The Annals of Statistics
1982, Vol. 10, No. 4, 1148–1162

ADAPTIVE PROCEDURES IN MULTIPLE DECISION PROBLEMS AND HYPOTHESIS TESTING[1]

BY ANDREW L. RUKHIN

Purdue University

Necessary and sufficient conditions for the existence of adaptive procedures for identification of one of several probability distributions or for testing a simple hypothesis against a simple alternative are obtained. By definition, adaptive procedures are required to exhibit the same asymptotic behavior for several parametric families as do the optimal (minimax) estimators for each of these families. The proofs are based on a multivariate version of Chernoff's theorem, providing asymptotic formulas for probabilities of large deviations for sums of i.i.d. random vectors. Some examples of adaptive procedures are considered, and the non-existence of such rules is established in certain situations.

1. Introduction. We start with the simple multiple decision problem where both the action space and the parameter space coincide and are finite, say, $\Theta = \{0, 1, \cdots, m\}$, $m \geq 1$. Thus a family of (different) probability distributions $\mathscr{P} = \{P_0, \cdots, P_m\}$ over a space $\mathscr{X}$ is given, and statistical inference about the finite-valued parameter is desired on the basis of a random sample $\mathbf{x} = (x_1, \cdots, x_n)$ which is obtained as realizations of a random variable X having one of these distributions.

If $\delta(\mathbf{x})$ is an estimator of this parameter, then the probability of incorrect decision $P_\theta(\delta(\mathbf{x}) \neq \theta)$ is the most important characteristic of the procedure δ. The asymptotic behavior of this probability for the minimax estimator δ^* has been studied by many scholars (see Bahadur, 1971, Krafft and Puri, 1974, Ghosh and Subramanyam, 1975). The main result here has the form

$$\lim_{n\to\infty}\max_\theta P_\theta^{1/n}(\delta^*(\mathbf{x}) \neq \theta) = \max_{\eta\neq\theta}\inf_{t\geq 0} E^{P_\theta} p_\eta^t(X) p_\theta^{-t}(X) = \max_{\eta\neq\theta}\inf_{t\geq 0}\int_{\mathscr{X}} p_\eta^t(x) p_\theta^{1-t}(x)\, d\mu(x) = \rho(\mathscr{P}), \tag{1.1}$$

where p_θ is the probability density of the distribution P_θ, $\theta \in \Theta$, with respect to a σ-finite measure μ on $\mathscr{X}$. Notice that $\rho(\mathscr{P}) < 1$, since all elements of $\mathscr{P}$ are distinct. It follows from (1.1) that for any procedure δ

$$\liminf_{n\to\infty}\max_\theta P_\theta^{1/n}(\delta(\mathbf{x}) \neq \theta) \geq \rho(\mathscr{P}). \tag{1.2}$$

A parallel result holds for hypothesis testing problems. Let us consider the case of testing a simple hypothesis P against the simple alternative Q. It is known (cf. Chernoff, 1956, Bahadur, 1971) that if $\alpha_n = \alpha_n(\phi^*)$ denotes the minimal size of the most powerful test $\phi^*(\mathbf{x})$ of this hypothesis which has a fixed power β, $0 < \beta < 1$, then

$$\alpha_n^{1/n} = \{E^P\phi^*(\mathbf{x})\}^{1/n} \to \exp\{-K(Q, P)\}. \tag{1.3}$$

Here $K(Q, P) = E^Q \log\{(dQ/dP)(X)\}$ is the Kullback-Leibler information number (Kullback, 1959). Moreover, for any test ϕ of the same or larger power

$$\liminf_{n\to\infty}\alpha_n^{1/n}(\phi) \geq \exp\{-K(Q, P)\}. \tag{1.4}$$

Received December 1980; revised May 1982.

[1] Research supported by NSF Grant MCS 78-02300.

AMS 1970 *subject classifications.* Primary 62F35; secondary 62F12, 62F05, 60F10.

Key words and phrases. Multiple decision problem with finite parameter space, testing of simple hypothesis, probability of incorrect decision, adaptive procedures, multivariate Chernoff's theorem.

Formula (1.3) has been generalized to the case of testing a hypothesis consisting of a finite number of distributions by Plachky and Steinebach (1977).

The proofs of both mentioned results are closely related to Chernoff's theorem (Chernoff, 1952, Bahadur, 1971) and can be obtained with its help.

Formulas (1.3) and (1.4) as well as (1.2) and (1.1) lead to the following question. What are suitable conditions on two pairs of distributions (P_1, Q_1) and (P_2, Q_2) such that there exists an "adaptive" test ϕ possessing the following properties? Its power as a test of P_1 against Q_1 and as a test of P_2 versus Q_2 is equal to a fixed number β, $0 < \beta < 1$ and its level behaves asymptotically as that of the most powerful test for both testing problems, i.e., for $i = 1, 2$

$$E^{Q_i}\phi(\mathbf{x}) = \beta$$

and

$$\{E^{P_i}\phi(\mathbf{x})\}^{1/n} \to \exp\{-K(Q_i, P_i)\}.$$

In this paper we obtain a necessary condition and a sufficient condition for the existence of such an adaptive test. These conditions can be interpreted as an expression for the degree of closeness between (P_1, Q_1) and (P_2, Q_2) in terms of an information type divergence.

In the multiple decision problem we will be interested in conditions on two (or several) parametric families $\mathscr{P}_1 = \{P_\theta^{(1)}, \theta \in \Theta\}$ and $\mathscr{P}_2 = \{P_\theta^{(2)}, \theta \in \Theta\}$ under which there exists an adaptive estimator δ, i.e. such that for $i = 1, 2$

$$\max_\theta \{P_\theta^{(i)}(\delta(x) \neq \theta)\}^{1/n} \to \rho(\mathscr{P}_i).$$

In other terms an adaptive estimator δ serves both families $\mathscr{P}_1$ and $\mathscr{P}_2$ in an asymptotically optimal (minimax) way. The existence of such an estimator in the case when θ is a real location parameter, and asymptotic optimality is defined by means of asymptotic variance, has been established in various settings (see Beran, 1974, Sacks, 1975, Stone 1975).

We remark that if in the definition of adaptive estimator one replaces the probability of incorrect decision by an arbitrary risk, $R(\theta, \delta) = E_\theta W(\theta, \delta(\mathbf{x}))$, where $W(\theta, \theta) = 0$ and $W(\theta, \eta) \neq 0$, $\theta \neq \eta$, all results of this article remain valid. Our results also hold for the Bayesian setting of the problem of adaptive estimation. Indeed if π_θ are positive prior probabilities then

$$\liminf_{n\to\infty} \{\textstyle\sum_\theta \pi_\theta P_\theta(\delta(\mathbf{x}) \neq \theta)\}^{1/n} = \liminf_{n\to\infty} \max_\theta P_\theta^{1/n}(\delta(\mathbf{x}) \neq \theta).$$

Therefore instead of minimax estimators one can speak about Bayes estimators and replace in the definition of an adaptive procedure the maximum of the risk by the Bayes risk.

In Section 3, we give a necessary condition and a sufficient condition for the existence of adaptive procedures in multiple decision problems and hypothesis testing problems. These conditions are obtained by studying most powerful tests and minimax estimators for the model described by a mixture of experiments defined by families $\mathscr{P}_1$ and $\mathscr{P}_2$. This study is performed in Section 2. The basic mathematical tool needed is a multivariate version of Chernoff's theorem, providing an asymptotic formula for probabilities of large deviations of sums of i.i.d. random vectors. In Section 4, we illustrate the necessary condition and the sufficient condition for the existence of adaptive procedures by several examples. In typical cases, adaptive estimators do not exist if the distributions $P_\theta^{(i)}$ and $P_\eta^{(k)}$, $i \neq k$, $\theta \neq \eta$ are more "similar" than the distributions $P_\theta^{(i)}$ and $P_\eta^{(i)}$.

2. The asymptotic behavior of optimal procedures for mixtures. In this section we will be interested in the asymptotic behavior of statistical procedures based on a likelihood function of the form $\sum_{k=1}^{\ell} w_k \prod_1^n f_k(x_j)$ where f_k, $k = 1, \cdots, \ell$ are probability densities, and w_k are positive weights. We start with the following key result.

LEMMA. *Let c_n, $n = 1, 2, \cdots$ be a sequence of positive numbers such that $n^{-1}\log c_n$ converges to a finite limit L. Then if f_i, g_i, $i = 1, \cdots, \ell$, are strictly positive probability densities, $w_1, \cdots, w_\ell \geq 0$, $w_1 + \cdots + w_\ell = 1$ and for all positive probabilities $q_1, \cdots, q_\ell$ and $i = 1, \cdots, \ell$,*

$$\Pr[\textstyle\sum_1^\ell q_k \log\{f_i(X)/g_k(X)\} > L] > 0, \tag{2.1}$$

then

$$\begin{aligned}&\Pr{}^{1/n}\{\textstyle\sum_1^\ell w_k \prod_1^n f_k(x_j) \geq c_n \sum_1^\ell w_k \prod_1^n g_k(x_j)\}\\ &\quad\to \max_{1\leq k\leq \ell} \inf_{s_1,\cdots,s_\ell\geq 0} e^{-(s_1+\cdots+s_\ell)L} E f_k^{s_1+\cdots+s_\ell}(X) \textstyle\prod_1^\ell g_i^{-s_i}(X).\end{aligned} \tag{2.2}$$

