_{y\to\infty} \frac{\frac{1}{b}K_b^{(2)}(y)}{K_b^{(2)}(y)} \le \frac{1}{b}. \end{aligned}$$

Taking $b \uparrow b_\mu$ we get

$$\limsup_{y\to\infty}\left[-\frac{\kappa'_\mu(y)}{\kappa_\mu(y)}\right] \le \frac{1}{b_\mu}$$

which completes the proof if $b_\mu = \infty$ as $\kappa'_\mu(y) \le 0$. If $b_\mu < \infty$ we get from (16)

$$\liminf_{y\to\infty}\left[-\frac{\kappa'_\mu(y)}{\kappa_\mu(y)}\right] \ge \liminf_{y\to\infty}\frac{\frac{1}{b_\mu}K_b^{(2)}(y)}{K_b^{(2)}(y)} \ge \frac{1}{b_\mu}$$

which completes the proof in the case $b_\mu < \infty$. □

The previous Lemma is now applied to get conditions under which the Bayes selection procedures are nontrivial. Let the measure μ be defined by

$$\mu(B) = \int_B \theta \, \mathrm{d}G_i(\theta).$$

The assumption (7) implies that μ is a finite measure. Furthermore,

$$b_\mu = b_G.$$

By the definition of ψ_{i1} and ψ_{i2} in (11) and (12), respectively

$$\frac{\kappa'_\mu(y)}{\kappa_\mu(y)} = -\frac{\psi_{i1}}{\psi_{i2}}.$$

This yields

$$\lim_{y\downarrow 0}\frac{\kappa'_\mu(y)}{\kappa_\mu(y)} = -\frac{1}{\mathbb{E}\Theta_i}$$

and by Lemma 1

$$\lim_{y\to\infty}\frac{\kappa'_\mu(y)}{\kappa_\mu(y)} = -\frac{1}{b_{G_i}}.$$

As the Bayes selection rule d_i^0 is nontrivial if and only if the function m_i has a zero on $(0,\infty)$ we get from (13) the following statement.

Proposition 2. If the distribution G_i is nondegenerate and has finite second moment then the Bayes selection rule d_i^0 is nontrivial if and only if

$$\mathbb{E}\Theta_i < \theta_0 < b_{G_i}. \tag{17}$$

Now we discuss the cases in which condition (17) is not fulfilled. If $\theta_0 \le \mathbb{E}\Theta_i \le b_{G_i}$ then the decreasing function $\theta_0\frac{\psi_{i1}}{\psi_{i2}} - 1$ is less or equal to zero. Hence

$$m_i(y_i) = \psi_{i2}(y_i)\left(\theta_0\frac{\psi_{i1}(y_i)}{\psi_{i2}(y_i)} - 1\right) \le 0$$

and $d_i^0(y_i) \equiv 1$. Conversely, if $\mathbb{E}\Theta_i \le b_{G_i} \le \theta_0$ then by Lemma 1

$$\begin{aligned}\left(\theta_0 \frac{\psi_{i1}(y_i)}{\psi_{i2}(y_i)} - 1\right) &\ge \lim_{y_i \to \infty} \left(\theta_0 \frac{\psi_{i1}(y_i)}{\psi_{i2}(y_i)} - 1\right) \\ &= \frac{\theta_0}{b_{G_i}} - 1 \ge 0\end{aligned}$$

so that

$$m_i(y_i) = \psi_{i2}(y_i)\left(\theta_0 \frac{\psi_{i1}(y_i)}{\psi_{i2}(y_i)} - 1\right) \ge 0$$

and $d_i^0(y_i) \equiv 0$. This is also intuitively clear as for $b_{G_i} \le \theta_0$ the random variable Θ_i is bounded above by θ_0.

We illustrate the condition (17) by an example.

Let

$$\gamma_{\alpha_i,\beta_i}(s) = I_{[0,\infty)}(s) \frac{\beta_i^{\alpha_i}}{\Gamma(\alpha_i)} s^{\alpha_i - 1} e^{-\beta_i s}$$

$\alpha_i > 1, \beta_i > 0$, be the density of a gamma distribution. We assume that $\frac{1}{\Theta_i}$ has the density $\gamma_{\alpha_i,\beta_i}$ or equivalent that Θ_i has the density $\frac{1}{t^2}\gamma_{\alpha_i,\beta_i}(\frac{1}{t})$. Then

$$f_i(y_i) = \mathbb{E}\frac{1}{\Theta_i} e^{-\frac{y_i}{\Theta_i}} = \frac{\alpha_i \beta_i^{\alpha_i}}{(y_i + \beta_i)^{\alpha_i + 1}}.$$

Similarly,

$$\psi_{i1}(y_i) = \mathbb{E}e^{-\frac{y_i}{\Theta_i}} = \frac{\beta_i^{\alpha_i}}{(y_i + \beta_i)^{\alpha_i}}$$

and

$$\psi_{i2}(y_i) = \mathbb{E}\Theta_i e^{-\frac{y_i}{\Theta_i}} = \frac{\beta_i^{\alpha_i}}{(\alpha_i - 1)(y_i + \beta_i)^{\alpha_i - 1}}.$$

