ADAPTIVE PROCEDURES FOR A FINITE NUMBER OF PROBABILITY DISTRIBUTION FAMILIES

by

Andrew L. Rukhin*
Purdue University, W. Lafayette, IND.

Mimeograph Series #81-15

Department of Statistics
Division of Mathematical Sciences

May 1981

*This work was supported by the National Science Foundation grant # MCS 7802300.

ABSTRACT

The estimation problem of a finite-valued parameter on the basis of a random sample of increasing size is considered. We derive a necessary and sufficient condition for the existence of an estimator asymptotically fully efficient (adaptive) for several distributions families. An example of one-parameter exponential family is considered.

AMS 1980 subject classifications: Primary 62F10, Secondary 62F12, 62B10.

Key words and phrases: Probability of incorrect decision, asymptotic efficiency, adaptive procedures, maximum likelihood estimator, information divergence.

ADAPTIVE PROCEDURES FOR A FINITE NUMBER

OF PROBABILITY DISTRIBUTION FAMILIES

By Andrew L. Rukhin*

Purdue University, W. Lafayette, IND.

1. *Introduction.* Let $\mathcal{P} = (P_1,\dots,P_m)$ be a family of m different probability distributions, and let $\underline{x} = (x_1,\dots,x_n)$ be a sequence of independent random variables having common distribution P_θ for some $\theta = 1,\dots,m$. On the basis of the random sample $\underline{x}$ statistical inference about the finite-valued parameter θ is desired.

If $\delta = \delta(\underline{x})$ is an estimator of this parameter, then we shall use the probability of incorrect decision, $P_\theta(\delta\neq\theta)$, as the risk function of δ. The asymptotic behavior of this risk has been studied by Kraff and Puri [7] who showed that if δ^* is an asymptotically minimax procedure then

$$(1.1) \qquad \lim_{n\to\infty} \max_{\theta} P_\theta^{1/n}(\delta^*\neq\theta) = \max_{\eta\neq\theta} \inf_{s\geq 0} E_\theta p^s(X,\eta)p^{-s}(X,\theta)$$

$$= \max_{\eta\neq\theta} \inf_{s\geq 0} \int p^s(x,\eta)p^{1-s}(x,\theta)d\mu(x) = \rho(\mathcal{P}),$$

where $p(x,\theta)$ is the probability density of the distribution P_θ with respect to a measure μ.

Notice that the quantity $\inf_{s>0} E_\theta p^s(X,\eta)p^{-s}(X,\theta)$ represents the Chernoff's function for the likelihood ratio and gives the asymptotics for the probability $P_\theta^{1/n}(\prod_1^n p(x_j,\eta) > \prod_1^n p(x_j,\theta))$ as the sample size n tends to infinity (see Bahadur [1], Chernoff [3], [4]).

*This work was supported by the National Science Foundation grant number MCS 7802300.

Clearly $0 \leq \rho(\mathcal{P}) < 1$, since all members of $\mathcal{P}$ are distinct, and $\rho(\mathcal{P}) = 0$ if and only if all distributions P_θ are mutually singular. Thus $\rho(\mathcal{P})$ can be interpreted as an information divergence of elements of $\mathcal{P}$. See Vajda [10] for further properties of $\rho(\mathcal{P})$. Assume now that the distributions P_θ involve a nuisance parameter α which also takes a finite number of values. Thus one has families $\mathcal{P}_\alpha = \{P_\theta^{(\alpha)}, \theta=1,\ldots,m\}, \alpha = 1, \ldots,\ell$. If δ is an estimator of the parameter θ, then we denote

$$(1.2) \qquad R(\alpha,\delta) = \lim_{n\to\infty} \inf[\max_\theta P_\theta^{(\alpha)}(\delta\neq\theta)]^{1/n}.$$

Then for any δ

$$R(\alpha,\delta) \geq \rho(\mathcal{P}_\alpha) = \rho_\alpha,$$

and a procedure δ_a is called *adaptive* if $R(\alpha,\delta_a) = \rho_\alpha$ for all α.

In other terms an adaptive estimator is asymptotically fully efficient under any of the families $\mathcal{P}_\alpha$, $\alpha = 1,\ldots,\ell$. In this paper we obtain a necessary and sufficient condition for the existence of an adaptive procedure. Roughly speaking, an adaptive stimator exists if and only if the members of different families $\mathcal{P}_\alpha$ and $\mathcal{P}_\beta$, $\alpha \neq \beta$, are not more similar than the elements of one of these families.