PROOF. We prove the Lemma for $\ell = 2$. Notice that

$$\begin{aligned}&\Pr\{2\max[w_1 \textstyle\prod_1^n f_1(x_j), w_2 \prod_1^n f_2(x_j)] \geq c_n \max[w_1 \prod_1^n g_1(x_j), w_2 \prod_1^n g_2(x_j)]\}\\ &\quad\geq \Pr\{w_1 \textstyle\prod_1^n f_1(x_j) + w_2 \prod_1^n f_2(x_j) \geq c_n[w_1 \prod_1^n g_1(x_j) + w_2 \prod_1^n g_2(x_j)]\}\\ &\quad\geq \Pr\{\max[w_1 \textstyle\prod_1^n f_1(x_j), w_2 \prod_1^n f_2(x_j)] \geq 2c_n \max[w_1 \prod_1^n g_1(x_j), w_2 \prod_1^n g_2(x_j)]\}.\end{aligned} \tag{2.3}$$

One has

$$\begin{aligned}&\max[\Pr\{\textstyle\prod_1^n f_1(x_j) \geq 2c_n \prod_1^n g_1(x_j), \prod_1^n f_1(x_j) \geq 2c_n w_1^{-1} w_2 \prod_1^n g_2(x_j)\},\\ &\quad \Pr\{\textstyle\prod_1^n f_2(x_j) \geq 2c_n w_1 w_2^{-1} \prod_1^n g_1(x_j), \prod_1^n f_2(x_j) \geq 2c_n \prod_1^n g_2(x_j)\}]\\ &\leq \Pr\{\max[w_1 \textstyle\prod_1^n f_1(x_j), w_2 \prod_1^n f_2(x_j)] \geq 2c_n \max[w_1 \prod_1^n g_1(x_j), w_2 \prod_1^n g_2(x_j)]\}\\ &\leq \Pr\{\textstyle\prod_1^n f_1(x_j) \geq 2c_n \prod_1^n g_1(x_j), \prod_1^n f_1(x_j) \geq 2c_n w_1^{-1} w_2 \prod_1^n g_2(x_j)\}\\ &+ \Pr\{\textstyle\prod_1^n f_2(x_j) \geq 2c_n w_1 w_2^{-1} \prod_1^n g_1(x_j), \prod_1^n f_2(x_j) \geq 2c_n \prod_1^n g_2(x_j)\},\end{aligned}$$

so that

$$\begin{aligned}&\Pr{}^{1/n}\{\max[w_1 \textstyle\prod_1^n f_1(x_j), w_2 \prod_1^n f_2(x_j)] \geq 2c_n \max[w_1 \prod_1^n g_1(x_j), w_2 \prod_1^n g_2(x_j)]\}\\ &\quad\sim \max[\Pr{}^{1/n}\{\textstyle\prod_1^n f_1(x_j) \geq 2c_n \prod_1^n g_1(x_j) \prod_1^n f_1(x_j) \geq 2c_n w_1^{-1} w_2 \prod_1^n g_2(x_j)\},\\ &\qquad \Pr{}^{1/n}\{\textstyle\prod_1^n f_2(x_j) \geq 2c_n w_1 w_2^{-1} \prod_1^n g_1(x_j), \prod_1^n f_2(x_j) \geq 2c_n \prod_1^n g_2(x_j)\}].\end{aligned}$$

To find the asymptotic behavior of these latter probabilities we use a two-dimensional version of Chernoff's theorem; see Groeneboom, Oosterhoff and Ruymgaart (1979), Bahadur and Zabell (1979) or Bartfai (1978). According to this theorem, if (Y_1, Z_1), (Y_2, Z_2), $\cdots$ is a sequence of i.i.d. random vectors in $\mathbb{R}^2$, then

$$\Pr{}^{1/n}(n^{-1} \textstyle\sum_1^n Y_j \geq y + \alpha_n, n^{-1} \sum_1^n Z_j \geq z + \beta_n) \to \inf_{s,t\geq 0} e^{-sy-tz} E e^{sY_1+tZ_1}. \tag{2.4}$$

Here $\alpha_n \to 0$, $\beta_n \to 0$, y and z are real numbers such that

$$\Pr(sY_1 + tZ_1 \geq sy + tz) > 0 \tag{2.5}$$

for all nonnegative s, t, $(s, t) \neq (0, 0)$. Condition (2.5) guarantees the continuity in y and z of the right-hand side of (2.4). It implies that (y, z) is an inner point of the set A_p from the condition of Theorem 5.1 of Groeneboom et al (1979).

We apply this theorem with $y = z = \lim n^{-1}\log c_n = L$ and $Y_j = \log\{f_1(x_j)/g_1(x_j)\}$, $Z_j = \log\{f_1(x_j)/g_2(x_j)\}$, or for $Y_j = \log\{f_2(x_j)/g_1(x_j)\}$, $Z_j = \log\{f_2(x_j)/g_2(x_j)\}$. In both of these cases, Condition (2.5) is met because of Assumption (2.1).

Thus

$$\begin{aligned}&\Pr{}^{1/n}\{\textstyle\prod_1^n f_1(x_j) \geq 2c_n \prod_1^n g_1(x_j), \prod_1^n f_1(x_j) \geq 2c_n w_1^{-1} w_2 \prod_1^n g_2(x_j)\}\\ &\qquad\to \inf_{s,t\geq 0} e^{-(s+t)L} E f_1^{s+t}(X) g_1^{-s}(X) g_2^{-t}(X),\end{aligned}$$

and

$$\Pr^{1/n}\{\prod_1^n f_2(x_j) \geq 2c_n w_1 w_2^{-1} \prod_1^n g_1(x_j), \prod_1^n f_2(x_j) \geq 2c_n \prod_1^n g_2(x_j)\}$$
$$\to \inf_{s,t\geq 0} e^{-(s+t)L} E f_2^{s+t}(X) g_1^{-s}(X) g_2^{-t}(X).$$

Therefore

$$\Pr^{1/n}\{\max[w_1 \prod_1^n f_1(x_j), w_2 \prod_1^n f_2(x_j)] \geq 2c_n \max[w_1 \prod_1^n g_1(x_j), w_2 \prod_1^n g_2(x_j)]\}$$
$$\to \max_{i=1,2} \inf_{s,t\geq 0} e^{-(s+t)L} E f_i^{s+t}(X) g_1^{-s}(X) g_2^{-t}(X),$$

and the left-hand side of (2.3) has the same limiting value. Thus the Lemma is proven.

REMARK 1. If, say, $\ell = 2$ and

$$\limsup \frac{\Pr\{\prod_1^n f_2(x_j) \geq 2c_n w_1 w_2^{-1} \prod_1^n g_1(x_j), \prod_1^n f_2(x_j) \geq 2c_n \prod_1^n g_2(x_j)\}}{\Pr\{\prod_1^n f_1(x_j) \geq 2c_n \prod_1^n g_1(x_j), \prod_1^n f_1(x_j) \geq 2c_n w_1^{-1} w_2 \prod_1^n g_2(x_j)\}} < \infty,$$

then condition (2.1) for $i = 1$ can be omitted. Indeed in this case

$$\lim_{n\to\infty} \Pr^{1/n}\{\max[w_1 \prod_1^n f_1(x_j), w_2 \prod_1^n f_2(x_j)] \geq 2c_n \max[w_1 \prod_1^n g_1(x_j), w_2 \prod_1^n g_2(x_j)]\}$$
$$= \lim_{n\to\infty} \Pr^{1/n}\{\prod_1^n f_1(x_j) \geq 2c_n \prod_1^n g_1(x_j), \prod_1^n f_1(x_j) \geq 2c_n w_1^{-1} w_2 \prod_1^n g_2(x_j)\}.$$

If for some $a, b \geq 0$, $a + b = 1$

$$P\left[a \log\left\{\frac{f_1(X)}{g_1(X)}\right\} + b \log\left\{\frac{f_2(X)}{g_2(X)}\right\} \leq L\right] = 1,$$

then

$$0 \leq \inf_{s,t\geq 0} e^{-(s+t)L} E f_2^{s+t}(X) g_1^{-s}(X) g_2^{-t}(X) \leq \inf_{s\geq 0} e^{-sL} E f_2^{s}(X) g_1^{-as}(X) g_2^{-bs}(X) = 0.$$

Therefore

$$\max_{k=1,2} \inf_{s,t\geq 0} e^{-(s+t)L} E f_k^{s+t}(X) g_1^{-s}(X) g_2^{-t}(X) = \inf_{s,t\geq 0} e^{-(s+t)L} E f_1^{s+t}(X) g_1^{-s}(X) g_2^{-t}(X),$$

and the assertion of the Lemma holds true.

REMARK 2. Under the conditions of the Lemma

$$\begin{aligned}(2.6)\quad &\Pr^{1/n}\{w_1 \prod_1^n f_1(x_j) + w_2 \prod_1^n f_2(x_j) \geq c_n[w_1 \prod_1^n g_1(x_j) + w_2 \prod_1^n g_2(x_j)]\}\\ &\sim \Pr^{1/n}\{w_1 \prod_1^n f_1(x_j) + w_2 \prod_1^n f_2(x_j) > c_n[w_1 \prod_1^n g_1(x_j) + w_2 \prod_1^n g_2(x_j)]\}.\end{aligned}$$

We recall now the following classical result concerning the estimation with zero-one loss of a finite parameter θ in a family of mutually absolutely continuous distributions Q_θ (see Wald, 1950, pages 125–128). There exists a least favorable distribution with positive probabilities such that the corresponding Bayes procedure δ^* is minimax and $Q_\theta(\delta^*(\mathbf{x}) \neq \theta)$ does not depend on θ.

Let $\mathscr{P}_k = \{P_\theta^{(k)}, \theta \in \Theta\}$, $P_\theta^{(k)} \neq P_\eta^{(k)}$ for $\theta \neq \eta$, $k = 1, \cdots, \ell$, be ℓ parametric families given on X with densities $p_k(\cdot, \theta)$. Also let $w_1, \cdots, w_\ell$ be positive weights $w_1 + \cdots + w_\ell = 1$ and assume all measures $P_\theta^{(k)}$ to be equivalent. The next result gives an asymptotic formula for the probability of incorrect decision for a minimax procedure δ^* based on an observation $\mathbf{x}$ from a distribution Q_θ which has the density of the form $\sum_k w_k \prod_1^n p_k(x_j, \theta)$.

