ADAPTIVE PROCEDURES IN MULTIPLE DECISION PROBLEMS AND HYPOTHESIS TESTING*

by

Andrew L. Rukhin
Purdue University

Department of Statistics
Division of Mathematical Sciences
Mimeograph Series #80-30

September 1980

*Research supported by NSF Grant MCS 78-02300

ADAPTIVE PROCEDURES IN MULTIPLE DECISION PROBLEMS AND HYPOTHESIS TESTING

by

Andrew L. Rukhin
Purdue University

ABSTRACT

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. These procedures exhibit the same asymptotic behavior for several parametric families as optimal (minimax) estimators for each of these families. The proofs are based on the multivariate version of Chernoff's theorem which gives 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.

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.

1. INTRODUCTION

We start with the simple multiple decision problem where both the action space and the parameter space coincide and are finite, say, $\Theta = \{0,1,\ldots,m\}$, $m \geq 1$. Thus a family of (different) probability distributions $\mathcal{P} = \{P_0,\ldots,P_m\}$ over a space $\mathcal{X}$ is given, and statistical inference about the finite-valued parameter is desired on the basis of a random sample $\underline{x} = (x_1,\ldots,x_n)$.

If $\delta(\underline{x})$ is an estimator of this parameter, then the probability of incorrect decision $P_\theta(\delta(\underline{x}) \neq \theta)$ is the most important characteristic of the procedure δ. The asymptotic behavior of this probability for minimax estimator δ^* was studied by many scholars (see Bahadur (1960), Krafft and Puri (1974), Ghosh and Subramanyam (1975)). The main result here has the form

$$\lim_{n\to\infty} \max_\theta P_\theta^{1/n}(\delta^*(\underline{x})\neq\theta) = \max_{\eta\neq\theta} \inf_{t\geq 0} E^{P_\theta} p_\eta^t(X)p_\theta^{-t}(X)$$

$$= \max_{\eta\neq\theta} \inf_{t\geq 0} \int_{\mathcal{X}} p_\eta^t(x)p_\theta^{1-t}(x)d\mu(x) = \rho(\mathcal{P}), \quad (1.1)$$

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

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

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

$$\alpha_n^{1/n} = [E^P \varphi^*(\underline{x})]^{1/n} \to \exp\{-K(Q,P)\} \ . \qquad (1.3)$$

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

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

Formula (1.3) has been generalized to the case of testing a finite hypothesis versus a finite alternative 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

$$E^{Q_i}\varphi(\underline{x}) = \beta$$

and

$$[E^{P_i}\varphi(\underline{x})]^{1/n} \to \exp\{-K(Q_i,P_i)\} \ .$$

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 $\mathcal{P}_1 = \{P_\theta^{(1)}, \theta \in \Theta\}$ and

$\mathcal{P}_2 = \{P_\theta^{(2)}, \theta \in \Theta\}$ under which these exists an adaptive estimator δ, i.e. such that for i=1,2

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

In other terms an adaptive estimator δ serves both families $\mathcal{P}_1$ and $\mathcal{P}_2$ 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 different settings (see Beran (1974), Sacks (1975), Stone (1975)).

The definition of adaptive estimator and all results of this article are valid if one replaces the probablity of incorrect decision by arbitrary risk, $R(\theta,\delta) = E_\theta W(\theta,\delta(\underline{x}))$, where $W(\theta,\theta) = 0$ and $W(\theta,\eta) \neq 0$, $\theta \neq \eta$.

In Section 3 we give the necessary condition and the 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 densities of $P_\theta^{(1)}$ and $P_\theta^{(2)}$. This study is performed in Section 2. The basic mathematical tool needed is multivariate Chernoff's theorem, which provides 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 (the gap between which seems to be difficult to fill) for the existence of adaptive procedures by several examples. Typically adaptive estimators do not exist if the measures $P_\theta^{(i)}$ 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 $w_1 \prod_1^n f_1(x_j)+w_2 \prod_1^n f_2(x_j)$ where f_1 and f_2 are probability densities, and $w_1+w_2 = 1$. We start with the following key result.

LEMMA. Let c_n, n=1,2,... 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=1,2, are strictly positive probability densities, w_1 and w_2 are positive numbers, $w_1+w_2 = 1$, and for all a, b $\geq$ 0, a+b = 1

$$\Pr\left(a \log \frac{f_1(X)}{g_1(X)} + b \log \frac{f_1(X)}{g_2(X)} > L\right) > 0 \tag{2.1}$$

and

$$\Pr\left(a \log \frac{f_2(X)}{g_1(X)} + b \log \frac{f_2(X)}{g_2(X)} > L\right) > 0 , \tag{2.2}$$

then

$$\Pr^{1/n}\left(w_1 \prod_1^n f_1(x_j)+w_2 \prod_1^n f_2(x_j) \geq c_n[w_1 \prod_1^n g_1(x_j)+w_2 \prod_1^n g_2(x_j)]\right)$$

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

Proof. Notice first of all that

$$\Pr(2 \max[w_1 \prod_1^n f_1(x_j), w_2 \prod_1^n f_2(x_j)] \geq c_n \max[w_1 \prod_1^n g_1(x_j), w_2 \prod_1^n g_2(x_j)])$$

$$\geq \Pr(w_1 \prod_1^n f_1(x_j)+w_2 \prod_1^n f_2(x_j) \geq c_n[w_1 \prod_1^n g_1(x_j)+w_2 \prod_1^n g_2(x_j)])$$

$$\geq \Pr(\max[w_1 \prod_1^n f_1(x_j), w_2 \prod_1^n f_2(x_j)] \geq 2c_n \max[w_1 \prod_1^n g_1(x_j), w_2 \prod_1^n g_2(x_j)]). \tag{2.3}$$

