KYBERNETIKA — VOLUME 35 (1999), NUMBER 3, PAGES 309–332

EXPONENTIAL RATES FOR THE ERROR PROBABILITIES IN SELECTION PROCEDURES

FRIEDRICH LIESE AND KLAUS J. MIESCKE

For a sequence of statistical experiments with a finite parameter set the asymptotic behavior of the maximum risk is studied foı the problem of classification into disjoint subsets. The exponential rates of the optimal decision rule is determined and expressed in terms of the normalized limit of moment generating functions of likelihood ratios. Necessary and sufficient conditions for the existence of adaptive classification rules in the sense of Rukhin [11] are given. The results are applied to the problem of the selection of the best population. Exponential families are studied as a special case, and an example for the normal case is included.

1. INTRODUCTION

The Bayes and maximum error probabilities in the problem of testing a simple null hypothesis versus a simple alternative, or more generally in a multiple decision problem, tend to zero with an exponential rate of convergence for increasing sample size. Pioneering work has been done in the paper by Chernoff [3]. Here and in papers by Krafft and Plachky [6], Krafft and Puri [7] and several other authors the i. i. d. case is treated. A more general version of the so called Chernoff theorem can be found in Vajda [15].

The classification problem is a multiple decision problem with given distributions $P_1, \ldots, P_k$ of the k populations. In the first part of this paper we deal with a more general question. Suppose we are given a family of distributions Q_ϑ, $\vartheta \in \Theta$, and $\Theta = \Theta_1 \cup \ldots \cup \Theta_k$ is a partition of the set Θ. After taking an observation we want to decide to which of the families $(Q_\vartheta)_{\vartheta \in \Theta_i}$ the distribution of the data belongs to. We are interested in asymptotic results. Therefore it is assumed that a sequence $Q_{n,\vartheta}$ is given. The special case of i. i. d. observations corresponds to $Q_{n,\vartheta} = Q_\vartheta \times \ldots \times Q_\vartheta = Q_\vartheta^n$. In general the asymptotic behavior of the Q_ϑ^n will be described through the requirement that the normalized logarithms of the moment generating functions converge to some function which is automatically convex. In generalization of the results of Krafft and Puri [7] we calculate the exponential rate of the error probabilities of the minimax decision rule for the classification into subsets. For this problem the distributions $Q_{n,\vartheta}$ are not completely known in many situations. In general they depend also on some nuisance parameter. The question which naturally arises is whether there exists a decision rule not depending on α, which achieves the same exponential rate of the error probabilities as given by the optimum decision rule for known α. This problem of the existence of such adaptive decisions for the problem of classification is studied in the second part of the paper. The results there are generalizations of Rukhin [11].

Suppose $P_{\eta_1}, \ldots, P_{\eta_k}$ are distributions depending on some real valued parameter η, for populations $\pi_1, \ldots, \pi_k$, respectively. The population with the largest η-value is called the best population. First we assume that the set $\{\eta_1, \ldots, \eta_k\}$ is known but we do not know which population belongs to which η-value. Taking observations from each population we want to select the population with the largest η-value. If Θ is the set of all permutations of $(1, \ldots, k)$ and $\Theta_i \subset \Theta$ the set of permutations such that the population which is at position i has the largest η-value, then we can see that the problem of selecting the best population can be reduced to the problem of classification into subsets. In this way we apply the results from Sections 2 and 3 to the selection problem. Especially we investigate the case where independent observations of size n_i are taken from population π_i. Explicit expressions are given for the exponential rate of the probabilitiy of selecting a false population. If the distributions P_{η_i} belong to some exponential family we compare the exponential rate of the best selection procedure for known η_i with the rate of the so called natural selection rule based on sample means of the statistic which generates the exponential family. It turns out that the natural selection rule is asymptotically optimal w..r.t. the exponential rates iff the two smallest sample sizes n_{i_1} and n_{i_2}, say, are asymptotically the same, i.e. it holds $n_{i_1}/n_{i_2} \to 1$. Furthermore, for populations from an exponential family this condition is necessary and sufficient for the existence of an adaptive selection rule which in this case is the natural rule based on the sample means.

2. CLASSIFICATION INTO SUBSETS

Let Θ be a finite nonempty set with N elements, and let $\Theta_1, \ldots, \Theta_k$ be a decomposition into disjoint subsets. We write $\vartheta_1 \equiv \vartheta_2$ if ϑ_1 and ϑ_2 belong to the same subset Θ_i. For any $\vartheta \in \Theta$ we set $\langle\vartheta\rangle = i$ if $\vartheta \in \Theta_i$. Hence $\vartheta_1 \equiv \vartheta_2$ iff $\langle\vartheta_1\rangle = \langle\vartheta_2\rangle$.

Suppose we are given a family of distributions $Q_\vartheta, \vartheta \in \Theta$, defined on the sample space $(\mathcal{X}, \mathcal{A})$. We will study the following problem. Taking an observation $x \in \mathcal{X}$ we have to decide to which family $(Q_\vartheta)_{\vartheta \in \Theta_i}$ the distribution of the corresponding distribution belongs to. $\mathbb{D} = \{1, \ldots, k\}$ is our decision space and a randomized decision $q = (q_1, \ldots, q_k)$ is a vector of measurable functions $q_i : \mathcal{X} \to [0,1]$ such that $\sum_{i=1}^k q_i(x) = 1$ for every $x \in \mathcal{X}$. $q_i(x)$ is the probability to decide for the subset Θ_i if x is observed. We use the 0-1 loss function $L : \Theta \times \mathbb{D} \to [0, \infty)$ defined by

$$L(\vartheta, i) = \begin{cases} 0 & \text{if} \quad \langle\vartheta\rangle = i \\ 1 & \text{else.} \end{cases}$$

The risk of decision q is then given by

$$R(\vartheta, q) = 1 - \int q_{\langle\vartheta\rangle} \, \mathrm{d}Q_\vartheta .$$

We examine the asymptotic behavior of the risk for increasing sample size. To this end, we assume that a sequence of experiments $E_n = (\mathcal{X}_n, \mathcal{A}_n, Q_{n,\vartheta}, \vartheta \in \Theta)$, $n = 1, 2, \ldots$ is given. Furthermore let $\pi(\vartheta), \vartheta \in \Theta$, be positive prior weights. Given a sequence $q_n = (q_{n,1}, \ldots, q_{n,k})$ of decisions we introduce the maximum risk $R_m(q_n)$ by $R_m(q_n) = \max_{\vartheta \in \Theta} R(\vartheta, q_n)$ and the Bayes risk by $R_\pi(q_n) = \sum_{\vartheta \in \Theta} \pi(\vartheta) R(\vartheta, q_n)$. Set $R_n^m = \min_{q_n} R_m(q_n)$, $R_n^\pi = \min_{q_n} R_\pi(q_n)$. The question now is how fast both R_n^m and R_n^π tend to zero if $n \to \infty$.

This problem was studied in the i. i. d. case $(\mathcal{X}^n, \mathcal{A}^n, Q_\vartheta^n)$ by Krafft and Puri [7] if the subsets Θ_i are singletons. If furthermore $k = 2$, i. e. $\Theta = \{\vartheta_1\} \cup \{\vartheta_2\}$ then we have a simple hypotheses testing problem.

In the i. i. d. case the exponential rate of convergence to zero of R_n^m, R_n^π was obtained by Chernoff [3, 4], by Krafft and Plachky [6] and other authors. A Chernoff type theorem for an increasing sequence of sub-σ-algebras, i. e. $E_n = (\mathcal{X}, \mathcal{A}_n, Q_{n,\vartheta}, \vartheta \in \Theta)$ where $Q_{n,\vartheta}$ is the restriction of Q_ϑ to $\mathcal{A}_n$, was obtained in Vajda [15].

Similar as in the above mentioned papers we use the concept of Hellinger integral to characterize relations between different distributions. Let P, Q be distributions defined on $(\mathcal{X}, \mathcal{A})$ and suppose P and Q are equivalent ($P \sim Q$), i.e. $P \ll Q$ and $Q \ll P$. Then $H_s(P, Q) = \int \left(\frac{\mathrm{d}P}{\mathrm{d}Q}\right)^s \mathrm{d}Q$ is called *Hellinger integral of order s*.

We summarize some well-known properties of Hellinger integrals important in the sequel. For proofs we refer to Liese and Vajda [9]. It holds $0 \le H_{s_1}(P, Q) \le 1 \le H_{s_2}(P, Q) < \infty$ if $s_1 \in [0, 1]$, $s_2 \notin [0, 1]$. Furthermore $H_0(P, Q) = H_1(P, Q) = 1$. $G(s) = \ln H_s(P, Q)$ is a convex function taking values in $(-\infty, +\infty]$.

We characterize the asymptotic behavior of the $Q_{n,\vartheta}$ from the experiments E_n with the help of Hellinger integrals. The next assumptions are fundamental for all further investigations.

Assumption 1. All distributions in the sequence of experiments

$$E_n = (\mathcal{X}_n, \mathcal{A}_n, Q_{n,\vartheta}, \vartheta \in \Theta)$$

are equivalent ($Q_{n,\vartheta_1} \sim Q_{n,\vartheta_2}$, for every $\vartheta_1, \vartheta_2 \in \Theta$) and there exists a sequence $c_n \to \infty$ so that for every $\vartheta_1, \vartheta_2 \in \Theta$, $-\infty < s < \infty$

$$\lim_{n \to \infty} \frac{1}{c_n} \ln H_s(Q_{n,\vartheta_1}, Q_{n,\vartheta_2}) = G(\vartheta_1, \vartheta_2, s) \tag{1}$$

exists, takes values in $(-\infty, +\infty]$, and for $\vartheta_1 \not\equiv \vartheta_2$, $G(\vartheta_1, \vartheta_2, s)$ is not identical zero in the interval $0 \le s \le 1$.

Note that by the convexity of $G(\vartheta_1, \vartheta_2, s)$ the set $\{s : G(\vartheta_1, \vartheta_2, s) < \infty\}$ is an interval with the interior, say, (a_1, a_2). Furthermore, by $0 \le H_s(Q_{n,\vartheta_1}, Q_{n,\vartheta_2}) \le 1$ for $0 \le s \le 1$ it holds $G(\vartheta_1, \vartheta_2, s) < \infty$ for $0 \le s \le 1$ if Assumption 1 is fulfilled.

Assumption 2. For every $\vartheta_1, \vartheta_2 \in \Theta$, $G(\vartheta_1, \vartheta_2, s)$ is continuous in $[0, 1]$ and continuously differentiable in (0,1).

Assumption 3. For every $\vartheta_1, \vartheta_2 \in \Theta$ it holds $[0,1] \subseteq (a_1, a_2)$ and $G(\vartheta_1, \vartheta_2, s)$ is continuously differentiable in (a_1, a_2). Furthermore $\lim_{s \downarrow a_1} G'(\vartheta_1, \vartheta_2, s) = -\infty$, $\lim_{s \uparrow a_2} G'(\vartheta_1, \vartheta_2, s) = \infty$.

