PLISKA

STUDIA MATHEMATICA BULGARICA

ПЛИСКА

БЪЛГАРСКИ МАТЕМАТИЧЕСКИ СТУДИИ

Pliska Stud. Math. Bulgar. **13** (2000), 91-115

PLISKA
STUDIA MATHEMATICA
BULGARICA

MAKING MULTIPLE DECISIONS ADAPTIVELY

Andrew L. Rukhin

The asymptotic behavior of multiple decision procedures is studied when the underlying distributions depend on an unknown nuisance parameter. An adaptive procedure must be asymptotically optimal for each value of this nuisance parameter, and it should not depend on its value. A necessary and sufficient condition for the existence of such a procedure is derived. Several examples are investigated in detail, and possible lack of adaptation of the traditional overall maximum likelihood rule is discussed.

1 Multiple Decision Problem and Adaptation

Let $x = (x_1, x_2, \cdots, x_n)$ be a random sample drawn from one of different probability distributions $F_1, \ldots, F_g$ with densities $f_1, \ldots, f_g$. We will assume that these distributions are mutually absolutely continuous with respect to some σ-finite measure μ, so that all densities f_i can be chosen to be positive on the same set. In some situations one can find prior probabilities $\lambda_1, \ldots, \lambda_g$ of the sample distributions $P_i = F_i \otimes \cdots \otimes F_i,\ i = 1, \ldots, g$. The performance of a multiple decision rule $\delta(x)$ taking values in the set $\{1, \ldots, g\}$ is traditionally measured by the error probabilities $P_i(\delta \neq i)$ with the Bayes risk $\sum_i \lambda_i P_i(\delta \neq i)$ or by the minimax risk $\max_i P_i(\delta \neq i)$.

For two probability distributions P and Q let

$$H_s(P, Q) = \log E^Q \Big[\frac{d\,P}{d\,Q}(X)\Big]^s \tag{1}$$

be the logarithm of the Hellinger type integral.

Obviously H_s is a convex analytic function of s defined on an interval containing the closed interval $[0, 1]$ with the derivative at $s = 0$ being $-K(Q, P)$ and at $s = 1$ equal to $K(P, Q)$. Here

$$K(P, Q) = E^P \log\Big[\frac{d\,P}{d\,Q}(X)\Big] \tag{2}$$

is the (Kullback–Leibler) information number. Hence the minimum of H_s is attained in the interval $(0,1)$.

These quantities specify the exponential rate of the risk decay in the multiple decision problem (see Chernoff, 1956, Renyi, 1970, Krafft and Puri, 1974).

Theorem 1 *Assume that $\lambda_i > 0$ for all i. Then for any multiple decision rule $\delta = \delta(x)$ based on the random sample $x = (x_1, x_2, \cdots, x_n)$ from the family $\mathcal{P} = \{P_i = F_i \otimes \cdots \otimes F_i,\ i = 1, \ldots, g\}$ one has*

$$\liminf \frac{1}{n} \log \max_i P_i(\delta \neq i) = \liminf \frac{1}{n} \log \sum_i \lambda_i P_i(\delta \neq i)$$

$$\geq \max_{i \neq k} \inf_{s>0} H_s(F_i, F_k) = \max_{i \neq k} \inf_{0<s<1} H_s(F_i, F_k) = \rho(\mathcal{P}). \tag{3}$$

For the Bayes rule

$$\{\delta_B(x) = i\} = \left\{ \prod_{j=1}^n \lambda_i f_i(x_j) = \max_k \prod_{j=1}^n \lambda_k f_k(x_j) \right\}$$

or the maximum likelihood rule

$$\{\hat{\delta}(x) = i\} = \left\{ \prod_{j=1}^n f_i(x_j) = \max_k \prod_{j=1}^n f_k(x_j) \right\}$$

(3) is the equality.

In this paper we examine a version of the classical multiple decision problem in which probability distributions P_i of a random sample are not known exactly, but only up to a (nuisance) parameter α taking values in a set $\mathcal{A}$. In other words, a collection of probability distribution families

$$\mathcal{P}_\alpha = (P_1^\alpha, \ldots, P_g^\alpha) \qquad \alpha \in \mathcal{A}$$

with $P_i^\alpha = F_i^\alpha \otimes \cdots \otimes F_i^\alpha$ is supposed to be given.

For example, a repeated message may be sent through the one of noisy channels indexed by α. The goal of the statistician is to recover the message, no matter which channel has been used. One can also think about $\mathcal{A}$ as of the set of individuals with different handwritings or as of the set of possible handwritings. A text formed by a sequence of written letters is to be recognized independently of the individual who wrote them.

Thus, independently of the true value of the nuisance parameter α, one would like to use an efficient rule for the sample $x = (x_1, \ldots, x_n)$. This objective is formalized with help of Theorem 1 by the following definition.

A rule δ_a is called *adaptive* if for all α

$$\liminf \frac{1}{n} \log \max_i P_i^\alpha(\delta_a \neq i) = \max_{i \neq k} \inf_{s>0} H_s(F_i^\alpha, F_k^\alpha) = \rho(\mathcal{P}_\alpha) = \rho_\alpha$$

and δ_a does not depend on α. In the Bayes setting with fixed prior probabilities λ_i this definition should be modified by replacing the minimax risk by the Bayes risk.

Thus, in the presence of unknown nuisance parameters an adaptive rule must exhibit the same asymptotic optimal behavior as when these parameters were given. Of course, it should not depend on these unknown parameters.

The concept of adaptation for a continuous parameter was introduced by Stein (1956). A survey of the work in this area of semiparametric inference can be found in the monograph by Bickel et al (1993).

Notice that sometimes adaptive rules exist, and sometimes they do not exist. For example, let $g = 2$, $\mathcal{A} = \{1, 2\}$, $F_1^1 = N(-1, 1)$ and $F_2^1 = N(1, 1)$. If for the second value of the nuisance parameter $\alpha = 2$, $F_1^2 = N(-2, 1)$ and $F_2^2 = N(2, 1)$, then the rule $\tilde{\delta}$ such that $\{\tilde{\delta} = 1\} = \{x : \overline{x} < 0\}$ is the Bayes procedure against the uniform prior and, as such, is fully asymptotically efficient for any α. However, if for $\alpha = 2$, $F_1^2 = N(2, 1)$ and $F_2^2 = N(-2, 1)$, then any multiple decision rule must be very confused about the true distribution. Indeed, when $\alpha = 1$, the negative values of $\overline{x}$ are indicative of the first distribution in our family, and when $\alpha = 2$ the situation is quite opposite. Thus it is intuitively clear and will be proven later that an adaptive procedure for such families cannot exist.

We derive the existence condition and the form of adaptive procedures assuming for simplicity that $\mathcal{A}$ is a finite set, say, $\mathcal{A} = \{1, \ldots, A\}$.

Let for finite measures F and G

$$\rho(F, G) = \inf_{s>0} H_s(F, G),$$

where H_s is defined by (1). We rescale the original distributions F_i^α as follows

$$\tilde{F}_i^\alpha = e^{-\rho_\alpha} F_i^\alpha.$$

If ρ_α is interpreted as the degree of difficulty of the α-th classification problem, "easier" families $\mathcal{P}_\alpha$ (with large in absolute value quantity ρ_α) are getting larger weights.

Notice that for any α $\max_{i \neq k} \rho(\tilde{F}_i^\alpha, \tilde{F}_k^\alpha) = 0$. Indeed

$$H_s(\tilde{F}_i^\alpha, \tilde{F}_k^\alpha) = \log e^{-\rho_\alpha} + H_s(F_i^\alpha, F_k^\alpha),$$

so that

$$\max_{i \neq k} \rho(\tilde{F}_i^\alpha, \tilde{F}_k^\alpha) = -\rho_\alpha + \max_{i \neq k} \inf_{s>0} H_s(F_i^\alpha, F_k^\alpha) = -\rho_\alpha + \rho_\alpha = 0.$$

It turns out that an adaptive classification procedure exists if and only if

$$\max_{\alpha \neq \beta} \max_{i \neq k} \rho(\tilde{F}_i^\alpha, \tilde{F}_k^\beta) \leq 0$$

or

$$\max_{\alpha,\beta} \max_{i\neq k} \inf_{s>0} \left[H_s(F_i^\alpha, F_k^\beta) - s\rho_\alpha - (1-s)\rho_\beta\right]$$

$$= \max_{\alpha} \max_{i\neq k} \inf_{s>0} \left[H_s(F_i^\alpha, F_k^\alpha) - \rho_\alpha\right] = 0. \tag{4}$$

This fact has been proved in Rukhin (1984); a more detailed proof is given in Section 2.

The heuristic interpretation of (4) is that an adaptive procedure exists if and only if any $\mathcal{P}_\alpha$ is at least as difficult as problems formed by $P_i^\alpha, P_k^\beta,\ i \neq k, \alpha \neq \beta$.

Also, as will be proven in Theorem 2, the following rule

$$\{\delta_a(x) = i\} = \{\max_\alpha \prod_{j=1}^n e^{-\rho_\alpha} f_i^\alpha(x_j) = \max_k \max_\alpha \prod_{j=1}^n e^{-\rho_\alpha} f_k^\alpha(x_j)\} \tag{5}$$

is adaptive if there are adaptive rules. Observe that δ_a is a Bayes procedure, but the corresponding prior probabilities for the nuisance parameter α, which are proportional to $\exp\{-n\rho_\alpha\}$, heavily depend on the sample size n.

One of the traditional ways to eliminate a nuisance parameter is by using the uniform (noninformative) prior for this parameter. As we shall see, in our problem this method may lead to non-adaptive procedures. Indeed, the resulting "naive" overall maximum likelihood classification rule

$$\{\delta_0(x) = i\} = \{\max_\alpha \prod_{j=1}^n f_i^\alpha(x_j) = \max_k \max_\alpha \prod_{j=1}^n f_k^\alpha(x_j)\}$$

may not be adaptive when (4) holds. As a matter of fact, a similar conclusion holds for any prior probabilities for α which do not depend on the sample size.

2 Adaptation condition: proof and corrolaries

We prove in this section the main result from Rukhin (1984) in a more direct and illuminating fashion.

Theorem 2 *An adaptive classification procedure exists if and only if the inequality (4) holds. If (4) holds, then the procedure (5) is adaptive.*

The proof of this Theorem is based on the following Lemmas.

