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.


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 $P = (P_1, \ldots, P_m)$ be a family of $m$ different probability distributions, and let $x = (x_1, \ldots, x_n)$ be a sequence of independent random variables having common distribution $P_\theta$ for some $\theta = 1, \ldots, m$. On the basis of the random sample $x$ statistical inference about the finite-valued parameter $\theta$ is desired.

If $\delta = \delta(x)$ 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 $\delta$. The asymptotic behavior of this risk has been studied by Kraff and Puri [7] who showed that if $\delta^*$ is an asymptotically minimax procedure then

$$\lim_{n \to \infty} \max_{\theta \neq \theta_0} P_\theta^{1/n}(\delta \neq \theta_0) = \max_{\theta \neq \theta_0} \inf_{s \geq 0} E_\theta p^S(x, \eta)p^{-S}(x, \theta)$$

$$= \max_{\theta \neq \theta_0} \inf_{s \geq 0} \int p^S(x, \eta)p^{1-s}(x, \theta)d\mu(x) = \rho(p),$$

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

Notice that the quantity $\inf_{s \geq 0} \int p^S(x, \eta)p^{-S}(x, \theta)d\mu(x)$ represents the Chernoff's function for the likelihood ratio and gives the asymptotics for the probability $P_\theta^{1/n}(\prod_{j=1}^n p(x_j, \eta) > \prod_{j=1}^n p(x_j, \theta))$ 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 $0 \leq \rho(P) < 1$, since all members of $P$ are distinct, and
\( \rho(P) = 0 \) if and only if all distributions $P_\theta$ are mutually singular. Thus
\( \rho(P) \) can be interpreted as an information divergence of elements of $P$.
See Vajda [10] for further properties of $\rho(P)$. Assume now that the distributions $P_\theta$ involve a nuisance parameter $\alpha$ which also takes a finite
number of values. Thus one has families $P_\alpha = \{ P_\theta(\alpha), \theta = 1, \ldots, m \}, \alpha = 1, \ldots, \ell$. If $\delta$ is an estimator of the parameter $\theta$, then we denote

\[
R(\alpha, \delta) = \lim_{n \to \infty} \inf \left[ \max_{\theta} \left. p_\theta^{(\alpha)}(\delta \neq \theta) \right] \right]^{1/n}.
\]

Then for any $\delta$

\[
R(\alpha, \delta) \geq \rho(P_\alpha) = \rho_\alpha,
\]
and a procedure $\delta_\alpha$ is called adaptive if $R(\alpha, \delta_\alpha) = \rho_\alpha$ for all $\alpha$.

In other terms an adaptive estimator is asymptotically fully effi-
cient under any of the families $P_\alpha, \alpha = 1, \ldots, \ell$. In this paper we obtain
a necessary and sufficient condition for the existence of an adaptive pro-
cedure. Roughly speaking, an adaptive estimator exists if and only if the
members of different families $P_\alpha$ and $P_\beta, \alpha \neq \beta$, are not more similar than
the elements of one of these families.

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

\[
\lim_{n \to \infty} \inf \left[ \sum_{\theta} u_\theta p_\theta^{(\alpha)}(\delta \neq \theta) \right]^{1/n} = R(\alpha, \delta),
\]
the results of this paper are true if in the definition of an adaptive pro-
cedure 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 \( \omega(\theta, \theta) = 0 \) and \( \omega(\theta, d) > 0 \) 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 \( \delta_\alpha \), 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 \( \max_k \sum_{j=1}^{n} c_k n \rho_k(x_j, \theta) \), where \( c_k, k = 1, \ldots, \ell \) are given constants and \( \rho_k(x, \theta) \) is the density of \( \rho_\theta^{(k)} \). 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 \( v_1, v_2, \ldots, v_\ell \) and \( k = 1, \ldots, \ell \)

\[
(2.1) \quad \Pr\{ \sum_r \max_\rho \left[ \log(f_k(x))/g_\rho(x) \right] - c_r + c_k > 0 \} > 0.
\]