Consequently, $-\frac{\psi_{i1}(y_i)}{\psi_{i2}(y_i)} = \frac{1-\alpha_i}{y_i+\beta_i}$ and

$$m_i(y_i) = \frac{\beta_i^{\alpha_i}}{(y_i + \beta_i)^{\alpha_i - 1}} \left(\frac{\theta_0}{y_i + \beta_i} - \frac{1}{\alpha_i - 1}\right)$$

so that the zero η_{0i} of m_i is $\eta_{i0} = \theta_0(\alpha_i - 1) - \beta_i$. Otherwise, $\mathbb{E}\Theta_i = \frac{\beta_i}{\alpha_i - 1}$ This means that the zero η_{0i} is positive if the expectation $\mathbb{E}\Theta_i$ is smaller than the critical value θ_0 and this is the statement of Proposition 2 if we take into account that $b_{G_i} = \infty$.

To apply the selection rule d_i^0 from (9) we need the zero η_{i0} of $m_i(y_i) = \theta_0\psi_{i1} - \psi_{i2}$. But ψ_{i1}, ψ_{i2} as well as η_{i0} depend on the unknown prior distribution G_i. Otherwise, the unknown function $m_i = \theta_0\psi_{i1} - \psi_{i2}$ is the expectation of a function of the observable data Y_i. Indeed, we get from (11) and (12)

$$m_i(y_i) = \mathbb{E}h(Y_i - y_i) \tag{18}$$

where $h(t) = (\theta_0 - t)I_{[0,\infty)}(t)$. The relation (18) connects the unknown function m_i with the observable data Y_i and is the key of the empirical Bayes methods in our model. Assume we have data from the past which can be taken into the decision. More precisely, we assume that for $i = 1, \dots, k$ the random variables $X_{i1}, \dots, X_{in}$ are i.i.d. with common density given by

$$f_i(y_i) = \int_0^\infty \frac{1}{\theta_i} e^{-\frac{y_i}{\theta_i}} \, \mathrm{d}G_i(\theta_i), \tag{19}$$

which is the density of the actual observation Y_i. The relations (11) and (12) show that

$$\begin{aligned} \widehat{m}_{in}(y) &= \frac{1}{n}\sum_{\ell=1}^{n}(\theta_0 + y - X_{i\ell})I_{[y,\infty)}(X_{i\ell}) \\ &= \frac{1}{n}\sum_{\ell=1}^{n} h(X_{i\ell} - y) \end{aligned} \tag{20}$$

is an unbiased and consistent estimator for the unknown function $m_i(y) = \theta_0\psi_{i1}(y) - \psi_{i2}(y)$. Using this fact Gupta and Liang [10] introduced an empirical Bayes selection procedure d_n by setting

$$d_{in}(\underline{X}_{in}, y_i) = \begin{cases} 1 & \text{if } \widehat{m}_{in}(y) < 0 \\ 0 & \text{otherwise.} \end{cases}$$

Our approach is slightly different. We have already seen that under the condition (17) the function $m_i(y)$ has a unique zero, say η_{i0}, and it holds $m_i(y) > 0$ for $0 \le y < \eta_{i0}$ and $m_i(y) < 0$ for $y > \eta_{i0}$. Using the historical data $\underline{X}_{in} = (X_{i1}, \dots, X_{in})$ from population π_i we construct a consistent estimator $\widehat{\eta}_{in}(\underline{X}_i)$ for η_{i0} and introduce the modified Gupta–Liang rule by

$$d^*_{in}(\underline{X}_{in}, y_i) = \begin{cases} 0 & \text{if } 0 \le y < \widehat{\eta}_{in} \\ 1 & \text{otherwise.} \end{cases} \tag{21}$$

Set $d^*_n(\underline{X}_n) = (d^*_{1n}(\underline{X}_{1n}), \dots, d^*_{kn}(\underline{X}_{kn}))$ for $\underline{X}_n = (\underline{X}_{1n}, \dots, \underline{X}_{kn})$.

Gupta and Liang [10] studied the rate of convergence to zero of $\mathbb{E}\mathcal{R}_n$ and proved that under some assumptions $\mathbb{E}\mathcal{R}_n = O(\frac{1}{n})$. But it seems to be extremely hard to get the constant $\lim_{n\to\infty} n\mathbb{E}\mathcal{R}_n$ which is necessary to characterize the efficiency of the Gupta–Liang rule. In this paper we establish the limit distribution of $n\mathcal{R}_n$ to characterize the efficiency of the selection rule d^*_{in}.