Lemma 1 *Let a and b be arbitrary real numbers. Then for any $i \neq k, \alpha, \beta$ and for any procedure δ the following inequality holds*

$$\left[\liminf \frac{1}{n} \log P_i^\alpha(\delta \neq i) + a\right] \vee \left[\liminf \frac{1}{n} \log P_k^\beta(\delta \neq k) + b\right]$$

$$\geq \inf_{s>0} [H_s\left(F_k^\beta, F_i^\alpha\right) + sb + (1-s)a] \vee \inf_{s>0} [H_s\left(F_i^\alpha, F_k^\beta\right) + sa + (1-s)b].$$

PROOF. Clearly the procedure δ_1

$$\delta_1(x) = \begin{cases} i, & \exp(na)\prod_{j=1}^n f_i^\alpha(x_j) > \exp(nb)\prod_{j=1}^n f_k^\beta(x_j) \\ k, & \text{otherwise} \end{cases}$$

minimizes the sum $\exp(na)P_i^\alpha(\delta(x) \neq i) + \exp(nb)P_k^\beta(\delta(x) \neq k)$. Also

$$P_i^\alpha(\delta_1(x) \neq i) = P_i^\alpha(\delta_1(x) = k)$$

$$\geq P_i^\alpha\left(\exp(nb)\prod_{j=1}^n f_k^\beta(x_j) \geq \exp(na)\prod_{j=1}^n f_i^\alpha(x_j)\right),$$

and the Chernoff theorem (see Bahadur, 1971) shows that

$$\liminf_{n\to\infty} \frac{1}{n}\log e^{na}P_i^\alpha\left(\frac{1}{n}\sum_{j=1}^n \log\frac{f_k^\beta}{f_i^\beta}(x_j) \geq a-b\right)$$

$$= \inf_{s>0}\left[H_s\left(F_k^\beta, F_i^\alpha\right) + (1-s)a + sb\right].$$

A similar inequality for $P_k^\beta(\delta_1 \neq k)$ concludes the proof of Lemma 1. □

The next result deals with the Bayes procedure, δ_b, based on the profile likelihood function, $\max_\alpha \exp(nb_\alpha)\prod_{j=1}^n f_i^\alpha(x_j)$ for some fixed constants b_α. More precisely, let for fixed positive prior probabilities $\lambda_k, k = 1, \ldots, A$

$$\{\delta_b(x_1, \ldots, x_n) = i\}$$

$$= \{\lambda_i \max_\alpha \exp(nb_\alpha)\prod_{j=1}^n f_i^\alpha(x_j) = \max_k \lambda_k \max_\alpha \exp(nb_\alpha)\prod_{j=1}^n f_k^\alpha(x_j)\}.$$

Lemma 2 *Assume that* $F_i^\alpha \neq F_k^\beta$ *for* $(i,\alpha) \neq (k,\beta)$ *and*

(6) $$\max_{\alpha,\beta}\max_{i\neq k}\left[b_\alpha - b_\beta - K\left(F_i^\alpha, F_k^\beta\right)\right] < 0.$$

Then for any α *and* i

$$\lim_{n\to\infty} n^{-1}\log P_i^\alpha\left(\delta_b(x_1,\ldots,x_n) \neq i\right) = r_\alpha^i(b_1,\ldots,b_A)$$

(7) $$= \max_{k:k\neq i}\max_\beta \inf_{s_1,\ldots,s_A\geq 0}\left\{\sum_{\gamma=1}^A s_\gamma(b_\beta - b_\gamma) + \log E_i^\alpha \prod_\gamma\left[\frac{f_k^\beta}{f_i^\gamma}(X)\right]^{s_\gamma}\right\}.$$

and for any α

(8) $$\lim_{n\to\infty} n^{-1}\max_i \log P_i^\alpha\left(\delta_b(x_1,\ldots,x_n) \neq i\right) = R_\alpha(b_1,\ldots,b_A),$$

where

$$R_\alpha(b_1,\dots,b_A) = \max_i r^i_\alpha(b_1,\dots,b_A)$$

$$= \max_{k\neq i}\max_\beta \inf_{s_1,\dots,s_A\geq 0}\left\{\sum_{\gamma=1}^A s_\gamma(b_\beta-b_\gamma)+\log E_i^\alpha \prod_\gamma\left[\frac{f_k^\beta}{f_i^\gamma}(X)\right]^{s_\gamma}\right\}.$$

For any procedure δ

$$\max_\alpha\left[b_\alpha + n^{-1}\liminf_{n\to\infty}\max_i \log P_i^\alpha\left(\delta(x_1,\dots,x_n)\neq i\right)\right]$$

$$\geq \max_\alpha\Big[b_\alpha + R_\alpha(b_1,\dots,b_A)\Big]. \tag{9}$$

Proof. One has

$$P_i^\alpha\left(\delta_b(x_1,\dots,x_n)\neq i\right)$$

$$= P_i^\alpha\left(\lambda_k e^{nb_\beta}\prod_{j=1}^n f_k^\beta(x_j)\geq \lambda_i e^{nb_\gamma}\prod_{j=1}^n f_i^\gamma(x_j) \text{ for some } \beta, k\neq i \text{ and all } \gamma\right)$$

$$\leq \sum_\beta\sum_{k:k\neq i} P_i^\alpha\left(\lambda_k \exp(nb_\beta)\prod_{j=1}^n f_k^\beta(x_j)\geq \lambda_i \exp(nb_\gamma)\prod_{j=1}^n f_i^\gamma(x_j) \text{ for all } \gamma\right)$$

$$\leq A(g-1)$$

$$\times \max_\beta\max_{k:k\neq i} P_i^\alpha\left(\lambda_k \exp(nb_\beta)\prod_{j=1}^n f_k^\beta(x_j)\geq \lambda_i \exp(nb_\gamma)\prod_{j=1}^n f_i^\gamma(x_j) \text{ for all } \gamma\right).$$

Also

$$P_i^\alpha\left(\delta_b(x_1,\dots,x_n)\neq i\right)$$

$$\geq \max_\beta\max_{k:k\neq i} P_i^\alpha\left(\lambda_k \exp(nb_\beta)\prod_{j=1}^n f_k^\beta(x_j)\geq \lambda_i \exp(nb_\gamma)\prod_{j=1}^n f_i^\gamma(x_j) \text{for all } \gamma\right).$$

Therefore, for a fixed α,

$$\lim_{n\to\infty} n^{-1}\max_\beta\max_{k:k\neq i}\log P_i^\alpha\left(\delta_b(x_1,\dots,x_n)\neq i\right)$$

$$= \lim_{n\to\infty} n^{-1}\max_\beta\max_{k:k\neq i}\log P_i^\alpha\left(\frac{1}{n}\sum_{j=1}^n \log\frac{f_k^\beta}{f_i^\gamma}(x_j)\geq b_\gamma - b_\beta \text{ for all } \gamma\right)$$

$$= \max_{k:k\neq i}\max_\beta \inf_{s_1,\dots,s_A\geq 0}\left\{\sum_{\gamma=1}^A s_\gamma(b_\beta-b_\gamma)+\log E_i^\alpha \prod_\gamma\left[\frac{f_k^\beta}{f_i^\gamma}(X)\right]^{s_\gamma}\right\}.$$

The last formula follows from the condition (6) and the multivariate Chernoff theorem (see Groeneboom, Oosterhoff and Ruymgaart, 1979). Thus, (7) is established, and the formula (8) easily follows.

Since δ_b is the Bayes rule with respect to the density proportional to $\max_\alpha \exp(nb_\alpha)\prod_{j=1}^n f_i^\alpha(x_j)$, for any rule δ

$$A\sum_i \lambda_i \max_\alpha \left[e^{nb_\alpha} P_i^\alpha\left(\delta(x_1,\ldots,x_n)\neq i\right)\right]$$

$$\geq \sum_i \lambda_i \int\cdots\int_{\{\delta\neq i\}} \max_\alpha \left[\exp(nb_\alpha)\prod_{j=1}^n f_i^\alpha(x_j)\right] d\mu(x_1)\cdots d\mu(x_n)$$

$$\geq \sum_i \lambda_i \int\cdots\int_{\{\tilde\delta_b\neq i\}} \max_\alpha \left[\exp(nb_\alpha)\prod_{j=1}^n f_i^\alpha(x_j)\right] d\mu(x_1)\cdots d\mu(x_n)$$

$$\geq \sum_i \lambda_i \max_\alpha \left[e^{nb_\alpha} P_i^\alpha\left(\tilde\delta_b(x_1,\ldots,x_n)\neq i\right)\right].$$

Therefore for any δ, (9) holds. □

Lemma 3 *If an adaptive procedure exists, then for all* $b_1,\ldots,b_A$

$$\max_\alpha[b_\alpha+\rho_\alpha]\geq\max_\alpha[b_\alpha+R_\alpha(b_1,\ldots,b_A)].$$

If with some $b_1,\ldots,b_A$

$$\rho_\alpha\geq R_\alpha(b_1,\ldots,b_A)$$

for $\alpha=1,\ldots,A$, *then an adaptive procedure exists. This inequality holds if*

$$\max_{k\neq i}\max_\beta\min_\gamma\inf_{s\geq0}\left\{\log E_i^\alpha\left[\frac{f_k^\beta}{f_i^\gamma}(X)\right]^s - s(b_\gamma-b_\beta)\right\}\leq\rho_\alpha.$$

Proof. If δ is an adaptive procedure, then applying Lemma 2 one deduces for any $b_1,\ldots,b_A$

$$\max_\alpha[b_\alpha+\rho_\alpha]\geq\max_\alpha[b_\alpha+R_\alpha(b_1,\ldots,b_A)].$$

Also Lemma 2 implies that

$$R_\alpha(b_1,\ldots,b_A)\geq\rho_\alpha$$

for $\alpha=1,\ldots,A$. According to the second condition of Lemma 3

$$R_\alpha(b_1,\ldots,b_A)=\rho_\alpha,$$

so that because of (8) δ_b must be adaptive. Since

$$R_\alpha(b_1,\ldots,b_A) \le \max_{k\ne i}\max_{\beta}\min_{\gamma}\inf_{s\ge 0}\left\{\log E_i^\alpha\left[\frac{f_k^\beta}{f_i^\gamma}(X)\right]^s - s(b_\gamma - b_\beta)\right\},$$

the last conclusion of Lemma 3 follows. $\square$

Proof.[of Theorem] Assume first that an adaptive procedure δ exists. Lemma 1 with $a=-\rho_\alpha$ and $b=-\rho_\beta$ implies

$$0 \ge \left[\liminf \frac{1}{n}\log P_i^\alpha(\delta\ne i) - \rho_\alpha\right] \vee \left[\liminf \frac{1}{n}\log P_k^\beta(\delta\ne k) - \rho_\beta\right]$$

$$\ge \inf_{s>0}\left[H_s\left(F_k^\beta,F_i^\alpha\right) - s\rho_\alpha - (1-s)\rho_\beta\right] \vee \inf_{s>0}\left[H_s\left(F_i^\alpha,F_k^\beta\right) - s\rho_\beta - (1-s)\rho_\alpha\right],$$

so that (4) is satisfied.