Then

\[
\lim_{n \to \infty} \frac{1}{n} \max_k \left[ e^{c_k n \rho_k(x_j)} \right] > \max_k \left[ e^{c_k n \rho_k(x_j)} \right]
\]

\[
= \lim_{n \to \infty} \frac{1}{n} \max_k \left[ e^{c_k n \rho_k(x_j)} \right] > \max_k \left[ e^{c_k n \rho_k(x_j)} \right]
\]

\[
= \max_{1 \leq k \leq \ell} \inf_{s_1, \ldots, s_\ell \geq 0} \exp \left\{ \sum_{r} s_r (c_k - c_r) \rho_r(x) \right\} \left( \prod_{r} \rho_r(x) \right)^{-s_r}(x).
\]
Proof. For any fixed $r$, $r = 1, \ldots, \ell$

\[
\Pr\{e_{1}^{c_{1}} e_{2}^{c_{2}} \cdots e_{n}^{c_{n}} f_{r}(x_{j}) \geq \max_{k} e_{1}^{c_{1}} e_{2}^{c_{2}} \cdots e_{n}^{c_{n}} g_{k}(x_{j})\} \\
\leq \Pr\{\max_{k} e_{1}^{c_{1}} e_{2}^{c_{2}} \cdots e_{n}^{c_{n}} f_{k}(x_{j}) \geq \max_{k} e_{1}^{c_{1}} e_{2}^{c_{2}} \cdots e_{n}^{c_{n}} g_{k}(x_{j})\} \\
\leq \sum_{i} \Pr\{e_{i}^{c_{i}} e_{2}^{c_{2}} \cdots e_{n}^{c_{n}} f_{i}(x_{j}) \geq \max_{k} e_{1}^{c_{1}} e_{2}^{c_{2}} \cdots e_{n}^{c_{n}} g_{k}(x_{j})\} \\
\leq \ell \max_{k} \Pr\{e_{1}^{c_{1}} e_{2}^{c_{2}} \cdots e_{n}^{c_{n}} f_{k}(x_{j}) \geq \max_{k} e_{1}^{c_{1}} e_{2}^{c_{2}} \cdots e_{n}^{c_{n}} g_{k}(x_{j})\}.
\]

It follows that

\[
\Pr^{1/n}\{\max_{k} e_{1}^{c_{1}} e_{2}^{c_{2}} \cdots e_{n}^{c_{n}} f_{k}(x_{j}) \geq \max_{k} e_{1}^{c_{1}} e_{2}^{c_{2}} \cdots e_{n}^{c_{n}} g_{k}(x_{j})\} \\
\leq \max_{k} \Pr^{1/n}\{e_{1}^{c_{1}} e_{2}^{c_{2}} \cdots e_{n}^{c_{n}} f_{k}(x_{j}) \geq \max_{i} e_{1}^{c_{1}} e_{2}^{c_{2}} \cdots e_{n}^{c_{n}} g_{i}(x_{j})\} \\
= \max_{k} \Pr^{1/n}\{e_{1}^{c_{1}} e_{2}^{c_{2}} \cdots e_{n}^{c_{n}} f_{k}(x_{j}) \geq e_{1}^{c_{1}} e_{2}^{c_{2}} \cdots e_{n}^{c_{n}} g_{i}(x_{j}), i = 1, \ldots, \ell\} \\
= \max_{k} \Pr^{1/n}\{n^{-1} \prod_{i} \log(f_{k}(x_{j})/g_{i}(x_{j})) \geq c_{i} - c_{k}, i = 1, \ldots, \ell\}.
\]

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 $c_{1}, \ldots, c_{\ell}$, $1 \leq i, k \leq \ell$

\[
\rho_{i,k}(c_{1}, \ldots, c_{\ell}) = \max_{\theta \neq \theta_{0}} \inf_{s_{1} = 0} \exp\sum_{r = 1}^{\infty} s_{r}(c_{i} - c_{r}) E_{\theta}^{(k)} p_{i}^{\sum_{r}^{s_{r}} r(x_{n}) p_{r}^{\sum_{r}^{s_{r}} r(x_{\theta})}.}
\]
Notice that for \( \ell = 1 \), \( \rho_{ij} = \rho(P) \). As we shall see, the quantities \( \rho_{ij} \) 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 \( \rho_{ik} \) 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 \( \rho_{ik} \) 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

\[
\rho_{ik}(c_1, \ldots, c_\ell) \leq \min\{1, \exp|c_i - c_k|\}.
\]