Applying (8) we get a representation for the conditional regret risk of the selection rule $\widehat{d}_n$

$$\mathcal{R}_n = \sum_{i=1}^{k} \int_0^\infty d^*_{in}(\underline{X}_i, y_i) m_i(y_i) \, \mathrm{d}y_i - \int_0^\infty d^0_i(y_i) m_i(y_i) \, \mathrm{d}y_i .$$

Using (8), (9) and (21) we arrive at

$$\begin{aligned}\mathcal{R}_n &= \int_{\widehat{\eta}_{in}}^{\infty} m_i(y_i)\,\mathrm{d}y_i - \int_{\eta_{io}}^{\infty} m_i(y_i)\,\mathrm{d}y_i \\ &= M_i(\eta_{io}) - M_i(\widehat{\eta}_{in}) \qquad (22)\end{aligned}$$

where

$$M_i(y_i) = \int_0^{y_i} m_i(s)\,\mathrm{d}s. \qquad (23)$$

3. ESTIMATION OF THE ZERO OF AN UNKNOWN FUNCTION

To prepare the main results we study the problem of estimating the zero of an unknown function m. Despite the fact that the unknown function m is continuous we have to deal with estimators for m which are not continuous functions. To overcome this difficulty we smooth the functions by integrating and construct estimators of the maximum point of the integrated function. Moreover, this approach allows us to apply the well-developed theory of M-estimators.

The function h appearing in (18) belongs to the class $\mathcal{H}$ of functions h: $\mathbb{R}_1 \to \mathbb{R}_1$ which may be written in the form

$$h(x) = \sum_{i=1}^{r} c_i I_{[x_i,\infty)}(x) + g(x) \qquad (24)$$

with some $r \in \{0, 1, \dots\}$, some real numbers c_i, nonnegative x_i and a Lipschitz continuous function g which vanishes on $(-\infty, 0]$. If $r = 0$ then the corresponding sum is supposed to be zero. Lipschitz continuity means the existence of a constant L such that $|g(y) - g(x)| \le L|y - x|$ for every $x, y \in \mathbb{R}_1$. Put $c = \sum_{i=1}^{r} |c_i|$ then

$$|h(x)| \le c + L|x|. \qquad (25)$$

We use the notation $a \wedge b = \min\{a, b\}$ and set $H(x,t) = 0$ for $t \le 0$ and put for $t > 0$

$$\begin{aligned}H(x,t) &= \int_0^t h(x-s)\,\mathrm{d}s \qquad (26)\\ &= \int_0^{t\wedge x} h(x-s)\,\mathrm{d}s\end{aligned}$$

where the last equality follows from $h(t) = 0$ for $t < 0$. Using the notation $\gamma_h = \sum_{i=1}^{r} |c_i| + \frac{1}{2}L$ we see that for $x \ge 0$

$$|H(x,t)| \le \gamma_h(x + x^2). \qquad (27)$$

If X is a non-negative random variable with $\mathbb{E}X^2 < \infty$ then

$$\mathbb{E}H(X,t) \le \gamma_h \left[(\mathbb{E}X^2)^{1/2} + \mathbb{E}X^2\right]$$

so that

$$M(t) = \mathbb{E}H(X,t) \tag{28}$$

is a bounded function

$$M(t) \leq \gamma_h \left[(\mathbb{E}X^2)^{1/2} + \mathbb{E}X^2\right]. \tag{29}$$

If h is a pure jump function $I_{[x_0,\infty)}$ then the corresponding function M is closely related to the cumulative distribution function $F(t) = \mathbb{P}(X < t)$ of X. In this case

$$M(t) = \int_0^t (1 - F(x_0 - s))\,\mathrm{d}s.$$

This shows that the function M has a derivative at t from the right given by $D^+M(t) = 1 - F(x_0 - t)$ whereas the derivative from the left is given by $D^-M(t) = 1 - F(x_0 - t - 0)$. If F is continuous then $D^+M(t) = D^-M(t)$ so that M is differentiable with derivative

$$\begin{aligned} M'(t) &= \mathbb{E}h(X-t) \\ &= m(t). \end{aligned} \tag{30}$$

If $h \in \mathcal{H}$ has no jump part then an application of the theorem of Lebesgue shows that (30) is valid, too. This means that for every $h \in \mathcal{H}$ and every r.v. with finite second moment and continuous c.d.f. F the statement (30) holds. The continuity of F implies the continuity of $m(t)$. Hence by (30)

$$M(t) = \int_0^t m(s)\,\mathrm{d}s = \int_0^t \mathbb{E}h(X-s))\,\mathrm{d}s = \mathbb{E}\int_0^t h(X-s))\,\mathrm{d}s. \tag{31}$$

Now we assume that there is some $\eta_0 > 0$ such that

$$m(t) = M'(t) > 0 \text{ for } t < \eta_0 \text{ and } m(t) = M'(t) < 0 \text{ for } t > \eta_0. \tag{32}$$

