Title

On the Minimax Approach to Overcoming Prior Uncertainty and Application to Pattern Recognition Problems

Author
Alexander Tartakovsky

Publication Date
2011-10-24

The usual consistent estimator of $\Delta\rho_i(\vec{\mu})$ is

$$\Delta\hat{\rho}_i^{(n)}(\vec{\mu}) = \frac{1}{n}\sum_{\nu=1}^{n}\xi_i(X_\nu),$$

where $\{X_\nu,\ \nu = 1,\dots,n\}$ is a sequence of random variables with probability distribution $\bar{P}_i(dx)$. Evidently, a consistent estimator of $F(\vec{\mu})$ is $\hat{F}_n(\vec{\mu}) = \max_i \Delta\hat{\rho}_i^{(n)}(\vec{\mu})$. This method can reduce computational time by the factor of tens and hundreds.

References

[1] Repin, V.G. and Tartakovsky, G.P. (1978). *Statistical Synthesis Under Prior Uncertainty and Adaptation of Information Systems.* Sovetskoe Radio, Moscow (In Russian).

[2] Wald, A. (1950). *Statistical Decision Functions.* John Wiley & Sons, Inc., New York.

[3] Tartakovsky, A.G. (1991). *Sequential Methods in the Theory of Information Systems.* Radio i Svyaz, Moscow (In Russian).

[4] Tartakovskii, G.P. (1984). Multiple-alternative adaptive pattern recognition with parameter estimation, when non-single-valued decisions are allowed. *Problems of Information Transmission*, **20**, 44-51.

[5] Verdenskaya, N.V. and Tartakovskii, A.G. (1991). Asymptotically optimal sequential testing multiple hypotheses for nonhomogeneous Gaussian processes in an asymmetric situation. *Theory of Probability and Its Application* **36** 536–547.

where $F(\vec{\mu}) = \max_{i\in\{1,\ldots,k\}} \Delta\rho\left(D_0(x,\Pi(\vec{\mu})),\vec{p}_i\right),\ k \le N$ and $D_0(x,\Pi(\vec{\mu}))$ is a Bayes test with respect to the *a priori* distribution

$$\Pi(\vec{\mu}) = \left(\sum_{i\in I^*} \mu_i p_{1i}, \ldots, \sum_{i\in I^*} \mu_i p_{ki}\right).$$

Thus the problem of the search for a least favorable distribution $\vec{\mu}^*$ is reduced to minimization of the function $F(\vec{\mu})$. If this function is continuous one can use standard numerical minimization methods (e.g., gradient or other minimum–seaking techniques). However, if the sets where likelihood ratios are constant have non-zero probability (i.e., the distributions of likelihood ratios are discontinuous), then, as we could see, the function $F(\vec{\mu})$ in general is discontinuous and, moreover, its values at the points $\vec{\mu}_r$ which correspond to the randomization, generally speaking, are not equal to the limit of $F(\vec{\mu})$ as we tend to these points in any direction. In this case we cannot apply standard gradient methods as well as the others without special modification. The minimization of $F(\vec{\mu})$ should include, therefore, a preliminary stage of the search for all values of $\vec{\mu}$ for which the Bayes test $D_0(x,\Pi(\vec{\mu}))$ requires randomization, a special investigation of $F(\vec{\mu})$ at these points and finally the search for minimum at the continuity points of $F(\vec{\mu})$ by standard methods.

A second question of practical importance is the calculation of the function $F(\vec{\mu})$. Sometimes this includes a calculation of complicated integrals in high-dimensional spaces. It may be more efficient to compute $F(\vec{\mu})$ by the Monte Carlo method. In fact,

$$\Delta\rho_i(\vec{\mu}) = \int_{\mathcal{X}} \sum_{l=1}^{N}\sum_{j=1}^{M} \delta_j^0(x,\Pi(\vec{\mu}))\tilde{L}_{lj} p_{il} dP_l(x)$$

$$= \int_{\mathcal{X}} \xi_i(x) d\bar{P}_i(x) = \bar{E}_i \xi_i(X),$$

where $\bar{P}_i(dx) = \sum_{l=1}^{N} P_l(dx) p_{il}$ is the mixture of probability measures with weights $\{p_{il}\}$, $\bar{E}_i$ is the corresponding expectation,

$$\xi_i(x) = \sum_{l=1}^{N}\sum_{j=1}^{M} \delta_j^0(x,\Pi(\vec{\mu}))\tilde{L}_{lj}\pi_l^{(i)}(x),$$

$\pi_l^{(i)}(x) = P_l(dx)p_{il}/\bar{P}_i(dx)$ is the corresponding *a posteriori* probability, $\tilde{L}_{lj} = L_{lj} - \min_s L_{sj}$.

Consequently, the distribution function $F_0(y)$ of statistic $Y(X)$ is given by

$$F_0(y) = \prod_{i=1}^{N} \Phi\left(\frac{\ln(C^*/\alpha_i) + m_i(T)}{\sqrt{2m_i(T)}}\right),$$

where $m_i(T) = \frac{1}{2}\int_0^T \tilde{A}_i^2(t)dt$. If again $\alpha_i = 1/N$ and $m_i(T) = m(T)$, then equation (83) for the threshold C^* has the form

$$\frac{N}{2\sqrt{\pi m(T)}} \int_0^{C^*} \Phi^{N-1}\left(\frac{\ln(Ny) + m(T)}{\sqrt{2m(T)}}\right) \exp\left\{-\frac{[\ln(Ny) + m(T)]^2}{4m(T)}\right\} dy$$

$$= \frac{L_0}{L}\left[1 - \Phi\left(\frac{\ln(NC^*) + m(T)}{\sqrt{2m(T)}}\right)\right]$$

or after manipulations

$$\frac{\exp\{-m(T)\}}{\sqrt{2\pi}} \int_{-\infty}^{h} \exp(\sqrt{2m(T)}y)\Phi^{N-1}(y)\exp(-y^2/2)dy = \frac{L_0}{L}[1 - \Phi(h)].$$

This equation can be simplified in the case of large values of $m(T)$.