THEOREM 2.1. *If all densities $p_k(\cdot, \theta)$, $k = 1, \cdots, \ell$ are positive and δ^* is a minimax estimator, then*

$$\begin{aligned}(2.7)\quad &\lim_{n\to\infty} \max_\theta [\textstyle\sum_k w_k P_\theta^{(k)}(\delta^*(\mathbf{x}) \neq \theta)]^{1/n} = \lim_{n\to\infty} \max_{\theta,k} [P_\theta^{(k)}(\delta^*(\mathbf{x}) \neq \theta)]^{1/n}\\ &= \max_{1\leq i,k\leq \ell} \max_{\theta\neq\eta} \inf_{s_1,\dots,s_\ell\geq 0} E_\theta^{(k)}[p_i(X,\eta)]^{\sum_1^\ell s_r} \textstyle\prod_{r=1}^\ell p_r^{-s_r}(X,\theta),\end{aligned}$$

where $E_\theta^{(k)}$ stands for expected value with respect to $P_\theta^{(k)}$.

PROOF. We prove Theorem 2.1 only in the case $\ell = 2$. The general case is quite similar.

Let $\hat{\delta}$ be a maximum likelihood estimator. We shall see that if $\hat{\delta}$ is not uniquely defined, i.e. when ties occur, then the asymptotic behavior of this estimator does not depend on the way in which these ties are broken. Thus for $k = 1, 2$

$$\begin{aligned} P_\theta^{(k)}(\hat{\delta}(\mathbf{x}) \neq \theta) &= P_\theta^{(k)}(w_1 \textstyle\prod_1^n p_1(x_j, \eta) + w_2 \prod_1^n p_2(x_j, \eta) > w_1 \prod_1^n p_1(x_j, \theta) + w_2 \prod_1^n p_2(x_j, \theta) \\ &\qquad \text{for some} \quad \eta \neq \theta) \\ &\leq \textstyle\sum_{\eta:\eta\neq\theta} P_\theta^{(k)}\{w_1 \prod_1^n p_1(x_j, \eta) + w_2 \prod_1^n p_2(x_j, \eta) > w_1 \prod_1^n p_1(x_j, \theta) \\ &\qquad + w_2 \textstyle\prod_1^n p_2(x_j, \theta)\} \\ &\leq (m-1) \max_{\eta:\eta\neq\theta} P_\theta^{(k)}\{w_1 \textstyle\prod_1^n p_1(x_j, \eta) + w_2 \prod_1^n p_2(x_j, \eta) \\ &\qquad > w_1 \textstyle\prod_1^n p_1(x_j, \theta) + w_2 \prod_1^n p_2(x_j, \theta)\}. \end{aligned}$$

Also

$$\begin{aligned} P_\theta^{(k)}(\hat{\delta}(\mathbf{x}) \neq \theta) \geq \max_{\eta:\eta\neq\theta} P_\theta^{(k)}\{w_1 \textstyle\prod_1^n p_1(x_j, \eta) + w_2 \prod_1^n p_2(x_j, \eta) \\ > w_1 \textstyle\prod_1^n p_1(x_j, \theta) + w_2 \prod_1^n p_2(x_j, \theta)\}. \end{aligned}$$

Thus because of our Lemma,

$$\begin{aligned} \lim_{n\to\infty}[P_\theta^{(k)}(\hat{\delta}(\mathbf{x}) \neq \theta)]^{1/n} &= \rho_k(\theta) \\ &= \max_{\eta:\eta\neq\theta}\max_{i=1,2}\inf_{s,t\geq 0} E_\theta^{(k)} p_i^{s+t}(X, \eta) p_1^{-s}(X, \theta) p_2^{-t}(X, \theta). \end{aligned} \tag{2.8}$$

Notice that conditions (2.1) and (2.2) are satisfied since because of the equivalence of our distributions

$$P_\theta^{(k)}\{a \log[p_i(X, \eta)/p_1(X, \theta)] + b \log[p_i(X, \eta)/p_2(X, \theta)] > 0\} > 0$$

if and only if

$$P_\eta^{(i)}\{a \log[p_i(X, \eta)/p_1(X, \theta)] + b \log[p_i(X, \eta)/p_2(X, \theta)] > 0\} > 0.$$

The latter inequality must hold since for all $a, b \geq 0$, $a + b > 0$,

$$E_\eta^{(i)}\{a \log[p_i(X, \eta)/p_1(X, \theta)] + b \log[p_i(X, \eta)/p_2(X, \theta)]\} > 0.$$

It follows from (2.8) that

$$\begin{aligned} \lim_{n\to\infty}[w_1 P_\theta^{(1)}(\hat{\delta}(\mathbf{x}) \neq \theta) + w_2 P_\theta^{(2)}(\hat{\delta}(\mathbf{x}) \neq \theta)]^{1/n} &= \max_{k=1,2}\lim_{n\to\infty}[P_\theta^{(k)}(\hat{\delta}(\mathbf{x}) \neq \theta)]^{1/n} \\ &= \max(\rho_1(\theta), \rho_2(\theta)). \end{aligned}$$

Thus if δ^* is a minimax procedure then

$$\limsup \max_\theta[w_1 P_\theta^{(1)}(\delta^*(\mathbf{x}) \neq \theta) + w_2 P_\theta^{(2)}(\delta^*(\mathbf{x}) \neq \theta)]^{1/n} \leq \max_{k=1,2}\max_\theta \rho_k(\theta). \tag{2.9}$$

To complete the proof assume for sake of concreteness that $\max_\theta \rho_1(\theta) \geq \max_\theta \rho_2(\theta)$ and

$$\max_\theta \rho_1(\theta) = \max_{i=1,2}\inf_{s,t\geq 0} E_\xi^{(1)} p_i^{s+t}(X, \zeta) p_1^{-s}(X, \xi) p_2^{-t}(X, \xi).$$

In this case we define the prior distribution λ to be concentrated on $\{\xi, \zeta\}$, $\lambda(\xi) = \lambda(\zeta) = ½$ and let δ_B be the corresponding Bayes estimator. Then for any procedure δ

$$\begin{aligned} &\max_\theta[w_1 P_\theta^{(1)}(\delta(\mathbf{x}) \neq \theta) + w_2 P_\theta^{(2)}(\delta(\mathbf{x}) \neq \theta)] \\ &\quad \geq ½\{w_1[P_\xi^{(1)}(\delta(\mathbf{x}) \neq \xi) + P_\zeta^{(1)}(\delta(\mathbf{x}) \neq \zeta)] + w_2[P_\xi^{(2)}(\delta(\mathbf{x}) \neq \xi) + P_\zeta^{(2)}(\delta(\mathbf{x}) \neq \zeta)]\} \\ &\quad \geq ½\{w_1[P_\xi^{(1)}(\delta_B(\mathbf{x}) \neq \xi) + P_\zeta^{(1)}(\delta_B(\mathbf{x}) \neq \zeta)] + w_2[P_\xi^{(2)}(\delta_B(\mathbf{x}) \neq \xi) + P_\zeta^{(2)}(\delta_B(\mathbf{x}) \neq \zeta)]\}. \end{aligned}$$

Using the Lemma again we see that

$[P_\xi^{(k)}(\delta_B(\mathbf{x}) \neq \xi)]^{1/n} = [P_\xi^{(k)}(\delta_B(\mathbf{x}) = \zeta)]^{1/n}$

$= [P_\xi^{(k)}\{w_1 \prod_1^n p_1(x_j, \zeta) + w_2 \prod_1^n p_2(x_j, \zeta) > w_1 \prod_1^n p_1(x_j, \xi) + w_2 \prod_1^n p_2(x_j, \xi)\}]^{1/n}$

$\to \max_{\iota=1,2}\inf_{s,t\geq 0} E_\xi^{(k)} p_\iota^{s+t}(X, \zeta) p_1^{-s}(X, \xi) p_2^{-t}(X, \xi).$

Also

$$[P_\zeta^{(k)}(\delta_B(\mathbf{x}) \neq \zeta)]^{1/n} \to \max_{\iota=1,2}\inf_{s,t\geq 0} E_\zeta^k p^{s+t}(X, \xi) p_1^{-s}(X, \zeta) p_2^{-t}(X, \zeta)$$

and for any δ

$\lim_{n\to\infty}\inf \max_\theta[w_1 P_\theta^{(1)}(\delta(\mathbf{x}) \neq \theta) + w_2 P_\theta^{(2)}(\delta(\mathbf{x}) \neq \theta)]^{1/n}$

$\geq \max_{k=1,2}\max\{\lim_{n\to\infty}[P_\xi^{(k)}(\delta_B(\mathbf{x}) \neq \xi)]^{1/n}, \lim_{n\to\infty}[P_\zeta^{(k)}(\delta_B(\mathbf{x}) \neq \zeta)]^{1/n}\}$

$= \max\{\lim_{n\to\infty}[P_\xi^{(1)}(\delta_B(\mathbf{x}) \neq \xi)]^{1/n}, \lim_{n\to\infty}[P_\zeta^{(1)}(\delta_B(\mathbf{x}) \neq \zeta)]^{1/n}\}$

$= \max_{k=1,2}\max_\theta \rho_k(\theta).$

This inequality combined with (2.9) proves the Theorem.