Now suppose that (4) holds. We prove that the procedure (5), which coincides with δ_b when $b_\alpha=-\rho_\alpha$, is adaptive.

Indeed the formula (8) of Lemma 2 shows that

$$\lim_{n\to\infty} n^{-1}\max_\alpha\left[-n\rho_\alpha + \max_i \log P_i^\alpha(\delta_b\ne i)\right] = \max_\alpha\left[-\rho_\alpha + R_\alpha(-\rho_1,\ldots,-\rho_A)\right]$$

$$= \max_\alpha\left\{-\rho_\alpha + \max_{k\ne i}\max_\beta\min_\gamma\inf_{s\ge 0}\left(s(\rho_\gamma-\rho_\beta) + \log E_i^\alpha\left[\frac{f_k^\beta}{f_i^\gamma}(X)\right]^s\right)\right\}$$

$$\le \max_{\alpha,\beta}\max_{k\ne i}\inf_{s\ge 0}\left\{s(\rho_\alpha-\rho_\beta) - \rho_\alpha + \log E_i^\alpha\left[\frac{f_k^\beta}{f_i^\alpha}(X)\right]^s\right\} = 0$$

with the last equality following from (4).

Thus for all α

$$R_\alpha(-\rho_1,\ldots,-\rho_A) \le \rho_\alpha.$$

Because of Lemma 3, Theorem 2 is proven. $\square$

The condition (4) can be reformulated in the following way.

Corollary 1 *An adaptive procedure exists if and only if for all $\alpha\ne\beta$ and all $i\ne k$*

$$\inf_{0<s<1}\left[H_s(F_i^\alpha,F_k^\beta) - s(\rho_\alpha-\rho_\beta)\right] \le \rho_\beta. \tag{10}$$

In other words an adaptive procedure exists if and only if

$$\max_{\alpha:\rho_\alpha\ge\rho_\beta}\max_{k\ne i}\inf_{0<s<1}\left[H_s(F_i^\alpha,F_k^\beta) - s(\rho_\alpha-\rho_\beta)\right] \le \rho_\beta.$$

PROOF. Clearly, if an adaptive rule exists, then according to (4) the condition (10) also holds.

If (10) is valid, then the unique infimum of the convex function of s, $H_s(F_i^\alpha, F_k^\beta) - s(\rho_\alpha - \rho_\beta)$, must be attained in the open interval $(0,1)$. Indeed the value of this function at $s = 1$ is $\rho_\beta - \rho_\alpha > \rho_\beta$.

Then

$$\inf_{s>0} \left[H_s(F_i^\alpha, F_k^\beta) - s(\rho_\alpha - \rho_\beta) \right] = \inf_{0<s<1} \left[H_s(F_i^\alpha, F_k^\beta) - s(\rho_\alpha - \rho_\beta) \right],$$

so that (4) follows.

The second condition of Corollary 1 implies (10), so that it is necessary and sufficient for adaptation. □

Theorem 3 *Let prior probabilities for the nuisance parameter α be proportional to $\exp(nb_\alpha)$ for fixed constants $b_1, \ldots, b_A$ satisfying (6). Then the Bayes estimator $\tilde{\delta}_b$ against the prior for (α, i) corresponding to the product of such probabilities and positive (independent of n) prior probabilities λ_i for $i = 1, \ldots, g$*

$$\{\tilde{\delta}_b(x_1, \ldots, x_n) = i\}$$

$$= \{\lambda_i \sum_\alpha \exp(nb_\alpha) \prod_{j=1}^n f_i^\alpha(x_j) = \max_k \lambda_k \sum_\alpha \exp(nb_\alpha) \prod_{j=1}^n f_k^\alpha(x_j)\},$$

as well as the estimator δ_b, is adaptive if and only if for all α

$$\max_{k\neq i} \max_{\beta:\beta\neq\alpha} \inf_{s_1,\ldots,s_A\geq 0} \left\{ \log E_i^\alpha \prod_\gamma \left[\frac{f_k^\beta}{f_i^\gamma}(X) \right]^{s_\gamma} + \sum_{\gamma=1}^A s_\gamma(b_\beta - b_\gamma) \right\} \leq \rho_\alpha. \tag{11}$$

PROOF. According to Lemma 2 adaptation for δ_b takes place if and only if

$$\max_{k\neq i} \max_{\beta} \inf_{s_1,\ldots,s_A\geq 0} \left\{ \log E_i^\alpha \prod_\gamma \left[\frac{f_k^\beta}{f_i^\gamma}(X) \right]^{s_\gamma} + \sum_{\gamma=1}^A s_\gamma(b_\beta - b_\gamma) \right\} \leq \rho_\alpha.$$

For $\beta = \alpha$ this condition holds automatically since

$$\max_{k\neq i} \inf_{s_1,\ldots,s_A\geq 0} \left\{ \log E_i^\alpha \prod_\gamma \left[\frac{f_k^\alpha}{f_i^\gamma}(X) \right]^{s_\gamma} + \sum_{\gamma=1}^A s_\gamma(b_\alpha - b_\gamma) \right\}$$

$$\leq \max_{k\neq i} \inf_{s\geq 0} \log E_i^\alpha \left[\frac{f_k^\alpha}{f_i^\alpha}(X) \right]^s = \rho_\alpha.$$

Thus condition (11) is necessary and sufficient for adaptation of δ_b.

An analysis of the proof of Lemma 2 shows that it goes through for $\tilde{\delta}_b$ with $\max_\alpha$ replaced by $\sum_\alpha$. □

Corollary 2 *If an adaptive procedure exists, then for all* $b_1, \dots, b_A$

$$\max_\alpha [b_\alpha + R_\alpha(b_1, \dots, b_A)] = \max_\alpha [b_\alpha + \rho_\alpha].$$

This follows directly from the first part of Lemma 3.

Corollary 3 *If for some* $\alpha \neq \beta$ *and* $i \neq k$, $F_i^\alpha = F_k^\beta$, *then an adaptive procedure cannot exist.*

Indeed in this situation

$$R_\alpha(0, \dots, 0) \geq \max_\gamma \inf_{s_1, \dots, s_A \geq 0} \log E_i^\alpha \prod_\gamma \left[\frac{f_i^\alpha}{f_i^\gamma}(X) \right]^{s_\gamma} = 0.$$

According to Corollary 2 the existence of an adaptive rule would imply

$$\max_\alpha \rho_\alpha \geq \max_\alpha R_\alpha(0, \dots, 0) = 0,$$

which is impossible.

The Corollary 3 supports the heuristic interpretation of (4) according to which an adaptive procedure exists if and only if the distributions from any $\mathcal{P}_\alpha$ are "at least as close" as the distributions P_i^α and P_k^β, $i \neq k, \alpha \neq \beta$.

Corollary 4 *If for all* $\alpha \neq \beta$

$$\max_{k \neq i} \inf_{s > 0} \left[H_s(F_k^\beta, F_i^\alpha) - s\rho_\beta - (1-s)\rho_\alpha \right] \leq 0 \wedge (\rho_\alpha - \rho_\beta + b_\alpha - b_\beta),$$

then δ_b *is adaptive.*

Proof. Under the condition of this Corollary an adaptive procedure exists. If $b_\alpha - b_\beta \geq \rho_\beta - \rho_\alpha$, then by (4)

$$\max_{k \neq i} \inf_{s \geq 0} \left[H_s\left(F_k^\beta, F_i^\alpha\right) - s(b_\alpha - b_\beta) \right]$$

$$\leq \max_{k \neq i} \inf_{s \geq 0} \left[H_s\left(F_k^\beta, F_i^\alpha\right) + s(\rho_\alpha - \rho_\beta) \right] \leq \rho_\alpha.$$

If $b_\alpha - b_\beta < \rho_\beta - \rho_\alpha$, then the condition of Corollary 4 shows that

$$\rho_\alpha - \rho_\beta + b_\alpha - b_\beta \geq \max_{k \neq i} \inf_{s > 0} \left[H_s(F_k^\beta, F_i^\alpha) - s\rho_\beta - (1-s)\rho_\alpha \right]$$

$$= \max_{i \neq k} \inf_{0 < s < 1} \left[H_s(F_i^\alpha, F_k^\beta) - s(\rho_\alpha - \rho_\beta) - \rho_\beta \right].$$

The last identity follows from Corollary 1. Therefore

$$\rho_\beta \geq \rho_\alpha + b_\alpha - b_\beta \geq \max_{k \neq i} \inf_{0 < s < 1} \left[H_s(F_k^\alpha, F_i^\beta) + sb_\alpha - sb_\beta \right].$$

In other terms (11) holds with $\gamma = \alpha$ and δ_b is adaptive. $\square$

3 Consistency Property and Example

Here we look at the consistency property of δ_b, i.e. at the conditions under which for any α and i

$$\lim_{n\to\infty} P_i^{\alpha}\left(\delta_b(x_1,\ldots,x_n)=i\right)=1.$$

Theorem 4 *Under condition (6) the procedure δ_b is consistent. The estimator δ_0 is consistent whenever $F_i^{\alpha}\neq F_k^{\beta}$ for $\alpha\neq\beta$, $i\neq k$.*

Proof. According to Theorem 3 the procedure δ_b is consistent if for any α the left-hand side of (11) is negative. This means that for all $\beta\neq\alpha$ and $i\neq k$ there exists γ such that the derivative of the function of s in (11), which vanishes at $s=0$, is negative at this point. This condition is implied by (6) which is always satisfied if $b_\alpha\equiv 0$. □

If the adaptation condition (4) is valid, the rule (5) is adaptive and automatically consistent. Moreover, under this condition the procedure (5) exhibits a good behavior even for small sample sizes. In particular, it often outperforms δ_0 in terms of error probabilities. The drawback of δ_a is that when (4) is violated, it may not even be consistent. In contrast, because of Thereom 4, the rule δ_0 is always consistent, although it may not be adaptive.