This condition together with (31) show that the function $M(t)$ has a unique maximum at η_0. Sometimes we need a quadratic behavior of M at η_0, i. e. we suppose that there is some $c > 0$ and some sufficiently small α_0 such that

$$M(\eta_0) - M(t) \geq \frac{c}{2}(t - \eta_0)^2 \qquad \text{for } |t - \eta_0| \leq \alpha_0. \tag{33}$$

We assume that $X_1, X_2, \ldots, X_n$ are independent copies of X. To estimate η_0 we define for $t \geq 0$

$$\widehat{M}_n(t) = \frac{1}{n}\sum_{i=1}^{n} H(X_i, t) \tag{34}$$

and notice that in view of (31) $\widehat{M}_n(t)$ is an unbiased estimator for $M(t)$. Furthermore, the stochastic process $\widehat{M}_n$ is pathwise continuous and every path is constant for all sufficiently large t which follows from (26) and $h(x) = 0$ for $x < 0$. Consequently $\widehat{M}_n(t)$ attains its maximum on $[0, \infty)$, say at $\widehat{\eta}_n$. The next theorem gives the asymptotic distribution of $\widehat{\eta}_n$. In Theorem 3 and in the following statements we denote the distribution of the r.v. Y by $\mathcal{L}(Y)$ and the weak convergence by $\Rightarrow$.

Theorem 3. Assume $X_1, X_2, \dots, X_n$ are non-negative, i.i.d. random variables, the common c.d.f. is continuous and $\mathbb{E}X_1 < \infty$. Suppose $h \in \mathcal{H}$ and the condition (32) is fulfilled. If $\widehat{M}_n$ is defined by (34), $\widehat{\eta}_n \in \operatorname{argmax} \widehat{M}_n$ and M from (31) is twice continuously differentiable in a neighborhood of η_0 and $c = -M''(\eta_0) > 0$ then

$$\mathcal{L}(\sqrt{n}(\widehat{\eta}_n - \eta_0)) \Rightarrow \mathbb{N}\left(0, \frac{\sigma^2}{c^2}\right)$$

where $\sigma^2 = \mathbb{V}(h(X_1 - \eta_0))$ is the variance of $h(X_1 - \eta_0)$.

Using the Taylor expansion of M at η_0 one obtains the following statement.

Corollary 4. Under the assumptions of Theorem 1 it holds

$$\mathcal{L}(n(M(\eta_0) - M(\widehat{\eta}_n)) \Rightarrow \mathcal{L}\left(\frac{\sigma^2}{2c}\chi_1^2\right)$$

where χ_1^2 has a χ^2-distribution with one degree of freedom.

4. APPLICATION TO THE SELECTION PROBLEM

We assume to have available i.i.d historical data $X_{i1}, \dots, X_{in}$ with common density (19) for every population π_i. Set $h(t) = (\theta_0 - t)I_{[0,\infty)}(t)$. Then by (10), $m_i(y_i) = \mathbb{E}h(X_{1i} - y_i) = \theta_0\psi_{i1}(y_i) - \psi_{i2}(y_i)$ and under the assumption (17) the function $m_i(y_i)$ has a unique zero, say η_{i0}. Set

$$M_i(t) = \int_0^t m_i(s)\,\mathrm{d}s. \tag{35}$$

Then

$$M_i(t) = \mathbb{E}\int_0^t h(X_{i1} - s))\,\mathrm{d}s.$$

In accordance with (34) for $t \geq 0$ we introduce the estimator $\widehat{M}_{in}(t)$ for $M_i(t)$ by

$$\begin{aligned}
\widehat{M}_{in}(t) &= \frac{1}{n}\sum_{i=1}^{n} H(X_{ij}, t) \\
&= \frac{1}{n}\sum_{i=1}^{n}\int_0^t h(X_{ij} - s)\,\mathrm{d}s \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(\theta_0 - X_{ij} + \frac{1}{2}(X_{ij} \wedge t)\right)(X_{ij} \wedge t).
\end{aligned}$$

Suppose $\widehat{\eta}_{in} \in \operatorname{argmax} \widehat{M}_{in}$ and define the empirical Bayes selection rule by

$$d_{in}^*(y_i) = \begin{cases} 0 & \text{if } y_i \leq \widehat{\eta}_{in} \\ 1 & \text{if } y_i > \widehat{\eta}_{in} \end{cases} \tag{36}$$

$$d_n^* = (d_{1n}^*, \dots, d_{kn}^*).$$

Then by (22) and (35) the conditional regret risk of d_n^* is

$$\mathcal{R}_n = \sum_{i=1}^{k} (M_i(\eta_{i0}) - M_i(\widehat{\eta}_{in})). \tag{37}$$

The next theorem is the main result of this paper. It gives the asymptotic distribution of $\mathcal{R}_n$ and is a direct consequence of Corollary 4.