COROLLARY 2.1. *Under the assumptions of Theorem* 2.1 *for* $k = 1, \cdots, \ell$

$$\rho(\mathscr{P}_k) \leq \max_\iota \max_{\eta\neq\theta} \inf_{s_1,\ldots,s_\ell\geq 0} E_\theta^{(k)}[p_\iota(X, \eta)]^{s_1+\cdots+s_\ell} \prod_{r=1}^{\ell} p_r^{-s_r}(X, \theta). \tag{2.10}$$

Now we assume that $m = 2$ and the hypothesis testing problem is considered. Let $\phi^*(\mathbf{x})$ be a most powerful test of the simple hypothesis $w_1 \prod_1^n p_1(x_j) + w_2 \prod_1^n p_2(x_j)$ against the simple alternative $w_1 \prod_1^n q_1(x_j) + w_2 \prod_1^n q_2(x_j)$.

THEOREM 2.2. *Assume that ϕ^* has a fixed power β (independent of the sample size n) and let α_n^* denote its level. Then*

$$\alpha_n^{*\,1/n} \to \max_{1\leq\iota,k\leq 2}\inf_{s,t\geq 0} e^{-(s+t)K} E^{P_\iota} q_k^{s+t}(X) p_1^{-s}(X) p_2^{-t}(X), \tag{2.11}$$

where $K = \max(K_1, K_2)$,

$$K_\iota = \min\left[E^{Q_\iota}\log\left\{\frac{q_i}{p_2}(X)\right\}, E^{Q_\iota}\log\left\{\frac{q_\iota}{p_1}(X)\right\}\right], \qquad i = 1, 2, \tag{2.12}$$

and it is assumed that $K_1 > K_2$ implies $1 > w_1 > \beta$, and $K_1 < K_2$ implies $1 > w_1 = 1 - w_2 > \beta$.

PROOF. It is well known that a most powerful test ϕ^* of our hypothesis is given by formula

$$\phi^*(\mathbf{x}) = \begin{cases} 1 & w_1 \prod_1^n q_1(x_j) + w_2 \prod_1^n q_2(x_j) > c_n[w_1 \prod_1^n p_1(x_j) + w_2 \prod_1^n p_2(x_j)], \\ \gamma_n & w_1 \prod_1^n q_1(x_j) + w_2 \prod_1^n q_2(x_j) = c_n[w_1 \prod_1^n p_1(x_j) + w_2 \prod_1^n p_2(x_j)], \\ 0 & w_1 \prod_1^n q_1(x_j) + w_2 \prod_1^n q_2(x_j) < c_n[w_1 \prod_1^n p_1(x_j) + w_2 \prod_1^n p_2(x_j)], \end{cases}$$

with some cor ants $c_n > 0$, $0 \leq \gamma_n \leq 1$. Thus

$$w_1 E^{Q_1}\phi^*(\mathbf{x}) + w_2 E^{Q_2}\phi^*(\mathbf{x}) = \beta$$

and

$$w_1 E^{P_1}\phi^*(\mathbf{x}) + w_2 E^{P_2}\phi^*(\mathbf{x}) = \alpha_n^*.$$

It follows that

$\sum_{\iota=1,2} w_\iota Q_\iota\{w_1 \prod_1^n q_1(x_j) + w_2 \prod_1^n q_2(x_j) > c_n[w_1 \prod_1^n p_1(x_j) + w_2 \prod_1^n p_2(x_j)]\} \leq \beta$

$\leq \sum_{\iota=1,2} w_\iota Q_\iota\{w_1 \prod_1^n q_1(x_j) + w_2 \prod_1^n q_2(x_j) \geq c_n[w_1 \prod_1^n p_1(x_j) + w_2 \prod_1^n p_2(x_j)]\}.$

As in the proof of the Lemma one has for $i = 1, 2$

$$Q_i(2\max[w_1 \prod_1^n q_1(x_j), w_2 \prod_1^n q_2(x_j)] \geq c_n \max[w_1 \prod_1^n p_1(x_j), w_2 \prod_1^n p_2(x_j)])$$
$$\geq Q_i([w_1 \prod_1^n q_1(x_j) + w_2 \prod_1^n q_2(x_j)] \geq c_n[w_1 \prod_1^n p_1(x_j) + w_2 \prod_1^n p_2(x_j)])$$
$$\geq Q_i(\max[w_1 \prod_1^n q_1(x_j), w_2 \prod_1^n q_2(x_j)] \geq 2c_n \max[w_1 \prod_1^n p_1(x_j), w_2 \prod_1^n p_2(x_j)]),$$

so that

$$\sum_{i=1,2} w_i[Q_i(\prod_1^n q_1(x_j) \geq 2c_n \prod_1^n p_1(x_j), \prod_1^n q_1(x_j) \geq 2c_n w_1^{-1} w_2 \prod_1^n p_2(x_j))$$
$$+ Q_i(\prod_1^n q_2(x_j) \geq 2c_n w_1 w_2^{-1} \prod_1^n p_1(x_j), \prod_1^n q_2(x_j) \geq 2c_n \prod_1^n p_2(x_j))] \leq \beta \tag{2.13}$$
$$\leq \sum_{i=1,2} w_i[Q_i(2\prod_1^n q_1(x_j) \geq c_n \prod_1^n p_1(x_j), 2\prod_1^n q_1(x_j) \geq c_n w_1^{-1} w_2 \prod_1^n p_2(x_j))$$
$$+ Q_i(2\prod_1^n q_2(x_j) \geq c_n w_1 w_2^{-1} \prod_1^n p_1(x_j), 2\prod_1^n q_2(x_j) \geq c_n \prod_1^n p_2(x_j))].$$

Let $Y_j = \log[q_1(x_j)/p_1(x_j)]$, $U_j = \log[q_1(x_j)/p_2(x_j)]$, $V_j = \log[q_2(x_j)/p_1(x_j)]$, $W_j = \log[q_2(x_j)/p_2(x_j)]$, $y_j = j^{-1}\log(c_j/2)$, $u_j = j^{-1}\log(w_1^{-1}w_2)$, $v_j = y_j + j^{-1}\log(w_1 w_2^{-1})$, $j = 1, 2, \cdots$.

Since $n^{-1}\sum_1^n Y_j$ converges in Q_i-probability to $E^{Q_i}Y_1$,

$$\lim_{n\to\infty}\inf Q_i(\sum_1^n Y_j \geq ny_n, \sum_1^n U_j \geq nu_n) = 0,$$

if $y = \lim_{n\to\infty}\sup y_n = \lim_{n\to\infty}\sup u_n > \min(E^{Q_i}Y_1, E^{Q_i}U_1)$, and

$$\lim_{n\to\infty}\sup Q_i(\sum_1^n Y_j \geq ny_n, \sum_1^n U_j \geq nu_n) = 1,$$

if

$$u = \lim_{n\to\infty}\inf y_n = \lim_{n\to\infty}\inf u_n < \min(E^{Q_i}Y_1, E^{Q_i}U_1).$$

Since at least one of the probabilities in the right hand-side of (2.13) does not tend to zero we conclude that

$$y \leq \max_{i=1,2}[\min(E^{Q_i}Y_1, E^{Q_i}U_1), \min(E^{Q_i}V_1, E^{Q_i}W_1)] = \max(K_1', K_2'),$$

where

$$K_i' = \max[\min\{E^{Q_i}\log(q_1/p_1), E^{Q_i}\log(q_1/p_2)\}, \quad \min\{E^{Q_i}\log(q_2/p_1)E^{Q_i}\log(q_2/p_2)\}].$$

We prove that $K_i' = K_i$, where K_i is as defined in (2.12). Let us show for instance that

$$\min\{E^{Q_1}\log(q_1/p_1), E^{Q_1}\log(q_1/p_2)\} \geq \min\{E^{Q_1}\log(q_2/p_1), E^{Q_1}\log(q_2/p_2)\}. \tag{2.14}$$

If $E^{Q_1}\log(p_2/p_1) \leq 0$, then $E^{Q_1}\log(q_2/p_1) \leq E^{Q_1}\log(q_2/p_2)$ and $E^{Q_1}\log(q_1/p_1) \leq E^{Q_1}\log(q_1/p_2)$. But $E^{Q_1}\log(q_2/p_1) \leq E^{Q_1}\log(q_1/p_1)$ so that in this situation (2.14) is true.

The case $E^{Q_1}\log(p_2/p_1) > 0$ can be treated analogously. Moreover

$$\min\{E^{Q_2}\log(q_2/p_1), E^{Q_2}\log(q_2/p_2)\} \geq \min\{E^{Q_2}\log(q_1/p_1), E^{Q_2}\log(q_1/p_2)\},$$

so that $K_i' = K_i$, $i = 1, 2$.

It follows from (2.13) that

$$\sum_{i=1,2} w_i \lim_{n\to\infty}\sup[Q_i(\sum_1^n Y_j \geq nu, \sum_1^n U_j \geq nu) + Q_i(\sum_1^n V_j \geq nu, \sum_1^n W_j \geq nu)] \leq \beta.$$

If $u < K$ and, say, $K_1 > K_2$, then

$$\lim_{n\to\infty}\sup Q_1(\sum_1^n Y_j \geq nu, \sum_1^n U_j \geq nu) = 1,$$

which is impossible because of (2.13).