Therefore, in practice one may want to use intermediate weights b_α, $\rho_\alpha\leq b_\alpha\leq 0$, to combine consistency and adaptation. Indeed, if with some constants b_α

$$\max_{k\neq i}\inf_{s>0}\left[H_s(F_k^{\beta},F_i^{\alpha})-s\rho_\beta-(1-s)\rho_\alpha\right]\leq 0\wedge(\rho_\alpha-\rho_\beta+b_\alpha-b_\beta),$$

for all $\alpha\neq\beta$, then according to Corollary 4 the corresponding rule δ_b will be adaptive (and automatically consistent).

The following example shows that there may be no adaptive classification rule when the families are formed by shifts of different, symmetric about zero distributions.

Let $g=2$, $\mathcal{A}=\{1,2\}$, $F_1^1=N(-1,1)$ and $F_2^1=N(1,1)$. For the second value of the nuisance parameter $\alpha=2$, let F_1^2 and F_2^2 be double exponential distributions with the means $-\mu$ and μ and the same scale parameter σ, i.e. the densities have the form

$$f_1^2(x)=\frac{1}{2\sigma}\exp\left\{-\frac{|x+\mu|}{\sigma}\right\}$$

and

$$f_2^2(x)=\frac{1}{2\sigma}\exp\left\{-\frac{|x-\mu|}{\sigma}\right\}.$$

In this situation $\rho_1=-1/2$ and $\rho_2=-\mu/\sigma+\log\left(1+\mu/\sigma\right)$. Also

$$H_s(F_1^1,F_2^2)=H_s(F_2^1,F_1^2)=H_{1-s}(F_2^2,F_1^1)=H_{1-s}(F_1^2,F_2^1)$$

$$=\log\left(\frac{1}{(2\pi)^{s/2}(2\sigma)^{1-s}}\int\left\{-\frac{sx^2}{2}-\frac{(1-s)|x-\mu-1|}{\sigma}\right\}\,dx\right).$$

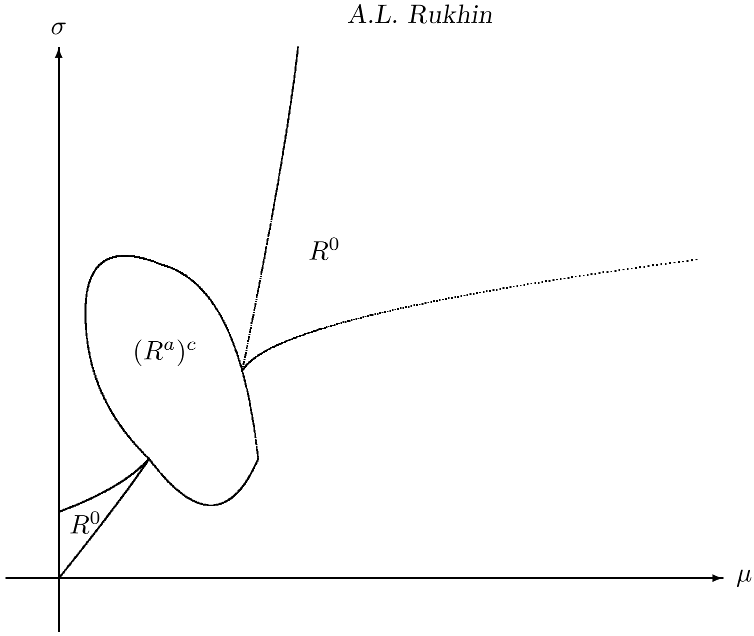

Figure 1: The region (μ, σ), for which an adaptive estimator exists.

The numerical evaluation of these integrals leads to Figure 1 showing the set R^a of pairs (μ, σ), where an adaptive estimator exists.

It also depicts the region R^0 obtained from Theorem 3 where the estimator δ_0 is adaptive. In this situation the set

$$\left\{(\mu, \sigma) : \inf_{s>0} \log E_1^1 \left[\frac{f_2^2}{f_1^2}\right]^s \leq \rho_1\right\}$$

is empty, and

$$\left\{(\mu, \sigma) : \inf_{s>0} \log E_2^2 \left[\frac{f_1^1}{f_2^1}\right]^s \leq \rho_2\right\} = \left\{(\mu, \sigma) : \frac{\mu}{\sigma} \leq c_o\right\}.$$

This example shows that the adaptation condition in the classification problem is different from that in the point estimation problem where there exists an adaptive estimator of the center of symmetry for any density with finite Fisher information (see Bickel et al, 1993).

4 Adaptation for Exponential Families

Let now the distributions F_i^α be members of a p-parameter exponential family, that is, let the densities f_i^α with respect to measure μ have the form

$$f_i^\alpha(u) = \exp\{\theta_i^\alpha \cdot u - \chi(\theta_i^\alpha)\}, \quad i = 1, \ldots, g.$$

Since the distributions P_i^α are supposed to be different, the common support of these distributions contains at least two points, and the function χ is strictly convex over the natural parameter space $\Theta = \{\theta : \chi(\theta) < \infty\}$ (which is a convex subset of R^p).

One has

$$H_s\left(e^a F_i^\alpha, e^b F_k^\beta\right) = \log \int \exp\{s[a + \theta_i^\alpha \cdot u - \chi(\theta_i^\alpha)] + (1-s)[b + \theta_k^\beta \cdot u - \chi(\theta_k^\beta)]\}\, d\mu(u)$$

$$= \chi\left(s\theta_i^\alpha + (1-s)\theta_k^\beta\right) + s[a - \chi(\theta_i^\alpha)] + (1-s)[b - \chi(\theta_k^\beta)].$$

In particular,

$$\rho_\alpha = \max_{i \neq k} \inf_{s>0} \left[\chi\left(s\theta_i^\alpha + (1-s)\theta_k^\alpha\right) - s\chi\left(\theta_i^\alpha\right) - (1-s)\chi\left(\theta_k^\beta\right)\right]. \tag{12}$$

By differentiating the left-hand side of (12) one notices that for a fixed α and $i \neq k$, the unique minimum is attained at $s = s_{ik}^\alpha$ such that

$$(\theta_i^\alpha - \theta_k^\alpha) \cdot \chi'\left(s\theta_i^\alpha + (1-s)\theta_k^\alpha\right) = \chi\left(\theta_i^\alpha\right) - \chi\left(\theta_k^\alpha\right).$$

Let

$$\sigma_{ik}^\alpha = s_{ik}^\alpha \theta_i^\alpha + (1 - s_{ik}^\alpha)\theta_k^\alpha,$$

so that

$$(\theta_i^\alpha - \theta_k^\alpha) \cdot \chi'\left(\sigma_{ik}^\alpha\right) = \chi\left(\theta_i^\alpha\right) - \chi\left(\theta_k^\alpha\right).$$

Then

$$\rho_\alpha = \max_{i \neq k} \left[\chi\left(\sigma_{ik}^\alpha\right) - s_{ik}^\alpha \chi\left(\theta_i^\alpha\right) - (1 - s_{ik}^\alpha)\chi\left(\theta_k^\alpha\right)\right]$$

$$= \max_{i \neq k} \left[\chi\left(\sigma_{ik}^\alpha\right) - \chi\left(\theta_k^\alpha\right) - \left(\sigma_{ik}^\alpha - \theta_k^\alpha\right)\chi'\left(\sigma_{ik}^\alpha\right)\right]. \tag{13}$$

According to Corollary 1, an adaptive rule exists if and only if for any β

$$\max_{\alpha:\rho_\alpha \geq \rho_\beta} \max_{i \neq k} \inf_{s>0} \Big[\chi(s\theta_i^\alpha + (1-s)\theta_k^\beta) - s\chi(\theta_i^\alpha) - (1-s)\chi(\theta_k^\beta)$$

$$- s\rho_\alpha - (1-s)\rho_\beta\Big] \leq 0. \tag{14}$$

Throughout this section we assume that $\theta_i^\alpha \neq \theta_k^\beta$ when $(i, \alpha) \neq (k, \beta)$. Then the minimum in (14) is attained at $s = s_{ik}^{\alpha\beta}$ such that

$$\left(\theta_i^\alpha - \theta_k^\beta\right) \cdot \chi'\left(s\theta_i^\alpha + (1-s)\theta_k^\beta\right) = \chi\left(\theta_i^\alpha\right) + \rho_\alpha - \chi\left(\theta_k^\beta\right) - \rho_\beta$$

provided that such a value exists. This value belongs to the interval $(0, 1)$ if and only if $\sigma_{ik}^{\alpha\beta} = s_{ik}^{\alpha\beta}\theta_i^\alpha + \left(1 - s_{ik}^{\alpha\beta}\right)\theta_k^\beta$ belongs to the segment connecting θ_i^α and θ_k^β, and under this condition δ_a is a consistent procedure.

Since
$$(\theta_i^\alpha - \theta_k^\beta) \cdot \chi^{'}\left(\sigma_{ik}^{\alpha\beta}\right) = \chi\left(\theta_i^\alpha\right) + \rho_\alpha - \chi\left(\theta_k^\beta\right) - \rho_\beta,$$
one has $\chi^{'}\left(\sigma_{ik}^{\alpha\beta}\right) = \chi^{'}\left(\sigma_{ki}^{\beta\alpha}\right)$, so that $\sigma_{ik}^{\alpha\beta} = \sigma_{ki}^{\beta\alpha}$.

Using these formulas one derives the following version of (14): for any β

$$\max_{\alpha:\rho_\alpha\geq\rho_\beta}\max_{i\neq k}\Big[\chi\left(\sigma_{ik}^{\alpha\beta}\right) - \chi\left(\theta_k^\beta\right) - \left(\sigma_{ik}^{\alpha\beta} - \theta_k^\beta\right)\cdot\chi^{'}\left(\sigma_{ik}^{\alpha\beta}\right)\Big] \leq \rho_\beta$$

and $\sigma_{ik}^{\alpha\beta}$ is a convex combination of θ_i^α and θ_k^β.

To make the adaptation condition more explicit we define for a fixed θ the function $W_\theta(t)$, $t \in \Theta$,
$$W_\theta(t) = \chi(t) - \chi(\theta) - (t-\theta)\cdot\chi^{'}(t).$$
Then $W_\theta(t)$ is a unimodal (quasi-concave) non-positive function of t, $W_\theta(\theta) = 0$. To see this let $y = \chi^{'}(t)$, so that y is the expected value of the distribution with natural parameter t. For a fixed θ with $h(y) = (\chi^{'})^{-1}(y)$, the function $\mathbf{H}(y) = W_\theta(h(y))$ is a concave function of y. Indeed, $\mathbf{H}^{'}(y) = \theta - h(y)$, and $\mathbf{H}^{''}(y) = -h^{'}(y)$ is a non-positive definite matrix. Function $\mathbf{H}$ has the unique maximum when $h(y) = \theta$, i.e. when $\chi^{'}(\theta) = y$, or when $t = \theta$.