Theorem 1. Suppose that Assumptions 1 and 2 are fulfilled for the experiments $E_n = (\mathcal{X}_n, \mathcal{A}_n, Q_{n,\vartheta}, \vartheta \in \Theta)$. Then for any sequence q_n of randomized decisions

$$\begin{aligned} \liminf_{n\to\infty} \frac{1}{c_n} \ln R_m(q_n) &= \liminf_{n\to\infty} \frac{1}{c_n} \ln R_\pi(q_n) \\ &\geq \max_{\vartheta_1 \not\equiv \vartheta_2} \inf_{0<s<1} G(\vartheta_1, \vartheta_2, s). \end{aligned}$$

Proof. Fix $\vartheta_1 \not\equiv \vartheta_2$ and put $\psi_n = 1 - q_{n,(\vartheta_1)}$. ψ_n is a test for $H_0 : Q_{n,\vartheta_1}$ versus $H_A : Q_{n,\vartheta_2}$ with

$$\begin{aligned} R(\vartheta_1, q_n) &= \int (1 - q_{n,(\vartheta_1)})\, \mathrm{d}Q_{n,\vartheta_1} \\ &= \int \psi_n \, \mathrm{d}Q_{n,\vartheta_1} \end{aligned} \tag{2}$$

and

$$\begin{aligned} R(\vartheta_2, q_n) &= \int (1 - q_{n,(\vartheta_2)})\, \mathrm{d}Q_{n,\vartheta_2} \\ &= \int (1 - \psi_n)\, \mathrm{d}Q_{n,\vartheta_2}. \end{aligned} \tag{3}$$

Let φ_n denote the Bayes test with prior $\frac{1}{2}, \frac{1}{2}$. Then

$$\frac{1}{2}(R(\vartheta_1, q_n) + R(\vartheta_2, q_n)) \geq \frac{1}{2} \int \varphi_n \, \mathrm{d}Q_{n,\vartheta_1} + \frac{1}{2} \int (1 - \varphi_n)\, \mathrm{d}Q_{n,\vartheta_2}.$$

Consequently by Lemma 3 in the Appendix with $a = 0$

$$\begin{aligned} &\liminf_{n\to\infty} \frac{1}{c_n} \ln \left[\max(R(\vartheta_1, q_n), R(\vartheta_2, q_n))\right] \\ &\quad\geq \liminf_{n\to\infty} \frac{1}{c_n} \ln \left[\frac{1}{2}(R(\vartheta_1, q_n) + R(\vartheta_2, q_n))\right] \\ &\quad\geq \inf_{0<s<1} G(\vartheta_1, \vartheta_2, s). \end{aligned}$$

Hence

$$\liminf_{n\to\infty} \frac{1}{c_n} \ln R_m(q_n) \geq \max_{\vartheta_1 \not\equiv \vartheta_2} \inf_{0<s<1} G(\vartheta_1, \vartheta_2, s).$$

Put $\pi' = \min\{\pi(\vartheta), \vartheta \in \Theta\}$ and $\pi'' = \max\{\pi(\vartheta), \vartheta \in \Theta\}$. Then

$$R_\pi(q_n) \leq k\pi'' R_m(q_n) \tag{4}$$

$$R_m(q_n) \leq \frac{1}{\pi'} R_\pi(q_n), \tag{5}$$

which completes the proof. □

Now we will construct decision rules which asymptotically attain the optimal rate in Theorem 1. To this end, we eliminate nuisance parameters by taking the maximum of likelihood, i.e. by using

$$L_{n,i} = \max_{\vartheta\in\Theta_i} f_{n,\vartheta},$$

where $f_{n,\vartheta}$ is the density of $Q_{n,\vartheta}$ w.r.t. some σ-finite measure μ_n on $(\mathcal{X}_n, \mathcal{A}_n)$. Let $A_n(x) \subseteq \{1,\ldots,k\}$ with $A_n(x) = \{i : L_{n,i}(x) = \max_{1\le j\le k} L_{n,j}(x)\}$, and denote by $q_n^0(x) = (q_{n,1}^0(x),\ldots,q_{n,k}^0(x))$ the uniform distribution on $A_n(x)$.

Theorem 2. Under the assumptions of Theorem 1 it holds that

$$\lim_{n\to\infty}\frac{1}{c_n}\ln R_m(q_n^0) = \lim_{n\to\infty}\frac{1}{c_n}\ln R_\pi(q_n^0) = \max_{\vartheta_1\not\equiv\vartheta_2}\inf_{0<s<1} G(\vartheta_1,\vartheta_2,s).$$

Proof. In view of inequalities (4), (5), we have to consider only $R_m(q_n^0)$. For fixed $\vartheta\in\Theta$ we have

$$\begin{aligned}
R(\vartheta, q_n^0) &\le \sum_{j\neq\langle\vartheta\rangle} Q_{n,\vartheta}(L_{n,j}\ge L_{n,\langle\vartheta\rangle})\\
&\le \sum_{j\neq\langle\vartheta\rangle} Q_{n,\vartheta}(L_{n,j}\ge f_{n,\vartheta})\\
&\le \sum_{j\neq\langle\vartheta\rangle}\sum_{\vartheta'\in\Theta_j} Q_{n,\vartheta}(f_{n,\vartheta'}\ge f_{n,\vartheta})\\
&\le kN\max_{\vartheta_1\not\equiv\vartheta_2}\left(Q_{n,\vartheta_1}(f_{n,\vartheta_1}\le f_{n,\vartheta_2}) + Q_{n,\vartheta_2}(f_{n,\vartheta_2}< f_{n,\vartheta_1})\right)
\end{aligned}$$

where N is the cardinality of Θ. The rest follows from Lemma 3 in the Appendix with $\pi_{1,n} = \pi_{2,n} = \frac{1}{2}$. □

Now we investigate the asymptotic behavior of the risks if the weights are allowed to depend on n. To be more precise we suppose that $\pi_n(\vartheta), \vartheta\in\Theta$, are non-negative numbers so that $a(\vartheta) = \lim_{n\to\infty}\frac{1}{c_n}\ln\pi_n(\vartheta)$ exists for every $\vartheta\in\Theta$. Set

$$R_{\pi_n}(q_n) = \sum_{\vartheta\in\Theta} R(\vartheta, q_n)\pi_n(\vartheta)$$

and

$$R_{\pi_n}^m(q_n) = \max_{\vartheta\in\Theta}[\pi_n(\vartheta)R(\vartheta,q_n)].$$

In order to construct a weighted maximum likelihood rule we set

$$L_{\pi_n,i} = \max_{\vartheta\in\Theta_i}[\pi_n(\vartheta) f_{n,\vartheta}]$$

and introduce $q_{\pi_n}^0$ analogously to q_n^0 where $L_{\pi_n,i}$ is used in A_n instead of $L_{n,i}$. Set $\Gamma(\vartheta_1,\vartheta_2,s) = sa(\vartheta_1) + (1-s)a(\vartheta_2) + G(\vartheta_1,\vartheta_2,s)$.

Theorem 3. If Assumptions 1 and 3 are fulfilled then for every sequence q_n

$$\liminf_{n\to\infty} \frac{1}{c_n} \ln R^m_{\pi_n}(q_n) = \liminf_{n\to\infty} \frac{1}{c_n} \ln R_{\pi_n}(q_n) \geq \max_{\vartheta_1 \not\equiv \vartheta_2} \inf_{0<s<1} \Gamma(\vartheta_1, \vartheta_2, s).$$

Proof. The proof of Theorem 3 is completely analogous to the proofs of Theorems 1 and 2 if we apply Lemma 4 instead of Lemma 3 and the inequalities

$$R_{\pi_n}(q_n) \leq N R^m_{\pi_n}(q_n)$$

and

$$R^m_{\pi_n}(q_n) \leq R_{\pi_n}(q_n)$$

instead of (4) and (5) respectively. □

3. ADAPTIVE CLASSIFICATION RULES

Suppose that we are given experiments $E_n = (\mathcal{X}_n, \mathcal{A}_n, Q_\vartheta \in \Theta)$ and that we want to classify the distribution of the data into the sets $\Theta_1, \ldots, \Theta_k$. In many situations the distributions $Q_{n,\vartheta}$ depend on some unknown nuisance parameter $\alpha \in \Lambda$. We introduce the experiments

$$\mathcal{E}_n = (\mathcal{X}_n, \mathcal{A}_n, Q^\alpha_{n,\vartheta}, (\alpha, \vartheta) \in \Lambda \times \Theta).$$

Let μ_n be a σ-finite dominating measure and denote by $f^\alpha_{n,\vartheta}$ the density of $Q^\alpha_{n,\vartheta}$ w.r.t. μ_n. We suppose that for every $\alpha, \beta \in \Lambda$ Assumption 1 is fulfilled for $Q^\alpha_{n,\vartheta_1}, Q^\beta_{n,\vartheta_2}$ instead of $Q_{n,\vartheta_1}, Q_{n,\vartheta_2}$. Set

$$G(\alpha, \beta, \vartheta_1, \vartheta_2, s) = \lim_{n\to\infty} \frac{1}{c_n} \ln H_s(Q^\alpha_{n,\vartheta_1}, Q^\beta_{n,\vartheta_2}),$$

and

$$G(\alpha, \beta) = \max_{\vartheta_1 \not\equiv \vartheta_2} \inf_{0<s<1} G(\alpha, \beta, \vartheta_1, \vartheta_2, s).$$

By Theorem 1 we get for every sequence of decision rules q_n and every $\alpha \in \Lambda$

$$\liminf_{n\to\infty} \frac{1}{c_n} \ln \left(\max_{\vartheta\in\Theta} R(\alpha, \vartheta, q_n) \right) \geq G(\alpha, \alpha). \tag{6}$$

If we fix $\alpha \in \Lambda$ and apply the maximum likelihood rule then this rule attains equality in (6). But, unfortunately, the maximum likelihood rule depends on the unknown nuisance parameter α. Similar as in Rukhin ([11, 12, 13]) we ask for the existence of such rules which do not depend on $\alpha \in \Lambda$ but attain the lower bound in (6) for every fixed $\alpha \in \Lambda$.

Definition 1. A sequence of classification rules q_n is called *adaptive* if

$$\lim_{n\to\infty} \frac{1}{c_n} \ln \left(\max_{\vartheta\in\Theta} R(\alpha,\vartheta,q_n) \right) = G(\alpha,\alpha)$$

for every $\alpha \in \Lambda$.