THEOREM 1. Let \( \hat{\delta} \) be an asymptotically minimax estimator of \( \theta \) based on the likelihood function \( \max[e^{i=1} \prod_{j=1}^{n} p_i(x_j, \theta)] \). Then

\[
\lim_{n \to \infty} \max_{k} e^{c_k n} \max_{\theta} p_\theta(k(\hat{\delta} \neq \theta))] / n = \max_{1 \leq i, k \leq \ell} e^{c_k \rho_{ik}(c_1, \ldots, c_\ell)}.
\]

Proof. Let \( \hat{\delta} \) be the maximum likelihood estimator of \( \theta \) based on

\[
\max[e^{i=1} \prod_{j=1}^{n} p_i(x_j, \theta)] = \pi_\theta(x, c_1, \ldots, c_\ell) = \pi_\theta(x).
\]

Thus \( \hat{\delta} = \theta \) if for \( \eta \neq \theta \)

\[
\pi_{\theta}(x) > \pi_{\eta}(x).
\]

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

\[
p_{\theta}(k(\pi_{\theta}(x) < \pi_{\eta}(x)) \leq p_{\theta}(k(\hat{\delta} \neq \theta))
\]

\[
\leq (m-1) \max_{\eta: \eta \neq \theta} p_{\theta}(k(\pi_{\theta}(x) < \pi_{\eta}(x))).
\]
Therefore because of our Lemma

\[(2.2) \quad \lim_{n \to \infty} \left[ p^{(k)}(\hat{\delta} \neq \theta) \right]^{1/n}\]

\[= \lim_{n \to \infty} \left[ \max_{n : n \neq \theta} p^{(k)}(\pi_0(\hat{x}) < \pi_n(x)) \right]^{1/n}\]

\[= \max_{n : n \neq \theta} \max_i \inf_{s_1, \ldots, s_{\ell} \geq 0} \exp\left( \sum r \left( c_i - c_r \right) \right) \exp(\varepsilon)_{pi} r(X, n) \Pi_p r(X, \theta) \]

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

\[E_{i}^{(i)} \left( \sum_{r} v_r \log(p_i(X, n)/p_r(X, \theta)) > 0, \right.\]

so that

\[p_{n}^{(i)} \left( \sum_{r} v_r \log(p_i(X, n)/p_r(X, \theta)) > 0 \right) > 0,\]

which is equivalent to the inequality

\[p_{\theta}^{(k)} \left( \sum_{r} v_r \log(p_i(X, n)/p_r(X, \theta)) > 0 \right) > 0.\]

If \( \hat{\delta} \) is a minimax procedure then

\[(2.3) \quad \max_{k} \left[ e^{c_k n} \max_{\theta \neq \hat{\delta}} p^{(k)}(\hat{\delta} \neq \theta) \right] \leq \max_{\theta \neq \{\hat{\delta} \neq \theta\}} \int \cdots \int_{\pi_0(\hat{x})} \mu(\hat{x}) \]

\[\leq \max_{\theta \neq \{\hat{\delta} \neq \theta\}} \int \cdots \int_{\pi_0(\hat{x})} \mu(\hat{x}) \leq \sum_{k} e^{c_k n} \max_{\theta \neq \hat{\delta}} p^{(k)}(\hat{\delta} \neq \theta) \]

and

\[(2.4) \quad \lim_{n \to \infty} \max_{k} \left[ e^{c_k n} \max_{\theta \neq \hat{\delta}} p^{(k)}(\hat{\delta} \neq \theta) \right]^{1/n} = \lim_{n \to \infty} \max_{k} \left[ e^{c_k n} \max_{\theta \neq \hat{\delta}} p^{(k)}(\hat{\delta} \neq \theta) \right]^{1/n}\]

\[\leq \lim_{n \to \infty} \max_{k} \left[ e^{c_k n} \max_{\theta \neq \hat{\delta}} p^{(k)}(\hat{\delta} \neq \theta) \right]^{1/n}.\]
\[ = \max_{i,k} e^{c_k \rho_{ik}(c_1, \ldots, c_\ell)}. \]