One has

$$\Pr(\max[w_1 \prod_1^n f_1(x_j), w_2 \prod_1^n f_2(x_j)] \geq 2c_n \max[w_1 \prod_1^n g_1(x_j), w_2 \prod_1^n g_2(x_j)])$$

$$= \Pr(\prod_1^n f_1(x_j) \geq 2c_n \prod_1^n g_1(x_j), \prod_1^n f_1(x_j) \geq 2c_n w_1^{-1} w_2 \prod_1^n g_2(x_j))$$

$$+ \Pr(\prod_1^n f_2(x_j) \geq 2c_n w_1 w_2^{-1} \prod_1^n g_1(x_j), \prod_1^n f_2(x_j) \geq 2c_n \prod_1^n g_2(x_j)) ,$$

so that

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

$$\sim \max[\Pr^{1/n}(\prod_1^n f_1(x_j) \geq 2c_n \prod_1^n g_1(x_j), \prod_1^n f_1(x_j) \geq 2c_n w_1^{-1} w_2 \prod_1^n g_2(x_j)) ,$$

$$\Pr^{1/n}(\prod_1^n f_2(x_j) \geq 2c_n w_1 w_2^{-1} \prod_1^n g_1(x_j), \prod_1^n f_2(x_j) \geq 2c_n \prod_1^n g_2(x_j))] .$$

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

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

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

$$\Pr(sY+tZ>sy+tz) > 0 \tag{2.5}$$

for all nonnegative s, t, $(s,t) \neq (0,0)$.

(The latter condition 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 $\{v \in R^2, \inf[Eq(X)\log q(X): \int xq(x)dx \geq v] < \infty\}$, which is demanded in 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 assumptions (2.1) and (2.2).

Thus

$$\Pr{}^{1/n}\Big(\prod_1^n f_1(x_j) \geq 2c_n \prod_1^n g_1(x_j),\ \prod_1^n f_1(x_j) \geq 2_n w_1^{-1} w_2 \prod_1^n g_2(x_j)\Big)$$

$$\to \inf_{s,t\geq 0} e^{-(s+t)L} Ef_1^{s+t}(X) g_1^{-s}(X) g_2^{-t}(X) ,$$

and

$$\Pr{}^{1/n}\Big(\prod_1^n f_2(x_j) \geq 2c_n w_1 w_2^{-1} \prod_1^n g_1(x_j),\ \prod_1^n f_2(x_j) \geq 2c_n \prod_1^n g_2(x_j)\Big)$$

$$\to \inf_{s,t\geq 0} e^{-(s+t)L} Ef_2^{s+t}(X) g_1^{-s}(X) g_2^{-t}(X).$$

Therefore

$$\Pr{}^{1/n}\Big(\max[w_1 \prod_1^n f_1(x_j), w_2 \prod_1^n f_2(x_j)] \geq 2c_n \max[w_1 \prod_1^n g_1(x_j), w_2 \prod_1^n g_2(x_j)]\Big)$$

$$\to \max_{i=1,2} \inf_{s,t\geq 0} e^{-(s+t)L} Ef_i^{s+t}(X) g_1^{-s}(X) g_2^{-t}(X),$$

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

Remark 2.1. The generalization of the Lemma to the case of arbitrary finite mixture of positive densities is straightforward. Multivariate Chernoff's theorem shows that if $w_1,\ldots,w_\ell > 0$, $w_1+\ldots+w_\ell = 1$, then

$$\Pr{}^{1/n}\left(\sum_{k=1}^{\ell} w_k \prod_1^n f_k(x_j) \geq c_n \sum_{k=1}^{\ell} w_k \prod_1^n g_k(x_j)\right)$$

$$\to \max_{1\leq k\leq \ell} \quad \inf_{s_1,\ldots,s_\ell \geq 0} e^{-(s_1+\ldots+s_\ell)L} E f_k^{s_1+\ldots+s_\ell}(X) \prod_{i=1}^{\ell} g_i^{-s_i}(X).$$

Here $L = \lim n^{-1} \log c_n$ and the probabilities $\Pr\left(\sum_k \alpha_k \log \frac{f_i(X)}{g_k(X)} > L\right)$ are assumed to be positive for all $\alpha_1,\ldots,\alpha_\ell \geq 0$, $\alpha_1+\ldots+\alpha_\ell = 1$.

Remark 2: If, say,

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

then condition (2.2) can be omitted.

Indeed in this case

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

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

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

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

then

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

$$\leq \inf_{s\geq 0} e^{-sL} E f_2^{s}(X) g_1^{-as}(X) g_2^{-bt}(X) = 0 ,$$

and the assertion of the Lemma holds true.

Notice also that under the conditons of the Lemma

$$Pr^{1/n}(w_1 \prod_1^n f_1(x_j)+w_2 \prod_1^n f_2(x_j)\geq c_n[w_1 \prod_1^n g_1(x_j)+w_2 \prod_1^n g_2(x_j)])$$

$$\sim Pr^{1/n}(w_1 \prod_1^n f_1(x_j)+w_2 \prod_1^n f_2(x_j)>c_n[w_1 \prod_1^n g_1(x_j)+w_2 \prod_1^n g_2(x_j)]) \ .$$

Now let $\mathcal{P}_k = \{P_\theta^{(k)}, \theta \in \Theta\}, P_\theta^{(k)} \neq P_\eta^{(k)}$ for $\theta \neq \eta$, $k=1,\ldots,\ell$ be ℓ parametric families given on $\mathcal{X}$. Also let $w_1,\ldots,w_\ell$ be positive probabilities, $w_1+\ldots+w_\ell = 1$ and assume all measures $P_\theta^{(k)}$ to be equivalent. The next result gives an asymptotic formula for the probability of incorrect decision for minimax procedure δ^* based on an observation $\underline{x} = (x_1,\ldots,x_n)$ with a density $\sum_k w_k \prod_1^n p_k(x_j,\theta)$, where $p_k(\cdot,\theta)$ denotes the density of $P_\theta^{(k)}$.