Theorem 5. Assume the G_i are nondegenerate, have finite second moment and the assumption (17) is fulfilled for $i = 1, \dots, k$. If the empirical Bayes selection rule is defined by (36) then for the distribution of the conditional regret risk (37), the following result holds

$$\mathcal{L}(n\mathcal{R}_n) \Rightarrow_{n\to\infty} \mathcal{L}\left(\sum_{i=1}^{k} \kappa_i \chi_i^2\right)$$

where the $\chi_1^2, \dots, \chi_k^2$ are i.i.d. with common χ^2-distribution with one degree of freedom and $\kappa_i = \frac{1}{2}(-m_i'(\eta_{i0}))^{-1}\mathbb{V}((\theta_0 - (Y_{i1} - \eta_{i0}))I_{[0,\infty)}(Y_{i1} - \eta_{i0}))$.

When we apply the empirical Bayes selection rule (36) we have the risk $\mathbb{R}(d_n^*)$ introduced in (1). In order to characterize the information contained in the data from the past we may consider the probability

$$\mathbb{P}(\mathbb{R}(d_n^*) > \mathbb{R}(d_G^0) + \varepsilon)).$$

Proposition 6. Under the assumptions of Theorem 5 it holds

$$\lim_{n\to\infty} \mathbb{P}\left(\mathbb{R}(d_n^*) > \mathbb{R}(d_G^0) + \frac{t}{n}\right) = 1 - H(t)$$

where $H(t)$ is the c.d.f. of $\sum_{i=1}^{k} \kappa_i \chi_i^2$.

5. PROOFS

The proof of Theorem 3 is divided into different steps. In the beginning we prove the consistency of $\widehat{\eta}_n$. The next step is to establish the $\sqrt{n}$-consistency which is followed by the proof of the asymptotic normality of $\widehat{\eta}_n$. Due to the lack of smoothness of h we cannot apply the standard Taylor expansion. Instead we apply ideas from empirical process theory for which we refer to van der Vaart and Wellner [20].

Lemma 7. Assume $h \in \mathcal{H}$ and $\mathbb{E}X_1^2 < \infty$. If the condition (32) is fulfilled then for $\widehat{\eta}_n \in \operatorname{argmax} \widehat{M}_n$ holds

$$\widehat{\eta}_n \to_{\mathbb{P}} \eta_0$$

where $\to_{\mathbb{P}}$is the symbol for stochastic convergence.

Proof. We use the notation $\left\|\widehat{M}_n - M\right\| = \sup_{0\le t<\infty} |\widehat{M}_n(t) - M(t)|$. It holds

$$\begin{aligned}\{|\widehat{\eta}_n - \eta_0| > \varepsilon\} &\subseteq \left\{\sup_{|t-\eta_0|>\varepsilon} \widehat{M}_n(t) > \widehat{M}_n(\eta_0)\right\}\\ &\subseteq \left\{\sup_{|t-\eta_0|>\varepsilon} M(t) + \left\|\widehat{M}_n - M\right\| \ge M(\eta_0) - \left\|\widehat{M}_n - M\right\|\right\}\\ &\subseteq \left\{\left\|\widehat{M}_n - M\right\| \ge \frac{1}{2}[M(\eta_0) - \sup_{|t-\eta_0|>\varepsilon} M(t)]\right\}.\end{aligned}$$

Due to the assumption (32) the function M is strictly increasing for $0 \le t \le \eta_0$ and strictly decreasing for $\eta_0 \le t < \infty$. Hence $M(\eta_0) - \sup_{|t-\eta_0|>\varepsilon} M(t) > 0$. Therefore it remains to show that

$$\left\|\widehat{M}_n - M\right\| \to_{\mathbb{P}} 0.$$

The inequality (27) implies that for every $\varepsilon > 0$ there is some T_ε such that

$$\mathbb{E}\int_{T_\varepsilon}^{\infty} |h(X_1 - s)|\,\mathrm{d}s < \varepsilon.$$

By inequality (25) for $0 \le s \le t$

$$\mathbb{E}\sup_{|s-t|<\delta}\left|\frac{1}{n}\sum_{i=1}^{n}\int_s^t h(X_i - \tau)\,\mathrm{d}\tau\right| \le \delta(c + L\mathbb{E}|X_1|).$$

Choose $0 = t_0 < \ldots < t_N = T_\varepsilon$ such that $|t_k - t_{k-1}| \le \varepsilon\,[c + L\mathbb{E}|X_1|]^{-1} =: \delta$. Then

$$\begin{aligned}\mathbb{E}\sup_{0\le t<\infty} |\widehat{M}_n(t) - M(t)| &\le \mathbb{E}\max_{0\le k\le N} |\widehat{M}_n(t_k) - M(t_k)|\\ &\quad +2\mathbb{E}\max_{k,t_{k-1}\le t\le t_k} |\widehat{M}_n(t_k) - \widehat{M}_n(t)|\\ &\quad +2\mathbb{E}\int_{T_\varepsilon}^{\infty} |h(X_1 - s)|\,\mathrm{d}s\\ &\le \mathbb{E}\max_{0\le k\le N} |\widehat{M}_n(t_k) - M(t_k)| + 4\varepsilon.\end{aligned}$$

To complete the proof we have only to note that $\mathbb{E}|\widehat{M}_n(t_k) - M(t_k)| \to_{n\to\infty} 0$ by the law of large numbers. □

To continue the proof of Theorem 3 we need a special case of a fluctuation inequality established in Ibragimov and Has'minskii [11].