Similarly as in Rukhin and Vajda [14], where the case of subsets Θ_i each consisting of one element have been treated, we now derive necessary and sufficient conditions for the existence of adaptive classification rules. Put

$$\Gamma_{\alpha,\beta}(a,b,\vartheta_1,\vartheta_2) = \inf_{0<s<1} [G(\alpha,\beta,\vartheta_1,\vartheta_2,s) + sa + (1-s)\,b].$$

Lemma 1. Assume that Assumption 1 is fulfilled for every $\alpha, \beta \in \Lambda$ and q_n is any sequence of selection rules. If Assumption 2 is fulfilled then

$$\liminf_{n\to\infty} \frac{1}{c_n} \ln\left[\max(R(\alpha,\vartheta_1,q_n), R(\beta,\vartheta_2,q_n))\right] \geq \inf_{0<s<1} G(\alpha,\beta,\vartheta_1,\vartheta_2,s).$$

If even the stronger Assumption 3 is fulfilled and

$$\lim_{n\to\infty} \frac{1}{c_n} \ln \pi_{1,n} = a, \; \lim_{n\to\infty} \frac{1}{c_n} \ln \pi_{2,n} = b,$$

then

$$\liminf_{n\to\infty} \frac{1}{c_n} \ln\left[\max(\pi_{1,n} R(\alpha,\vartheta_1,q_n), \pi_{2,n} R(\beta,\vartheta_2,q_n))\right] \geq \Gamma_{\alpha,\beta}(a,b,\vartheta_1,\vartheta_2).$$

Remark. If $a = b$ then $\Gamma_{\alpha,\beta}(a,b,\vartheta_1,\vartheta_2) = a + \inf_{0<s<1} G(\alpha,\beta,\vartheta_1,\vartheta_2,s)$. Therefore the first and the second statement coincides in this case.

Proof. Denote by $q_{n,i}$, $1 \leq i \leq k$, the components of the vector q_n and fix $\vartheta_1 \not\equiv \vartheta_2$. Put $\psi_n = 1 - q_{n,\langle\vartheta_1\rangle}$. Then ψ_n is a test for $H_0 : Q^{\alpha}_{n,\vartheta_1}$ versus H_A: $Q^{\beta}_{n,\vartheta_2}$ with

$$\begin{aligned} R(\alpha,\vartheta_1,q_n) &= \int [1 - q_{n,\langle\vartheta_1\rangle}]\, \mathrm{d}Q^{\alpha}_{n,\vartheta_1} = \int \psi_n \, \mathrm{d}Q^{\alpha}_{n,\vartheta_1}, \\ R(\beta,\vartheta_2,q_n) &= \int [1 - q_{n,\langle\vartheta_2\rangle}]\, \mathrm{d}Q^{\beta}_{n,\vartheta_2} = \int \sum_{i\neq\langle\vartheta_2\rangle} q_{n,i} \, \mathrm{d}Q^{\beta}_{n,\vartheta_2} \\ &\geq \int q_{n,\langle\vartheta_1\rangle} \, \mathrm{d}Q^{\beta}_{n,\vartheta_2} = \int [1-\psi_n] \, \mathrm{d}Q^{\beta}_{n,\vartheta_2}. \end{aligned}$$

Let φ_n denote the Bayes test for $H_0 : Q^{\alpha}_{n,\vartheta_1}$ versus $H_A : Q^{\beta}_{n,\vartheta_2}$ with prior weights $\pi_{1,n}, \pi_{2,n}$. Then

$$\begin{aligned} &2\max(\pi_{1,n} R(\alpha,\vartheta_1,q_n), \pi_{2,n} R(\beta,\vartheta_2,q_n)) \\ &\quad\geq \pi_{1,n} \int \varphi_n \, \mathrm{d}Q^{\alpha}_{n,\vartheta_1} + \pi_{2,n} \int (1-\varphi_n) \, \mathrm{d}Q^{\beta}_{n,\vartheta_2}. \end{aligned} \tag{7}$$

The first statement now follows from Lemma 3 in the Appendix by putting $\pi_{1,n} = \pi_{2,n} = 1/2$, where the second statement is a direct consequence of Lemma 4 in the Appendix.

Assume now that Λ is finite with L elements and eliminate the nuisance parameter $\alpha \in \Lambda$ by taking the maximum over $\alpha \in \Lambda$. Set

$$\begin{aligned} M_{n,i}(x) &= \max_{\alpha\in\Lambda,\vartheta\in\Theta_i} \left[f_{n,\vartheta}^{\alpha}(x)\exp\{-c_n G(\alpha,\alpha)\}\right], \\ A_n(x) &= \left\{ i : M_{n,i}(x) = \max_{1\le j\le k} M_{n,j}(x)\right\}, \end{aligned}$$

and denote by $q_n^1(x) = (q_{n,1}^1(x), \ldots, q_{n,k}^1(x))$ the uniform distribution on A_n.

Lemma 2. If the Assumptions 1 and 3 are fulfilled and Λ is finite then

$$\begin{aligned} &\limsup_{n\to\infty}\left[\frac{1}{c_n}\ln R(\alpha,\vartheta,q_n^1) - G(\alpha,\alpha)\right] \\ \le\; &\max_{\beta\in\Lambda}\max_{\vartheta,\vartheta':\vartheta'\not\equiv\vartheta} \Gamma_{\alpha,\beta}\left(-G(\alpha,\alpha), -G(\beta,\beta),\vartheta,\vartheta'\right). \end{aligned}$$

Proof. By the definition of q_n^1 we get

$$\begin{aligned} &R\left(\alpha,\vartheta,q_n^1\right) \\ &\le \sum_{j\ne\langle\vartheta\rangle} Q_{n,\vartheta}^{\alpha}(M_{n,\langle\vartheta\rangle} \le M_{n,j}) \\ &\le \sum_{j\ne\langle\vartheta\rangle} Q_{n,\vartheta}^{\alpha}\left(\exp\{-c_n G(\alpha,\alpha)\} f_{n,\vartheta}^{\alpha} \le M_{n,j}\right) \\ &\le \sum_{\beta\in\Lambda}\sum_{j\ne\langle\vartheta\rangle}\sum_{\vartheta'\in\Theta_j} Q_{n,\vartheta}^{\alpha}\left(\exp\{-c_n G(\alpha,\alpha)\} f_{n,\vartheta}^{\alpha} \le \exp\{-c_n G(\beta,\beta)\} f_{n,\vartheta'}^{\beta}\right) \\ &\le kNL \max_{\vartheta'\not\equiv\vartheta}\max_{\beta\in\Lambda} Q_{n,\vartheta}^{\alpha}\left(\exp\{-c_n G(\alpha,\alpha)\} f_{n,\vartheta}^{\alpha} \le \exp\{-c_n G(\beta,\beta)\} f_{n,\vartheta'}^{\beta}\right). \end{aligned}$$

Set $\pi_{1,n} = \exp\{-c_n G(\alpha,\alpha)\}$, $\pi_{2,n} = \exp\{-c_n G(\beta,\beta)\}$ and denote by φ_n the Bayes test for $H_0 : Q_{n,\vartheta}^{\alpha}$ versus $H_1 : Q_{n,\vartheta'}^{\beta}$ with relative weights $\pi_{1,n}, \pi_{2,n}$, respectively. Then

$$\begin{aligned} &\exp\{-c_n G(\alpha,\alpha)\}\, Q_{n,\vartheta}^{\alpha}\left(\exp\{-c_n G(\alpha,\alpha)\} f_{n,\vartheta}^{\alpha} \le \exp\{-c_n G(\beta,\beta)\} f_{n,\vartheta'}^{\beta}\right) \\ &\quad\le \pi_{1,n}\int \varphi_n\, \mathrm{d}Q_{n,\vartheta}^{\alpha} + \pi_{2,n}\int (1-\varphi_n)\, \mathrm{d}Q_{n,\vartheta'}^{\beta}. \end{aligned}$$

To complete the proof we have only to apply Lemma 4 in the Appendix. □

Now we are ready to formulate a nessesary and sufficient condition for the existence of adaptive classification rules. This condition corresponds to Theorem 2.1 in Rukhin [13].

Theorem 4. Suppose that Assumptions 1 and 3 are fulfilled for every $\alpha, \beta \in \Lambda$. If an adaptive classification rule exists then

$$\inf_{0<s<1} [G(\alpha,\beta,\vartheta_1,\vartheta_2,s) - sG(\alpha,\alpha) - (1-s)\,G(\beta,\beta)] \leq 0 \tag{8}$$

for all α, $\beta \in \Lambda$, ϑ_1, $\vartheta_2 \in \Theta$ with $\vartheta_1 \not\equiv \vartheta_2$. Conversely, if (8) is fulfilled and Λ is finite then the weighted maximum likelihood classification rule q_n^1 is adaptive.

Proof. Suppose an adaptive rule q_n^a exists. Put $q_n = q_n^a$, $\pi_{1,n} = \exp\{-c_n G(\alpha,\alpha)\}$ and $\pi_{2,n} = \exp\{-c_n G(\beta,\beta)\}$. Then by the second statement in Lemma 1

$$\begin{aligned}
&\inf_{0<s<1}[G(\alpha,\beta,\vartheta_1,\vartheta_2,s) - sG(\alpha,\alpha) - (1-s)\,G(\beta,\beta)] \\
=\;& \Gamma_{\alpha,\beta}(-G(\alpha,\alpha), -G(\beta,\beta), \vartheta_1, \vartheta_2) \\
\leq\;& \liminf_{n\to\infty} \frac{1}{c_n} \ln\left[\max(\pi_{1,n}R(\alpha,\vartheta_1,q_n^a), \pi_{2,n}R(\beta,\vartheta_2,q_n^a))\right] \\
\leq\;& \liminf_{n\to\infty} \frac{1}{c_n} \max\left[\max_{\vartheta_1\in\Theta} \ln R(\alpha,\vartheta_1,q_n^a) - G(\alpha,\alpha), \max_{\vartheta_2\in\Theta} \ln R(\beta,\vartheta_2,q_n^a) - G(\beta,\beta)\right] \\
\leq\;& 0,
\end{aligned}$$

by the adaptivity of q_n^a. Conversely, let (8) be fulfilled. Then by Lemma 2,

$$\limsup_{n\to\infty}\left[\frac{1}{c_n} \ln \max_{\vartheta\in\Theta} R(\alpha,\vartheta,q_n^1)\right] \leq G(\alpha,\alpha),$$

which completes the proof. □

Remark. Condition (8) has an intuitive interpretation. As the Hellinger integral $H_s(P,Q)$ of order $0 < s < 1$ of the distributions P and Q is between 0 and 1 the expression $-\ln H_s(P,Q)$ is non-negative, and it can be shown that $-\ln H_s(P,Q)$ is small iff the variational distance between P and Q is small. For details we refer to Liese and Vajda [9]. Hence $-G(\alpha,\beta,\vartheta_1,\vartheta_2,s)$ is an asymptotic measure for the distance between Q^α_{n,ϑ_1} and Q^β_{n,ϑ_2}. Note that condition (8) is equivalent to

$$\inf_{0<s<1} [-sG(\alpha,\alpha) - (1-s)\,G(\beta,\beta) - (-G(\alpha,\beta,\vartheta_1,\vartheta_2,s))] \leq 0$$

for every α, $\beta \in \Lambda, \vartheta_1$, $\vartheta_2 \in \Theta$, $\vartheta_1 \not\equiv \vartheta_2$, which says that if an adaptive rule exists then asymptotically the distance between Q^α_{n,ϑ_1} and Q^β_{n,ϑ_2} is not smaller than the minimum distances within the models $Q^\alpha_{n,\vartheta}$, $\vartheta \in \Theta$ and $Q^\beta_{n,\vartheta}$, $\vartheta \in \Theta$.