Observe that $W_\theta(t) = -K(F_t, F_\theta)$ with K defined by (2), so that the function W and the following conditions formulated in its terms have distinct information-theoretic interpretation. Also $W_{\theta_k^\beta}\left(\sigma_{ik}^\beta\right) = W_{\theta_i^\beta}\left(\sigma_{ik}^\beta\right)$.

Notice that the inequality $W_\theta(t) \leq W_\theta(u)$ means that

$$W_t(u) = \chi(u) - \chi(t) - (u-t)\cdot\chi^{'}(u) \geq (t-\theta)\cdot[\chi^{'}(u) - \chi^{'}(t)].$$

Since $W_t(u) \leq 0$, one obtains

$$(t-\theta)\cdot[\chi^{'}(u) - \chi^{'}(t)] \leq 0. \tag{15}$$

Moreover, if $(t-\theta)\cdot[\chi^{'}(u) - \chi^{'}(t)] = 0$, then $W_t(u) = 0$ and $t = u$.

According to (13)
$$\rho_\alpha = \max_{i\neq k} W_{\theta_i^\alpha}\left(\sigma_{ik}^\alpha\right).$$

Also the consistency condition means that for any $\alpha \neq \beta$

$$\max_{i\neq k} W_{\theta_k^\beta}\left(\theta_i^\alpha\right) < \rho_\beta - \rho_\alpha. \tag{16}$$

The adaptation condition can be rewritten in terms of the functions W_θ as the combination of (16) and the following inequalities

$$\max_{\alpha:\rho_\alpha\geq\rho_\beta}\max_{i\neq k} W_{\theta_k^\beta}\left(\sigma_{ik}^{\alpha\beta}\right) \leq \rho_\beta = \max_{i\neq k} W_{\theta_k^\beta}\left(\sigma_{ik}^\beta\right). \tag{17}$$

Under the consistency condition $\sigma_{ik}^{\alpha\beta} = s_{ik}^{\alpha\beta}\theta_i^\alpha + \left(1 - s_{ik}^{\alpha\beta}\right)\theta_k^\beta$ with $0 < s_{ik}^{\alpha\beta} < 1$. Therefore, by unimodality $W_{\theta_k^\beta}\left(\sigma_{ik}^{\alpha\beta}\right) > W_{\theta_k^\beta}\left(\theta_i^\alpha\right)$. Because of (17)

$$W_{\theta_k^\beta}\left(\theta_i^\alpha\right) < \rho_\beta, \tag{18}$$

which implies (16).

Thus, the following result was established.

Proposition 1 *The adaptation condition holds if and only if the inequalities (18) and (17) are valid.*

Observe that

$$W_{\theta_k^\beta}\left(\sigma_{ik}^{\alpha\beta}\right) - W_{\theta_i^\alpha}\left(\sigma_{ik}^{\alpha\beta}\right) = \rho_\beta - \rho_\alpha,$$

which directly shows that $W_{\theta_k^\beta}\left(\sigma_{ik}^{\alpha\beta}\right) \le W_{\theta_i^\alpha}\left(\sigma_{ik}^{\alpha\beta}\right)$ if and only if $\rho_\alpha \ge \rho_\beta$.

Assume that $g = A = 2$. Then the inequalities (17) and (18) can be written in a more specific form: if $\rho_1 \le \rho_2$

$$W_{\theta_1^1}\left(\sigma_{12}^{12}\right) \vee W_{\theta_2^1}\left(\sigma_{21}^{12}\right) \vee W_{\theta_1^1}\left(\theta_2^2\right) \vee W_{\theta_2^1}\left(\theta_1^2\right) \le \rho_1,$$

and if $\rho_2 \le \rho_1$,

$$W_{\theta_1^2}\left(\sigma_{21}^{12}\right) \vee W_{\theta_2^2}\left(\sigma_{12}^{12}\right) \vee W_{\theta_1^2}\left(\theta_2^1\right) \vee W_{\theta_2^2}\left(\theta_1^1\right) \le \rho_2.$$

Proposition 2 1. *Let*

$$R_0 = \{\rho_1 < \rho_2, W_{\theta_1^1}\left(\theta_2^2\right) \vee W_{\theta_2^1}\left(\theta_1^2\right) < \rho_1\} \bigcup \{\rho_2 \le \rho_1, W_{\theta_1^2}\left(\theta_2^1\right) \vee W_{\theta_2^2}\left(\theta_1^1\right) < \rho_2\}.$$

Then $(\theta_1^2, \theta_2^2) \in R_0$ *is a necessary adaptation condition, in which case*

$$\left(\theta_1^1 - \theta_2^2\right) \cdot \chi'\left(\theta_2^2\right) \le \left(\theta_1^1 - \theta_2^2\right) \cdot \chi'\left(\sigma_{12}^1\right),$$

$$\left(\theta_2^1 - \theta_1^2\right) \chi'\left(\theta_1^2\right) \le \left(\theta_2^1 - \theta_1^2\right) \cdot \chi'\left(\sigma_{12}^1\right),$$

$$\left(\theta_1^1 - \theta_2^2\right) \cdot \chi'\left(\sigma_{12}^2\right) \le \left(\theta_1^1 - \theta_2^2\right) \cdot \chi'\left(\theta_1^1\right)$$

and

$$\left(\theta_2^1 - \theta_1^2\right) \cdot \chi'\left(\sigma_{12}^2\right) \le \left(\theta_2^1 - \theta_1^2\right) \cdot \chi'\left(\theta_1^2\right).$$

2. *Assume that* $\theta_1^1 - \theta_2^2 = \zeta\left(\theta_1^2 - \theta_2^1\right)$ *with* $\zeta > 0$. *Provided that* $\chi\left(\theta_1^1\right) \ne \chi\left(\theta_2^1\right)$, *an adaptive procedure exists if and only if*

$$\frac{\theta_1^1 - \theta_2^1}{\chi\left(\theta_1^1\right) - \chi\left(\theta_2^1\right)} = \frac{\theta_1^2 - \theta_2^2}{\chi\left(\theta_1^2\right) - \chi\left(\theta_2^2\right)}, \tag{19}$$

in which case

$$\left[\chi\left(\theta_1^1\right) - \chi\left(\theta_2^1\right)\right]\left[\chi\left(\theta_1^2\right) - \chi\left(\theta_2^2\right)\right] > 0,$$

$$\zeta = \frac{\chi\left(\theta_1^1\right) - \chi\left(\theta_2^2\right) + \rho_1 - \rho_2}{\chi\left(\theta_1^2\right) - \chi\left(\theta_2^1\right) + \rho_2 - \rho_1},$$

$$\left(\theta_1^1 - \theta_2^2\right) \cdot \left[\chi^{'}\left(\theta_1^1\right) - \chi^{'}\left(\theta_2^1\right)\right] \geq 0$$

and

$$\left(\theta_1^1 - \theta_2^2\right) \cdot \left[\chi^{'}\left(\theta_1^2\right) - \chi^{'}\left(\theta_2^2\right)\right] \geq 0.$$

Proof.

1. According to (18) in the region R_0^c adaptation cannot happen.

Because of (15) the inequality $W_{\theta_1^1}\left(\theta_2^2\right) \leq W_{\theta_1^1}\left(\sigma_{12}^1\right)$ implies

$$\left(\theta_2^2 - \theta_1^1\right) \cdot \left[\chi^{'}\left(\sigma_{12}^1\right) - \chi^{'}\left(\theta_2^2\right)\right] \leq 0.$$

Similarly, the inequality $W_{\theta_2^2}\left(\theta_1^1\right) \leq W_{\theta_2^2}\left(\sigma_{12}^2\right)$ shows that

$$\left(\theta_1^1 - \theta_2^2\right) \cdot \left[\chi^{'}\left(\sigma_{12}^2\right) - \chi^{'}\left(\theta_1^1\right)\right] \leq 0$$

and the other inequalities in 1. follows in the same way.

2. As in 1. the inequality $W_{\theta_1^1}\left(\sigma_{12}^{12}\right) \leq W_{\theta_1^1}\left(\sigma_{12}^1\right)$ leads to

$$\left(\sigma_{12}^{12} - \theta_1^1\right) \cdot \left[\chi^{'}\left(\sigma_{12}^1\right) - \chi^{'}\left(\sigma_{12}^{12}\right)\right] \leq 0.$$

According to the adaptation condition σ_{12}^{12} is a convex combination of θ_1^1 and θ_2^2. Thus,

$$\left(\theta_2^2 - \theta_1^1\right) \cdot \chi^{'}\left(\sigma_{12}^1\right) \leq \left(\theta_2^2 - \theta_1^1\right) \cdot \chi^{'}\left(\sigma_{12}^{12}\right) = \chi\left(\theta_2^2\right) - \chi\left(\theta_1^1\right) + \rho_2 - \rho_1.$$

Similarly the inequalities $W_{\theta_2^1}\left(\sigma_{21}^{12}\right) \leq W_{\theta_2^1}\left(\sigma_{12}^1\right)$, $W_{\theta_1^2}\left(\sigma_{21}^{12}\right) \leq W_{\theta_1^2}\left(\sigma_{12}^2\right)$ and $W_{\theta_2^2}\left(\sigma_{12}^{12}\right)$ $\leq W_{\theta_2^2}\left(\sigma_{12}^2\right)$ imply that

$$\left(\theta_1^2 - \theta_2^1\right) \cdot \chi^{'}\left(\sigma_{12}^1\right) \leq \left(\theta_1^2 - \theta_2^1\right) \cdot \chi^{'}\left(\sigma_{21}^{12}\right),$$

$$\left(\theta_2^1 - \theta_1^2\right) \cdot \chi^{'}\left(\sigma_{12}^2\right) \leq \left(\theta_2^1 - \theta_1^2\right) \cdot \chi^{'}\left(\sigma_{21}^{12}\right),$$

and

$$\left(\theta_1^1 - \theta_2^2\right) \cdot \chi^{'}\left(\sigma_{12}^2\right) \leq \left(\theta_1^1 - \theta_2^2\right) \cdot \chi^{'}\left(\sigma_{12}^{12}\right).$$