Lemma 8. Assume $[a,b]$ is a finite interval and $V(t), a \le t \le b$ is a continuous stochastic process. If $\mathbb{E}|V(s)|^2 \le \alpha$ and $\mathbb{E}|V(t) - V(s)|^2 \le \alpha|t-s|^2$ for every $a \le s, t \le b$ then there is a universal constant ρ such that

$$\mathbb{E} \sup_{|s-t|\le\delta} |V(t) - V(s)| \le \rho\sqrt{\alpha\delta}. \tag{38}$$

We fix $h \in \mathcal{H}$, $0 < a < \eta_0 < b$ and study the stochastic process

$$V_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [h(X_i - t) - \mathbb{E}h(X_i - t)]\,. \tag{39}$$

Using the inequality (25) we see that

$$\mathbb{E}|V_n(s)|^2 \le \mathbb{E}(c + L|X_1|)^2.$$

Assume now in addition that h is Lipschitz continuous. Then

$$\mathbb{E}|V_n(t) - V_n(s)|^2 \le L^2|t-s|^2.$$

Setting $\alpha = \max(\mathbb{E}(c + L|X_1|)^2, L^2)$ we arrive at

$$\mathbb{E} \sup_{|s-t|\le\delta} |V_n(t) - V_n(s)| \le \rho\sqrt{\alpha\delta} \tag{40}$$

and

$$\begin{aligned} \mathbb{E} \sup_{a\le t\le b} |V_n(t)| &\le \mathbb{E}|V_n(\eta_0)| + \mathbb{E} \sup_{|s-t|\le b-a} |V_n(t) - V_n(s)| \\ &\le c(a,b,h) \end{aligned} \tag{41}$$

where $c(a,b,h) = \sqrt{\alpha} + \rho\sqrt{\alpha(b-a)}$. Now we suppose that $h = I_{[x_0,\infty)}(x)$ is a pure jump function. Then

$$V_n(t) = \sqrt{n}(F(x_0 + t) - F_n(x_0 + t)) \tag{42}$$

where $F_n(s) = \frac{1}{n}\sum_{i=1}^{n} I_{(-\infty,s)}(X_i)$ and $F(s) = \mathbb{P}(X_1 < s)$ are the empirical and the cumulative distribution function, respectively. Dvoretzky, Kiefer and Wolfowitz [3] proved that there is a universal constant K such that

$$\mathbb{P}(\sqrt{n}\,\|F_n - F\| \ge x) \le K \exp\left\{-2x^2\right\}.$$

Hence

$$\mathbb{E}\sup_t |V_n(t)| = \int_0^\infty \mathbb{P}(\sup_t |V_n(t)| \ge x)\,\mathrm{d}x \le \int_0^\infty K \exp\left\{-2x^2\right\}\mathrm{d}x.$$

As the set of all $h \in \mathcal{H}$ for which $\sup_n \mathbb{E}\sup_{a\le t\le b} |V_n(t)| < \infty$ holds is a linear subset we see that for every $h \in \mathcal{H}$ and every finite interval $[a,b]$ we find some $c(a,b,h)$ such that

$$\sup_n \mathbb{E} \sup_{a\le t\le b} |V_n(t)| \le c(a,b,h). \tag{43}$$

Now we study the process $\widehat{M}_n$ in a neighborhood of η_0. We have for every $\delta > 0$ and $t \geq \eta_0$

$$\begin{aligned}
&\mathbb{E} \sup_{\eta_0 \leq t \leq \eta_0+\delta} \sqrt{n}|(\widehat{M}_n(t) - \widehat{M}_n(\eta_0)) - (M(t) - M(\eta_0))| \\
= &\ \mathbb{E} \sup_{\eta_0 \leq t \leq \eta_0+\delta} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_{\eta_0}^{t} (h(X_i - s) - \mathbb{E}h(X_i - s))\, \mathrm{d}s \right| \\
\leq &\ \mathbb{E} \sup_{\eta_0 \leq t \leq \eta_0+\delta} \left| \int_{\eta_0}^{t} V_n(s)\, \mathrm{d}s \right| \leq \int_{\eta_0}^{\eta_0+\delta} \mathbb{E} \sup_{\eta_0 \leq t \leq \eta_0+\delta} |V_n(t)|\, \mathrm{d}s \\
\leq &\ c(a, b, h)\delta.
\end{aligned}$$

A similar inequality holds for $\eta_0 - \delta \leq t \leq \eta_0$. Consequently

$$\mathbb{E} \sup_{|t-\eta_0| \leq \delta} |(\widehat{M}_n(t) - \widehat{M}_n(\eta_0)) - (M(t) - M(\eta_0))| \leq \frac{c(a, b, h)}{\sqrt{n}} \delta.$$

Now we are ready to apply Theorem 3.2.5 in van der Vaart and Wellner [20].