4. APPLICATION TO SELECTING THE BEST POPULATION

Suppose we have independent samples from populations $\pi_1, \ldots, \pi_k$, where the distributions depend on a parameter. Let π_i have the parameter $\eta_i \in (a,b)$. We assume that the set $\{\eta_1,\ldots,\eta_k\}$ is known, but not which population belongs to which η-value. The task now is to select that population with the largest η-value. For every

$1 \le i \le k$, $n = 1, 2, \dots$ let $(R_{i,n}, \mathcal{R}_{i,n})$ be the sample space of a sample from population π_i with distribution P_{i,n,η_i}. The samples from $\pi_1, \dots, \pi_k$ are assumed to be independent, but we do not assume that the observations from π_i are independent. Moreover the sample sizes are allowed to be different for different populations. Let Θ denote the set of all permutations of $(1, \dots, k)$ where for $\vartheta \in \Theta$, $\vartheta : (1, \dots, k) \mapsto (i_1, \dots, i_k)$, we set $\vartheta(\eta_j) = \eta_{i_j}$ and $Q_{n,\vartheta} = P_{1,n,\vartheta(\eta_1)} \times \dots \times P_{k,n,\vartheta(\eta_k)}$. We may assume w.l.g. that $\eta_1 \le \dots \le \eta_k$. In addition we assume that $\eta_{k-1} < \eta_k$. Put $\Theta_j = \{\vartheta : \vartheta(j) = k\}$. For $\vartheta \in \Theta_j$ the population π_j has the largest η-value. Introduce the experiment E_n by

$$E_n = \left(\prod_{i=1}^{k} R_{i,n}, \ \bigotimes_{i=1}^{k} \mathcal{R}_{i,n}, Q_{n,\vartheta}, \vartheta \in \Theta \right). \tag{9}$$

The problem of selecting the best population, i.e. with the largest η-value, is thus reduced to the decision for one of the subsets $\Theta_1, \dots, \Theta_k$. We assume that for every η_i, η_j the distributions $P_{i,n,\eta_i}, P_{j,n,\eta_j}$ satisfy the Assumptions 1 and 2 if we substitute Q_{n,ϑ_1} and Q_{n,ϑ_2} by P_{i,n,η_i} and P_{j,n,η_j}, respectively. Put

$$\lim_{n \to \infty} \frac{1}{c_n} \ln H_s(P_{i,n,\eta}, P_{i,n,\eta'}) = G_i(\eta, \eta', s),$$

where $\eta, \eta' \in \{\eta_1, \dots, \eta_k\}$. In the sequel we will apply repeatedly the following well known property of Hellinger integrals:

$$H_s(P_1 \times \dots \times P_m, Q_1 \times \dots \times Q_m) = \prod_{i=1}^{m} H_s(P_i, Q_i). \tag{10}$$

Let ϑ_1, $\vartheta_2 \in \Theta$ with $\vartheta_1(l) = i_l$ and $\vartheta_2(l) = j_l$. Then this property implies that the family $Q_{n,\vartheta} = P_{1,n,\vartheta(\eta_1)} \times \dots \times P_{k,n,\vartheta(\eta_k)}$ satisfies Assumptions 1 and 2 and it holds

$$G(\vartheta_1, \vartheta_2, s) = \lim_{n \to \infty} \frac{1}{c_n} \ln H_s(Q_{n,\vartheta_1}, Q_{n,\vartheta_2}) = \sum_{l=1}^{k} G_l(\eta_{i_l}, \eta_{j_l}, s). \tag{11}$$

In order to apply the earlier results we have to find $\max_{\vartheta_1 \not\equiv \vartheta_2} \inf_{0<s<1} G(\vartheta_1, \vartheta_2, s)$. Since $G_i(\eta, \eta', s) \le 0$ for $0 < s < 1$ we have to take the maximum only over such permutations which coincide at a maximum number of $k-2$ arguments. For any $\vartheta_1 \not\equiv \vartheta_2$ the indices of populations with parameter η_k are different. Hence we have to take into consideration only permutations $(i_1, \dots, i_k), (j_1, \dots, j_k)$ for which there are $l, m \in \{1, \dots, k\}$, $l \ne m$ with $i_l = j_m = k$, $i_m = j_l$, $i_r = j_r$, $r \ne l$, $r \ne m$. Put $t = i_m = j_l$. Then by the above arguments

$$\max_{\vartheta_1 \not\equiv \vartheta_2} \inf_{0<s<1} G(\vartheta_1, \vartheta_2, s) = \max_{l,m: l \ne m; t: t<k} \inf_{0<s<1} \left[G_l(\eta_k, \eta_t, s) + G_m(\eta_t, \eta_k, s) \right]. \tag{12}$$

Let us study the rate of convergence to zero of the risks of decisions for the sequence of experiments E_n from (9). By construction, $\vartheta^{-1}(k)$ is the index of the population with the largest η-value. We set $L(\vartheta, i) = 0$ if $i = \vartheta^{-1}(k)$, and $L(\vartheta, i) = 1$, else. Let

$q_{i,n} : \prod_{i=1}^{k}(R_{i,n}, \mathcal{R}_{i,n}) \mapsto [0,1]$ be measurable with $\sum_{i=1}^{k} q_{i,n} = 1$, where $q_{i,n}$ is the probability to decide in favor of population π_i. The risk is given by

$$R(\vartheta, q_n) = 1 - \int q_{\vartheta^{-1}(k),n} \, \mathrm{d}Q_{n,\vartheta} = 1 - Q_{n,\vartheta}(CS, q_n),$$

where CS stands for correct selection and $q_n = (q_{1,n}, \ldots, q_{k,n})$. Denote by $g_{i,n,\eta}$ the density of $P_{i,n,\eta}$ w. r. t. to some σ-finite measure $\nu_{i,n}$, and put $\mu_n = \nu_{1,n} \times \ldots \times \nu_{k,n}$. Then

$$f_{n,\vartheta} = \frac{\mathrm{d}Q_{n,\vartheta}}{d\mu_n} = \prod_{i=1}^{k} g_{i,n,\vartheta(\eta_i)}.$$

Consider again the maximum likelihood selection rule q_n^0 introduced in Theorem 2. The following Theorem is a direct consequence of Theorems 1 and 2.

Theorem 5. Suppose that for every $\eta, \eta' \in \{\eta_1, \ldots, \eta_k\}$ the distributions $P_{i,n,\eta}$, $P_{i,n,\eta'}, \eta \neq \eta'$, satisfy Assumptions 1 and 2. Then for every sequence q_n of selection rules

$$\begin{aligned} \liminf_{n \to \infty} \frac{1}{c_n} \ln \left(\max_{\vartheta \in \Theta} (1 - Q_{n,\vartheta}(CS, q_n)) \right) \\ \geq \max_{l,m: l \neq m; t: t<k} \inf_{0<s<1} \left[G_l(\eta_k, \eta_t, s) + G_m(\eta_t, \eta_k, s) \right], \end{aligned} \tag{13}$$

where the lower bound is attained by the maximum likelihood selection rule q_n^0.

We now assume that the samples from the populations consist of independent observations distributed according to $P_{\eta_1}, \ldots, P_{\eta_k}$. Let the sample sizes be $n_1, \ldots, n_k$. Put $n = n_1 + \ldots + n_k$ and suppose $\frac{n_i}{n} \underset{n \to \infty}{\longrightarrow} \gamma_i$, where $0 < \gamma_i < 1$. We have $P_{i,n,\eta_i} = P_{\eta_i}^{n_i}$. Set $G(\eta, \eta', s) = \ln H_s(P_\eta, P_{\eta'})$. If $c_n = n$ is chosen, then by (10)

$$G_i(\eta, \eta', s) = \gamma_i \ln(H_s(P_\eta, P_{\eta'})) = \gamma_i G(\eta, \eta', s), \tag{14}$$

$$G(\vartheta_1, \vartheta_2, s) = \sum_{l=1}^{k} \gamma_l G(\eta_{i_l}, \eta_{j_l}, s) \tag{15}$$

and by (12)

$$\max_{\vartheta_1 \neq \vartheta_2} \inf_{0<s<1} G(\vartheta_1, \vartheta_2, s) = \max_{l \neq m; t<k} \inf_{0<s<1} \left[\gamma_l G(\eta_k, \eta_t, s) + \gamma_m G(\eta_t, \eta_k, s) \right]. \tag{16}$$

Consider the case of asymptotically equal sample sizes, i. e. $\gamma_1 = \ldots = \gamma_k = \frac{1}{k}$. Because of $H_s(P, Q) = H_{1-s}(Q, P)$ we have $G(\eta, \eta', s) = G(\eta', \eta, 1-s)$. The convex function $G(\eta, \eta', s) + G(\eta', \eta, s)$ is symmetric w. r. t. $s = 1/2$. Hence

$$\max_{\vartheta_1 \neq \vartheta_2} \inf_{0<s<1} G(\vartheta_1, \vartheta_2, s) = \frac{2}{k} \max_{r<k} G\left(\eta_k, \eta_r, \frac{1}{2}\right). \tag{17}$$

If $P_\eta \neq P_{\eta'}$ for every $\eta \neq \eta'$, then the family $P_{i,n,\eta} = P_{\eta_i}^{n_i}$ satisfies the assumptions of Theorem 5.

Corollary. Suppose $P_{\eta_i} \neq P_{\eta_j}$ for $i \neq j$, $\frac{n_i}{n} \mapsto \gamma_i$, and the observations are taken independently from the populations. Then for every sequence of selection rules q_n

$$\begin{aligned}
&\liminf_{n\to\infty} \frac{1}{n} \ln \left(\max_{\vartheta \in \Theta} (1 - Q_{n,\vartheta}(CS, q_n)) \right) \\
&\geq \max_{l \neq m; t < k} \inf_{0<s<1} [\gamma_l G(\eta_k, \eta_t, s) + \gamma_m G(\eta_t, \eta_k, s)] , \qquad (18)
\end{aligned}$$

where the maximum likelihood rule q_n^0 attains the lower bound. If $\gamma_1 = \ldots = \gamma_k = \frac{1}{k}$, then the right hand term in (18) is simplified to $\frac{2}{k} \max_{t<k} G(\eta_k, \eta_t, \frac{1}{2})$.