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

$$\lim_{n\to\infty} \max_\theta [\sum_k w_k P_\theta^{(k)}(\delta^*(\underline{x})\neq\theta)]^{1/n}$$

$$= \lim_{n\to\infty} \max_{\theta,k} [P_\theta^{(k)}(\delta^*(\underline{x})\neq\theta)]^{1/n}$$

$$= \max_{1\leq i,k\leq\ell} \ \max_{\theta\neq\eta} \ \inf_{s_1,\ldots,s_\ell\geq 0} E_\theta^{(k)}[p_i(X,\eta)]^{\sum_1^\ell s_r} \prod_{r=1}^{\ell} p_r^{-s_r}(X,\theta), \qquad (2.7)$$

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 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

$$P_\theta^{(k)}(\hat{\delta}(\underline{x})\neq\theta)$$

$$= P_\theta^{(k)}\Big(w_1 \prod_1^n p_1(x_j,\eta)+w_2 \prod_1^n p_2(x_j,\eta)>w_1 \prod_1^n p_1(x_j,\theta)+w_2 \prod_1^n p_2(x_j,\theta) \text{ for some } \eta \neq \theta\Big)$$

$$\leq \sum_{\eta:\eta\neq\theta} P_\theta^{(k)}\Big(w_1 \prod_1^n p_1(x_j,\eta)+w_2 \prod_1^n p_2(x_j,\eta)>w_1 \prod_1^n p_1(x_j,\theta)+w_2 \prod_1^n p_2(x_j,\theta)\Big)$$

$$\leq (m-1) \max_{\eta:\eta\neq\theta} P_\theta^{(k)}\Big(w_1 \prod_1^n p_1(x_j,\eta)+w_2 \prod_1^n p_2(x_j,\eta)>w_1 \prod_1^n p_1(x_j,\theta)+w_2 \prod_1^n p_2(x_j,\theta)\Big).$$

Also

$$P_\theta^{(k)}(\hat{\delta}(\underline{x})\neq\theta)$$

$$\geq \max_{\eta:\eta\neq\theta} P_\theta^{(k)}\Big(w_1 \prod_1^n p_1(x_j,\eta)+w_2 \prod_1^n p_2(x_j,\eta)>w_1 \prod_1^n p_1(x_j,\theta)+w_2 \prod_1^n p_2(x_j,\theta)\Big).$$

Thus because of our Lemma

$$\lim_{n\to\infty}[P_\theta^{(k)}(\hat{\delta}(\underline{x})\neq\theta)]^{1/n}$$

$$= \rho_k(\theta)$$

$$= \max_{\eta:\eta\neq\theta} \max_{i=1,2} \inf_{s,t\geq 0} E_\theta^{(k)} p_i^{s+t}(X,\eta)p_1^{-s}(X,\theta)p_2^{-t}(X,\theta). \tag{2.8}$$

Notice that conditions (2.1) and (2.2) are satisfied since

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

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

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

It follows from (2.8) that

$$\lim_{n\to\infty}[w_1P_\theta^{(1)}(\hat{\delta}(\underline{x})\neq\theta)+w_2P_\theta^{(2)}(\hat{\delta}(\underline{x})\neq\theta)]^{1/n} = \max(\rho_1(\theta),\rho_2(\theta)).$$

Thus if δ^* is a minimax procedure then

$$\lim\sup \max_\theta[w_1P_\theta^{(1)}(\delta^*(\underline{x})\neq\theta)+w_2P_\theta^{(2)}(\delta^*(\underline{x})\neq\theta)]^{1/n} \leq \max_{k=1,2} \max_\theta \rho_k(\theta). \quad (2.9)$$

Assume now for concreteness sake 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 $\{\xi,\zeta\},\lambda(\xi) = \lambda(\zeta) = 1/2$ and let δ_B be the corresponding Bayes estimator. Then for any procedure δ

$$\max_\theta[w_1P_\theta^{(1)}(\delta(\underline{x})\neq\theta)+w_2P_\theta^{(2)}(\delta(\underline{x})\neq\theta)]$$

$$\geq 2^{-1}[w_1[P_\xi^{(1)}(\delta(\underline{x})\neq\xi)+P_\zeta^{(1)}(\delta(\underline{x})\neq\zeta)]+w_2[P_\xi^{(2)}(\delta(\underline{x})\neq\xi)+P_\zeta^{(2)}(\delta(\underline{x})\neq\zeta)]]$$

$$\geq 2^{-1}[w_1[P_\xi^{(1)}(\delta_B(\underline{x})\neq\xi)+P_\zeta^{(1)}(\delta_B(\underline{x})\neq\zeta)]+w_2[P_\xi^{(2)}(\delta_B(\underline{x})\neq\xi)+P_\zeta^{(2)}(\delta_B(\underline{x})\neq\zeta)]].$$

Using the Lemma again we see that

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

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

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

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

Also

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

and for any δ

$$\liminf_{n\to\infty}\ \max_\theta[w_1P_\theta^{(1)}(\delta(\underline{x})\neq\theta)+w_2P_\theta^{(2)}(\delta(\underline{x})\neq\theta)]^{1/n}$$

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

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

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

This inequality combined with (2.8) proves the Theorem.