A result similar to (1.1) holds as well if δ^* is the Bayes estimator with respect to positive prior probabilities u_θ, $\theta = 1,\ldots,m$, and $\max_\theta P_\theta(\delta^*\neq\theta)$ is replaced by the Bayes risk $\sum_\theta u_\theta P_\theta(\delta^*\neq\theta)$. Since

$$\lim_{n\to\infty} \inf[\sum_\theta u_\theta P_\theta^{(\alpha)}(\delta \neq \theta)]^{1/n} = R(\alpha,\delta),$$

the results of this paper are true if in the definition of an adaptive procedure maximum of the risk is replaced by the Bayes risk. Moreover, one can also substitute zero-one loss by a more general loss function $W(\theta,d)$

such that $W(\theta,\theta) = 0$ and $W(\theta,d) > 0$ for $\theta \neq d$. (See Ghosh and Subramanyam [5].)

The existence of adaptive procedures is related to a more general problem of the form of minimax estimators for the new risk function $R(\alpha,\delta)/\rho_\alpha$. It is easy to see that δ_a, if exists, is minimax for this risk. We determine a minimax estimator in the general situation, i.e., when an adaptive procedure may not exist. We also evaluate the quantity $v = \inf_\delta \max_\alpha R(\alpha,\delta)/\rho_\alpha$, which represents the value of the corresponding game.

2. *The Asymptotical Behavior of Minimax Estimators.* In this section we study the asymptotical behavior of minimax procedures based on likelihood function of the form $\max_k[e^{c_k n} \prod_1^n p_k(x_j,\theta)]$, where c_k, $k = 1,\ldots,\ell$ are given constants and $p_k(x,\theta)$ is the density of $P_\theta^{(k)}$. We start with the following basic result.

Lemma. Let $x_1,x_2,\ldots$ be a sequence of i.i.d. random variables and let f_k,g_k, $k = 1,\ldots,\ell$ be positive measurable functions such that for any nonnegative $v_1,\ldots,v_\ell$ and $k = 1,\ldots,\ell$

$$\Pr\{\sum_r v_r[\log(f_k(X)/g_r(X)) - c_r + c_k] > 0\} > 0. \tag{2.1}$$

Then

$$\lim_{n\to\infty} \Pr^{1/n}\{\max_k[e^{c_k n} \prod_1^n f_k(x_j)] \geq \max_k[e^{c_k n} \prod_1^n g_k(x_j)]\}$$

$$= \lim_{n\to\infty} \Pr^{1/n}\{\max_k[e^{c_k n} \prod_1^n f_k(x_j)] > \max_k[e^{c_k n} \prod_1^n g_k(x_j)]\}$$

$$= \max_{1\leq k\leq \ell} \inf_{s_1,\ldots,s_\ell\geq 0} \exp\{\sum_r s_r(c_k - c_r)\} E f_k^{\Sigma s_r}(X) \prod_r g_r^{-s_r}(X).$$

Proof. For any fixed r, r = 1,...,ℓ

$$\Pr\{e^{c_r n}\prod_1^n f_r(x_j) \geq \max_k[e^{c_k n}\prod_1^n g_k(x_j)]\}$$

$$\leq \Pr\{\max_k[e^{c_k n}\prod_1^n f_k(x_j)] \geq \max_k[e^{c_k n}\prod_1^n g_k(x_j)]\}$$

$$\leq \sum_i \Pr\{e^{c_i n}\prod_1^n f_i(x_j) \geq \max_k[e^{c_k n}\prod_1^n g_k(x_j)]\}$$

$$\leq \ell \max_k \Pr\{e^{c_i n}\prod_1^n f_i(x_j) \geq \max_k[e^{c_k n}\prod_1^n g_k(x_j)]\}.$$