5. SELECTIONS UNDER UNEQUAL SAMPLE SIZES IN EXPONENTIAL FAMILIES

We now restrict our considerations to exponential families. Let $(R, \mathcal{R})$ be a measurable space, $T : R \mapsto \mathbb{R}_1$ a measurable function and ν a σ-finite measure on $(R, \mathcal{R})$. The set

$$I_T = \left\{ \eta : \int \exp\{\eta T\} \, d\nu < \infty \right\}$$

is an interval. We assume that the interior of I_T is nonempty and denote it by (a_1, a_2). Put $K(\eta) = \ln \int e^{\eta T} \, d\nu$ and

$$P_\eta(A) = \int_A \exp\{\eta T - K(\eta)\} \, d\nu. \qquad (19)$$

Suppose the family $(P_\eta)_{\eta \in I_T}$ is nontrivial in the sense that $\nu \circ T^{-1}$ is not concentrated at one point. Then $P_\eta \neq P_{\eta'}$ for $\eta \neq \eta'$ and K is strictly convex. The definition of H_s yields for $\eta_1, \eta_2 \in I_T$

$$H_s(P_{\eta_1}, P_{\eta_2}) = \exp\{-D(\eta_1, \eta_2, s)\} \qquad (20)$$

where $D(\eta_1, \eta_2, s) = sK(\eta_1) + (1-s) K(\eta_2) - K(s\eta_1 + (1-s)\eta_2)$. Suppose now that the populations $\pi_1, \ldots, \pi_k$ have distributions $P_{\eta_1}, \ldots, P_{\eta_k}$, $a_1 < \eta_1 \leq \ldots \leq \eta_{k-1} < \eta_k < a_2$ and n_i independent observations are taken from π_i where $\frac{n_i}{n} \mapsto \gamma_i$. Denote by $\gamma' \leq \gamma''$ the two smallest γ_i-values which will be very crucial in the following. Since K is convex the function $\eta \mapsto D(\eta, \eta', s)$ is nonincreasing for $\eta < \eta'$ and nondecreasing for $\eta > \eta'$. Hence $G(\eta, \eta', s) = \ln H_s(P_\eta, P_{\eta'})$ and

$$\begin{aligned}
&\max_{l \neq m; t<k} \inf_{0<s<1} [\gamma_l G(\eta_k, \eta_t, s) + \gamma_m G(\eta_t, \eta_k, s)] \\
&= - \min_{l \neq m; t<k} \sup_{0<s<1} [\gamma_l D(\eta_k, \eta_t, s) + \gamma_m D(\eta_t, \eta_k, s)] \\
&= - \min_{l \neq m} \sup_{0<s<1} [\gamma_l D(\eta_k, \eta_{k-1}, s) + \gamma_m D(\eta_{k-1}, \eta_k, s)] \\
&= - \sup_{0<s<1} [\gamma' D(\eta_k, \eta_{k-1}, s) + \gamma'' D(\eta_{k-1}, \eta_k, s)] \\
&=: -M(\gamma', \gamma'', \eta_{k-1}, \eta_k). \qquad (21)
\end{aligned}$$

Proposition 1. For independent observations from populations, whose distributions belong to the exponential family (19) and have parameters $\eta_1 \leq \ldots \leq \eta_{k-1} < \eta_k$, it holds

$$\liminf_{n\to\infty} \frac{1}{n} \ln \left(\max_{\vartheta}(1 - Q_{n,\vartheta}(CS, q_n)) \right) \geq -M(\gamma', \gamma'', \eta_{k-1}, \eta_k)$$

for any sequence of selection rules q_n, and the maximum likelihood rule q_n^0 attains the lower bound.

The statement of Proposition 1 is a direct consequence of the Corollary of Theorem 5.

Suppose the r. v. $X_{i,j}, j = 1, \ldots, n_i, i = 1, \ldots, k$ with values in $(R, \mathcal{R})$ are independent and $X_{i,j} \sim P_{\eta_i}$. Put $\bar{T}_{i,n} = \frac{1}{n_i}\sum_{j=1}^{n_i} T(X_{i,j})$ and set $A_n = \{i : \bar{T}_{i,n} = \max_{1\leq l \leq k} \bar{T}_{l,n}\}$. Then the natural selection rule, which selects in terms of the largest of the values $\bar{T}_{i,n}$, is given by the uniform distribution $\tilde{q}_n = (\tilde{q}_{n,1}, \ldots, \tilde{q}_{n,k})$ on A_n. If $Q_{n,\vartheta}$ is the true distribution then $\pi_{\vartheta^{-1}(k)}$ has the largest η-value η_k and $\pi_{\vartheta^{-1}(l)}$ has parameter η_l. Hence

$$\begin{aligned}
& \max_{l\neq k} Q_{n,\vartheta}\left(\bar{T}_{\vartheta^{-1}(k),n} < \bar{T}_{\vartheta^{-1}(l),n}\right) \\
\leq \;& Q_{n,\vartheta}\left(\bar{T}_{\vartheta^{-1}(k),n} < \max_{l\neq k} \bar{T}_{\vartheta^{-1}(l),n}\right) \\
\leq \;& 1 - Q_{n,\vartheta}(CS, \tilde{q}_n) \\
\leq \;& Q_{n,\vartheta}\left(\bar{T}_{\vartheta^{-1}(k),n} \leq \max_{l\neq k} \bar{T}_{\vartheta^{-1}(l),n}\right) \\
\leq \;& (k-1) \max_{l\neq k} Q_{n,\vartheta}\left(\bar{T}_{\vartheta^{-1}(k),n} \leq \bar{T}_{\vartheta^{-1}(l),n}\right).
\end{aligned} \tag{22}$$

Lemma 5 yields

$$\begin{aligned}
& \lim_{n\to\infty} \frac{1}{n} \ln Q_{n,\vartheta}\left(\bar{T}_{\vartheta^{-1}(k),n} < \bar{T}_{\vartheta^{-1}(l),n}\right) \\
&= \lim_{n\to\infty} \frac{1}{n} \ln Q_{n,\vartheta}\left(\bar{T}_{\vartheta^{-1}(k),n} \leq \bar{T}_{\vartheta^{-1}(l),n}\right) \\
&= \left(\gamma_{\vartheta^{-1}(k)} + \gamma_{\vartheta^{-1}(l)}\right) D\left(\eta_l, \eta_k, \frac{\gamma_{\vartheta^{-1}(k)}}{\gamma_{\vartheta^{-1}(l)} + \gamma_{\vartheta^{-1}(k)}}\right).
\end{aligned}$$

Consequently,

$$\begin{aligned}
& \lim_{n\to\infty} \frac{1}{n} \ln \max_{\vartheta} (1 - Q_{n,\vartheta}(CS, \tilde{q}_n)) \\
&= \max_{\vartheta} \lim_{n\to\infty} \frac{1}{n} \ln (1 - Q_{n,\vartheta}(CS, \tilde{q}_n)) \\
&= -\min_{\vartheta} \min_{l\neq k} \left(\gamma_{\vartheta^{-1}(k)} + \gamma_{\vartheta^{-1}(l)}\right) D\left(\eta_l, \eta_k, \frac{\gamma_{\vartheta^{-1}(k)}}{\gamma_{\vartheta^{-1}(k)} + \gamma_{\vartheta^{-1}(l)}}\right).
\end{aligned}$$

To evaluate the term on the right hand side we note that

$$D(\eta_l, \eta_k, s) \geq D(\eta_{k-1}, \eta_k, s)$$

for every $0 \leq s \leq 1$. Hence

$$\begin{aligned}
&\min_{\vartheta} \min_{l \neq k} (\gamma_{\vartheta^{-1}(k)} + \gamma_{\vartheta^{-1}(l)}) D\left(\eta_l, \eta_k, \frac{\gamma_{\vartheta^{-1}(k)}}{\gamma_{\vartheta^{-1}(k)} + \gamma_{\vartheta^{-1}(l)}}\right) \\
&\quad \geq \min_{\vartheta} \min_{l \neq k} (\gamma_{\vartheta^{-1}(k)} + \gamma_{\vartheta^{-1}(l)})\, D\left(\eta_{k-1}, \eta_k, \frac{\gamma_{\vartheta^{-1}(k)}}{\gamma_{\vartheta^{-1}(k)} + \gamma_{\vartheta^{-1}(l)}}\right) \\
&\quad = \min_{i \neq j} (\gamma_i + \gamma_j)\, D\left(\eta_{k-1}, \eta_k, \frac{\gamma_i}{\gamma_i + \gamma_j}\right) \\
&\quad = \min_{\vartheta} (\gamma_{\vartheta^{-1}(k)} + \gamma_{\vartheta^{-1}(k-1)})\, D\left(\eta_{k-1}, \eta_k, \frac{\gamma_{\vartheta^{-1}(k)}}{\gamma_{\vartheta^{-1}(k)} + \gamma_{\vartheta^{-1}(k-1)}}\right) \\
&\quad \geq \min_{\vartheta} \min_{l} (\gamma_{\vartheta^{-1}(k)} + \gamma_{\vartheta^{-1}(l)})\, D\left(\eta_{k-1}, \eta_k, \frac{\gamma_{\vartheta^{-1}(k)}}{\gamma_{\vartheta^{-1}(k)} + \gamma_{\vartheta^{-1}(l)}}\right)
\end{aligned}$$

and

$$\begin{aligned}
&\lim_{n \to \infty} \frac{1}{n} \ln \max_{\vartheta} (1 - Q_{n,\vartheta}(CS, \tilde{q}_n)) \\
&\quad = \; -\min_{i \neq j} (\gamma_i + \gamma_j)\, D\left(\eta_{k-1}, \eta_k, \frac{\gamma_i}{\gamma_i + \gamma_j}\right) \\
&\quad =: \; -\min_{i \neq j} L(\gamma_i, \gamma_j, \eta_{k-1}, \eta_k).
\end{aligned}$$

Thus we obtained

Theorem 6. If $X_{i,j} \sim P_{\eta_i}$, and all P_{η_i} belong to the exponential family (19), then

$$\lim_{n \to \infty} \frac{1}{n} \ln \max_{\vartheta \in \Theta} (1 - Q_{n,\vartheta}(CS, \tilde{q}_n)) = -\min_{i \neq j} L(\gamma_i, \gamma_j, \eta_{k-1}, \eta_k).$$

Consider again γ', γ'', the two smallest γ-values. From the first statement in Lemma 7 in the Appendix we get

$$M(\gamma', \gamma'', \eta_{k-1}, \eta_k) \geq \min_{i \neq j} L(\gamma_i, \gamma_j, \eta_{k-1}, \eta_k), \tag{23}$$

which says that for known η_i the maximum likelihood selection rule is at least as good as the natural rule. Furthermore, by the second statement in Lemma 7, equality holds in (23) iff $\gamma' = \gamma''$. Thus we have proved the following statement.

Proposition 2. Suppose the conditions of Proposition 1 are fulfilled and denote by $\tilde{q}_n$ the natural selection rule. Then

$$\lim_{n\to\infty} \frac{1}{n} \ln \max_{\vartheta\in\Theta}(1 - Q_{n,\vartheta}(CS, \tilde{q}_n)) \geq \lim_{n\to\infty} \frac{1}{n} \ln \max_{\vartheta\in\Theta}(1 - Q_{n,\vartheta}(CS, q_n^0))$$

where equality holds iff $\gamma' = \gamma''$.