Therefore $u \leq K$ and

$$\lim_{n\to\infty} n^{-1} \log c_n = \lim_{n\to\infty} y_n = \lim_{n\to\infty} u_n = \lim_{n\to\infty} v_n = K.$$

Now we study the asymptotic behavior of the level α_n^* of test ϕ^*. Observe that

$$\sum_{i=1,2} w_i P_1(w_1 \prod_1^n q_1(x_j) + w_2 \prod_1^n q_2(x_j) > c_n[w_1 \prod_1^n p_1(x_j) + w_2 \prod_1^n p_2(x_j)]) \le \alpha_n^*$$
$$\le \sum_{i=1,2} w_i P_i(w_1 \prod_1^n q_1(x_j) + w_2 \prod_1^n q_2(x_j) \ge c_n[w_1 \prod_1^n p_1(x_j) + w_2 \prod_1^n p_2(x_j)]).$$

Since all measures P_i, Q_i $i = 1,2$ are equivalent and for a, $b \ge 0$, $a + b = 1$

$$aE^{Q_i} \log(q_i/p_1) + bE^{Q_i}\log(q_i/p_2) > K_i,$$

one deduces

$$P_i(a \log(q_i/p_1) + b \log(q_i/p_2) > K_i) > 0.$$

If for $k \ne i$ and all a, $b \ge 0$, $a + b = 1$,

$$P_i(a \log(q_k/p_1) + b \log(q_k/p_2) > K_i) > 0,$$

then we can use the Lemma to derive the following limiting relation

$$p_i^{1/n}\{w_1 \prod_1^n q_1(x_j) + w_2 \prod_1^n q_2(x_j) \ge c_n[w_1 \prod_1^n p_1(x_j) + w_2 \prod_1^n p_2(x_j)]\}$$
$$\to \max_{k=1,2} \inf_{s,t\ge 0} e^{-(s+t)K} E^{P_i} q_k^{s+t}(X) p_1^{-s}(X) p_2^{-t}(X).$$

If, say, $K_1 > K_2$ and for $i = 1, 2$, $a + b = 1$

$$P_i(a \log(q_2/p_1) + b \log(q_2/p_2) > K_1) = 0,$$

then for all sufficiently large n

$$P_i\{2 \prod_1^n q_2(x_j) \ge c_n w_1 w_2^{-1} \prod_1^n p_1(x_j),\ 2 \prod_1^n q_2(x_j) \ge c_n \prod_1^n p_2(x_j)\}$$
$$\le P_i\{2 \prod_1^n q_1(x_j) \ge c_n w_1 w_2^{-1} \prod_1^n p_1(x_j),\ 2 \prod_1^n q_1(x_j) \ge c_n \prod_1^n p_2(x_j)\}.$$

Remark 1 shows that (2.11) holds in this case as well, and therefore Theorem 2.2 is proven.

COROLLARY 2.2. *If $K_1 \ge K_2$, then*

$$e^{-K(Q_1,P_1)} \le \max_{k=1,2} \inf_{s,t\ge 0} e^{-(s+t)K_1} E^{P_1} q_k^{s+t}(X) p_1^{-s}(X) p_2^{-t}(X).$$

Indeed it follows from proof of Theorem 2.2 that $K_1 > K_2$ implies $E^{Q_2}\phi^* \to 0$, $E^{P_2}\phi^* \to 0$ and $E^{Q_1}\phi^* \to w_1^{-1} < 1$. Thus ϕ^* is a test of hypothesis P_1 versus Q_1 of asymptotic power at least β. Because of (1.3)

$$e^{-K(Q_1,P_1)} \le \lim_{n\to\infty} (E^{P_1}\phi^*)^{1/n}.$$

But it also follows from the proof that

$$(E^{P_1}\phi^*)^{1/n} \to \max_{i=1,2} \inf_{s,t\ge 0} e^{-(s+t)K_1} E^{P_1} q_i^{s+t}(X) p_1^{-s}(X) p_2^{-t}(X).$$

If $K_1 = K_2$, then for $i = 1, 2$ $\liminf E^{Q_i}\phi^* > 0$, and

$$\max_{k=1,2} \inf_{s,t\ge 0} e^{-(s+t)K_i} E^{P_i} q_k^{s+t}(X) p_2^{-t}(X) = \lim_{n\to\infty} (E^{P_i}\phi^*)^{1/n} \ge e^{-K(Q_i,P_i)}.$$

3. Necessary and sufficient conditions for the existence of adaptive procedures. Let $\mathscr{P}_k = \{P^{(k)}, \theta \in \Theta\}$, $k = 1, \cdots, \ell$ be ℓ families given over the same space $\mathscr{X}$ and indexed by a finite parameter θ. An estimator $\delta(\mathbf{x})$ based on a random sample $\mathbf{x} = (x_1, \cdots, x_n)$ is said to be adaptive for these families if for all $k = 1, \cdots, \ell$

$$\max_\theta [P_\theta^{(k)}(\delta(\mathbf{x}) \ne \theta)]^{1/n} \to \rho(\mathscr{P}_k). \tag{3.1}$$

Here $\rho(\mathscr{P})$ is defined by formula (1.1) and (3.1) means that $\delta(\mathbf{x})$ is asymptotically minimax with respect to each family $\mathscr{P}_k$.

THEOREM 3.1. *If an adaptive estimator exists for families* $\mathscr{P}_k = \{P^{(k)}, \theta \in \Theta\}$ *with pairwise equivalent distributions then*

$$\max_{1\le k\le \ell}\rho(\mathscr{P}_k) \ge \max_{1\le\iota\ne k\le \ell}\max_{\theta\ne\eta}\inf_{s_1,\dots,s_\ell\ge 0}E_\theta^{(k)}[p_\iota(X,\eta)]^{\Sigma s_r}\prod_{r=1}^{\ell}p_r^{-s_r}(X,\theta), \tag{3.2}$$

where $p_\iota(x,\theta)$ *denotes the density of* $P_\theta^{(\iota)}$. *If for all* $k = 1, \cdots, \ell$

$$\rho(\mathscr{P}_k) \ge \max_{\iota:\iota\ne k}\max_{\theta\ne\eta}\inf_{s_1,\dots,s_\ell\ge 0}E_\theta^{(k)}[p_\iota(X,\eta)]^{\Sigma s_r}\prod_{r=1}^{\ell}p_r^{-s_r}(X,\theta), \tag{3.3}$$

then an adaptive estimator exists.

PROOF. Let $w_1, \cdots, w_\ell$ be positive probabilities. Also let $\delta^*(\mathbf{x})$ be the minimax estimator based on the density $\sum w_k\prod_1^n p_k(x_j,\theta)$. Then if δ is an adaptive estimator,

$$\max_{1\le k\le\ell}\max_\theta P_\theta^{(k)}(\delta(\mathbf{x})\ne\theta) \ge \max_\theta\sum w_k P_\theta^{(k)}(\delta(\mathbf{x})\ne\theta) \ge \max_\theta\sum w_k P_\theta^{(k)}(\delta^*(\mathbf{x})\ne\theta).$$

Theorem 2.1 and formula (3.1) imply that

$$\max_{1\le k\le \ell}\rho(\mathscr{P}_k) \ge \max_{1\le\iota,k\le \ell}\max_{\theta\ne\eta}\inf_{s_1,\dots,s_\ell\ge 0}E_\theta^{(k)}[p_\iota(X,\eta)]^{\Sigma s_r}\prod_{r=1}^{\ell}p_r^{-s_r}(X,\theta). \tag{3.4}$$

But

$$\max_{\theta\ne\eta}\inf_{s_1,\dots,s_\ell\ge 0}E_\theta^{(k)}[p_k(X,\eta)]^{\Sigma s_r}\prod_{r=1}^{\ell}p_r^{-s_r}(X,\theta) \tag{3.5}$$

$$\le \max_{\theta\ne\eta}\inf_{s\ge 0}E_\theta^{(k)}p_k^s(X,\eta)p_k^{-s}(X,\theta) = \rho(\mathscr{P}_k),$$

so that (3.4) is equivalent to (3.2).

If condition (3.3) is met then the estimator $\delta^*(\mathbf{x})$ is adaptive. Indeed it follows from the proof of Theorem 2.1 that

$$\max_\theta[P_\theta^{(k)}(\delta^*(\mathbf{x})\ne\theta)]^{1/n} \to \max_{1\le\iota\le\ell}\max_{\theta\ne\eta}\inf_{s_1,\dots,s_\ell\ge 0}E_\theta^{(k)}[p_\iota(X,\eta)]^{\Sigma s_r}\prod_{r=1}^{\ell}p_r^{-s_r}(X,\theta).$$

But because of (3.5) the latter relation implies that

$$\lim\max_\theta[P_\theta^{(k)}(\delta^*(\mathbf{x})\ne\theta)]^{1/n} \le \rho(\mathscr{P}_k),$$

so that δ^* is adaptive.

COROLLARY 3.1. *If an adaptive estimator exists then* (3.4) *is actually an equality, as follows from Corollary* 2.1.

COROLLARY 3.2. *If* $\ell = 2$ *and for some* $\theta\ne\eta$, $p_1(x,\theta) = p_2(x,\eta)$, *then adaptive estimators do not exist.*

This fact follows from the identity

$$\inf_{s,t\ge 0}E_\theta^{(1)}p_2^{s+t}(X,\eta)p_1^{-s}(X,\theta)p_2^{-t}(X,\theta) = \inf_{t\ge 0}E_\theta^{(1)}p_1^t(X,\theta)p_2^{-t}(X,\theta) = 1,$$

which is true since $E_\theta^{(1)}p^t(X,\theta)p_2^{-t}(X,\theta)$ is a convex function of t and its derivative at zero is positive:

$$E_\theta^{(1)}\log[p_1(X,\theta)/p_2(X,\theta)] > 0.$$

It follows from Theorem 2.1 (see (2.8)) that the maximum likelihood estimator $\hat\delta$ or the Bayes estimator δ_B corresponding to the prior concentrated at two parametric points $\{\xi, \zeta\}$ are asymptotically minimax for $\sum_{k=1}^{\ell}w_k\prod_1^n p_k(x_j,\theta)$ for any fixed positive weights $w_1, \cdots, w_\ell$. Therefore under Condition (3.3) both estimators are adaptive.

Another example of an adaptive procedure under (3.3) is the overall maximum likelihood estimator $\tilde\delta(\mathbf{x}) : \tilde\delta(\mathbf{x}) = \eta$ iff

$$\max_k\prod_1^n p_k(x_j,\eta) = \max_\theta\max_k\prod_1^n p_k(x_j,\theta)$$

with ties broken in any (random) way.