It follows that

$$\Pr^{1/n}\{\max_k[e^{c_k n}\prod_1^n f_k(x_j)] \geq \max_k[e^{c_k n}\prod_1^n g_k(x_j)]\}$$

$$\sim \max_k \Pr^{1/n}\{e^{c_k n}\prod_1^n f_k(x_j) \geq \max_i[e^{c_i n}\prod_1^n g_k(x_j)]\}$$

$$= \max_k \Pr^{1/n}\{e^{c_k n}\prod_1^n f_k(x_j) \geq e^{c_i n}\prod_1^n g_i(x_j),\ i=1,\ldots,\ell\}$$

$$= \max_k \Pr^{1/n}\{n^{-1}\prod_1^n \log(f_k(x_j)/g_i(x_j)) \geq c_i - c_k,\ i=1,\ldots,\ell\}.$$

The conclusion of Lemma now results from the multivariate version of Chernoff's Theorem (see Bartfai [2], Groeneboom, Oosterhoff and Ruymgaart [6] or Steinebach [9]).

The following quantities play a crucial and unheralded role in deciding the existence of adaptive procedures. Define for real $c_1,\ldots,c_\ell$, $1 \leq i, k \leq \ell$

$$\rho_{ik}(c_1,\ldots,c_\ell) = \max_{\theta\neq\eta}\ \inf_{s_1,\ldots,s_\ell\geq 0} \exp\{\sum_r s_r(c_i-c_r)\}E_\theta^{(k)} p_i^{\Sigma s_r}(X,\eta)\prod_r p_r^{-s_r}(X,\theta).$$

Notice that for $\ell = 1$, $\rho_{ij} = \rho(P)$. As we shall see, the quantities ρ_{ij} in general case preserve the interpretation of information divergence of families P_i and P_k in the configuration $\{P_r, r=1,\ldots,\ell\}$.

In the definition of ρ_{ik} we assume that all densities $p_r(x,\theta)$, $r = 1,\ldots,\ell$ are strictly positive. This condition is supposed to hold throughout this paper. Under this agreement ρ_{ik} is a continuous function of $c_1,\ldots,c_\ell$ on the set where it is finite. All these functions are translation invariant:

$$\rho_{ik}(c_1 + c,\ldots,c_\ell + c) = \rho_{ik}(c_1,\ldots,c_\ell);$$

and

$$\rho_{ik}(c_1,\ldots,c_\ell) \le \min\{1,\exp\{c_i - c_k\}\}.$$

THEOREM 1. <u>Let</u> δ^* <u>be an asymptotically minimax estimator of</u> θ <u>based on the likelihood function</u> $\max_i[e^{c_i n}\prod_1^n p_i(x_j,\theta)]$. <u>Then</u>

$$\lim_{n\to\infty} \max_k[e^{c_k n}\max_\theta P_\theta^{(k)}(\delta^* \ne \theta)]^{1/n} = \max_{1\le i,k\le \ell} e^{c_k}\rho_{ik}(c_1,\ldots,c_\ell).$$

<u>Proof.</u> Let $\hat\delta$ be the maximum likelihood estimator of θ based on $\max_i[e^{c_i n}\prod_1^n p_i(x_j,\theta)] = \pi_\theta(\underline{x},c_1,\ldots,c_\ell) = \pi_\theta(\underline{x})$. Thus $\hat\delta = \theta$ if for $\eta \ne \theta$

$$\pi_\theta(\underline{x}) > \pi_\eta(\underline{x}).$$

It is easy to see that the definition of $\hat\delta$, when this inequality is the equality for $\eta \ne \theta$, is immaterial in our asymptotical analysis. Also for any $\eta \ne \theta$

$$P_\theta^{(k)}(\pi_\theta(\underline{x}) < \pi_\eta(\underline{x})) \le P_\theta^{(k)}(\hat\delta \ne \theta)$$

$$\le (m-1)\max_{\eta:\eta\ne\theta} P_\theta^{(k)}(\pi_\theta(\underline{x}) < \pi_\eta(\underline{x})).$$

Therefore because of our Lemma

$$(2.2)\qquad \lim_{n\to\infty}[P_\theta^{(k)}(\hat\delta \neq \theta)]^{1/n}$$

$$= \lim_{n\to\infty}[\max_{\eta:\eta\neq\theta} P_\theta^{(k)}(\pi_\theta(\underline{x}) < \pi_\eta(\underline{x}))]^{1/n}$$

$$= \max_{\eta:\eta\neq\theta}\ \max_i\ \inf_{s_1,\ldots,s_\ell\geq 0} \exp\{\sum_r s_r(c_i-c_r)\}E_\theta^{(k)} p_i^{\Sigma s_r}(X,\eta)\Pi_r p_r^{-s_r}(X,\theta).$$