Now we examine whether for populations from an exponential family and non-equal sample sizes adaptive selection procedures do exist. Suppose $\alpha = (\alpha_1, \ldots, \alpha_k)$, $\beta = (\beta_1, \ldots, \beta_k)$ and assume ϑ_1, ϑ_2 are permutations with $\vartheta_1(l) = i_l$ and $\vartheta_2(l) = j_l$. Set $Q^{\alpha}_{n,\vartheta_1} = P^{n_1}_{\alpha_{i_1}} \times \ldots \times P^{n_k}_{\alpha_{i_k}}$ and $Q^{\beta}_{n,\vartheta_2} = P^{n_1}_{\beta_{j_1}} \times \ldots \times P^{n_k}_{\beta_{j_k}}$, where each P_η is from the family (19). Then by (20), analogously to (15),

$$G(\alpha, \beta, \vartheta_1, \vartheta_2, s) = -\sum_{l=1}^{k} \gamma_l D(\alpha_{i_l}, \beta_{j_l}, s). \tag{24}$$

Next, the parameters will be specialized. Put for $a_1 < t, x, \eta_0 < a_2$ and sufficiently small $h > 0$, $\alpha = (\eta_0, \ldots, \eta_0, t, t+h)$ and $\beta = (\eta_0, \ldots, \eta_0, x, x+h)$. Denote by $\gamma' \leq \gamma''$ the two smallest γ_i-values. Then by formula (24), with the same arguments which led to (12) and (21), we get

$$\begin{aligned} G(\alpha, \alpha) &= -\sup_{0<s<1} [\gamma' D(t, t+h, s) + \gamma'' D(t+h, t, s)] \\ G(\beta, \beta) &= -\sup_{0<s<1} [\gamma' D(x, x+h, s) + \gamma'' D(x+h, x, s)]. \end{aligned}$$

Let l, m be two indices such that $\gamma_l = \gamma', \gamma_m = \gamma''$ and choose two permutations ϑ_1^0 and ϑ_2^0 such that $\alpha_{i_l} = t+h, \alpha_{i_m} = t$ and $\beta_{j_l} = x, \beta_{j_m} = x+h$. Then by (24)

$$\begin{aligned} &\max_{\vartheta_1 \not\equiv \vartheta_2} \inf_{0<s<1} G(\alpha, \beta, \vartheta_1, \vartheta_2, s) \\ &\quad \geq \inf_{0<s<1} G(\alpha, \beta, \vartheta_1^0, \vartheta_2^0, s) \\ &\quad = -\sup_{0<s<1} [\gamma' D(t+h, x, s) + \gamma'' D(t, x+h, s)]. \end{aligned}$$

Suppose now an adaptive selection rule q_n^a exists. Then by the definition of the adaptivity

$$\limsup_{n\to\infty} \frac{1}{c_n} \ln R(\alpha, \vartheta_1^0, q_n^a) \leq \lim_{n\to\infty} \frac{1}{c_n} \ln \left(\max_{\vartheta\in\Theta} R(\alpha, \vartheta, q_n^a)\right) = G(\alpha, \alpha)$$

and analogously

$$\limsup_{n\to\infty} \frac{1}{c_n} \ln R(\beta, \vartheta_2^0, q_n^a) \leq G(\beta, \beta).$$

Hence

$$\limsup_{n\to\infty} \frac{1}{c_n} \ln \left[\max(R(\alpha, \vartheta_1^0, q_n^a), R(\beta, \vartheta_2^0, q_n^a))\right] \leq \max(G(\alpha, \alpha), G(\beta, \beta))$$

and by the first statement in Lemma 1

$$\max(G(\alpha,\alpha), G(\beta,\beta)) \geq \inf_{0<s<1} G(\alpha,\beta,\vartheta_1^0,\vartheta_2^0,s).$$

Hence

$$\begin{aligned}\min\Big(&\sup_{0<s_1<1}[\gamma' D(t,t+h,s_1)+\gamma'' D(t+h,t,s_1)],\\ &\qquad \sup_{0<s_2<1}[\gamma' D(x,x+h,s_2)+\gamma'' D(x+h,x,s_2)]\Big)\\ &\leq \sup_{0<s<1}[\gamma' D(t+h,x,s)+\gamma'' D(t,x+h,s)].\end{aligned}$$

Consequently, by $D\big(x,y,\frac{1}{2}\big) = D\big(y,x,\frac{1}{2}\big)$ we get with $\kappa = \frac{\gamma'}{\gamma'+\gamma''}$,

$$\begin{aligned}&\min\big(D(t,t+h,\tfrac{1}{2}), D(x,x+h,\ \tfrac{1}{2})\big)\\ &\leq \sup_{0<s<1}\left[\kappa D(t+h,x,s)+(1-\kappa)\, D(t,x+h,s)\right].\end{aligned}$$

Put $t = x+(1-2\kappa)\,h$. Then the last inequality for sufficiently small $h>0$ implies together with Lemma 6 that $\kappa = \frac{1}{2}$, i.e. $\gamma' = \gamma''$. Conversely, if $\gamma' = \gamma''$, then by Proposition 2 the natural selection rule attains for all parameter configurations the optimal exponential rate and is adaptive. To summarize, we have proved the following theorem.

Theorem 7. Under the assumptions of Proposition 1 an adaptive selection rule exists for the family $(P_\eta)_{a_1<\eta<a_2}$ iff the smallest two sample sizes are asymptotically equal, i.e. $\gamma' = \gamma''$. In this case the natural selection rule is adaptive.

Example. We now study the special case of normal populations and suppose that the r.v. $X_{i,j} \sim N(\mu_i,\sigma_i^2)$ $1 \leq i \leq n_i$; $i = 1,\ldots,k$, are independent. The variances $\sigma_1^2,\ldots,\sigma_k^2$ are supposed to be known and without loss of generality we assume $\mu_1 \leq \ldots \leq \mu_{k-1} < \mu_k$. $(\bar{X}_1,\ldots,\bar{X}_k)$ is a sufficient statistic with $\bar{X}_i \sim N(\mu_i,p_i), p_i = \frac{\sigma_i^2}{n_i}$. To study the problem of selection of the best population it is therefore enough to consider the experiment

$$E = (\mathbb{R}_k,\mathcal{B}_k,Q_\vartheta,\vartheta\in\Theta)$$

where $Q_\vartheta = \prod_{i=1}^k N(\mu_{\vartheta(i)},p_i)$ and Θ is the set of all permutations of $\{1,\ldots,k\}$. The set $\{\mu_1,\ldots,\mu_k\}$ is known and fixed. The asymptotic $p_i \to 0$ includes both the asymptotic for large sample sizes and the small variance asymptotic. To be more precise we assume that $p_{i,n} > 0$ $i = 1,\ldots,k$, are sequences such that $\lim_{n\to\infty}\frac{p_{i,n}}{p_n} = \rho_i > 0$ exist, where $p_n = \sum_{i=1}^k p_{i,n}$. Denote by $\rho' \geq \rho''$ the two largest ρ_i-values. We set $Q_{n,\vartheta} = \prod_{i=1}^k N(\mu_{\vartheta(i)},p_{i,n})$ and study the asymptotic behavior of the error probability of selection procedures in the experiments

$$E_n = (\mathbb{R}_k,\mathcal{B}_k,Q_{n,\vartheta},\vartheta\in\Theta). \tag{25}$$

To apply the results from Chapter 4 we need the Hellinger integral for two normal distributions. A simple calculation shows

$$H_s(N(\mu_1,\sigma^2),N(\mu_2,\sigma^2)) = \exp\left\{-\frac{1}{2}s(1-s)\,\frac{(\mu_1-\mu_2)^2}{\sigma^2}\right\}. \tag{26}$$

If $\vartheta_1(l) = i_l,\ \vartheta_2(l) = j_l$ then by (10)

$$\ln H_s(Q_{n,\vartheta_1},Q_{n,\vartheta_2}) = -\frac{1}{2}s(1-s)\sum_{l=1}^{k}\frac{(\mu_{i_l}-\mu_{j_l})^2}{p_{l,n}}. \tag{27}$$

Consequently, with $c_n = \frac{1}{p_n}$

$$\lim_{n\to\infty}\frac{1}{c_n}\ln H_s(Q_{n,\vartheta_1},Q_{n,\vartheta_2}) = -\frac{1}{2}s(1-s)\sum_{l=1}^{k}\frac{(\mu_{i_l}-\mu_{j_l})^2}{\rho_l} = G(\vartheta_1,\vartheta_2,s)$$

and

$$\begin{aligned}\max_{\vartheta_1\neq\vartheta_2}\inf_{0<s<1} G(\vartheta_1,\vartheta_2,s) &= -\frac{1}{8}\frac{(\mu_{k-1}-\mu_k)^2}{\rho'}-\frac{1}{8}\frac{(\mu_k-\mu_{k-1})^2}{\rho''}\\ &= -\frac{1}{8}\frac{\rho'+\rho''}{\rho'\rho''}(\mu_{k-1}-\mu_k)^2.\end{aligned}$$

Analogously to Theorem 5 we get from Theorem 1 and Theorem 2 that for every sequence of selection rules q_n for the experiments E_n in (25)

$$\liminf_{n\to\infty} p_n \ln\left(\max_{\vartheta\in\Theta}(1-Q_{n,\vartheta}(CS,q_n))\right) \geq -\frac{1}{8}\,\frac{\rho'+\rho''}{\rho'\rho''}\,(\mu_{k-1}-\mu_k)^2 \tag{28}$$

where the lower bound is attained by the maximum likelihood selection rule q_n^0.