Corollary 2.1. Under the assumptions of Theorem 2.1 for $k=1,\ldots,\ell$

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

Now we assume that $m=2$ and the hypothesis testing problem is considered. Let φ^* be the 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

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

where

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

$$K_i = \min(E^{Q_i}\log\frac{q_i}{p_o}(X),\ E^{Q_i}\log\frac{q_i}{p_1}(X)),\quad i=1,2, \qquad (2.12)$$

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 the most powerful test φ^* of our hypotheses is given by formula

$$\varphi^*(\underline{x}) = \begin{cases} 1 & w_1 \prod_1^n q_1(x_j)+w_2 \prod_1^n q_2(x_j)>c_n[w_1 \prod_1^n p_1(x_j)+w_2 \prod_1^n p_2(x_j)] \\ \gamma_n & \text{-"-} \qquad = \qquad \text{-"-} \\ 0 & \text{-"-} \qquad < \qquad \text{-"-} \end{cases}$$

with some constants $c_n > 0$, $0 \le \gamma_n \le 1$. Thus

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

and

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

It follows that

$$\sum_{i=1,2} w_i Q_i(w_1 \prod_1^n q_1(x_j)+w_2 \prod_1^n q_2(x_j)>c_n[w_1 \prod_1^n p_1(x_j)+w_2 \prod_1^n p_2(x_j)])$$

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

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

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

$$\ge Q_i([w_1 \prod_1^n q_1(x_j)+w_2 \prod_1^n q_2(x_j)]\ge c_n[w_1 \prod_1^n p_1(x_j)+w_2 \prod_1^n p_2(x_j)])$$

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

so that

$$\sum_{i=1,2} w_i[Q_i(\prod_1^n q_1(x_j)\geq 2c_n \prod_1^n p_1(x_j), \prod_1^n q_1(x_j)\geq 2c_n w_1^{-1} w_2 \prod_1^n p_2(x_j))$$

$$+ Q_i(\prod_1^n q_2(x_j)\geq 2c_n w_1 w_2^{-1} \prod_1^n p_1(x_j), \prod_1^n q_2(x_j)\geq 2c_n \prod_1^n p_2(x_j))]$$

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

$$+ Q_i(2\prod_1^n q_2(x_j)\geq c_n w_1 w_2^{-1} \prod_1^n p_1(x_j), 2\prod_1^n q_2(x_j)\geq c_n \prod_1^n p_2(x_j))] .$$

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

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

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

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

If $E^{Q_1} \log \frac{p_2}{p_1} \leq 0$, then

$$E^{Q_1} \log \frac{q_2}{p_1} \leq E^{Q_1} \log \frac{q_2}{p_2}$$

and

$$E^{Q_1} \log \frac{q_1}{p_1} \leq E^{Q_1} \log \frac{q_1}{p_2}.$$

But

$$E^{Q_1} \log \frac{q_2}{p_1} \leq E^{Q_1} \log \frac{q_1}{p_1},$$

so that in this situation (2.14) is true.

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

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

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

It follows from (2.13) that

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

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

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

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

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

$$\leq \alpha^* \leq \sum_{i=1,2} w_i P_i\Big(w_1 \prod_1^n q_1(x_j)+w_2 \prod_1^n q_2(x_j)\geq c_n\Big[w_1 \prod_1^n p_1(x_j)+w_2 \prod_1^n p_2(x_j)\Big]\Big) .$$

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

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

one deduces

$$P_i\Big(a \log \frac{q_i}{p_1} + b \log \frac{q_i}{p_2} > K_i\Big) > 0.$$

If for $k \neq i$ and all $a, b \geq 0$, $a+b = 1$,

$$P_i\Big(a \log \frac{q_k}{p_1} + b \log \frac{q_k}{p_2} > K_i\Big) > 0,$$

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

$$P_i^{1/n}\Big(w_1 \prod_1^n q_1(x_j)+w_2 \prod_1^n q_2(x_j)\geq c_n\Big[w_1 \prod_1^n p_1(x_j)+w_2 \prod_1^n p_2(x_j)\Big]\Big)$$

$$\to \max_{k=1,2} \inf_{s,t\geq 0} e^{-(s+t)K} E^{P_i} q_k^{s+t}(X) p_1^{-s}(X) p_2^{-t}(X).$$

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

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

then for all sufficiently large n

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

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

Remark 2 shows that (2.11) holds in this case as well, and therefore Theorem 2.2 is proven.

Corollary 2.2. If $K_1 \geq K_2$, then

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

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

$$e^{-K(Q_1,P_1)} \leq \lim_{n\to\infty}[E^{P_1}\varphi^*]^{1/n} .$$

But it also follows from the proof that

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

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

$$\max_{k=1,2} \inf_{s,t \geq 0} e^{-(s+t)K_i} E^{P_i} q_k^{s+t}(X) p_1^{-s}(X) p_2^{-t}(X) = \lim_{n\to\infty}(E^{P_i}\varphi^*)^{1/n} \geq e^{-K(Q_i,P_i)} .$$

3. NECESSARY AND SUFFICIENT CONDITIONS FOR THE EXISTENCE OF ADAPTIVE PROCEDURES

Let $\mathcal{P}_k = \{P_\theta^{(k)}, \theta \in \Theta\}$, $k=1,\ldots,\ell$ be ℓ families given over the same space $\mathcal{X}$ and indexed by a finite parameter θ. An estimator $\delta(\underline{x})$ based on a random sample $\underline{x} = (x_1,\ldots,x_n)$ is said to be adaptive for these families if for all $k=1,\ldots,\ell$

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

Here $\rho(\mathcal{P})$ is defined by formula (1.1) and (3.1) means that $\delta(\underline{x})$ is asymptotically minimax with respect to all families $\mathcal{P}_k$.

Theorem 3.1. If an adaptive estimator exists for families $\mathcal{P}_k = \{P_\theta^{(k)}, \theta \in \Theta\}$ with pairwise equivalent distribtuions then

$$\max_{1 \le k \le \ell} \rho(\mathcal{P}_k)$$

$$\ge \max_{1 \le i \ne k \le \ell} \max_{\theta \ne \eta} \inf_{s_1,\ldots,s_\ell \ge 0} E_\theta^{(k)} [p_i(X,\eta)]^{\Sigma s_r} \prod_{r=1}^{\ell} p_r^{-s_r}(X,\theta), \tag{3.2}$$

where $p_i(x,\theta)$ denotes the density of $P_\theta^{(i)}$. If for all $k=1,\ldots,\ell$

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

then an adaptive estimator exists.