Notice that the condition (2.1) of the Lemma is satisfied since for all nonnegative $v_1,\ldots,v_\ell$

$$E_\eta^{(i)}(\sum_r v_r \log(p_i(X,\eta)/p_r(X,\theta)) > 0,$$

so that

$$P_\eta^{(i)}(\sum_r v_r \log(p_i(X,\eta)/p_r(X,\theta)) > 0) > 0,$$

which is equivalent to the inequality

$$P_\theta^{(k)}(\sum_r v_r \log(p_i(X,\eta)/p_r(X,\theta)) > 0) > 0.$$

If δ_m is a minimax procedure then

$$(2.3)\qquad \max_k[e^{c_k n}\max_\theta P_\theta^{(k)}(\delta_m \neq \theta)] \leq \max_\theta \int_{\{\delta_m\neq\theta\}}\!\!\!\cdots\int \pi_\theta(\underline{x})d\mu(\underline{x})$$

$$\leq \max_\theta \int_{\{\hat\delta\neq\theta\}}\!\!\!\cdots\int \pi_\theta(\underline{x})d\mu(\underline{x}) \leq \sum_k e^{c_k n}\max_\theta P_\theta^{(k)}(\hat\delta \neq \theta)$$

and

$$(2.4)\qquad \lim_{n\to\infty}\max_k[e^{c_k n}\max_\theta P_\theta^{(k)}(\delta_m\neq\theta)]^{1/n} = \lim_{n\to\infty}\max_k[e^{c_k n}\max_\theta P_\theta^{(k)}(\delta^* \neq \theta)]^{1/n}$$

$$\leq \lim_{n\to\infty}\max_k[e^{c_k n}\max_\theta P_\theta^{(k)}(\hat\delta \neq \theta)]^{1/n}$$

$$= \max_{i,k} e^{c_k} \rho_{ik}(c_1,\ldots,c_\ell).$$

We prove now that (2.4) is actually the equality, i.e., that $\hat{\delta}$ is an asymptotically minimax procedure. For a fixed k, $1 \le k \le \ell$, let ξ and ζ be two different parametric values defined in the following way:

$$\max_i \rho_{ik}(c_1,\ldots,c_\ell)$$

$$= \max_i \inf_{s_1,\ldots,s_\ell \ge 0} \exp\{\sum_r s_r(c_i - c_r)\} E_\xi^{(k)} p_i^{\Sigma s_r}(X,\zeta) \prod_r p_r^{-s_r}(X,\xi).$$

Also let δ_B be the Bayes estimator for the prior distribution assigning weights 1/2 to ξ and ζ and the likelihood function $\pi_\theta(\underline{x})$. Then for any δ

$$\max_k[e^{c_k n} P_\xi^{(k)}(\delta_B \ne \xi)] + \max_k[e^{c_k n} P_\zeta^{(k)}(\delta_B \ne \zeta)]$$

$$\le \int_{\{\delta_B \ne \xi\}}\!\!\cdots\!\int \pi_\xi(\underline{x}) d\mu(\underline{x}) + \int_{\{\delta_B \ne \zeta\}}\!\!\cdots\!\int \pi_\zeta(\underline{x}) d\mu(\underline{x})$$

$$\le \int_{\{\delta \ne \xi\}}\!\!\cdots\!\int \pi_\xi(\underline{x}) d\mu(\underline{x}) + \int_{\{\delta \ne \zeta\}}\!\!\cdots\!\int \pi_\zeta(\underline{x}) d\mu(\underline{x})$$

$$\le \sum_k e^{c_k n}[P_\xi^{(k)}(\delta \ne \xi) + P_\zeta^{(k)}(\delta \ne \zeta)].$$

Thus

$$\liminf_{n\to\infty} \max_k[e^{c_k n} \max_\theta P_\theta^{(k)}(\delta \ne \theta)]^{1/n}$$

$$\ge \lim_{n\to\infty} \max\{\max_k[e^{c_k n} P_\xi^{(k)}(\delta_B \ne \xi)]^{1/n}, \max_k[e^{c_k n} P_\zeta^{(k)}(\delta_B \ne \zeta)]^{1/n}\}.$$