By combining these inequalities one obtains

$$\left(\theta_1^1 - \theta_2^2\right) \cdot \chi^{'}\left(\sigma_{12}^2\right) \leq \left(\theta_1^1 - \theta_2^2\right) \cdot \chi^{'}\left(\sigma_{12}^{12}\right) \leq \left(\theta_1^1 - \theta_2^2\right) \cdot \chi^{'}\left(\sigma_{12}^1\right)$$

and

$$\left(\theta_1^2 - \theta_2^1\right) \cdot \chi^{'}\left(\sigma_{12}^1\right) \leq \left(\theta_1^2 - \theta_2^1\right) \cdot \chi^{'}\left(\sigma_{21}^{12}\right) \leq \left(\theta_1^2 - \theta_2^1\right) \cdot \chi^{'}\left(\sigma_{12}^2\right).$$

Under the condition of linear dependence in Proposition 2 all these inequalities reduce to equalities, so that

$$\chi\left(\theta_1^1\right) - \chi\left(\theta_2^2\right) + \rho_1 - \rho_2 = \left(\theta_1^1 - \theta_2^2\right) \cdot \chi^{'}\left(\sigma_{12}^{12}\right)$$

$$= \left(\theta_1^1 - \theta_2^2\right) \cdot \chi^{'}\left(\sigma_{12}^1\right) = \left(\theta_1^1 - \theta_2^2\right) \cdot \chi^{'}\left(\sigma_{12}^2\right) = \left(\theta_1^1 - \theta_2^2\right) \cdot \chi^{'}\left(\sigma_{21}^{12}\right).$$

and

$$\chi\left(\theta_2^1\right) - \chi\left(\theta_1^2\right) + \rho_1 - \rho_2 = \left(\theta_2^1 - \theta_1^2\right) \cdot \chi^{'}\left(\sigma_{21}^{12}\right)$$

$$= \left(\theta_2^1 - \theta_1^2\right) \cdot \chi^{'}\left(\sigma_{12}^2\right) = \left(\theta_2^1 - \theta_1^2\right) \cdot \chi^{'}\left(\sigma_{12}^2\right) = \left(\theta_2^1 - \theta_1^2\right) \cdot \chi^{'}\left(\sigma_{12}^{12}\right).$$

According to derivation of (15), one must have

$$W_{\sigma_{12}^{12}}\left(\sigma_{12}^1\right) = W_{\sigma_{21}^{12}}\left(\sigma_{12}^1\right) = W_{\sigma_{12}^{12}}\left(\sigma_{12}^2\right) = W_{\sigma_{21}^{12}}\left(\sigma_{12}^2\right) = 0,$$

i.e.

$$\sigma_{12}^1 = \sigma_{12}^2 = \sigma_{21}^{12} = \sigma_{12}^{12} = \overline{\sigma}.$$

Therefore these convex combinations of the vectors $\theta_1^1, \theta_1^2, \theta_1^2, \theta_2^2$ coincide, while the vectors $\theta_1^1 - \theta_2^2$ and $\theta_1^2 - \theta_2^1$ are linearly dependent. Therefore $\theta_1^1 - \theta_2^1, \theta_1^1 - \theta_1^2$ and $\theta_1^1 - \theta_2^2$ are linearly dependent. It follows that $\theta_1^1 - \theta_2^1 = \kappa(\theta_1^2 - \theta_2^2)$. Since a convex combination of θ_1^2 and θ_2^2 coincides with a convex combination of θ_1^1 and θ_2^2, κ is positive. This fact and the identity

$$\chi\left(\theta_1^1\right) - \chi\left(\theta_2^1\right) = \left(\theta_1^1 - \theta_2^1\right) \cdot \chi^{'}\left(\overline{\sigma}\right)$$

$$= \kappa\left(\theta_1^2 - \theta_2^2\right) \cdot \chi^{'}\left(\overline{\sigma}\right) = \kappa\left[\chi\left(\theta_1^2\right) - \chi\left(\theta_2^2\right)\right]$$

show that (19) is true. If $\chi\left(\theta_1^1\right) = \chi\left(\theta_2^1\right)$, then necessarily $\chi\left(\theta_1^2\right) = \chi\left(\theta_2^2\right)$.

Under any of these conditions the rule

$$\{\delta_a(x_1, \ldots, x_n) = 1\} = \left\{\theta_1^1 \cdot \overline{x} - \chi\left(\theta_1^1\right) > \theta_2^1 \cdot \overline{x} - \chi\left(\theta_2^1\right)\right\}$$

$$= \left\{\theta_1^2 \cdot \overline{x} - \chi\left(\theta_1^2\right) > \theta_2^2 \cdot \overline{x} - \chi\left(\theta_2^2\right)\right\}$$

is the maximum likelihood procedure for both $\alpha = 1$ and $\alpha = 2$, so that it must be adaptive.

Since

$$\chi\left(\theta_2^2\right) - \chi\left(\theta_1^1\right) + \rho_2 - \rho_1 = \left(\theta_2^2 - \theta_1^1\right) \cdot \chi^{'}\left(\overline{\sigma}\right)$$

$$= \zeta\left(\theta_2^1 - \theta_1^2\right) \cdot \chi^{'}\left(\overline{\sigma}\right) = \zeta\left[\chi\left(\theta_2^1\right) - \chi\left(\theta_1^2\right) + \rho_1 - \rho_2\right],$$

the formula for ζ follows.

As in the proof of 1. one obtains from the fact that $\sigma_{12}^1 = \sigma_{12}^{12} = \overline{\sigma} = \sigma_{12}^2 = \sigma_{12}^{12}$

$$\left(\theta_2^2 - \theta_1^1\right) \cdot \chi^{'}\left(\theta_1^1\right) \leq \chi\left(\theta_2^2\right) - \chi\left(\theta_1^1\right) + \rho_2 - \rho_1$$

$$\leq \left(\theta_2^2 - \theta_1^1\right) \cdot \chi^{'}\left(\theta_2^2\right).$$

Similarly,

$$\left(\theta_2^1-\theta_1^2\right)\chi^{'}\left(\theta_1^2\right)\le\chi\left(\theta_2^1\right)-\chi\left(\theta_1^2\right)+\rho_1-\rho_2$$
$$\le\left(\theta_2^1-\theta_1^2\right)\cdot\chi^{'}\left(\theta_2^1\right).$$

The above formula for ζ establishes now two last inequalities of Proposition 2. □

The most illuminating form of the adaptation region from 2. of Proposition 2 is when $p=1$ (see Rukhin, 1997). In this case assuming that $\theta_1^1<\theta_2^2$ if $\theta_1^1<\theta_2^2,\theta_1^2<\theta_2^1$, adaptation occurs if and only if (19) is valid with a non-negative right-hand side. If $\theta_1^1>\theta_2^2,\theta_1^2>\theta_2^1$, no adaptive rule exists.

To find the adaptation region R^0 for the rule δ_0 one can use Theorem 3 with $b=0$. A simpler form of this region holds when $p=1$. Then (cf Rukhin, 1982) provided that $\theta_i^\alpha\ne\theta_i^\beta$ for $\alpha\ne\beta$,

$$\max_{i\ne k}\max_{\beta,\beta\ne\alpha}\inf_{s_\gamma\ge 0,\gamma=1,\ldots,A}\left\{\log E_i^\alpha\prod_\gamma\left[\frac{f_k^\beta}{f_i^\gamma}(X)\right]^{s_\gamma}\right\}$$
$$=\max_{i\ne k}\max_{\beta,\beta\ne\alpha}\min_\gamma\inf_{s\ge 0}\left\{\chi\left(\theta_i^\alpha+s(\theta_k^\beta-\theta_i^\gamma)\right)-\chi\left(\theta_i^\alpha\right)-s\left[\chi\left(\theta_k^\beta\right)-\chi\left(\theta_i^\gamma\right)\right]\right\}.$$

Thus when $p=1$, (11) is also necessary for adaptation.

Let $\tau_{ik}^{\alpha\beta}$ belonging to the interval with the end-points θ_i^α and θ_k^β, be defined by the formula

$$(\theta_i^\alpha-\theta_k^\beta)\cdot\chi^{'}\left(\tau_{ik}^{\alpha\beta}\right)=\chi\left(\theta_i^\alpha\right)-\chi\left(\theta_k^\beta\right), \tag{20}$$

According to (20), $\tau_{ik}^{\alpha\beta}=\tau_{ki}^{\beta\alpha}$, $\tau_{ik}^{\alpha\alpha}=\sigma_{ik}^\alpha$ and $W_{\theta_k^\beta}\left(\tau_{ik}^{\alpha\beta}\right)=W_{\theta_i^\alpha}\left(\tau_{ik}^{\alpha\beta}\right)$.

Also

$$\left(\theta_i^\alpha-\theta_k^\beta\right)\cdot\left(\chi^{'}(\tau_{ik}^{\alpha\beta})-\chi^{'}(\sigma_{ik}^{\alpha\beta})\right)=\rho_\alpha-\rho_\beta.$$

Thus, in this case the adaptation condition for δ_0 can be written in the following form. For any β

$$\max_{i\ne k}\max_{\alpha:\rho_\alpha\ge\rho_\beta}\min_\gamma\inf_{s>0}\left[\chi(\theta_k^\beta+s(\theta_i^\alpha-\theta_k^\gamma))-\chi(\theta_k^\beta)-s\chi(\theta_i^\alpha)+s\chi(\theta_k^\gamma)\right]\le\rho_\beta.$$

The infimum above is attained when $\theta_k^\beta+s(\theta_i^\alpha-\theta_k^\gamma)=\tau_{ik}^{\alpha\gamma}$, which corresponds to positive s if and only if

$$\left(\tau_{ik}^{\alpha\gamma}-\theta_k^\beta\right)(\theta_i^\alpha-\theta_k^\gamma)>0. \tag{21}$$

Denote by $\Gamma_{ik}^{\alpha\beta}$ the set of all γ for which (21) holds. Then always $\beta\in\Gamma_{ik}^{\alpha\beta}$.

Thus δ_0 is adaptive if and only if

$$\max_{\alpha:\rho_\alpha\ge\rho_\beta}\max_{i\ne k}\min_{\gamma\in\Gamma_{ik}^{\alpha\beta}}W_{\theta_k^\beta}(\tau_{ik}^{\alpha\gamma})\le\rho_\beta=\max_{i\ne k}W_{\theta_k^\beta}\left(\sigma_{ik}^\beta\right). \tag{22}$$

When $A=2$, for $\gamma=\alpha$, (21) takes the form

$$\left(\sigma_{ik}^\alpha-\theta_k^\beta\right)(\theta_i^\alpha-\theta_k^\alpha)>0,$$

and for $\gamma=\beta$ it holds automatically.