Clearly

$$\max_\theta[P_\theta^{(\iota)}(\tilde{\delta}(\mathbf{x})\theta)]^{1/n} \sim \max_{\theta\neq\eta}\{P^{(\iota)}(\max \prod_1^n p_k(x_j,\eta) > \max \prod_1^n p_k(x_j,\theta))\}^{1/n}$$

$$\to \max_{1\le k\le \ell}\max_{\theta\neq\eta}\inf_{s_1,\dots,s_\ell\ge 0}E^{(\iota)}[p_k(X,\eta)]^{\sum_1^\ell s_r}\prod_{r=1}^{\ell} p_r^{-s_r}(X,\theta) = \max_\theta \rho_\iota(\theta),$$

so that $\tilde{\delta}$ is adaptive.

Thus we can formulate the following result.

THEOREM 3.2. *If Condition* (3.3) *is satisfied then the following estimators are adaptive*:

(i) *maximum likelihood estimators* $\hat{\delta}(\mathbf{x})$ *defined by*

$$\hat{\delta}(\mathbf{x}) = \eta \text{ iff } \sum_1^\ell w_k \prod_1^n p_k(x_j,\eta) = \max_\theta \sum_1^\ell w_k \prod_1^n p_k(x_j,\theta),$$

where $w_1, \cdots, w_\ell$ *are fixed positive weights*;

(ii) *overall maximum likelihood estimator defined by*

$$\hat{\delta}(\mathbf{x}) = \eta \text{ iff } \max_k \prod_1^n p_k(x_j,\eta) = \max_\theta \max_k \prod_1^n p_k(x_j,\theta).$$

Now let us consider hypothesis testing problems and adaptive tests.

THEOREM 3.3. *If an adaptive test of hypothesis* P_1 *versus* Q_1 *and* P_2 *versus* Q_2 *exists, then*

$$\max(e^{-K(Q_1,P_1)}, e^{-K(Q_2,P_2)}) \ge \max_{\iota\neq k}\inf_{s,t\ge 0} e^{-(s+t)K}E^{P_\iota}q_k^{s+t}(X)p_1^{-s}(X)p_2^{-t}(X), \tag{3.6}$$

where $K = \max(K_1, K_2)$ *and* K_ι, $i = 1, 2$ *are given by* (2.12).

If $K_1 = K_2 = K$ *and for* $i = 1, 2$

$$\max_{k:k\neq\iota}\inf_{s,t\ge 0}e^{-(s+t)K}E^{P_\iota}q_k^{s+t}(X)p_1^{-s}(X)p_2^{-t}(X) \le e^{-K(Q_\iota,P_\iota)}, \tag{3.7}$$

then an adaptive test exists.

PROOF. Let ϕ be an adaptive test for hypotheses P_1 against Q_1 and P_2 against Q_2 of power β. Then

$$w_1E^{Q_1}\phi + w_2E^{Q_2}\phi = \beta,$$

so that ϕ as a test of hypothesis $w_1P_1 + w_2P_2$ against $w_1Q_1 + w_2Q_2$ has power β for any positive $w_1, w_2, w_1 + w_2 = 1$. Therefore

$$\begin{aligned}
&\max(e^{-K(Q_1,P_1)}, e^{-K(Q_2,P_2)}) \\
&\quad = \max(\lim_{n\to\infty}(E^{P_1}\phi)^{1/n}, \lim_{n\to\infty}(E^{P_2}\phi)^{1/n}) \\
&\quad = \lim_{n\to\infty}(w_1E^{P_1}\phi + w_2E^{P_2}\phi)^{1/n} \ge \lim_{n\to\infty}(w_1E^{P_1}\phi^* + w_2E^{P_2}\phi^*)^{1/n} \\
&\quad = \max_{\iota,k}\inf_{s,t>0}e^{-(s+t)K}E^{P_\iota}q_k^{s+t}(X)p_1^{-s}(X)p_2^{-t}(X).
\end{aligned} \tag{3.8}$$

Here ϕ^* is the optimal test of hypothesis $w_1P_1 + w_2P_2$ versus $w_1Q_1 + w_2Q_2$, and the weights w_1, w_2 are assumed to satisfy the condition of Theorem 2.2.

Also for $i = 1, 2$

$$\begin{aligned}
\inf_{s,t\ge 0}e^{-(s+t)K}E^{P_\iota}q_\iota^{s+t}(X)p_1^{-s}(X)p_2^{-t}(X) &\le \inf_{s,t\ge 0}e^{-(s+t)K_\iota}E^{P_\iota}q_\iota^{s+t}(X)p_2^{-t}(X) \\
&\le \inf_{s\ge 0}e^{-sK_\iota}E^{P_\iota}q_\iota^{s}(X)p_\iota^{-s}(X) = e^{-K(Q_\iota,P_\iota)}.
\end{aligned} \tag{3.9}$$

Therefore (3.8) implies (3.6).

Now assume that $K_1 = K_2 = K$. Let

$$\phi_i(\mathbf{x}) = \begin{cases} 1 & w_1 \prod_1^n q_1(x_j) + w_2 \prod_1^n q_2(x_j) > c_n^{(i)}[w_1 \prod_1^n p_1(x_j) + w_2 \prod_1^n p_2(x_j)], \\ \gamma_n^{(i)} & w_1 \prod_1^n q_1(x_j) + w_2 \prod_1^n q_2(x_j) = c_n^{(i)}[w_1 \prod_1^n p_1(x_j) + w_2 \prod_1^n p_2(x_j)], \\ 0 & w_1 \prod_1^n q_1(x_j) + w_2 \prod_1^n q_2(x_j) < c_n^{(i)}[w_1 \prod_1^n p_1(x_j) + w_2 \prod_1^n p_2(x_j)], \end{cases}$$

where $c_n^{(i)}$ and $\gamma_n^{(1)}$, $0 \le \gamma_n^{(i)} \le 1$ are chosen in such a way that for $i = 1, 2$

$$E^{Q_i}\phi_i(\mathbf{x}) = \beta.$$

It follows from the proof of Theorem 2.2 that $\lim\{n^{-1}\log c_n^{(i)}\} = K$. Define a new test $\phi : \phi = \phi_1$ if $c_n^{(1)} \le c_n^{(2)}$ and $\phi = \phi_2$ otherwise. Then $E^{Q_i}\phi(\mathbf{x}) \ge \beta$ and

$$\lim_{n\to\infty}\{E^{P_i}\phi(\mathbf{x})\}^{1/n} = \max_{1\le i,k\le 2}\inf_{s,t\ge 0} e^{-(s+t)K} E^{P_i} q_k^{s+t}(X) p_1^{-s}(X) p_2^{-t}(X).$$

Because of (3.9) the latter relation implies that

$$\lim_{n\to\infty}\{E^{P_i}\phi(\mathbf{x})\}^{1/n} \le e^{-K(Q_i,P_i)},$$

which proves the Theorem.

COROLLARY 3.2. *If an adaptive test exists then* (3.8) *is actually an equality.*

4. Examples. In this Section we illustrate Theorem 3.1 by two examples, assuming for simplicity that $\Theta = \{0, 1\}$.

EXAMPLE 1. *One-parameter exponential families.* Let measures $P_\theta^{(k)}$ be defined over an abstract space $\mathscr{X}$ and let their densities with respect to some σ-finite measure μ be of the form

$$p_k(x, \theta) = \{C(\alpha_k(\theta))\}^{-1}\exp\{\alpha_k(\theta)v(x)\},$$

$\alpha_k(\theta) \ne \alpha_i(\theta)$ for $k \ne i$. Here $C(\alpha) = \int_{\mathscr{X}} e^{\alpha v(x)}\, d\mu(x)$ and α belongs to the natural parameter space, which is, of course, an interval with endpoints α_-, α_+. We assume that the common support of all measures $P_\theta^{(k)}$ has at least two points. It is well known that in this case $f(\alpha) = \log C(\alpha)$ is a strictly convex function. We define for $k = 1, \cdots, \ell$

$$H(\alpha_k(\theta), \alpha_k(\eta)) = \inf_{0<s<1} \int p_k^{1-s}(x, \theta) p_k^s(x, \eta)\, d\mu(x) \tag{4.1}$$
$$= \exp \inf_{0<s<1}[f(\alpha_k(\eta)s + \alpha_k(\theta)(1-s)) - sf(\alpha_k(\eta)) - (1-s)f(\alpha_k(\theta))],$$

so that

$$\rho(\mathscr{P}_k) = \exp\{\max_{\eta\ne\theta} H(\alpha_k(\theta), \alpha_k(\eta))\}.$$

To check the conditions of Theorem 3.1 assume that $\ell = 2$. Then we have to evaluate

$$\inf_{s,t\ge 0} E_\theta^{(1)} p_2^{s+t}(x, \eta) p_1^{-s}(X, \theta) p_2^{-t}(X, \theta)$$
$$= \exp\{\inf_{s,t\ge 0}[f(\alpha_2(\eta)(s+t) + \alpha_1(\theta)(1-s) - \alpha_2(\theta)t) - (s+t)f(\alpha_2(\eta)) - (1-s)f(\alpha_1(\theta)) + tf(\alpha_2(\theta))]\} \tag{4.2}$$

and

$$\inf_{s,t\ge 0} E_\theta^{(2)} p_1^{s+t}(X, \eta) p_1^{-s}(X, \theta) p_2^{-t}(X, \theta)$$
$$= \exp\{\inf_{s,t\ge 0}[f(\alpha_1(\eta)(s+t) + \alpha_2(\theta)(1-t) - \alpha_1(\theta)s) - (s+t)f(\alpha_1(\eta)) - (1-t)f(\alpha_2(\theta)) + sf(\alpha_1(\theta))]\} \tag{4.3}$$

for $\theta \neq \eta$.