Again our Lemma entails that

$$\lim_{n\to\infty}[P_\xi^{(k)}(\delta_B \ne \xi)]^{1/n} = \lim_{n\to\infty}[P_\zeta^{(k)}(\delta_B \ne \zeta)]^{1/n}$$

$$= \lim_{n\to\infty}[P_\zeta^{(k)}(\pi_\xi(\underline{x}) > \pi_\zeta(\underline{x}))]^{1/n}$$

$$= \max_i \rho_{ik}(c_1,\ldots,c_\ell).$$

Hence for any asymptotically minimax procedure δ^*

$$\lim_{n\to\infty} \max_k [e^{c_k n} \max_\theta P_\theta^{(k)}(\delta^* \neq \theta)]^{1/n} \geq \max_{i,k} e^{c_k} \rho_{ik}(c_1,\ldots,c_\ell).$$

This inequality combined with (2.4) proves Theorem 1.

Corollary 1. For $k = 1,\ldots,\ell$

$$\rho_{kk}(c_1,\ldots,c_\ell) \leq \rho_k \leq \max_i \rho_{ik}(c_1,\ldots,c_\ell).$$

The first of these inequalities follows from the definition of ρ_{kk} and ρ_k; the second is direct consequence of (2.2).

3. *The Existence of Adaptive Procedures.* We prove in this section our main results.

THEOREM 2. If an adaptive procedure exists then for all real $c_1,\ldots,c_\ell$

$$\max_k e^{c_k} \rho_k \geq \max_{i,k} e^{c_k} \rho_{ik}(c_1,\ldots,c_\ell). \tag{3.1}$$

If for some $c_1,\ldots,c_\ell$

$$\rho_k = \max_i \rho_{ik}(c_1,\ldots,c_\ell), \quad k = 1,\ldots,\ell, \tag{3.2}$$

then an adaptive estimator exists.

Proof. Let δ_m be a minimax estimator for the likelihood function $\pi_\theta(\underline{x})$ from Theorem 1. If an adaptive estimator δ_a exists then one has as in (2.3)

$$\max_k [e^{c_k n} \max_\theta P_\theta^{(k)}(\delta_m \neq \theta)] \leq \sum_k e^{c_k n} \max_\theta P_\theta^{(k)}(\delta_a \neq \theta),$$

so that

$$\lim_{n\to\infty} \max_k [e^{c_k n} \max_\theta P_\theta^{(k)}(\delta_m \neq \theta)]^{1/n} \leq \max_k \lim_{n\to\infty} [e^{c_k n} \max_\theta P_\theta^{(k)}(\delta_a \neq \theta)]^{1/n}$$

$$= \max_k e^{c_k} \rho_k .$$

This inequality and Theorem 1 imply (3.9).

If (3.2) holds then according to (2.2) the maximum likelihood estimator $\hat{\delta}$ based on $\pi_\theta(\underline{x})$ is adaptive.

Corollary 2. If an adaptive procedure exists then (3.1) is actually an equality.

This fact follows from Corollary 1.

Corollary 3. If for some $i \neq k$ and $\theta \neq \eta$, $p_i(x,\eta) = p_k(x,\theta)$ for all x, then there is no adaptive estimator.

Indeed in this case

$$\rho_{ik}(0,\ldots,0) \geq \inf_{s_1,\ldots,s_\ell \geq 0} E_\theta^{(k)} p_k^{\sum_r s_r}(X,\theta) \prod_r p_r^{-s_r}(X,\theta) = 1,$$

since every partial derivative of the latter function at the origin is nonnegative:

$$E_\theta^{(k)} \log[p_k(X,\theta)/p_r(X,\theta)] \geq 0,$$

and its infimum in the region $s_1 \geq 0,\ldots,s_\ell \geq 0$ is attained at zero. Therefore

$$\max_k \rho_k < \max_{i,k} \rho_{ik}(0,\ldots,0) = 1,$$

and adaptive procedure cannot exist.