In order to study the asymptotic of the error probabilities of the natural selection rule we recall to the well known fact that

$$\lim_{t\to\infty} P(Z>t)\left[\frac{1}{\sqrt{2\pi}t}e^{-\frac{t^2}{2}}\right]^{-1} = 1$$

if $Z\sim N(0,1)$. This implies

$$\lim_{x\to\infty}\frac{1}{x}\ln P(Z>a\sqrt{x}) = -\frac{1}{2}a^2. \tag{29}$$

Denote by $Y_i : \mathbb{R}_k \to \mathbb{R}_1$ the projections to the coordinates. Then $Y_i \sim N(\mu_{\vartheta(i)}, p_{i,n})$ in the experiments E_n. The natural selection rule $\tilde{q}_n$ selects the population π_i with $Q_{n,\vartheta}$-probability one iff $Y_i > \max_{j\neq i} Y_j$. Hence

$$\begin{aligned}\max_{1\leq l\leq k-1} Q_{n,\vartheta}(Y_{\vartheta(k)} < Y_{\vartheta(l)}) &\leq 1 - Q_{n,\vartheta}(CS,\tilde{q}_n)\\ &\leq (k-1)\max_{1\leq l\leq k-1} Q_{n,\vartheta}(Y_{\vartheta(k)} < Y_{\vartheta(l)})\end{aligned}$$

and by (29) and similar considerations as before Theorem 6

$$\lim_{n\to\infty} p_n \ln \max_{\vartheta}(1 - Q_{n,\vartheta}(CS, \hat{q}_n)) = -\frac{1}{2}\frac{1}{\rho' + \rho''}(\mu_{k-1} - \mu_k)^2 . \tag{30}$$

To compare the exponential rates of error probabilities of the maximum likelihood rule and the natural rule we note that

$$(\rho'\rho'')^{1/2} \le \frac{1}{2}(\rho' + \rho'')$$

and consequently

$$\frac{1}{2}\frac{1}{\rho' + \rho''} \le \frac{1}{8}\frac{\rho' + \rho''}{\rho'\rho''} \tag{31}$$

where the sign of equality holds iff $\rho' = \rho''$. This means that in view of (28) and (30) the maximum likelihood selection rule is strictly better than the natural selection rule w. r. t. the exponential rates, except for $\rho' = \rho''$ where the rates coincide. To relate this statement to Theorem 6 we consider the large sample asymptotic $p_{i,n} = \frac{\sigma^2}{n_i}$ and note that for known σ^2 the family $N(\mu, \sigma^2)$,$-\infty < \mu < \infty$, is an exponential family with $K(\mu) = \frac{1}{2}\frac{\mu^2}{\sigma^2}$. Then $D(\mu_1, \mu_2, s) = \frac{1}{2}s(1-s)\,\frac{(\mu_1 - \mu_2)^2}{\sigma^2}$.

Note that

$$\rho_i = \lim_{n\to\infty} \frac{p_{i,n}}{p_n} = \lim_{n\to\infty} \frac{n}{n_i}\left(\sum_{i=1}^{k} \frac{n}{n_i}\right)^{-1} = \frac{\Gamma}{\gamma_i}$$

where $\Gamma = \sum_{i=1}^{k} \frac{1}{\gamma_i}$. From this we can see that

$$M(\gamma', \gamma'', \mu_{k-1}, \mu_k) = \frac{\Gamma}{8}\frac{\rho'\rho''}{\rho' + \rho''}\frac{(\mu_{k-1} - \mu_k)^2}{\sigma^2}$$

and

$$\min_{i\neq j} L(\gamma_i, \gamma_j, \mu_{k-1}, \mu_k) = \frac{\Gamma}{2}\frac{1}{\rho' + \rho''}\frac{(\mu_{k-1} - \mu_k)^2}{\sigma^2}.$$

Therefore inequality (31) implies (23) and from the relations (28) and (30) we get the statements of Theorem 6 and Proposition 2 for normal populations.

Now we examine under which conditions there exist asymptotically optimal selection rules which do not use the knowledge of $\{\mu_1, \ldots, \mu_k\}$ in the experiments (25). To study the problem of the existence of adaptive selection rules we assume that ϑ_1, ϑ_2 are permutations with $\vartheta_1(l) = i_l, \vartheta_2(l) = j_l$ and set $\alpha = (\alpha_1, \ldots, \alpha_k)$, $\alpha_1 \le \ldots \le \alpha_{k-1} < \alpha_k$, $\beta = (\beta_1, \ldots, \beta_k)$, $\beta_1 \le \ldots \le \beta_{k-1} < \beta_k$,

$$Q^{\alpha}_{n,\vartheta_1} = N(\alpha_{i_1}, p_{1,n}) \times \ldots \times N(\alpha_{i_k}, p_{k,n})$$
$$Q^{\beta}_{n,\vartheta_2} = N(\beta_{j_1}, p_{1,n}) \times \ldots \times N(\beta_{j_k}, p_{k,n}),$$

$c_n = \frac{1}{p_n}$. Then by (26)

$$G(\alpha, \beta, \vartheta_1, \vartheta_2, s) = \lim_{n\to\infty} \frac{1}{c_n} \ln H_s(Q^{\alpha}_{n,\vartheta_1}, Q^{\beta}_{n,\vartheta_2}) = -\sum_{l=1}^{k} \frac{1}{2}s(1-s)\,\frac{(\alpha_{i_l} - \beta_{j_l})^2}{\rho_l}$$

and

$$
\begin{aligned}
G(\alpha,\alpha) &= -\frac{1}{8}\frac{\rho'+\rho''}{\rho'\rho''}(\alpha_k-\alpha_{k-1})^2 \\
G(\beta,\beta) &= -\frac{1}{8}\frac{\rho'+\rho''}{\rho'\rho''}(\beta_k-\beta_{k-1})^2.
\end{aligned}
$$

Put for $h>0$ and $x>t$, $\alpha=(\mu_0,\ldots,\mu_0,t,t+h), \beta=(\mu_0,\ldots,\mu_0,x,x+h)$. Let l, m be two indices such that $\rho_l=\rho', \rho_m=\rho''$ and choose two permutations $\vartheta_1^0, \vartheta_2^0$ such that $\alpha_{i_l}=t+h, \alpha_{i_m}=t, \beta_{j_l}=x, \beta_{j_m}=x+h$.

Then

$$
G(\alpha,\alpha)=G(\beta,\beta)=-\frac{1}{8}\frac{\rho'+\rho''}{\rho'\rho''}h^2
$$

and

$$
\begin{aligned}
\inf_{0<s<1} G(\alpha,\beta,\vartheta_1^0,\vartheta_2^0,s) &= \inf_{0<s<1} -\frac{1}{2}s(1-s)\left[\frac{(t+h-x)^2}{\rho'}+\frac{(t-x-h)^2}{\rho''}\right] \\
&= -\frac{1}{8}\left[\frac{(t+h-x)^2}{\rho'}+\frac{(t-x-h)^2}{\rho''}\right] \\
&= -\frac{1}{8}\frac{\rho'+\rho''}{\rho'\rho''}(h^2+(t-x)^2)-\frac{1}{4}h\left(\frac{t-x}{\rho'}-\frac{t-x}{\rho''}\right).
\end{aligned}
$$

Suppose an adaptive selection rule exists. Then by the first statement in Lemma 1 and Definition 1

$$
\inf_{0<s<1} G(\alpha,\beta,\vartheta_1^0,\vartheta_2^0,s) \le \max\left(G(\alpha,\alpha),G(\beta,\beta)\right).
$$

Hence

$$
-\frac{1}{8}\frac{\rho'+\rho''}{\rho'\rho''}(t-x)^2 \le -\frac{1}{4}h\left(\frac{t-x}{\rho'}-\frac{t-x}{\rho''}\right).
$$

Because of $x>t$ we get

$$
\frac{\rho'+\rho''}{\rho'\rho''}(x-t)\ge 2h\frac{\rho'-\rho''}{\rho'\rho''}.
$$

As $\rho''\le\rho'$ this inequality for every $h>0$ and every x,t with $x>t$ implies $\rho'=\rho''$.

As we know already that for $\rho'=\rho''$ the natural rule is asymptotically optimal, we conclude that an adaptive selection rule for the experiments (25) exist iff $\rho'=\rho''$. For the large sample asymptotic $n_i\to\infty$ this statement is also a consequence of Theorem 7.

APPENDIX

In this Appendix, we collect and modify large deviation results for the log-likelihood. Although our statements below are presented for arbitrary sequences of distributions, some of the proofs are straightforward modifications of the proofs in the i. i. d. case, since only some asymptotic behavior of the Hellinger integral is used. Therefore, complete proofs are omitted for brevity, but they can be found in Liese and Miescke [8].

Let P and Q be distributions on $(\mathcal{X}, \mathcal{A})$ with densities f and g, resp., w. r. t. some σ-finite μ. Suppose $\pi_1, \pi_2 > 0$. Let φ be a Bayes test for $H_0 : P$ versus $H_A : Q$ with weights π_1, π_2. Then $\varphi = 1$ on $\{\pi_1 f < \pi_2 g\}$, $\varphi = 0$ on $\{\pi_1 f \geq \pi_2 g\}$, and $0 \leq \varphi \leq 1$ be any value on $\{\pi_1 f = \pi_2 g\}$. If $0 < s < 1$, then

$$\pi_1 \int \varphi \, \mathrm{d}P + \pi_2 \int (1 - \varphi) \, \mathrm{d}Q \leq \pi_1^s \pi_2^{1-s} H_s(P, Q). \tag{32}$$

This inequality has been established in Krafft and Plachky [6]. Suppose now we are given sequences P_n and Q_n of distributions, and $\pi_{i,n}$, i=1,2, with

$$\lim_{n\to\infty} \frac{1}{c_n} \ln \pi_{1,n} = a, \quad \lim_{n\to\infty} \frac{1}{c_n} \ln \pi_{2,n} = b.$$

Let φ_n be a sequence of Bayes tests for $H_0 : P_n$ versus $H_A : Q_n$ with weights $\pi_{1,n}$ and $\pi_{2,n}$. Then by (32), with, say $G(s) = \lim_{n\to\infty} (1/c_n) \ln H_s(P_n, Q_n)$,

$$\limsup_{n\to\infty} \left[\pi_{1,n} \int \varphi_n \, \mathrm{d}P_n + \pi_{2,n} \int (1 - \varphi_n) \, \mathrm{d}Q_n \right]^{\frac{1}{c_n}} \leq e^{sa+(1-s)b+G(s)}. \tag{33}$$

Under some mild conditions equality holds in (33). This statement is known as Chernoff's theorem which was originally proved for the i. i. d. case and for $\pi_{1,n} = \pi_1$, $\pi_{2,n} = \pi_2$ which implies $a = b = 0$. The case where P_n, Q_n are restrictions to an increasing sequence of sub-σ-algebras has been considered in Vajda [15]. An inspection of the proofs of Chernoff's theorem in Krafft and Plachky [6] and Vajda [15] shows that their arguments up to small modifications also work in the general situation. This leads to the following statement.

Lemma 3. Assume $P_n \sim Q_n$ and $\lim_{n\to\infty} \frac{1}{c_n} \ln H_s(P_n, Q_n) = G(s)$ exists for every $0 \leq s \leq 1$, where G is continuously differentiable in $(0, 1)$, continuous in $[0, 1]$ and not identical zero. If $a = b$ then

$$\lim_{n\to\infty} \left[\pi_{1,n} \int \varphi_n \, \mathrm{d}P_n + \pi_{2,n} \int (1 - \varphi_n) \, \mathrm{d}Q_n \right]^{\frac{1}{c_n}} = \exp\left\{ a + \inf_{0<s<1} G(s) \right\}.$$

The restriction to equivalent distributions ($P_n \sim Q_n$) is not necessary. But for simplicity we consider only this situation everywhere throughout the paper.

To deal with the case where $\pi_{1,n}$ and $\pi_{2,n}$ have different asymptotic behavior in the sense of exponential rates one needs additional assumptions on $G(s)$. Then arguments of the paper by Krafft and Plachky [6] may be applied.