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

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

Theorem 2.1 and formula (3.1) imply that

$$\max_{1\le k\le \ell} \rho(\mathcal{P}_k)$$

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

But

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

$$\le \max_{\theta\ne\eta} \inf_{s\ge 0} E_\theta^{(k)} p_k^s(X,\eta) p_k^{-s}(X,\theta) = \rho(\mathcal{P}_k) , \tag{3.5}$$

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

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

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

But because of (3.5) the latter relation implies that

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

so that δ^* is adaptive.

Corollary 3.1. If an adaptive estimator exists then (3.4) is actually an equality (this 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.

Indeed in this case

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

since $E_\theta^{(1)} p_1^t(X,\theta) p_2^{-t}(X,\theta)$ 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 $\sum_{k=1}^{\ell} w_k \prod_1^n p_k(x_j,\theta)$. Therefore under condition (3.3) both estimators are adaptive.

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

$$\max_\theta [P_\theta^{(i)}(\tilde{\delta}(\underline{x}) \neq \theta)]^{1/n} \sim \max_{\theta \neq \eta} [P_\theta^{(i)}(\max_k \prod_1^n p_k(x_j,\eta) > \max_k \prod_1^n p_k(x_j,\theta))]^{1/n}$$

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

$$= \max_\theta \rho_i(\theta) ,$$

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

Instead of δ_B in practice one would prefer to use a more reasonable prior distribution resulting, for instance, in a Bayes estimator δ_0 which has asymptotically constant risk under each of the parametric families $\mathcal{P}_k$, $k=1,\ldots,\ell$, i.e.

$$\lim_{n\to\infty} [P_\theta^{(k)}(\delta_0(\underline{x}) \neq \theta)]^{1/n} = \max_\theta \rho_k(\theta) . \quad (3.6)$$

These prior probabilities can be found in the following way.

Let

$$\lambda_\theta = \exp\{-nu_\theta\}\Big(\sum_{\eta \in \Theta} e^{-nu_\eta}\Big)^{-1} .$$

Then because of the Lemma

$$[P_\theta^{(k)}(\delta_o(\underline{x})\neq\theta)]^{1/n} \sim \max_{\eta:\eta\neq\theta} [P_\theta^{(k)}(\Sigma w_k \prod_1^n p_k(x_j,\eta)\lambda_\eta > \Sigma w_k \prod_1^n p_k(x_j,\theta)\lambda_\theta)]^{1/n}$$

$$\sim \max_{1\leq i\leq \ell} \max_{\eta:\eta\neq\theta} [P_\theta^{(k)}(\sum_1^n \log \frac{p_i(x_j,\eta)}{p_r(x_j,\theta)} > n(u_\eta - u_\theta),\ r=1,\ldots,\ell)]^{1/n}$$

$$\to \max_{1\leq i\leq \ell} \max_{\eta:\eta\neq\theta} \inf_{s_1,\ldots,s_\ell\geq 0} e^{\Sigma s_r(u_\theta-u_\eta)} E^{(k)} p_k^{\Sigma s_r}(X,\eta) \prod_1^\ell p_r^{-s_r}(X,\theta).$$

We show that one can find numbers u_θ, $\theta \in \Theta$ such that the latter quantity equals to $\max_\theta \rho_k(\theta)$.

Let $z_{\theta\eta}$ for $\theta \neq \eta$ be reals such that

$$\max_{i,k} \inf_{s_1,\ldots,s_\ell\geq 0} e^{\Sigma s_r z_{\theta\eta}} E_\theta^{(k)} p_i^{\Sigma s_r}(X,\eta) \prod_1^\ell p_r^{-s_r}(X,\theta) = \max_{i,\theta} \rho_i(\theta). \qquad (3.7)$$

Clearly all $z_{\theta\eta}$ are nonnegative and because of (3.6) one has for all θ

$$\max_{i,k} \inf_{s_1,\ldots,s_\ell\geq 0} e^{\Sigma s_r(u_\theta-u_\eta)} E_\theta^{(k)} p_i^{\Sigma s_r}(X,\eta) \prod_1^\ell p_r^{-s_r}(X,\theta)$$

$$\leq \max_{i,\theta} \rho_i(\theta)$$

$$= \max_{i,k} \inf_{s_1,\ldots,s_\ell\geq 0} e^{\Sigma s_r z_{\theta\eta}} E_\theta^{(k)} p_i^{\Sigma s_r}(X,\eta) \prod_1^\ell p_r^{-s_r}(X,\theta) .$$

Thus for all θ, η

$$u_\theta - u_\eta \le z_{\theta\eta}$$

and for each θ there exists η, $\eta \ne \theta$ with the property

$$u_\theta - u_\eta = z_{\theta\eta} .$$

These formulae mean that

$$u_\theta = \min_{\eta:\eta\ne\theta} [u_\eta + z_{\theta\eta}] . \tag{3.8}$$

These simultaneous equations can be solved in the following way.

Put $u_0 = 0$. We construct a solution of (3.8) in such way that $0 \le u_0 \le u_1 \le \cdots \le u_m$. Then (3.8) reduces to a recursive formula

$$u_i = \min_{k<i} [u_k + z_{ik}] .$$

To make such representation possible we have to assume only that for $i=1,\ldots,m$

$$\min_{k>i} z_{ik} > \min_{k<i} z_{ik} ,$$

which can be achieved by reparametrization of the elements of Θ . Clearly the resulting estimator $\delta_0(\underline{x})$ will be asymptotically minimax.

We summarize our results as follows.

Theorem 3.2. If condition (3.3) is satisfied then the following estimators are adaptive:

(i) maximum likelihood estimator $\hat{\delta}(\underline{x})$

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

(ii) overall maximum likelihood estimator

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

(iii) Bayes estimator $\delta_0(\underline{x})$ with asymptotically constant risk,

$$\delta_0(\underline{x}) = \eta \text{ iff } \sum_k w_k \prod_1^n p_k(x_j,\eta)\lambda_\eta = \max_\theta \sum_k w_k \prod_1^n p_k(x_j,\theta)\lambda_\theta ,$$

where $\lambda_\theta = e^{-nu_\theta}(\sum_\eta e^{-nu_\eta})^{-1}$, and constants u_θ satisfy equations (3.8) with coefficients $z_{\theta\eta}$ defined by (3.7).