Notice first of all that the vector of partial derivatives of the functions in (4.2) and (4.3) does not vanish in the open quadrant $\{(s, t), s > 0, t > 0\}$. (Actually only the subset of this region, where $\alpha_- < (s + t)\alpha_2(\eta) + (1 - s)\alpha_1(\theta) - t\alpha_2(\theta) < \alpha_+$, has to be considered.) Indeed, in the case of the function in (4.2), this vector vanishes if and only if

$$\{\alpha_2(\eta) - \alpha_1(\theta)\}f'(\alpha_1(\theta) + s\{\alpha_2(\eta) - \alpha_1(\theta)\} + t\{\alpha_2(\eta) - \alpha(\theta)\}) = f(\alpha_2(\eta)) - f(\alpha_1(\theta))$$

and

$$\{\alpha_2(\eta) - \alpha_2(\theta)\}f'(\alpha_1(\theta) + s\{\alpha_2(\eta) - \alpha_1(\theta)\} + t\{\alpha_2(\eta) - \alpha_2(\theta)\}) = f(\alpha_2(\eta)) - f(\alpha_2(\theta)),$$

which implies that

$$\{f(\alpha_2(\eta)) - f(\alpha_1(\theta))\}\{\alpha_2(\eta) - \alpha_1(\theta)\}^{-1} = \{f(\alpha_2(\eta)) - f(\alpha_2(\theta))\}\{\alpha_2(\eta) - \alpha_2(\theta)\}^{-1}. \tag{4.4}$$

Since $\{f(\alpha_2(\eta)) - f(\alpha)\}\{\alpha_2(\eta) - \alpha\}^{-1}$ as a function of α is strictly monotone, (4.4) means that $\alpha_1(\theta) = \alpha_2(\theta)$, which contradicts our assumption. Therefore

$$\begin{aligned}
&\inf_{s,t\geq 0}[f(\alpha_2(\eta)(s + t) + \alpha_1(\theta)(1 - s) - \alpha_2(\theta)t)\\
&\qquad - (s + t)f(\alpha_2(\eta)) - (1 - s)f(\alpha_1(\theta)) + tf(\alpha_2(\theta))]\\
&\quad = \min\{\inf_{s\geq 0}[f(\alpha_2(\eta)s + \alpha_1(\theta)(1 - s)) - sf(\alpha_2(\eta)) - (1 - s)f(\alpha_1(\theta))],\\
&\qquad \inf_{t\geq 0}[f(\alpha_1(\theta) + (\alpha_2(\eta) - \alpha_2(\theta))t) - f(\alpha_1(\theta)) - tf(\alpha_2(\eta)) + tf(\alpha_2(\theta))]\}.\\
&\quad = \min\{H(\alpha_2(\eta), \alpha_1(\theta)), H(\alpha_1(\theta), \alpha_2(\eta))\}.
\end{aligned}$$

Thus condition (3.3) of Theorem 3.1 means: for $k = 1, 2$

$$\max_{\theta\neq\eta}[H(\alpha_k(\theta), \alpha_k(\eta))] \geq \max_{i:i\neq k}\max_{\theta\neq\eta}H(\alpha_i(\theta), \alpha_k(\eta)).$$

The Condition (3.2) takes the form

$$\max_{i=1,2}\max_{\theta\neq\eta}H(\alpha_i(\theta), \alpha_i(\eta)) \geq \max_{i\neq k}\max_{\theta\neq\eta}H(\alpha_i(\theta), \alpha_i(\eta)).$$

These results can be easily extended to the case of arbitrary finite ℓ. We formulate them as

THEOREM 4.1. *Let*

$$p_k(x, \theta) = \{C(\alpha_k(\theta))\}^{-1}\exp\{\alpha_k(\theta)v(x)\}, \qquad k = 1, \cdots, \ell, \tag{4.5}$$

be densities of a one-parameter exponential family $\alpha_i(\theta) \neq \alpha_k(\theta)$ for $i \neq k$. Assume that the common support of p_k, $k = 1, \cdots, \ell$ contains more than one point. If an adaptive estimator exists, then

$$\max_{1\leq k\leq\ell}\max_{\theta\neq\eta}H(\alpha_k(\theta), \alpha_k(\eta)) \geq \max_{1\leq i\neq k\leq\ell}\max_{\theta\neq\eta}H(\alpha_i(\theta), \alpha_k(\eta)), \tag{4.6}$$

where $H(\alpha_i(\theta), \alpha_k(\eta))$ is defined in (4.1). *If for $k = 1, \cdots, \ell$*

$$\max_{\theta\neq\eta}H(\alpha_k(\theta), \alpha_k(\eta)) \geq \max_{i:i\neq k}\max_{\theta\neq\eta}H(\alpha_i(\theta), \alpha_k(\eta)). \tag{4.7}$$

then an adaptive estimator exits.

The heuristic interpretation of (4.6) is that an adaptive estimator cannot exist if all measures $P_\theta^{(k)}$ and $P_\eta^{(i)}$, $i \neq k$, $\theta \neq \eta$ are "closer to each other" than measures $P_\theta^{(k)}$ and $P_\eta^{(k)}$.

Theorem 4.1 contains many interesting cases.

(i) *Normal distributions with unknown means $\alpha_k(\theta)$ and known variance σ^2.* In this case $u(x) = x/\sigma^2$, $C(\alpha) = \exp\{\alpha^2/(2\sigma^2)\}$, $f(\alpha) = \alpha^2/(2\sigma^2)$. Easy calculation shows that $H(\alpha, \beta) = -(\alpha - \beta)^2/(8\sigma^2)$. According to Theorem 4.1 an adaptive estimator exists if for $k = 1, \cdots, \ell$

$$\min_{i:i\neq k}\min_{\theta\neq\eta}|\alpha_i(\theta) - \alpha_k(\eta)| \geq \min_{\theta\neq\eta}|\alpha_k(\theta) - \alpha_k(\eta)|. \tag{4.8}$$

If an adaptive estimator exists then

(4.9) $$\min_{1\le k\le \ell}\min_{\theta\neq\eta}|\alpha_\iota(\theta)-\alpha_\iota(\eta)|\le \min_{\iota\neq k}\min_{\theta\neq\eta}|\alpha_\iota(\theta)-\alpha_k(\eta)|.$$

This example shows that the sufficient condition of Theorem 3.1 is actually stronger than the necessary condition. If, say, $m=1$, $-\alpha_1(0)=\alpha_1(1)=\alpha>0$ and $\alpha_2(0)<0<\alpha_2(1)$ then Condition (4.9) means that $\alpha\le\min\{|\alpha_2(0)|,\alpha_2(1)\}$, but Condition (4.8) just means that $\alpha_2(1)=\alpha=-\alpha_2(0)$. Other examples of such kind can be obtained from (ii) and (iii).

(ii) *Scale parameter families.* Here we have densities of the form

$$p_k(x,\theta)=C\{\alpha(\theta)\}^a\exp\{-\alpha(\theta)|x|^{a-1}\},\qquad -\infty<x<\infty,$$

or of the form

$$p_k(x,\theta)=C\{\alpha(\theta)\}^a\exp\{-\alpha(\theta)x^{a-1}\},\qquad x\ge 0.$$

(These families include normal, exponential and double exponential distributions with unknown scale parameter.) In this case $C(\alpha)=\alpha^{-a}$, $f(\alpha)=-a\log\alpha,\alpha>0$, $a>0$. Also

$$H(\alpha,\beta)=a\inf_{s>0}[s\log\alpha+(1-s)\log\beta-\log\{\alpha s+(1-s)\beta\}]$$

$$=a\inf_{s>0}[s\log\gamma-\log(1-s+s\gamma)],$$

where $\gamma=a/b$. For $\gamma\neq 1$

$$\inf_{s>0}[s\log\gamma-\log(1-s+s\gamma)]=(\gamma-1-\log\gamma)(\gamma-1)^{-1}-\log\{(\gamma-1)/\log\gamma\}=h(\gamma).$$

It is easy to check that $h(\gamma^{-1})=h(\gamma)$ and that $h(\gamma)$ is a unimodal function which attains its maximum at $\gamma=1$ and is increasing for $0<\gamma<1$.

Therefore inequalities (4.7) in this case mean that

$$1<\max_{\theta\neq\eta}\max[\alpha_k(\theta)/\alpha_k(\eta),\alpha_k(\eta)/\alpha_k(\theta)]\le\max_{\iota\cdot\iota\neq k}\max_{\theta\neq\eta}\max[\alpha_\iota(\theta)/\alpha_k(\eta),\alpha_k(\eta)/\alpha_\iota(\theta)],$$

in which situation an adaptive estimator exists.