Lemma 4. Suppose $\lim_{n\to\infty} \frac{1}{c_n} \ln H_s(P_n, Q_n) = G(s)$ exists for every $-\infty < s < \infty$, and assume that the interior $I^0 = (a_1, a_2)$ of $I = \{s : G(s) < \infty\}$ contains $[0,1]$. If G is continuously differentiable in I^0, and if $\lim_{s\downarrow a_1} G'(s) = -\infty$ and $\lim_{s\uparrow a_2} G'(s) = \infty$, then

$$\begin{aligned}&\lim_{n\to\infty}\left[\pi_{1,n}\int \varphi_n \,\mathrm{d}P_n + \pi_{2,n}\int (1-\varphi_n)\,\mathrm{d}Q_n\right]^{\frac{1}{c_n}}\\ &= \exp\left\{\inf_{0<s<1}[as + (1-s)\,b + G(s)]\right\}.\end{aligned}$$

Now we formulate a large deviation result for the arithmetic mean of r. v. from an exponential family. To be more precise let $P_\eta, \eta \in I$, be the nontrivial exponential family (19) with natural parameter set I_T. Let $m_i \to \infty, i = 1,2$ so that $\kappa = \lim_{m\to\infty} \frac{m_1}{m}$ exists and satisfies $0 \le \kappa < 1$, where $m = m_1 + m_2$.

Suppose $\eta_i \in I_T^0 = (a_1, a_2)$ and that $X_{i,j}, j = 1, \ldots, m_i, i = 1,2$ are independent r. v. with $X_{i,j} \sim P_{\eta_i}$. Set $\bar{T}_{i,m} = \frac{1}{m_i}\sum_{i=1}^{m_i} T(X_{i,j})$ and

$$D(\eta_1, \eta_2, s) = sK(\eta_1) + (1-s)\,K(\eta_2) - K(s\eta_1 + (1-s)\eta_2).$$

The following large deviation result may be proved by the method of exponential centering which can be found in Bahadur [1] and Rüschendorf [10].

Lemma 5. If the r. v. $X_{i,j}$ are independent with $X_{i,j} \sim P_{\eta_i}$, $\eta_1 < \eta_2$, $\eta_i \in I_T^0$, $\frac{m_1}{m} \underset{m\to\infty}{\longrightarrow} \kappa$ and $0 < \kappa < 1$, then

$$\begin{aligned}&\lim_{m\to\infty} \frac{1}{m} \ln P(\bar{T}_{1,m} - \bar{T}_{2,m} > 0)\\ &= \lim_{m\to\infty} \frac{1}{m} \ln P(\bar{T}_{1,m} - \bar{T}_{2,m} \ge 0) = -D(\eta_1, \eta_2, \kappa).\end{aligned}$$

Next we formulate two technical lemmas on convex functions. Suppose K is a convex function defined on the interval I and K is twice continuously differentiable in the interior (a,b) of I. An application of l'Hospital's rule yields, for every fixed $x \in (a,b)$, and for every t_1, t_2,

$$\begin{aligned}\lim_{h\downarrow 0} \frac{2}{h^2} D(x + t_1 h, x + t_2 h, s) &= \lim_{h\downarrow 0} \frac{2}{h^2}[sK(x + t_1 h) + (1-s)\,K(x + t_2 h)\\ &\quad - K(s(x + t_1 h) + (1-s)(x + t_2 h))]\\ &= K''(x)s(1-s)(t_1 - t_2)^2, \qquad (34)\end{aligned}$$

where the limit is uniform for $0 \le s \le 1$.

Lemma 6. Suppose $x \in (a,b)$, $K''(x) > 0$, $h_n > 0$, and $h_n \downarrow 0$ for $n \to \infty$. If $0 < \kappa \le \frac{1}{2}$ and

$$\begin{aligned}&\min\left(D(x, x+h_n, \tfrac{1}{2}),\ D(x+(1-2\kappa)\,h_n x, x+2(1-\kappa)\,h_n, \tfrac{1}{2})\right)\\ &\quad\le \sup_{0<s<1}\,[\kappa D(x, x+2(1-\kappa)\,h_n, s) + (1-\kappa)\,D(x+h_n, x+(1-2\kappa)\,h_n, s)]\end{aligned}$$

for every n, then it follows that $\kappa = \frac{1}{2}$.

Proof. The relation (34) yields, uniformly in $0 \le s \le 1$

$$\begin{aligned}&\frac{2}{h_n^2}\,[\kappa D(x, x+2(1-\kappa)\,h_n, s) + (1-\kappa)\,D(x+h_n, x+(1-2\kappa)\,h_n, s)]\\ &\quad= 4\kappa(1-\kappa)\,K''(x)s(1-s) + o(1),\end{aligned}$$

and

$$\begin{aligned}&\frac{2}{h_n^2}\left[\min\left(D(x, x+h_n, \tfrac{1}{2})\right), D(x+(1-2\kappa)\,h_n x, x+2(1-\kappa)\,h_n, \tfrac{1}{2}))\right]\\ &\quad= \frac{1}{4}K''(x) + o(1).\end{aligned}$$

Hence, by the assumed inequalities, as $n \to \infty$,

$$\frac{1}{4}K''(x) \le \sup_{0<s<1} 4\kappa(1-\kappa)\,K''(x)s(1-s) = \kappa(1-\kappa)\,K''(x),$$

which implies $\kappa = \frac{1}{2}$, since $K''(x) > 0$. □

Lemma 7. Let $K : (a,b) \to \mathbb{R}$, be a convex function and assume that $a < x < y < b$ are fixed. If $\gamma_1, \ldots, \gamma_k$ are positive numbers, and $\gamma' \le \gamma''$ are the two smallest γ_i-values, then

$$\begin{aligned}&\min_{i\neq j}(\gamma_i+\gamma_j)D\left(x, y, \frac{\gamma_i}{\gamma_i+\gamma_j}\right)\\ &\quad\le (\gamma'+\gamma'')\sup_{0<s<1}\left[\frac{\gamma'}{\gamma'+\gamma''}D(x,y,s) + \frac{\gamma''}{\gamma'+\gamma''}D(x,y,1-s)\right]. \qquad (35)\end{aligned}$$

If $\gamma' = \gamma''$ then equality holds. Conversely, if for a strictly convex K equality holds in the above inequality, then $\gamma' = \gamma''$.

Sketch of proof: We have

$$\begin{aligned}&\min_{i\neq j}(\gamma_i+\gamma_j)D\left(x, y, \frac{\gamma_i}{\gamma_i+\gamma_j}\right)\\ &\quad\le (\gamma'+\gamma'')\min\left[D\left(x,y,\frac{\gamma'}{\gamma'+\gamma''}\right), D\left(x,y,\frac{\gamma''}{\gamma'+\gamma''}\right)\right]\\ &\quad\le (\gamma'+\gamma'')\left[\frac{\gamma'}{\gamma'+\gamma''}D\left(x,y,\frac{\gamma'}{\gamma'+\gamma''}\right) + \frac{\gamma''}{\gamma'+\gamma''}D\left(x,y,\frac{\gamma''}{\gamma'+\gamma''}\right)\right]\\ &\quad\le (\gamma'+\gamma'')\sup_{0<s<1}\left[\frac{\gamma'}{\gamma'+\gamma''}D(x,y,s) + \frac{\gamma''}{\gamma'+\gamma''}D(x,y,1-s)\right].\end{aligned}$$

Hence (35) is established. It is easy to see that $D(x,y,\alpha) \geq 2\alpha D(x,y,\frac{1}{2})$ for $x < y, \alpha \leq \frac{1}{2}$. Consequently, for $\gamma' = \gamma''$ and $\gamma_i \leq \gamma_j$

$$\begin{aligned}
&(\gamma_i + \gamma_j)\, D\left(x, y, \frac{\gamma_i}{\gamma_i + \gamma_j}\right) \\
\geq\ & 2\gamma_i D\left(x, y, \frac{1}{2}\right) \\
\geq\ & 2\gamma' D\left(x, y, \frac{1}{2}\right) \\
=\ & (\gamma' + \gamma'') \sup_{0<s<1} \left[\frac{\gamma'}{\gamma' + \gamma''} D(x,y,s), \frac{\gamma''}{\gamma' + \gamma''} D(x,y,1-s)\right]
\end{aligned}$$

as $D(x,y,s) + D(x,y,1-s)$ is symmetric w. r. t. $s_0 = \frac{1}{2}$. Hence for $\gamma' = \gamma''$ we have equality in (35).

Conversely, an analysis of the stability of the above inequalities shows that equality occurs only for $\gamma' = \gamma''$.

(Received May 28, 1997.)

REFERENCES

[1] R. R. Bahadur: Some Limit Theorems in Statistics. Soc. Indust. Appl. Math., Philadelphia 1971.

[2] I. A. Bucklew: Large Deviation Techniques in Decision, Simulation and Estimation. J. Wiley, New York 1990.

[3] H. Chernoff: A measure of asymptotic efficiency for tests of hypothesis based on the sum of observations. Ann. Math. Statist. *23* (1952), 493–507.

[4] H. Chernoff: Large sample theory: Parametric case. Ann. Math. Statist. *27* (1956), 1–22.

[5] A. Dembo and O. Zeitouni: Large sample Deviation Techniques and Applications. Jones and Bartlett, London 1993.

[6] O. Krafft and D. Plachky: Bounds for the power of likelihood ratio tests and their asymptotic and their asymptotic properties. Ann. Math. Statist. *41* (1970), 5, 1646–1664.

[7] O. Krafft and M. L. Puri: The asymptotic behaviour of the minimax risk for multiple decision problems. Sankhyā *36* (1974), 1, 1–12 .

[8] F. Liese and K. L. Miescke: Exponential Rates for the Error Probabilities in Selection Procedures. Preprint 96/5, FB Mathematik, Universität Rostock, Rostock 1996.

[9] F. Liese and I. Vajda: Convex Statistical Distances. Teubner, Leipzig 1987.

[10] L. Rüschendorf: Asymptotische Statistik. BGB Teubner, Stuttgart 1988.

[11] A. L. Rukhin: Adaptive procedure for a finite numbers of probability distributions. Statist. Decis. Theory Related Topics III. *2* (1982), 269–285.

[12] A. L. Rukhin: Adaptive classification procedures. J. Amer. Statist. Assoc. *79* (1984), 415–422.

[13] A. L. Rukhin: Adaptive testing of multiple hypotheses for stochastic processes. Stochastic Anal. Appl. *12* (1994), 5, 567–582.

[14] A. L. Rukhin and I. Vajda: Adaptive decision making for stochastic processes. J. Statist. Plann. Inference *45* (1995), 313–329.

[15] I. Vajda: Theory of Statistical Inference and Information. Kluwer, Dordrecht 1989.

Prof. Dr. Friedrich Liese, Department of Mathematics, University of Rostock, Universitätplatz 1. 18051 Rostock. Germany.
e-mail: friedrich.liese@mathematik.uni-rostock.de

Prof. Dr. Klaus J. Miescke, Department of Statistics and Computer Science, University of Illinois at Chicago, 851 South Morgan Street, Chicago, Illinois 60607-7045. U.S.A.
e-mail: klaus@uic.edu