THEOREM 3. <u>An adaptive procedure exists if and only if for</u> $k = 1,\dots,\ell$

$$\rho_k = \rho_{kk}(-\log \rho_1,\dots,-\log \rho_\ell) \geq \max_{i:i\neq k} \rho_{ik}(-\log \rho_1,\dots,-\log \rho_\ell). \tag{3.3}$$

<u>Proof</u>. Denote $c_k^0 = -\log \rho_k$, $\gamma_k = \max_i \rho_{ik}(c_1^0,\dots,c_\ell^0)$, $k = 1,\dots,\ell$. Theorem 2 implies that if an adaptive procedure exists then

$$1 = \max_k e^{c_k^0}\rho_k \geq \max_k \gamma_k/\rho_k. \tag{3.4}$$

Because of Corollary 1

$$\rho_k \leq \gamma_k,$$

which together with (3.4) shows that $\rho_k = \gamma_k$, $k = 1,\dots,\ell$. Since $\rho_k \geq \rho_k(c_1^0,\dots,c_\ell^0)$ formula (3.3) is established.

If (3.3) holds, then an adaptive procedure exists according to (3.2), which proves Theorem 3.

Condition (3.3) means that for all k and some $\theta \neq \eta$ the infimum

$$\inf_{s_1,\dots,s_\ell\geq 0} \int \tilde{p}_k(x,\theta)\tilde{p}_k^{\Sigma s_r}(x,\eta) \prod_r \tilde{p}_r^{-s_r}(x,\theta)d\mu(x),$$

where $\tilde{p}_k(x,\theta) = p_k(x,\theta)/\rho_k$, is attained when $s_r = 0$ for $r \neq k$, and also for all $i \neq k$ and all $\theta \neq \eta$

$$\inf_{s_1,\dots,s_\ell\geq 0} \int \tilde{p}_k(x,\theta)\tilde{p}_i^{\Sigma s_r}(x,\eta) \prod_r \tilde{p}_r^{-s_r}(x,\theta)d\mu(x) \leq 1.$$

Note that for all $k = 1,\dots,\ell$

$$\max_{\theta\neq\eta} \inf_{s\geq 0} \int \tilde{p}_k^{1-s}(x,\theta)\tilde{p}_k^{s}(x,\eta)d\mu(x) = 1.$$

If condition (3.3) is satisfied then the maximum likelihood estimator $\hat{\delta}_0$ based on $\max_i \prod_1^n (p_i(x_j,\theta)/\rho_i)$ is adaptive. It is also minimax for the risk function $R(\alpha,\delta)/\rho_\alpha$: for any δ

$$1 = R(\alpha,\hat{\delta}_0)/\rho_\alpha \le \max_\alpha R(\alpha,\delta)/\rho_\alpha .$$

It follows from the proof of Theorem 1 (see (2.3)) that one has for all real $c_1,\ldots,c_\ell$ even if (3.3) is not met

$$\max_\alpha e^{c_\alpha} R(\alpha,\hat{\delta}_0) = \max_{\alpha,\beta} e^{c_\alpha} \rho_{\beta\alpha}(c_1,\ldots,c_\ell) \le \max_\alpha e^{c_\alpha} R(\alpha,\delta),$$

so that for any δ

$$\max_\alpha R(\alpha,\hat{\delta}_0)/\rho_\alpha \le \max_\alpha R(\alpha,\delta)/\rho_\alpha .$$

We have proved the following result.

THEOREM 4. <u>The maximum likelihood estimator</u> $\hat{\delta}_0$ <u>based on</u> $\max_i \prod_1^n (p_i(x_j,\theta)/\rho_i)$ <u>is adaptive if condition</u> (3.3) <u>is satisfied. This estimator is always minimax for the risk function</u> $R(\alpha,\delta)/\rho_\alpha$, <u>where</u> $R(\alpha,\delta)$ <u>is defined by</u> (1.2).

Because of Theorem 1 the value v of the game defined by the risk $R(\alpha,\delta)/\rho_\alpha$ has the form

$$v = \max_{i,k}[\rho_{ik}(-\log \rho_1,\ldots,-\log \rho_\ell)/\rho_k] \ge 1.$$

It is easy to see that $v = 1$ if and only if an adaptive procedure exists.