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

\[ \max_{i} \rho_{ik}(c_1, \ldots, c_\ell) \]

\[ = \max_{i} \inf_{s_1, \ldots, s_\ell \geq 0} \exp\{ \sum_i s_r(c_i-c_r)E_{\xi}(k)p_i r(X,\xi) + \sum_r s_r r(X,\xi) \}. \]

Also let \( \delta_B \) be the Bayes estimator for the prior distribution assigning weights \( 1/2 \) to \( \xi \) and \( \zeta \) and the likelihood function \( \pi_\theta(x) \). Then for any \( \delta \)

\[ \max_k \left[ e^{c_k n \rho(k)(\delta_B \neq \xi)} \right] + \max_k \left[ e^{c_k n \rho(k)(\delta_B \neq \zeta)} \right] \]

\[ \leq \int_{\{\delta_B \neq \xi\}} \pi_\xi(x)d\mu(x) + \int_{\{\delta_B \neq \zeta\}} \pi_\zeta(x)d\mu(x) \]

\[ \leq \int_{\{\delta \neq \xi\}} \pi_\xi(x)d\mu(x) + \int_{\{\delta \neq \zeta\}} \pi_\zeta(x)d\mu(x) \]

\[ \leq \sum_k e^{c_k n \left[ p_k \rho(k)(\delta \neq \xi) + p_k \rho(k)(\delta \neq \zeta) \right]}. \]

Thus

\[ \lim_{n \to \infty} \max_k \left[ e^{c_k n \rho(k)(\delta \neq \theta)} \right] \leq \frac{1}{n} \]

\[ \geq \lim_{n \to \infty} \max_k \left[ e^{c_k n \rho(k)(\delta_B \neq \xi)} \right]^{1/n}, \max_k \left[ e^{c_k n \rho(k)(\delta_B \neq \zeta)} \right]^{1/n}. \]

Again our Lemma entails that

\[ \lim_{n \to \infty} \left[ p_k \rho(k)(\delta_B \neq \xi) \right]^{1/n} = \lim_{n \to \infty} \left[ p_k \rho(k)(\delta_B \neq \zeta) \right]^{1/n} \]

\[ = \lim_{n \to \infty} \left[ p_k \rho(k)(\pi_\xi(x) > \pi_\zeta(x)) \right]^{1/n} \]
\[ = \max_{i} \rho_{ik}(c_1, \ldots, c_\ell). \]

Hence for any asymptotically minimax procedure \( \delta^* \)
\[ \lim_{n \to \infty} \max_{k \neq i, \ell} \frac{e^{c_k n}}{\max_{\theta} p_{\theta}(\delta^* \neq \theta)} \leq \sum_{i, k} \rho_{ik}(c_1, \ldots, c_\ell). \]

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 \( \rho_{kk} \) and \( \rho_k \); the second is a 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 \( c_1, \ldots, c_\ell \)
\[ \max_{k} e^{c_k} \rho_k \geq \max_{i, k} e^{c_k} \rho_{ik}(c_1, \ldots, c_\ell). \]

If for some \( c_1, \ldots, c_\ell \)
\[ \rho_k = \max_{i} \rho_{ik}(c_1, \ldots, c_\ell), \quad k = 1, \ldots, \ell, \]

then an adaptive estimator exists.

**Proof.** Let \( \delta_m \) be a minimax estimator for the likelihood function \( n_\theta(x) \) from Theorem 1. If an adaptive estimator \( \delta_a \) exists then one has as in (2.3)
\[ \max_{k} e^{c_k n} \max_{\theta} p_{\theta}(\delta_m \neq \theta) \leq \sum_{k} e^{c_k n} \max_{\theta} p_{\theta}(\delta_0 \neq \theta), \]
so that
\[
\lim_{n \to \infty} \max_k \left( e^{c_k n} \max_{\theta} p^{(k)}(\delta_m \neq \theta) \right)^{1/n} \leq \max_k \lim_{n \to \infty} e^{c_k n} \max_{\theta} p^{(k)}(\delta_a \neq \theta)^{1/n} = \max_k e^{c_k} p_k.
\]

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 \( \pi_0(x) \) 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 \( \theta \neq \eta \), \( p_{i}(x, n) = p_{k}(x, \theta) \) for all \( x \), then there is no adaptive estimator.