Lemma 9. Assume the assumptions of Lemma 7 are fulfilled then $\widehat{\eta}_n$ is $\sqrt{n}$–consistent. i. e. the sequence $\sqrt{n}(\widehat{\eta}_n - \eta_0)$ is stochastically bounded.

Proof. The estimator $\widehat{\eta}_n$ is consistent by Lemma 7. Set $\phi_n(\delta) = c(a, b, h)\delta$ and $r_n = \sqrt{n}$ to apply Theorem 3.2.5 in van der Vaart and Wellner [20]. □

The $\sqrt{n}$–consistency of $\widehat{\eta}_n$ allows us to introduce the local parameters ξ. Set

$$W_n(\xi) = \sum_{i=1}^{n} \left(H\left(X_i, \eta_0 + \frac{\xi}{\sqrt{n}} \right) - H(X_i, \eta_0) \right).$$

Then $\widehat{\xi}_n = \sqrt{n}(\widehat{\eta}_n - \eta_0)$ is a maximum point of W_n and due to the $\sqrt{n}$-consistency of $\widehat{\eta}_n$ the new sequence $\widehat{\xi}_n$ is stochastically bounded. We approximate $W_n(\xi)$ by a stochastic term linear in ξ and a nonstochastic term which is nonlinear in ξ. Introduce

$$\tilde{W}_n(\xi) = \sum_{i=1}^{n} (h(X_i - \eta_0) - \mathbb{E}h(X_i - \eta_0)) \frac{\xi}{\sqrt{n}} - \mathbb{E}(H\left(X_i, \eta_0 + \frac{\xi}{\sqrt{n}}) - H(X_i, \eta_0)) \right).$$

We show that for any fixed real numbers c_1, c_2

$$\sup_{c_1 \leq \xi \leq c_2} |W_n(\xi) - \tilde{W}_n(\xi)| \to_{\mathbb{P}} 0 \qquad \text{as } n \to \infty. \tag{44}$$

Indeed, for $\xi \geq 0$ by (39)

$$\begin{aligned}
|W_n(\xi) - \tilde{W}_n(\xi)| &= \left| \sqrt{n} \int_{\eta_0}^{\eta_0+\xi/\sqrt{n}} [V_n(s) - V_n(\eta_0)]\, \mathrm{d}s \right| \\
&\leq \xi \sup_{|s-\eta_0| \leq \frac{\xi}{\sqrt{n}}} |V_n(s) - V_n(\eta_0).
\end{aligned}$$

Using a similar representation for $\xi \leq 0$ we get

$$\sup_{c_1 \leq \xi \leq c_2} |W_n(\xi) - \tilde{W}_n(\xi)| \leq |\xi| \sup_{|s-\eta_0| \leq |\frac{\xi}{\sqrt{n}}|} |V_n(s) - V_n(\eta_0)| \tag{45}$$

To prove (44) we note that the set of all $h \in \mathcal{H}$ for which (44) holds, is a linear set. Therefore we have only to consider the special cases in which h is Lipschitz continuous or a pure jump function. If h fulfils a Lipschitz condition then (44) follows from (45) and (40). If h is a pure jump function then by (42) $V_n(t) = \sqrt{n}(F(x_0 + t) - F_n(x_0 + t))$. Let $\mathbb{B}$ be the Brownian bridge. The distribution of the stochastic processes V_n converge to the process $-B(F(x_0 + \cdot))$ which is continuous as $F(t) = \mathbb{P}(X_1 < t)$ is continuous by assumption. Consequently the sequence V_n is asymptotically equicontinuous, i. e.

$$\lim_{\delta \to 0} \lim_{n \to \infty} \mathbb{P}\left(\sup_{|s-t|<\delta} |V_n(t) - V_n(s)| > \varepsilon \right) = 0.$$

This statement and (45) yield (44). Now we evaluate the processes $\tilde{W}_n(\xi)$, $c_1 \leq \xi \leq c_2$. Recall that $M(s) = \mathbb{E}H(X_i, s)$ is differentiable and has a local maximum at η_0. This gives $M'(\eta_0) = \mathbb{E}h(X_i - \eta_0) = 0$. Then the twice differentiability of M at η_0 provides

$$\tilde{W}_n(\xi) = \xi \frac{1}{\sqrt{n}} \sum_{i=1}^{n} h(X_i - \eta_0) + M''(\eta_0)\frac{\xi^2}{2} + o(1).$$

Let Z be a random variable having a normal distribution $\mathbb{N}(0, \sigma^2)$, where $\sigma^2 = \mathbb{V}(h(X_i - \eta_0))$. Introduce the process W by $W(\xi) = \xi Z + M''(\eta_0)\frac{\xi^2}{2}$ and notice that W has a unique maximum at $\widehat{\xi} = -[M''(\eta_0)]^{-1}Z$. The central limit theorem and (44) show that the distributions of the processes $W_n(\xi)$ $c_1 \leq \xi \leq c_2$ converge weakly to the distribution of W. Now we apply the Argmax continuous mapping Theorem 3.2.2 in van der Vaart and Wellner [20] to see that the distributions of the maximum points $\widehat{\xi}_n = \sqrt{n}(\widehat{\eta}_n - \eta_0)$ of W_n converge to the distribution $\mathcal{L}(\widehat{\xi})$ of the maximum point $\widehat{\xi}$ of W. This completes the proof of Theorem 3.