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

$$\max(e^{-K(Q_1,P_1)}, e^{-K(Q_2,P_2)}) \geq \max_{i \neq k} \inf_{s,t \geq 0} e^{-(s+t)K} E^{P_i} q_k^{s+t}(X) p_1^{-s}(X) p_2^{-t}(X), \tag{3.9}$$

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

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

$$\max_{k:k \neq i} \inf_{s,t \geq 0} e^{-(s+t)K} E^{P_i} q_k^{s+t}(X) p_1^{-s}(X) p_2^{-t}(X) \leq e^{-K(Q_i,P_i)}, \tag{3.10}$$

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

$$w_1 E^{Q_1}\varphi + w_2 E^{Q_2}\varphi = \beta ,$$

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

$$\max(e^{-K(Q_1,P_1)}, e^{-K(Q_2,P_2)}) = \max(\lim_{n\to\infty}(E^{P_1}\varphi)^{1/n}, \lim_{n\to\infty}(E^{P_2}\varphi)^{1/n})$$

$$= \lim_{n\to\infty}(w_1 E^{P_1}\varphi + w_2 E^{P_2}\varphi)^{1/n}$$

$$\geq \lim_{n\to\infty}(w_1 E^{P_1}\varphi^* + w_2 E^{P_2}\varphi^*)^{1/n}$$

$$= \max_{i,k} \inf_{s,t\geq 0} e^{-(s+t)K} E^{P_i} q_k^{s+t}(X)p_1^{-s}(X)p_2^{-t}(X) .$$

(3.11)

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

Also for i=1,2

$$\inf_{s,t\geq 0} e^{-(s+t)K} E^{P_i} q_i^{s+t}(X)p_1^{-s}(X)p_2^{-t}(X)$$

$$\leq \inf_{s,t\geq 0} e^{-(s+t)K_i} E^{P_i} q_i^{s+t}(X)p_1^{-s}(X)p_2^{-t}(X)$$

$$\leq \inf_{s\geq 0} e^{-sK_i} E^{P_i} q_i^{s}(X)p_i^{-s}(X) = e^{-K(Q_i,P_i)} . \qquad (3.12)$$

Therefore (3.11) implies (3.9).

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

$$\varphi_i(\underline{x}) = \begin{cases} 1 & [w_1 \prod_1^n q_1(x_j)+w_2 \prod_1^n q_2(x_j)]>c_n^{(i)}[w_1 \prod_1^n p_1(x_j)+w_2 \prod_1^n p_2(x_j)] \\ \gamma_n^{(i)} & \quad - " - \qquad = \qquad - " - \\ 0 & \quad - " - \qquad < \qquad - " - \end{cases}$$

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

$$E^{Q_i}\varphi_i(\underline{x}) = \beta.$$

It follows from the proof of Theorem 2.2 that $\lim n^{-1} \log c_n^{(i)} = K$. Define a new test φ: $\varphi = \varphi_1$ if $c_n^{(1)} \leq c_n^{(2)}$ and $\varphi = \varphi_2$ otherwise.

Then

$$E^{Q_i}\varphi(\underline{x}) \geq \beta$$

and

$$\lim_{n\to\infty}[E^{P_i}\varphi(\underline{x})]^{1/n} = \max_{1\leq i,k\leq 2}\ \inf_{s,t\geq 0}\ e^{-(s+t)K}E^{P_i}q_k^{s+t}(X)p_1^{-s}(X)p_2^{-t}(X)\ .$$

Because of (3.12) the latter relation implies that

$$\lim_{n\to\infty}[E^{P_i}\varphi(\underline{x})]^{1/n} \leq e^{-K(Q_i,P_i)},$$

which proves the Theorem.

Corollary 3.2. If an adaptive test exists then (3.11) is actually an equality.

4. EXAMPLES

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

1°. One-parameter exponential families.

Let meausures $P_\theta^{(k)}$ be defined over an abstract space $\mathcal{X}$ and let their densities with respect to some σ-finite measure μ be of the form

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

$\alpha_k(\theta) \neq \alpha_i(\theta)$ for $k\neq i$. Here $C(\alpha) = \int_{\mathcal{X}} e^{\alpha v(x)}d\mu(x)$ and α belongs to the natural parameter space, which is, of course, an interval with end-points 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.

Now one has for $k=1,\dots,\ell$

$$\rho(\mathcal{P}_k) = \inf_{0<s<1} \int p_k^{1-s}(x,0)p_k^{s}(x,1)d\mu(x)$$

$$= \exp\{\inf_{0<s<1} [f(\alpha_k(1)s+\alpha_k(0)(1-s))-sf(\alpha_k(1))-(1-s)f(\alpha_k(0))]\}$$

$$= H(\alpha_i(0),\alpha_i(1)). \tag{4.1}$$

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

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

$$= \exp\{\inf_{s,t\geq 0} [f(\alpha_2(\eta)(s+t)+\alpha_1(\theta)(1-s)-\alpha_2(\theta)t)-(s+t)f(\alpha_2(\eta))$$

$$- (1-s)f(\alpha_1(\theta))+tf(\alpha_2(\theta))]\} \tag{4.2}$$

and

$$\inf_{s,t\geq 0} E_\theta^{(2)} p_1^{s+t}(X,\eta)p_1^{-s}(X,\theta)p_2^{-t}(X,\theta)$$

$$= \exp\{\inf_{s,t\geq 0} [f(\alpha_1(\eta)(s+t)+\alpha_2(\theta)(1-t)-\alpha_1(\theta)s)-(s+t)f(\alpha_1(\eta))$$

$$- (1-t)f(\alpha_2(\theta))+sf(\alpha_1(\theta))]\} \tag{4.3}$$

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

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

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

$$\inf_{s,t\geq 0} [f(\alpha_2(\eta)(s+t)+\alpha_1(\theta)(1-s)-\alpha_2(\theta)t)-(s+t)f(\alpha_2(\eta))-(1-s)f(\alpha_1(\theta))+tf(\alpha_2(\theta))]$$

$$= \min\{\inf_{s\geq 0} [f(\alpha_2(\eta)s+\alpha_1(\theta)(1-s))-sf(\alpha_2(\eta))-(1-s)f(\alpha_1(\theta))$$

$$\inf_{t\geq 0} [f(\alpha_1(\theta)+(\alpha_2(\eta)-\alpha_2(\theta))t)-f(\alpha_1(\theta))-tf(\alpha_2(\eta))+tf(\alpha_2(\theta))]\} .$$