Indeed in this case
\[
p_{i,k}(0, \ldots, 0) \geq \inf_{s_1, \ldots, s_{\ell} \geq 0} E^{(k)} \frac{\sum_{r} s_r x_r \theta}{p_k(x, \theta) \prod_{r} p_r(x, \theta) = 1},
\]

since every partial derivative of the latter function at the origin is nonnegative:
\[
E^{(k)} \log[p_k(x, \theta)/p_{r}(x, \theta)] \geq 0,
\]

and its infimum in the region \( s_1 \geq 0, \ldots, s_{\ell} \geq 0 \) is attained at zero. Therefore
\[
\max_k p_k < \max_{i,k} p_{i,k}(0, \ldots, 0) = 1,
\]
and adaptive procedure cannot exist.
THEOREM 3. An adaptive procedure exists if and only if for
\[ k = 1, \ldots, \ell \]
\[ \rho_k = \rho_{kk}(-\log \rho_1, \ldots, -\log \rho_\ell) \geq \max_{i: i \neq k} \rho_{ik}(-\log \rho_1, \ldots, -\log \rho_\ell). \tag{3.3} \]

Proof. Denote \( c_k^0 = -\log \rho_k \), \( \gamma_k = \max_i \rho_{ik}(c_i^0, \ldots, c_\ell^0), k = 1, \ldots, \ell. \)

Theorem 2 implies that if an adaptive procedure exists then
\[ \frac{c_k^0}{\rho_k} = \max \frac{\gamma_k/\rho_k}{k}. \tag{3.4} \]

Because of Corollary 1
\[ \frac{\rho_k}{\gamma_k}, \]

which together with (3.4) shows that \( \rho_k = \gamma_k, k = 1, \ldots, \ell. \) Since
\[ \rho_k \geq \rho_k(c_1^0, \ldots, c_\ell^0) \]

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
\[ \inf \int \tilde{p}_k(x, \theta) \frac{\sum s}{\rho_k} r(x, \eta) \prod \tilde{p}_r(x, \theta) du(x), \]
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 \( \theta \neq \eta \)
\[ \inf \int \tilde{p}_k(x, \theta) \frac{\sum s}{\rho_i} r(x, \eta) \prod \tilde{p}_r(x, \theta) du(x) \leq 1. \]

Note that for all \( k = 1, \ldots, \ell \)
\[ \max_{\theta \neq \eta} \inf_{s \geq 0} \int \tilde{p}_k^{-s}(x, \theta) \bar{p}_k^{-s}(x, \eta) du(x) = 1. \]

If condition (3.3) is satisfied then the maximum likelihood estimator
\[ \hat{\theta}_0 \] based on \( \max_{i=1}^n \frac{p_i(x_j, \theta)/\rho_i}{\rho_i} \) is adaptive. It is also minimax for the
risk function $R(\alpha, \delta) / \rho_\alpha$: for any $\delta$

$$1 = R(\alpha, \delta_0) / \rho_\alpha \leq \max_\alpha R(\alpha, \delta) / \rho_\alpha.$$ 

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

$$\max_\alpha e^{\alpha R(\alpha, \delta_0)} = \max_{\alpha, \beta} e^{\alpha R(\alpha, \beta)}(c_1, \ldots, c_\ell) \leq \max_\alpha e^{\alpha R(\alpha, \delta)},$$

so that for any $\delta$

$$\max_\alpha R(\alpha, \delta_0) / \rho_\alpha \leq \max_\alpha R(\alpha, \delta) / \rho_\alpha.$$ 

We have proved the following result.

**THEOREM 4.** The maximum likelihood estimator $\delta_0$ based on

$$\max_1^n (p_i(x_j, \theta) / \rho_i)$$

is adaptive if condition (3.3) is satisfied. This estimator is always minimax for the risk function $R(\alpha, \delta) / \rho_\alpha$, where $R(\alpha, \delta)$ is defined by (1.2).