It is worth noting that the estimator $\hat{\delta}_0$ is essentially different from the naive overall maximum likelihood estimator, i.e., from the maximum likelihood estimator based on $\max_i \prod_1^n p_i(x_j,\theta)$. In fact one can construct examples where the latter estimator is not adaptive but $\hat{\delta}_0$ is. Thus Theorem 4 suggests a method of elimination of the nuisance parameter α: one should use prior distribution for α with probabilities proportional to $1/\rho_\alpha$ to obtain a possibly adaptive rule.

4. *An Example.* Let distributions $P_\theta^{(k)}$ form one-parameter exponential family, i.e., their densities are of the form

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

where $v(x)$ is a real-valued statistic. As earlier we assume that all distributions $P_\theta^{(k)}$, $\theta = 1,\ldots,m$ are different so that the common support of all measures $P_\theta^{(k)}$ includes at least two points. Define

$$C(a) = \int \exp\{av(x)\}d\mu(x);$$

then the function $f(a) = \log C(a)$ is strictly convex. One has for $k = 1,\ldots,\ell$

$$\log \rho_k = \max_{\theta\neq\eta}\ \inf_{s\geq 0} \log \int p_k^{1-s}(x,\theta)p_k^s(x,\eta)d\mu(x) \tag{4.1}$$

$$= \max_{\theta\neq\eta}\ \min_{0\leq s\leq 1} [f(a_k(\theta)+s(a_k(\eta)-a_k(\theta)))-s[f(a_k(\eta))-f(a_k(\theta))]-f(a_k(\theta))]$$

$$= \min_{0\leq s\leq 1} [f(a_k(\theta_k)+s(a_k(\eta_k)-a_k(\theta_k)))-s[f(a_k(\eta_k))-f(a_k(\theta_k))]-f(a_k(\theta_k))].$$

Also

$$\log \rho_{ik}(c_1,\ldots,c_\ell) = \max_{\theta\neq\eta}\ \inf_{s_1,\ldots,s_\ell\geq 0} [f(a_k(\theta)+\sum_r s_r(a_i(\eta)-a_r(\theta)))$$

$$- \sum_r s_r[f(a_r(\eta))-f(a_r(\theta))+c_r-c_i]-f(a_k(\theta))].$$

We prove now that

$$\inf_{s_1,\ldots,s_\ell\geq 0} [f(a_k(\theta)+\sum_r s_r(a_i(\eta)-a_r(\theta))) \tag{4.2}$$

$$- \sum_r s_r[f(a_r(\eta))-f(a_r(\theta))+c_r-c_i]-f(a_k(\theta))]$$

$$= \min_{1\leq r\leq \ell}\ \inf_{s\geq 0}[f(a_k(\theta)+s(a_i(\eta)-a_r(\theta)))$$

$$- s[f(a_r(\eta))-f(a_r(\theta))+c_r-c_i]-f(a_k(\theta))].$$

Indeed if there exists a point $(s_1,\ldots,s_\ell)$ such that the vector of partial derivatives of the function in the left-hand side of (4.2) vanishes, then for $r = 1,..,\ell$

$$(a_i(\eta)-a_r(\theta))f'(a_k(\theta)+\sum_r s_r(a_i(\eta)-a_r(\theta))) = f(a_r(\eta))-f(a_r(\theta))+c_r-c_i.$$

Since f' is strictly monotone function, this formula entails

$$a_k(\theta) + \sum s_r(a_i(\eta) - a_r(\theta)) = a,$$

and the left-hand side of (4.2) is equal to

$$f(a) + \inf_{s_1,\ldots,s_\ell \geq 0} [- \sum_r s_r(f(a_r(\eta)) - f(a_r(\theta)) + c_r - c_i)] - f(a_k(\theta)).$$

The latter infimum is clearly attained when $s_r = 0$ for some r, i.e., on the boundary of the set $S = \{(s_1,\ldots,s_\ell), s_r \geq 0, r = 1,\ldots,\ell\}$. This is also true when the gradient of the function in (4.2) does not vanish in S. Repeating the previous argument one obtains (4.2). Denote

(4.3) $H_{ir}(a_k(\theta),a_r(\theta),a_i(\eta))$

$$= \inf_{s\geq 0}[f(a_k(\theta)+s(a_i(\eta)-a_r(\theta)))-s[f(a_r(\eta))-f(a_r(\theta))+\log(\rho_i/\rho_r)]-f(a_k(\theta))].$$

Then for $k = 1,\ldots,\ell$

(4.4) $$\log \rho_k = H_{kk}(a_k(\theta_k),a_k(\theta_k),a_k(\eta_k))$$

and we have proved that

$$\log \rho_{ik}(-\log \rho_1,\ldots,-\log \rho_\ell) = \max_{\theta\neq\eta} \min_r H_{ir}(a_k(\theta),a_r(\theta),a_i(\eta)).$$

These facts and Theorem 3 provide us with the following result.