We show now that

$$\inf_{t\geq 0} [f(\alpha_1(\theta)+(\alpha_2(\eta)-\alpha_2(\theta))t)-f(\alpha_1(\theta))-tf(\alpha_2(\eta))+tf(\alpha_2(\theta))]$$

$$\geq 0 \geq \inf_{s\geq 0} [f(\alpha_2(\eta)s+\alpha_1(\theta)(1-s))-sf(\alpha_2(\eta))-(1-s)f(\alpha_1(\theta))]$$

$$= \log H(\alpha_1(\theta),\alpha_2(\eta)) . \tag{4.5}$$

If the equation

$$(\alpha_2(\eta)-\alpha_2(\theta))f'(\alpha_1(\theta)+(\alpha_2(\eta)-\alpha_2(\theta))t) = f(\alpha_2(\eta))-f(\alpha_2(\theta))$$

has a nonnegative solution $t = t_0$, then

$$f(\alpha_1(\theta)+(\alpha_2(\eta)-\alpha_2(\theta))t_0)-f(\alpha_1(\theta))-t_0f(\alpha_2(\eta))+t_0f(\alpha_2(\theta))$$

$$= f(\alpha_1(\theta)+(\alpha_2(\eta)-\alpha_2(\theta))t_0)-f(\alpha_1(\theta))$$

$$-t_0(\alpha_2(\eta)-\alpha_2(\theta))f'(\alpha_1(\theta)+(\alpha_2(\eta)-\alpha_2(\theta))t_0) \geq 0 ,$$

which proves (4.5).

Observe now that since f' is strictly increasing, $\alpha_2(\eta)>\alpha_2(\theta)$ implies

$$f'(\alpha_+) > [f(\alpha_2(\eta))-f(\alpha_2(\theta))][\alpha_2(\eta)-\alpha_2(\theta)]^{-1} ,$$

and $\alpha_2(\eta) < \alpha_2(\theta)$ implies

$$f'(\alpha_-) < [f(\alpha_2(\theta))-f(\alpha_2(\eta))][\alpha_2(\theta)-\alpha_2(\eta)]^{-1} .$$

Thus (4.4) is always valid and condition (3.3) of Theorem 3.1 can be written as

$$\min_{i=1,2} [H(\alpha_i(\theta),\alpha_i(\eta)] \geq \max_{i\neq k} H(\alpha_i(\theta),\alpha_k(\eta)) ,$$

for $\theta \neq \eta$. The condition (3.2) takes the form

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

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

Theorem 4.1. Let for $k=1,\ldots,\ell$

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

be densities of one-parameter exponential family, $\alpha_i(\theta) \neq \alpha_k(\theta)$ for $i\neq k$. Assume that the common support of p_k, $k=1,\ldots,\ell$ 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 $k=1,\ldots,\ell$

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

then an adaptive estimator exists.

Theorem 4.1 contains many interesting particular cases.

(i) Densities of the form

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

or of the form

$$p_k(x,\theta) = C[\alpha(\theta)]^a \exp\{-\alpha(\theta)x^{a-1}\}, \quad x \geq 0 .$$

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

$$H(\alpha_1,\alpha_2) = \inf_{s\geq 0} \alpha_2^{as}\alpha_1^{a(1-s)}[\alpha_2 s+\alpha_1(1-s)]^{-a} = [\inf_{s\geq 0} \frac{\beta^s}{1+s(\beta-1)}]^a$$

$$= [\frac{\log \beta}{\beta-1} \exp(1-\frac{\log \beta}{\beta-1})]^a = h^a(\beta) ,$$

where $\beta = \alpha_2\alpha_1^{-1}$. It is easy to check that $h(\beta^{-1}) = h(\beta)$ and that $h(\beta)$ is a unimodal function which attains maximum at $\beta = 1$ and is increasing for $0 < \beta < 1$.

Therefore inequalities (4.7) in this case mean that

$$1 < \max_{\theta\neq\eta} \max[\alpha_k(\theta)/\alpha_k(\eta),\alpha_k(\eta)/\alpha_k(\theta)]$$

$$\leq \max_{i:i\neq k} \max_{\theta\neq\eta} \max[\alpha_i(\theta)/\alpha_k(\eta),\alpha_k(\eta)/\alpha_i(\theta)] ,$$

in which situation an adaptive estimator exists.

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

$$\max_{1\leq k\leq \ell} \max_{\theta\neq\eta} \max[\alpha_k(\theta)/\alpha_k(\eta),\alpha_k(\eta)/\alpha_k(\theta)]$$

$$> \max_{1\leq i\neq k\leq \ell} \max_{\theta\neq\eta} \max[\alpha_i(\theta)/\alpha_k(\eta),\alpha_k(\eta)/\alpha_i(\theta)].$$

The heuristic interpretation of this condition 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 $P_\eta^{(k)}$.