5 Further Examples

5.1 Poisson Family

The distribution F_i^α is a Poisson distribution with the parameter λ_i^α, $\lambda_i^\alpha > 0$. Thus

$$f_i^\alpha(u) = e^{-\lambda_i^\alpha} \frac{(\lambda_i^\alpha)^u}{u!}, \quad u = 0, 1 \ldots$$

and

$$\theta_i^\alpha = \log \lambda_i^\alpha,$$
$$\chi(\theta_i^\alpha) = \exp(\theta_i^\alpha) = \lambda_i^\alpha.$$

We use the mean value parameterization and employ the function

$$\mathbf{P}_\lambda(\eta) = \eta - \lambda - \eta \log \frac{\eta}{\lambda},$$

which replaces $W_\theta(t)$ in the adaptation conditions.

One has

$$\rho_\alpha = \max_{i \neq k} \mathbf{P}_{\lambda_i^\alpha}(\eta_{ik}^\alpha)$$

with

$$\eta_{ik}^\alpha = \exp(\sigma_{ik}^\alpha) = \frac{\lambda_i^\alpha - \lambda_k^\alpha}{\log \frac{\lambda_i^\alpha}{\lambda_k^\alpha}}.$$

Let

$$\eta_{ik}^{\alpha\beta} = \exp(\sigma_{ik}^{\alpha\beta}) = \frac{\lambda_i^\alpha - \lambda_k^\beta + \rho_\alpha - \rho_\beta}{\log \frac{\lambda_i^\alpha}{\lambda_k^\beta}}.$$

According to (16) and (17) an adaptive procedure exists if and only if

$$\max_{\alpha,\beta:\rho_\beta \geq \rho_\alpha} \max_{i \neq k} \left[\mathbf{P}_{\lambda_i^\alpha}\left(\lambda_k^\beta\right) - \rho_\alpha + \rho_\beta\right] < 0$$

and

$$\max_{\beta:\rho_\beta \geq \rho_\alpha} \max_{i \neq k} \mathbf{P}_{\lambda_i^\alpha}\left(\eta_{ik}^{\alpha\beta}\right) \leq \rho_\alpha.$$

According to (22) the adaptation region R^0 has the form

$$\max_{\beta:\rho_\beta \geq \rho_\alpha} \max_{i \neq k} \min_{\gamma:(v_{ik}^{\beta\gamma} - \lambda_k^\alpha)(\lambda_i^\beta - \lambda_k^\gamma) > 0} \mathbf{P}_{p_k^\alpha}\left(v_{ik}^{\beta\gamma}\right) \leq \rho_\alpha$$

with

$$v_{ik}^{\alpha\beta} = \exp(\tau_{ik}^{\alpha\beta}) = \frac{\lambda_i^\alpha - \lambda_k^\beta}{\log \frac{\lambda_i^\alpha}{\lambda_k^\beta}}.$$

The Figure 2 shows the adaptation regions for δ_a and δ_0 when $\lambda_1^1 = 1, \lambda_2^1 = 2$. The procedure δ_0 is not adaptive when δ_a is, for the values of λ_1^2, λ_2^2 from the shells in the lower left corner and in the upper right corner.

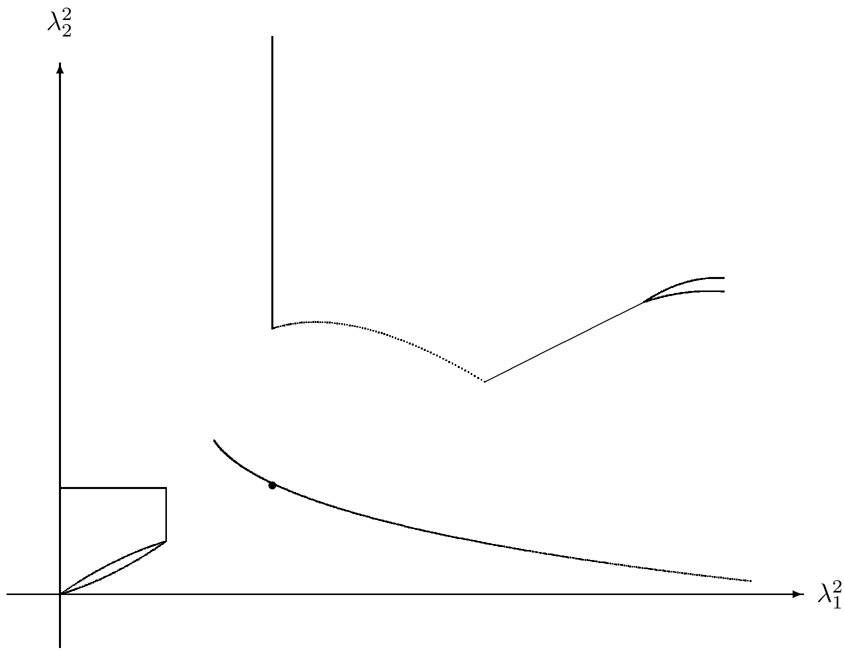

Figure 2: The adaptation regions for δ_0 and δ_a with Poisson distributions for $\lambda_1^1 = 1, \lambda_2^1 = 2$.

5.2 Binomial Family

The distribution F_i^α is a binomial distribution with the probability of success p_i^α, $0 < p_i^\alpha < 1$. Thus

$$f_i^\alpha(u) = \left(\begin{array}{c} N \\ u \end{array} \right) (p_i^\alpha)^u \, (1 - p_i^\alpha)^{N-u}\,, \quad u = 0, 1 \ldots, N$$

and

$$\theta_i^\alpha = \log \frac{p_i^\alpha}{1 - p_i^\alpha},$$

$$\chi(\theta_i^\alpha) = N \log(1 + \exp(\theta_i^\alpha)) = -N \log(1 - p_i^\alpha).$$

All adaptation conditions are the same for all values of N, which therefore can be taken to be equal to 1.

As in the previous example, it is more convenient to use the mean value parameterization and to work instead of

$$W_\theta(t) = \log \left(\frac{1 + \exp(\theta)}{1 + \exp(t)} \right) - (\theta - t) \frac{\exp(\theta)}{1 + \exp(\theta)}$$

with the classical entropy function

$$\mathbf{H}_p(q) = -\left[q \log \frac{q}{p} + (1-q)\log \frac{1-q}{1-p}\right]$$

for $t = \log q - \log(1-q)$ and $\theta = \log p - \log(1-p)$.

One has

$$\rho_\alpha = \max_{i \neq k} \mathbf{H}_{p_i^\alpha}\left(r_{ik}^\alpha\right)$$

with

$$r_{ik}^\alpha = \frac{\log \frac{1-p_i^\alpha}{1-p_k^\alpha}}{\log \frac{p_k^\alpha(1-p_i^\alpha)}{p_i^\alpha(1-p_k^\alpha)}}.$$

Let

$$r_{ik}^{\alpha\beta} = \frac{\log \frac{1-p_i^\alpha}{1-p_k^\beta} + \rho_\alpha - \rho_\beta}{\log \frac{p_k^\beta(1-p_i^\alpha)}{p_i^\alpha(1-p_k^\beta)}}.$$

According to (16) and (17) an adaptive procedure exists if and only if

$$\max_{\alpha,\beta:\rho_\beta \geq \rho_\alpha} \max_{i \neq k} \left[\mathbf{H}_{p_i^\alpha}\left(p_k^\beta\right) - \rho_\alpha + \rho_\beta\right] < 0$$

and

$$\max_{\beta:\rho_\beta \geq \rho_\alpha} \max_{i \neq k} \mathbf{H}_{p_i^\alpha}\left(r_{ik}^{\alpha\beta}\right) \leq \rho_\alpha.$$

If

$$q_{ik}^{\alpha\beta} = \frac{\exp(\tau_{ik}^{\alpha\beta})}{1+\exp(\tau_{ik}^{\alpha\beta})} = \frac{\log \frac{1-p_k^\beta}{1-p_i^\alpha}}{\log \frac{p_i^\alpha(1-p_k^\beta)}{p_k^\beta(1-p_i^\alpha)}},$$

the adaptation region R^0 because of (22) has the form

$$\max_{\beta:\rho_\beta \geq \rho_\alpha} \max_{i \neq k} \min_{\gamma:(q_{ik}^{\beta\gamma} - p_k^\alpha)(p_i^\beta - p_k^\gamma) > 0} \mathbf{H}_{p_k^\alpha}\left(q_{ik}^{\beta\gamma}\right) \leq \rho_\alpha.$$

The Figure 3 shows the adaptation region in the lower left corner and in the right upper corner when $p_1^1 = 1/3, p_2^1 = 2/3$.

5.3 Multivariate Normal Family

The distribution F_i^α is the multivariate normal with the mean η_i^α and the nonsingular covariance matrix R, the same for all i and α. Then

$$\theta_i^\alpha = R^{-1}\eta_i^\alpha$$

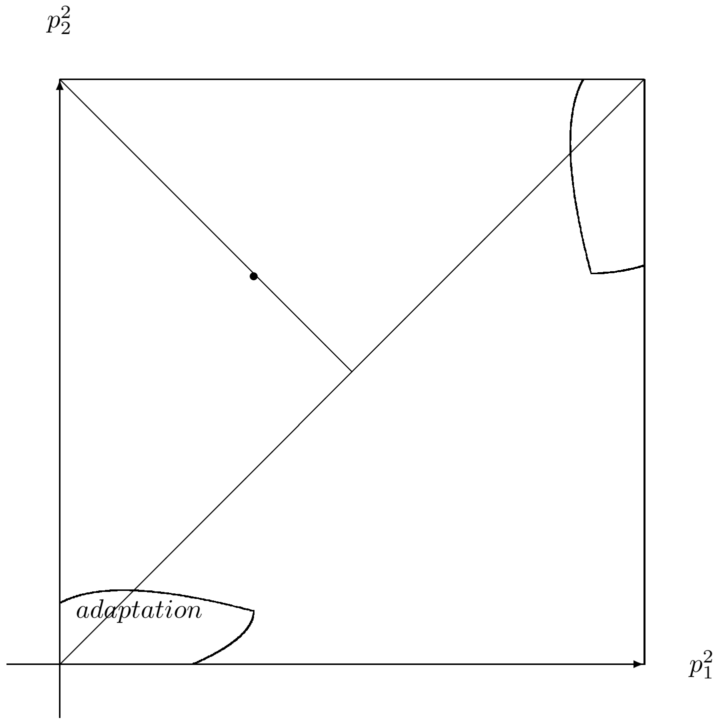

Figure 3: The adaptation regions for binomial distributions for $p_1^1 = 1/3, p_2^1 = 2/3$.