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

$$v = \max_{i, k} [p_i(x_j, \theta) / \rho_i] \geq 1.$$ 

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

It is worth noting that the estimator $\delta_0$ is essentially different from the naive overall maximum likelihood estimator, i.e., from the maximum likelihood estimator based on $\max_1^n p_i(x_j, \theta)$. In fact one can construct examples where the latter estimator is not adaptive but $\delta_0$ is.

Thus Theorem 4 suggests a method of elimination of the nuisance parameter $\alpha$: one should use prior distribution for $\alpha$ 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 $k = 1, \ldots, \ell$

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

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

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

Also

$$\log \rho_{1k}(c_1, \ldots, c_\ell) = \max_{\theta \neq n} \inf_{s_1, \ldots, s_\ell \geq 0} \left[ f(a_k(\theta)+\sum_r s_r(a_i(\eta)-a_r(\theta))) - \sum_r s_r[f(a_r(\eta))-f(a_r(\theta))+c_r-c_i]-f(a_k(\theta)) \right].$$

We prove now that

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

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

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 \( r = 1, \ldots, \ell \)
\[
(a_i(n) - a_r(\theta)) f'(a_k(\theta) + \sum_r s_r(a_i(n) - a_r(\theta))) = f(a_r(n)) - f(a_r(\theta)) + c_r - c_i.
\]
Since \( f' \) is strictly monotone function, this formula entails
\[
a_k(\theta) + \sum_r s_r(a_i(n) - a_r(\theta)) = a,
\]
and the left-hand side of (4.2) is equal to
\[
f(a) + \inf_{s_1, \ldots, s_\ell \geq 0} \left[ - \sum_r s_r(f(a_r(n)) - f(a_r(\theta)) + c_r - c_i) \right] - f(a_k(\theta)).
\]
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
\[
(4.3) \quad H_{ir}(a_k(\theta), a_r(\theta), a_i(n))
\]
\[
= \inf_{s > 0} [f(a_k(\theta) + s(a_i(n) - a_r(\theta)))] - s[f(a_r(n)) - f(a_r(\theta)) + \log(\rho_i/\rho_r)] - f(a_k(\theta)).
\]
Then for \( k = 1, \ldots, \ell \)
\[
(4.4) \quad \log \rho_k = H_{kk}(a_k(\theta_k), a_k(\theta_k), a_k(n_k))
\]
and we have proved that
\[
\log \rho_{ik}(-\log \rho_1, \ldots, -\log \rho_\ell) = \max_{\theta \neq \eta} \min_r H_{ir}(a_k(\theta), a_r(\theta), a_i(n)).
\]
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 \( \rho_k \) is determined by (4.3) and (4.4). An adaptive estimator of \( \theta \) 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
\[
H_{ir}(a_k(\theta), a_r(\theta), a_i(n)) \leq \log \rho_k,
\]
and for any \( k \) the inequality

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

holds for all \( r \neq k \) and all \( \theta_k, \eta_k \) defined by (4.1).

The last statement of Theorem 5 easily follows since the condition

\[
\rho_k = \rho_{kk}(-\log \rho_1, \ldots, -\log \rho_r)
\]

means that

\[
\max_{\theta \neq \eta} \min_r H_k(a_k(\theta), a_r(\theta), a_k(\eta)) = H_{kk}(a_k(\theta), a_k(\theta), a_k(\eta)) = \max_{\theta \neq \eta} H_k(a_k(\theta), a_k(\theta), a_k(\eta)).
\]

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

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

where

\[
\bar{v} = n^{-1} \sum_{i=1}^{n} v(x_j).
\]

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

\[
\max_k [a_k(\theta)f^*(a_r(\theta)) - f(a_k(\theta)) - \log \rho_k] > \max_k [a_k(n)f^*(a_r(\theta)) - f(a_k(n)) - \log \rho_k]
\]

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 \( \sigma^2 \). Then \( v(x) = x \),

\[
C(a) = \exp[a^2/(2\sigma^2)], f(a) = a^2/(2\sigma^2),
\]

and

\[
\rho_k = \max_{\theta \neq \eta} \exp\{-[a_k(\theta) - a_k(\eta)]^2/(4\sigma^2)\}.
\]
If \( \ell = 2, \theta = 1,2 \), then it can be deduced from Theorem 5 that an adaptive estimator of \( \theta \) 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