Theorem 5 is a direct consequence of Corollary 4 as the X_{ij} have the density f_i and consequently, the c.d.f. is continuous. The assumption (17) provides (32).

Proposition 6 follows directly from Theorem 5.

ACKNOWLEDGEMENTS

The authors are very grateful to a Referee for his suggestions and comments.

(Received March 28, 2000.)

REFERENCES

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

[2] J. J. Deely: Multiple Decision Procedures from Empirical Bayes Approach. Ph.D. Thesis (Mimeo. Ser. No. 45). Dept. Statist., Purdue Univ., West Lafayette, Ind. 1965.
[3] A. Dvoretzky, J. Kiefer, and J. Wolfowitz: Asymptotic minimax character of the sample distribution functions and of the classical multinomial estimator. Ann. Math. Statist. *27* (1956), 642–669.
[4] S. S. Gupta and S. Panchapakesan: Subset selection procedures: review and assessment. Amer. J. Management Math. Sci. *5* (1985), 235–311.
[5] S. S. Gupta and T. Liang: Empirical Bayes rules for selecting the best binomial population. In: Statistical Decision Theory and Related Topics IV (S. S. Gupta and J. O. Berger, eds.), Vol. 1. Springer–Verlag, Berlin 1986, pp. 213–224.
[6] S. S. Gupta and T. Liang: On empirical Bayes selection rules for sampling inspection. J. Statist. Plann. Inference *38* (1994), 43–64.
[7] S. S. Gupta, T. Liang, and R.-B. Rau: Empirical Bayes two stage procedures for selecting the best Bernoulli population compared with a control. In: Statistical Decision Theory and Related Topics V. (S. S. Gupta and J. O. Berger, eds.), Springer–Verlag, Berlin 1994, pp. 277–292.
[8] S. S. Gupta, T. Liang, and R.-B. Rau: Empirical Bayes rules for selecting the best normal population compared with a control. Statist. Decision *12* (1994), 125–147.
[9] S. S. Gupta and F. Liese: Asymptotic distribution of the random regret risk for selecting exponential populations. Technical Report #98-08C, Center for Statistical Decision Sciences and Dept. of Statistics, Purdue University 1998. (Revised March, 2000.)
[10] S. S. Gupta and T. Liang: Selecting good exponential populations compared with a control: nonparametric empirical Bayes approach. Sankhya, Ser. B *61* (1999), 289–304.
[11] I. A. Ibragimov and R. Z. Has'minskii: Statistical Estimation: Asymptotic Theory. Springer, New York 1981.
[12] N. L. Johnson, S. Kotz and N. Balakrishnan: Continuous Univariate Distributions, Vol. 1. Second edition. Wiley, New York 1994.
[13] J. Jurečková and P. K. Sen: Robust Statistical Procedures, Asymptotics and Interrelations. Wiley, New York 1996.
[14] F. Liese and I. Vajda: Consistency of M-estimates in general regression models. J. Multivariate Anal. *50* (1994), 93–114.
[15] J. Pfanzagl: On measurability and consistency of minimum contrast estimators. Metrika *14* (1969), 249–272.
[16] D. Pollard: Asymptotics for least absolute deviation regression estimators. Econometric Theory *7* (1991), 186–199.
[17] H. Robbins: An empirical Bayes approach to statistics. In: Proc. Third Berkeley Symp., Math. Statist. Probab. 1, Univ. of California Press 1956, pp. 157–163.
[18] G. R. Shorack and J. A. Wellner: Empirical Processes with Applications to Statistics. Wiley, New York 1986.
[19] A. Wald: Note on the consistency of the maximum likelihood estimate. Ann. Math. Statist. 20 (1949), 595–601.
[20] A. W. van der Vaart and J. A. Wellner: Weak Convergence and Empirical Processes (Springer Series in Statistics). Springer–Verlag, Berlin 1996.

Prof. Dr. Shanti S. Gupta, Department of Statistics, Purdue University, West Lafayette, IN 47907-1399. U. S. A.

Prof. Dr. Friedrich Liese, Department of Mathematics, University of Rostock, Universitätsplatz 1, D-18051 Rostock. Germany.

e-mails: sgupta@stat.purdue.edu, friedrich.liese@mathematik.uni-rostock.de