Also because of (4.6) an adaptive estimator does not exist if

$$\max_{1\le k\le\ell}\max_{\theta\neq\eta}\max[\alpha_k(\theta)/\alpha_k(\eta),\alpha_k(\eta)/\alpha_k(\eta)]$$
$$>\max_{1\le\iota\neq k\le\ell}\max_{\theta\neq\eta}\max[\alpha_\iota(\theta)/\alpha_k(\eta),\alpha_k(\eta)/\alpha_\iota(\theta)].$$

(iii) *Binomial distribution.* Here

$$p_k(x,\theta)=\binom{N}{x}\{p_k(\theta)\}^x\{1-p_k(\theta)\}^{n-x},\qquad x=0,\cdots,N,$$

$\alpha(\theta)=\log p(\theta)[\log\{1-p(\theta)\}]^{-1}$. Although this example is of the type treated in Theorem 4.1 it is more convenient to evaluate the function $H(p_k(0),p_k(1))$ directly:

$$\exp H(p_k(0),p_k(1))$$
$$=\inf_{s\ge0}\sum_{x=0}^{N}\binom{N}{x}[p_k(0)]^{sx}[q_k(0)]^{s(N-x)}[p_k(1)]^{(1-s)x}[q_k(1)]^{(1-s)(N-x)}$$
$$=\inf_{s\ge0}\{[p_k(0)]^s[p_k(1)]^{1-s}+[q_k(0)]^s[q_k(1)]^{1-s}\}^N=\rho(\mathscr{P}_k),$$

$k=1,\cdots,\ell,\qquad q_k(\theta)=1-p_k(\theta),\qquad \theta=0,1.$

For a fixed k let

$$H_k(\gamma)=\inf_{s\ge0}[p_k^{1-s}(1)\gamma^s+q_k^{1-s}(1)(1-\gamma)^s]$$

$0\le\gamma\le1$. The function $H_k(\gamma)$ is unimodal with a maximum at $\gamma=p_k(1)$. The condition

$$H(p_k(0),p_k(1))\ge\max_{\iota:\iota\neq k}H(p_\iota(0),p_k(1)),$$

which is equivalent to (3.3), means that

$$H_k(p_k(0)) \geq \max_{\iota:\iota\neq k} H_k(p_i(0)).$$

This condition, of course, signifies that $p_k(0)$ is "closer" to $p_k(1)$ than $p_i(0)$, $i \neq k$, and if this holds for all k, an adaptive estimator exists. If it exists, then

$$\max_k H_k(p_k(0)) \geq \max_{1<\iota\neq k<\ell} H_k(p_\iota(0)).$$

Similar examples can be given for Theorem 3.3.

EXAMPLE 2. *Location parameter families on a cyclic group.* Assume that $\mathscr{X} = \Theta = \{0, 1\}$, $p_k(x, \theta) = p_k(x - \theta)$, $k = 1, \cdots, \ell$, where difference $x - \theta$ is understood to be modulo two. Thus $p_k(0) + p_k(1) = 1$ and

$$\rho(\mathscr{P}_k) = \inf_{s\geq 0}\{p_k^s(1)p_k^{1-s}(0) + p_k^s(0)p_k^{1-s}(1)\} = 2\{p_k(0)p_k(1)\}^{1/2}.$$

Also if, say, $\ell = 2$, $\theta \neq \eta$

$$\begin{aligned}
&\inf_{s,t\geq 0} E_\theta^{(1)} p_2^{s+t}(X, \eta)p_1^{-s}(X, \theta)p_2^{-t}(X, \theta)\\
&= \inf_{s,t\geq 0}[p_2^{s+t}(\eta)p_1^{1-s}(\theta)p_2^{-t}(\theta) + p_2^{s+t}(\theta)p_1^{1-s}(\eta)p_2^{-t}(\eta)]\\
&= \min\{\inf_{s\in A}[p_1^{1-s}(\theta)p_2^s(\eta) + p_1^{1-s}(\eta)p_2^s(\theta)],\\
&\quad \inf_{s\in B} 2[p_1(\theta)p_1(\eta)]^{(1-s)/2}[p_2(\theta)p_2(\eta)]^{s/2}\},
\end{aligned}$$

where

$$A = \{s : s > 0, p_1^{1-s}(\theta)p_2^s(\eta) > p_1^{1-s}(\eta)p_2^s(\theta)\},$$

and B is the complement of A.

If $p_1(\theta) < p_1(\eta)$, then the set B contains zero and

$$\inf_{s\in B} 2[p_1(\theta)p_1(\eta)]^{(1-s)/2}[p_2(\theta)p_2(\eta)]^{s/2} \leq 2[p_1(\theta)p_1(\eta)]^{1/2} = \rho(\mathscr{P}_1).$$

If $p_1(\theta) > p_1(\eta)$ and $p_2(\eta)p_2(\theta) = p_2(0)p_2(1) < p_1(0)p_1(1)$, then

$$\inf_{s\in B} 2[p_1(\theta)p_1(\eta)]^{(\lambda-s)/2}[p_2(\theta)p_2(\eta)]^{s/2} = 0.$$

If $p_1(\theta) > p_1(\eta)$ and $p_2(0)p_2(1) > p_1(0)p_1(1)$, the set A contains the interval $[0, 1]$ and

$$\inf_{s\in A}[p_1^{1-s}(\theta)p_2^s(\eta) + p_1^{1-s}(\eta)p_2^s(\theta)] \leq \inf_{s\in B} 2[p_1(\theta)p_1(\eta)]^{(1-s)/2}[p_2(\theta)p_2(\eta)]^{s/2}.$$

Let for $0 < p < 1$

$$H_k(p) = \inf_{0\leq s\leq 1}[p_k^{1-s}(0)p^s + p_k^{1-s}(1)(1-p)^s].$$

Then $H_k(p)$ is a unimodal function with a unique maximum at $p = p_k(0)$, and it is increasing in the interval $(0, p_k(0))$. The inequality

$$\inf_{s\in A}[p_1^{1-s}(\theta)p_2^s(\eta) + p_1^{1-s}(\eta)_2^s(\theta)] \leq \rho(\mathscr{P}_k)$$

means that

$$H_1(p_2(1)) \leq H_1(p_1(1)).$$

Also if $p_2(\eta) < p_2(\theta)$

$$\begin{aligned}
&\inf_{s,t\geq 0} E_\theta^{(1)} p_2^{s+t}(X, \eta)p_1^{-s}(x, \theta)p_2^{-t}(x, \theta)\\
&\quad = \min\{\inf_{s\in B}[p_1^{1-s}(\theta)p_2^s(\eta) + p_1^{1-s}(\eta)p_2^s(\theta)], \inf_{s\in A} 2[p_1(\theta)p_1(\eta)]^{(1-s)/2}[p_2(\theta)p_2(\eta)]^{s/2}\}.
\end{aligned}$$

The latter quantity is less than $\rho(\mathscr{P}_1)$ if $p_1(\theta) > p_1(\eta)$ or if $p_1(\theta) < p_1(\eta)$ and $p_2(0)p_2(1) < p_1(0)p_1(1)$. When $p_1(\theta) < p_1(\eta)$ and $p_2(0)p_2(1) > p_1(0)p_1(1)$, this inequality means that

$$H(p_2(\eta)) \leq H_1(p_1(1)).$$

Thus

$$\rho(\mathscr{P}_1) < \max_{\theta\neq\eta}\inf_{s,t\geq 0} E_\theta^{(1)} p_2^{s+t}(X, \eta) p_1^{-s}(X, \theta) p_2^{-t}(X, \theta),$$

if $p_2(0)p_2(1) > p_1(0)p_1(1)$ and $p_1(1) > p_1(0)$, $p_2(0) > p_2(1)$, $H(p_2(1)) > H(p_1(1))$, or $p_1(0) > p_1(1)$, $p_2(1) > p_2(0)$, $H_1(p_2(1)) > H_1(p_1(1))$.

Because of the mentioned properties of the function H inequalities $p_1(1) > p_1(0)$, $p_2(0) > p_2(1)$ and $|p_2(0) - \frac{1}{2}| > |p_1(1) - \frac{1}{2}|$ (which is tantamount to $p_2(0)p_2(1) > p_1(0)p_1(1)$) imply that $H(p_2(0)) > H(p_1(1))$. Also inequalities $p_1(0) > p_1(1)$, $p_2(1) > p_2(0)$ and $|p_2(1) - \frac{1}{2}| > |p_1(1) - \frac{1}{2}|$ imply that $H(p_2(1)) > H_1(p_1(1))$.

Therefore, in general, an adaptive estimator exists if and only if $p_k(1) > p_1(0)$ $k = 1, \cdots, \ell$ or $p_k(1) < p_k(0)$, $k = 1, \cdots, \ell$. In these cases the estimator which takes the value corresponding to the minimal (maximal) observed frequency is adaptive.

Acknowledgments. The author is grateful to Professor Herman Rubin for an interesting discussion. Thanks are also due to Professor Leon Gleser, who read the manuscript and made several critical remarks, and to the referee for many helpful suggestions.

REFERENCES

BAHADUR, R. R. (1971). Some Limit Theorems in Statistics. Regional Conference Series in Applied Mathematics. Society for Industrial and Applied Mathematics, Philadelphia.

BAHADUR, R. R. and ZABELL, S. L. (1979). Large deviations of the sample mean in general vector spaces. *Ann. Probability* **7** 587–621.

BARTFAI, P. (1978). Large deviations of the sample mean in Euclidean spaces. Mimeograph Series No. 78-13, Statist. Dept., Purdue Univ.

BERAN, R. (1974). Asymptotically efficient adaptive rank estimates in location models. *Ann. Statist.* **2** 63–74.

CHERNOFF, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statist.* **23** 493–507.

CHERNOFF, H. (1956). Large sample theory-parametric case. *Ann. Math. Statist.* **27** 1–22.

GHOSH, J. K. and SUBRAMANYAM, K. (1975). Inference about separated families in large samples. *Sankhyā* **37** 502–513.

GROENEBOOM, P., OOSTERHOFF, J. and RUYMGAART, F. H. (1979). Large deviation theorems for empirical probability measures. *Ann. Probability* **7** 553–586.

KRAFFT, O. and PURI, M. L. (1974). The asymptotic behavior of the minimax risk for multiple decision problems. *Sankhyā* **36** 1–12.

KULLBACK, S. (1959). *Information and Statistics.* Wiley, New York.

PLACHKY, D. and STEINEBACH, J. (1977). A generalization of a result of Chernoff in large sample theory. *Math. Operationsforsch. Statist., Ser. Statistics* **8** 375–379.

SACKS, J. (1975). An asymptotically efficient sequence of estimators of a location parameter. *Ann. Statist.* **3** 285–298.

STONE, C. J. (1975). Adaptive maximum likelihood estimators of a location parameter. *Ann. Statist.* **3** 267–284.

WALD, A. (1950). *Statistical Decision Functions.* Wiley, New York.

DEPARTMENT OF STATISTICS
PURDUE UNIVERSITY
WEST LAFAYETTE, INDIANA 47907