(ii) Poisson distribution,

$$p_k(x,\theta) = e^{-\lambda(\theta)} \frac{[\lambda(\theta)]^x}{x!} , \quad x=0,1,\ldots$$

In this case $\alpha(\theta) = \log \lambda(\theta)$, $C(\alpha) = \exp\{e^{\alpha}\}$

$$H(\alpha_1,\alpha_2) = \log \inf_{s\geq 0}[e^{e^{\alpha_1 s+\alpha_2(1-s)}} - se^{\alpha_1} - (1-s)e^{\alpha_2}]$$

$$= \log[(\alpha_1-\alpha_2)^{-1}((e^{\alpha_1}-e^{\alpha_2})(1-\log((e^{\alpha_1}-e^{\alpha_2})/(\alpha_1-\alpha_2)))+\alpha_2 e^{\alpha_1}-\alpha_1 e^{\alpha_2})].$$

Theorem 4.1 gives a necessary and a sufficient condition for the existence of adaptive estimators in this situation.

(iii) Binomial distribution,

$$p_k(x,\theta) = \binom{N}{x}[p_k(\theta)]^x[1-p_k(\theta)]^{n-x}, \quad x=0,1,\dots,N ,$$

$\alpha(\theta) = \log p(\theta)[\log(1-p(\theta))]^{-1}$. 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:

$$H(p_k(0),p_k(1))$$

$$= \inf_{s\geq 0} \sum_{x=0}^{N} \binom{N}{x}[p_k(0)]^{sx}[q_k(0)]^{s(N-x)}[p_k(1)]^{(1-s)x}[q_k(1)]^{(1-s)(N-x)}$$

$$= \inf_{s\geq 0}\{[p_k(0)]^s[p_k(1)]^{1-s}+[q_k(0)]^s[q_k(1)]^{1-s}\}^N = \rho(\mathcal{P}_k)$$

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

For a fixed k let

$$H_k(\gamma) = \inf_{s\geq 0}[p_k^{1-s}(1)\gamma^s+q_k^{1-s}(1)(1-\gamma)^s]$$

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

$$H(p_k(0),p_k(1)) \geq \max_{i:i\neq k} H(p_i(0),p_k(1)) ,$$

which is equivalent to (3.3), means that

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

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

$$\max_k H_k(p_k(0)) \geq \max_{1\leq i\neq k\leq \ell} H_k(p_i(0)) .$$

2°. Location parameter families on a cyclic group.

Assume that $\mathcal{X} = \Theta = \{0,1\}$, $p_k(x,\theta) = p_k(x-\theta)$, $k=1,\ldots,\ell$, where difference $x-\theta$ is understood modulo two. Thus $p_k(0)+p_k(1) = 1$ and

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

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

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

$$= \inf_{s,t\geq 0} [p_2^{s+t}(\eta)p_1^{1-s}(\theta)p_2^{-t}(\theta)+p_2^{s+t}(\theta)p_1^{1-s}(\eta)p_2^{-t}(\eta)]$$

$$= \min\{ \inf_{s\in A} [p_1^{1-s}(\theta)p_2^s(\eta)+p_1^{1-s}(\eta)p_2^s(\theta)] ,$$

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

where A is a subset of the positive halfline where

$p_1^{1-s}(\theta)p_2^s(\eta) > p_1^{1-s}(\eta)p_2^s(\theta)$,and B is its complement.

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

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

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 interval $[0,1]$ and

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

Let for $0 < p < 1$

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

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

$$\inf_{s\in A} [p_1^{1-s}(\theta)p_2^s(\eta)+p_1^{1-s}(\eta)_2^s(\theta)] \le \rho(\wp_k)$$

means that

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

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

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

$$= \min\{\inf_{s\in B} [p_1^{1-s}(\theta)p_2^s(\eta)+p_1^{1-s}(\eta)p_2^s(\theta)] ,$$

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

The latter quantity is less than $\rho(\wp_1)$ 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

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

Thus

$$\rho(\wp_1) < \max_{\theta\ne\eta} \inf_{s,t\ge 0} E_\theta^{(1)}p_2^{s+t}(X,\eta)p_1^{-s}(X,\theta)p_2^{-t}(X,\theta) ,$$

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 $|p_2(0)-1/2| > |p_1(1)-1/2|$ (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 $|p_2(1)-1/2| > |p_1(1)-1/2|$ imply that $H(p_2(1)) > H_1(p_1(1))$.

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

Acknowledgement. The author is grateful to Professor Herman Rubin for many helpful suggestions.

REFERENCES

[1] Bahadur, R. R. (1971). Some limit theorems in statistics. Regional Conference Series in Applied Mathematics. Society for Industrial and Applied Mathematics, Philadelphia.

[2] Bahadur, R. R. and Zabell, S. L. (1979). Large deviations of the sample mean in general vector spaces. Ann. Probability 7, 587-621.

[3] Bartfai, P. (1978). Large deviations of the sample mean in Euclidean spaces. Mimeograph Series No. 78-13, Statist. Dept., Purdue University.

[4] Beran, R. (1974). Asymptotically efficient adaptive rank estimates in location models. Ann. Statist. 2, 63-74.

[5] 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.

[6] Chernoff, H. (1956). Large sample theory-parametric case. Ann. Math. Statist. 27, 1-22.

[7] Ghosh, J. K. and Subramanyam, K. (1975). Inference about separated families in large samples. Sankhyā, 37, 502-513.

[8] Groeneboom, P., Oosterhoff, J. and Ruymgaart, F. H. (1979). Large deviation theorems for empirical probability measures, Ann. Probability, 7, 553-586.

[9] Krafft, O. and Puri, M. L. (1974). The asymptotic behavior of the minimax risk for multiple decision problems. Sankhyā, 36, 1-12.

[10] Kullback, S. (1959). Information and Statistics. Wiley, New York.

[11] 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.

[12] Sacks, J. (1975). An asymptotically efficient sequence of estimators of a location parameter. Ann. Statist. 3, 285-298.

[13] Stone, C. J. (1975). Adaptive maximum likelihood estimators of a location parameter. Ann. Statist. 3, 267-284.