and

$$\chi(\theta_i^\alpha) = \frac{1}{2}\eta_i^\alpha \cdot R^{-1}\eta_i^\alpha = \frac{1}{2}\theta_i^\alpha \cdot R\theta_i^\alpha = \frac{1}{2}||\theta_i^\alpha||_R^2.$$

One has $W_\theta(t) = -||\theta - t||_R^2/2$, and according to (12),

$$\rho_\alpha = \frac{1}{2}\max_{i\neq k}\inf_{s>0}\left[||s\theta_i^\alpha + (1-s)\theta_k^\alpha||_R^2 - 2s||\theta_i^\alpha||_R^2 - 2(1-s)||\theta_k^\alpha||_R^2\right]$$

$$= -\frac{1}{8}\min_{i\neq k}||\theta_i^\alpha - \theta_k^\alpha||_R^2 = -\frac{1}{8}\min_{i\neq k}\left(\eta_i^\alpha - \eta_k^\alpha\right)\cdot R^{-1}\left(\eta_i^\alpha - \eta_k^\alpha\right). \tag{23}$$

Now we determine the explicit form of (17). For fixed $\alpha \neq \beta, i \neq k$ and $\theta_i^\alpha \neq \theta_k^\beta$

$$s_{ik}^{\alpha\beta} = \frac{\rho_\alpha - \rho_\beta + ||\theta_i^\alpha||_R^2/2 - ||\theta_k^\beta||_R^2/2 - \theta_k^\beta \cdot R\left(\theta_i^\alpha - \theta_k^\beta\right)}{||\theta_i^\alpha - \theta_k^\beta||_R^2}$$

$$= \frac{1}{2} + \frac{\rho_\alpha - \rho_\beta}{||\theta_i^\alpha - \theta_k^\beta||_R^2},$$

so that

$$1 - s_{ik}^{\alpha\beta} = \frac{1}{2} - \frac{\rho_\alpha - \rho_\beta}{||\theta_i^\alpha - \theta_k^\beta||_R^2}.$$

The condition $0 < s_{ik}^{\alpha\beta} < 1$, which is necessary for adaptation because of (10), means that δ_a is consistent and corresponds to the region R_0 discussed in Proposition 2. It can be rewritten in the form $s_{ik}^{\alpha\beta}\left(1 - s_{ik}^{\alpha\beta}\right) > 0$, which means that

$$2|\rho_\alpha - \rho_\beta| < ||\theta_i^\alpha - \theta_k^\beta||_R^2.$$

One has

$$W_{\theta_k^\beta}\left(\sigma_{ik}^{\alpha\beta}\right) = -\frac{||\theta_i^\alpha - \theta_k^\beta||_R^2}{2}\left[\frac{1}{2} - \frac{\rho_\alpha - \rho_\beta}{||\theta_i^\alpha - \theta_k^\beta||_R^2}\right],$$

so that if $\rho_\alpha \geq \rho_\beta$, condition (17) means that for $i \neq k$

$$\frac{||\theta_i^\alpha - \theta_k^\beta||_R}{2} - \frac{\rho_\alpha - \rho_\beta}{||\theta_i^\alpha - \theta_k^\beta||_R} \geq \sqrt{2|\rho_\beta|}.$$

This inequality signifies that for all $i \neq k$

$$(24) \quad ||\theta_i^\alpha - \theta_k^\beta||_R \geq \sqrt{2|\rho_\alpha|} + \sqrt{2|\rho_\beta|} = \frac{1}{2}\left[\min_{\ell \neq m} ||\theta_\ell^\alpha - \theta_m^\alpha||_R + \min_{\ell \neq m} ||\theta_\ell^\beta - \theta_m^\beta||_R\right].$$

Thus (24) provides the necessary and sufficient condition for the existence of an adaptive rule in this example.

Under this condition, the adaptive rule δ_a has the form

$$\{\delta_a(x_1, \ldots, x_n) = i\} = \{\min_\alpha \left[(\overline{x} - \eta_i^\alpha) \cdot R^{-1}(\overline{x} - \eta_i^\alpha) + 2\rho_\alpha\right]$$

$$= \min_{\alpha,k} \left[(\overline{x} - \eta_k^\alpha) \cdot R^{-1}(\overline{x} - \eta_k^\alpha) + 2\rho_\alpha\right]\}$$

with ρ_α given in (23).

The overall maximum likelihood rule δ_0 has the form

$$\{\delta_0(x_1, \ldots, x_n) = i\} = \{\min_\alpha \left[(\overline{x} - \eta_i^\alpha) \cdot R^{-1}(\overline{x} - \eta_i^\alpha)\right]$$

$$= \min_{\alpha,k} \left[(\overline{x} - \eta_k^\alpha) \cdot R^{-1}(\overline{x} - \eta_k^\alpha)\right]\}.$$

The adaptation region R^0 for the rule δ_0 is described by condition (11) according to which for any $\alpha \neq \beta$ and $i \neq k$

$$\min_{s_1,\ldots,s_A > 0}\left[||\theta_i^\alpha + \sum_m s_m(\theta_k^\beta - \theta_m^\gamma)||_R^2 + \sum_m s_m\left(||\theta_m^\gamma||_R^2 - ||\theta_k^\beta||_R^2\right) - ||\theta_i^\alpha||_R^2\right]$$

$$\leq 2\rho_\alpha. \tag{25}$$

Let V_{km}^{β} denote the Gram matrix corresponding to vectors $\theta_k^\beta - \theta_m^\gamma, \gamma = 1, \ldots, A$ with respect to the inner product determined by $||\cdot||_R^2$. Also denote by $v_{ik}^{\alpha\beta}$ the vector whose γ-th coordinate has the form $\theta_i^\alpha \cdot R\left(\theta_i^\gamma - \theta_k^\beta\right) + ||\theta_k^\beta||_R^2/2 - ||\theta_i^\alpha||_R^2/2$. Assuming that V_{ki}^{β} is an invertible matrix and the vector $\left[V_{ki}^{\beta}\right]^{-1} v_{ik}^{\alpha\beta}$ has non-negative coordinates, one obtains the adaptation condition for δ_0

$$\min_{k\neq i} \min_{\beta\neq\alpha} \left[v_{ik}^{\alpha\beta}\right]^T \cdot R \left[V_{ki}^{\beta}\right]^{-1} v_{ik}^{\alpha\beta} \geq -2\rho_\alpha. \tag{26}$$

If the vectors $\theta_k^\beta - \theta_i^\gamma, \gamma = 1, \ldots, A$ are linearly dependent and r denotes the rank of V_{ki}^{β}, then minimum in (25) is attained at r-dimensional boundary of the positive orthant $\{s_1, \ldots, s_A > 0\}$, and the condition (26) can be rewritten in terms of Gram matrices of linearly independent vectors $\theta_k^\beta - \theta_i^\gamma$.

5.4 Normal Family with Unknown Mean and Variance

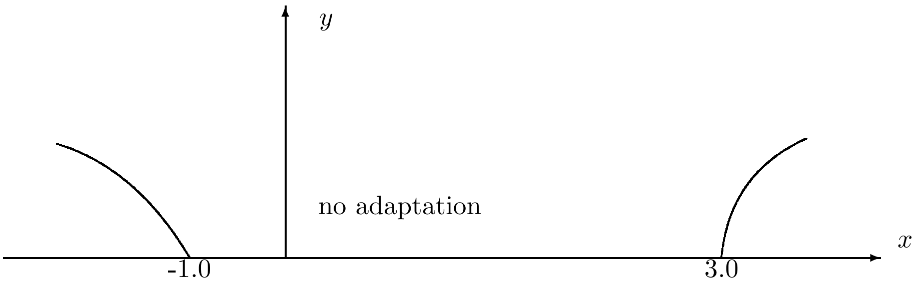

Figure 4: The adaptation regions for normal distributions.

The distribution F_i^α on the real line is normal with the mean η_i^α and the variance κ_i^α. The vector of natural parameters of the corresponding two-parameter exponential family with $u = (x, -x^2/2)^T$ has the form

$$\theta_i^\alpha = (v_i^\alpha, w_i^\alpha)^T = (\eta_i^\alpha/\kappa_i^\alpha, 1/\kappa_i^\alpha)^T,$$

so that with $\theta = (v, w)^T$

$$\chi(\theta) = \frac{1}{2}\left[\frac{v^2}{w} - \log w\right].$$

One has with $t = (t_1, t_2)^T$

$$2W_\theta(t) = \frac{t_1^2}{t_2} - \log t_2 - \frac{v^2}{w} + \log w - \frac{2t_1}{t_2}\left[t_1 - v\right] + \left[\frac{t_1^2}{t_2^2} + \frac{1}{t_2}\right](t_2 - w)$$

$$= -w\left(\frac{t_1}{t_2} - \frac{u}{w}\right)^2 - \frac{w}{t_2} + \log\frac{w}{t_2} + 1.$$

The Figure 4 shows the adaptation region for $\theta_1^1 = (-1,1)^T, \theta_2^1 = (1,1)^T, \theta_1^2 = (x,y)^T$ and $\theta_2^2 = (x-2,y)^T$, so that $\theta_1^1 - \theta_2^2 = \theta_2^1 - \theta_1^2$. Thus, in the notation of Proposition 2, $\zeta = -1$. In this situation $\sigma_{12}^1 = (0,1)^T$ with $\rho_1 = -1/2$.

Bibliography

[1] P. J. Bickel, C. A. J. Klaassen, Y. Ritov, J. A. Wellner. Efficient and Adaptive Estimation for Semiparametric Models. Johns Hopkins University Press, Baltimore, 1993.

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

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

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

[5] A. Renyi. On some problems of statistics from the point of view of information theory. In: Proceedings of the Colloquium on Information Theory, Vol. **2**, Budapest, Bolyai Math. Soc., 1970.

[6] A. L. Rukhin. Adaptive procedures for a finite number of probability distributions families. In: Proceedings of Third Purdue Symposium on Decision Theory (eds. S. Gupta, J. Berger), Academic Press, New York, 1982.

[7] A. L. Rukhin. Adaptive classification procedures. *Journ. Amer. Statist. Assoc.* **79** (1984), 415-422.

[8] A. L. Rukhin. Adaptive multiple decision procedures for exponential families. In: Advances in Statistical Decision Theory (eds. S. Panchapakesan, N. Balakrishnan), Birkhauser, Boston, 1997, 99-118.

[9] C. Stein. Efficient nonparametric testing and estimation. In: Proc. Third Berkeley Symp. Math. Statist. Probab. 1, University of California Press, 1956, 187-196.

Department of Mathematics & Statistics UMBC
Baltimore, MD 21250, USA
e-mail: `rukhin@math.umbc.edu`