THEOREM 5. Let for $k = 1,\ldots,\ell$

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

Then ρ_k is determined by (4.3) and (4.4). An adaptive estimator of θ exists if and only if for all $i \neq k$ and all $\theta \neq \eta$ there exists r, $1 \leq r \leq \ell$ such that

$$H_{ir}(a_k(\theta),a_r(\theta),a_i(\eta)) \leq \log \rho_k,$$

<u>and for any</u> k <u>the inequality</u>

$$H_{kr}(a_k(\theta_k),a_r(\theta_k),a_k(\eta_k)) \leq \log \rho_k$$

<u>holds for all</u> $r \neq k$ <u>and all</u> θ_k, η_k <u>defined by</u> (4.1).

The last statement of Theorem 5 easily follows since the condition

$$\rho_k = \rho_{kk}(-\log \rho_1,\ldots,-\log \rho_\ell)$$

means that

$$\max_{\theta\neq\eta} \min_{r} H_{kr}(a_k(\theta),a_r(\theta),a_k(\eta)) = H_{kk}(a_k(\theta_k),a_k(\theta_k),a_k(\eta_k))$$

$$= \max_{\theta\neq\eta} H_{kk}(a_k(\theta),a_k(\theta),a_k(\eta)).$$

The estimator $\hat{\delta}_0$ of Theorem 4 has the form

$$\{\delta_0=\theta\} = \{\max_k[a_k(\theta)\bar{v}-f(a_k(\theta))-\log \rho_k] > \max_k[a_k(\eta)\bar{v}-f(a_k(\eta))-\log \rho_k],\eta\neq\theta\}$$

where

$$\bar{v} = n^{-1}\sum_1^n v(x_j).$$

A simple necessary condition for the existence of an adaptive procedure is the consistency of $\hat{\delta}_0$ for any distribution $P_\theta^{(k)}$. Since under $P_\theta^{(k)}$ with probability one $\bar{v} \to f'(a_k(\theta))$, one concludes that the existence of an adaptive estimator implies that for $r = 1,\ldots,\ell$, $\theta \neq \eta$

$$\max_k[a_k(\theta)f'(a_r(\theta))-f(a_k(\theta))-\log \rho_k] > \max_k[a_k(\eta)f'(a_r(\theta))-f(a_k(\eta))-\log \rho_k]$$

As a specification of this example let us consider the case of normal densities $p_k(x,\theta)$ with unknown mean $a_k(\theta)$ and known variance σ^2. Then $v(x) = x$,

$$C(a) = \exp\{a^2/(2\sigma^2)\},\ f(a) = a^2/(2\sigma^2),$$

and

$$\rho_k = \max_{\theta\neq\eta} \exp\{-[a_k(\theta) - a_k(\eta)]^2/(8\sigma^2)\}.$$

If $\ell = 2$, $\theta = 1,2$, then it can be deduced from Theorem 5 that an adaptive estimator of θ exists if and only if

$$a_1(1) + a_1(2) = a_2(1) + a_2(2)$$

and differences $a_1(2) - a_1(1)$ and $a_2(2) - a_2(1)$ are of the same sign. In the latter case the estimator, which takes value 1 when $2\bar{x} < a_1(1) + a_1(2)$, is adaptive. (cf Laderman [8], Wald[11].)

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

[3] 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.

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

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

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

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

[8] Laderman, J. (1955). On the asymptotic behavior of decision procedures. *Ann. Math. Statist.* 26, 551-575.

[9] Steinebach, J. (1978). Convergence rates of large deviation probabilities in the multidimensional case. *Ann. Probability* 6, 751-759.

[10] Vajda, I. (1970). On the amount of information contained in a sequence of independent observations, *Kybernetika* (Academia) Praha, 6, 306-323.

[11] Wald, A. (1950). *Statistical Decision Functions.* New York: John Wiley.