5 Computational Aspects

In previous sections we have found the solutions of the minimax problems in closed form in the more or less general conditions when the set I^* where a least fovorable distribution is strictly positive can be determined in advance. However, in some cases this set is not known in advance and in order to find the structure of the MAR test and its parameters it is necessary to solve the system of equalities and inequalities (13) and (14) (or even (10) and (11)). The problem of Section 4.4 can serve as an example. The solution of this system which represents a least favorable distibution $\vec{\mu}^* = \{\mu_i^*\}_{i\in I^*}$ ($\mu_i^* > 0$ for $i \in I^*$, $\sum_{i\in I^*} \mu_i^* = 1$) in the general case can be written as follows

$$\vec{\mu}^* = \arg\min_{\vec{\mu}} F(\vec{\mu}),$$

is also weakly depends on N.

Suppose now that the components $X_i = X_i(t)$ represent continuous-time Gaussian processes observed on the interval $[0, T]$, $T < \infty$ with stochastic differentials

$$dX_i(t) = \begin{cases} [A_i(t) + \eta_i(t)]dt + dW_i(t), & \text{if } H_i \text{ is true,} \\ \eta_i(t)dt + dW_i(t), & \text{if } H_k,\ k \neq i \text{ or } H_0 \text{ are true,} \end{cases}$$

where $X_i(0) = 0$, $A_i(t)$ are some deterministic functions (signals), $\eta_i(t)$ are correlated Gaussian processes and $W_i(t)$ are standard Wiener processes (noises). In other words, the useful signal can present only in one of the N resolution elements (channels) or can absent in all channels (hypothesis H_0). The hypothesis H_i is identified, as before, with the presence of a signal in ith channel. The problem is to detect a signal and to indicate the number of a channel where a signal appears on the basis of observation of the process $X(t) = (X_1(t), \ldots, X_N(t))$ on the interval $[0, T]$. Let the noises $\eta_i(t) + W_i(t)$ and $\eta_k(t) + W_k(t)$ be mutually independent. Then the likelihood ratios (84)

$$\Lambda_i(X(t),\ 0 \le t \le T) = \Lambda_i(X_i(t),\ 0 \le t \le T) = \Lambda_i(T), \quad i = 1, \ldots, N$$

have the following representation [3], [5]

$$\Lambda_i(T) = \exp\left\{ \int_0^T \tilde{A}_i(t)[dX_i(t) - \hat{\eta}_i(t)dt] - \frac{1}{2}\int_0^T \tilde{A}_i^2(t)dt \right\},$$

where

$$\tilde{A}_i(t) = A_i(t) - \int_0^t r_i(t,u)A_i(u)du, \quad \hat{\eta}_i(t) = \int_0^t r_i(t,u)dX_i(u).$$

Here $r_i(t, u)$ is a response of a filter which carries out the optimal mean-square filtering of the process $\eta_i(t)$ in the mixture with white Gaussian noise $dW_i(t)$ ($r_i(t, u)$ is described by a well known Wiener–Hopf equation).

It could be shown [3] that under hypothesis H_0 the log-likelihood ratios $Z_i(t) = \ln \Lambda_i(t)$ are the processes with independent Gaussian increments and parameters

$$E_0 Z_i(T) = -\frac{1}{2}E_0[Z_i(T) - E_0 Z_i(T)]^2 = -\frac{1}{2}\int_0^T \tilde{A}_i^2(t)dt,$$

where E_0 denotes the expectation with respect to the measure P_0.

The MAR test is non-randomized and the threshold C^* is determined from equation (83) where $F_0(y)$ is defined by (86). Let the distribution of the location of a target be uniform ($\alpha_i = 1/N$) and the signal-to-noise ratios be the same for all elements ($q_i = q$). Then using (86) and (83) after simple manipulations we obtain the following equation for the constant $h = [N(1+q)C^*]^{-(1+1/q)}$

$$B_{v,N}(h) + \frac{L_0}{L}[1-(1-h)^N]\frac{\Gamma(N+v)}{\Gamma(N)\Gamma(1+v)} = 1, \tag{87}$$

where $v = 1/(1+q)$, $\Gamma(y)$ is the gamma function and

$$B_{\lambda_1,\lambda_2}(y) = \frac{\Gamma(\lambda_1)\Gamma(\lambda_2)}{\Gamma(\lambda_1+\lambda_2)}\int_0^y t^{\lambda_1-1}(1-t)^{\lambda_2-1}dt, \quad y \in [0,1]$$

is the beta distribution with the parameters λ_1, λ_2.

If $N \gg 1$, which is usually typical for modern radar systems, then equation (86) is simplified and takes the following form

$$\frac{L_0}{L}N^v[1-\exp(-hN)] + v\int_{Nh}^{\infty} t^{v-1}\exp(-t)dt \approx 1.$$

If, besides, $v \ll 1$ (i.e. $q \gg 1$) then we have

$$\frac{L_0}{L}N^v[1-\exp(-hN)] + \Gamma(1+v) - (Nh)^v \approx 1.$$

It follows from this that asymptotically for large values of signal-to-noise ratio ($v \to 0$) the threshold C^* weakly depends on the number of resolution elements. In fact, for $q \gg 1$ the equation becomes

$$\frac{L_0}{L}N^{1/q}\left[1-\exp\left(-\frac{1}{qC^*}\right)\right] + \Gamma\left(1+\frac{1}{1+q}\right) - \left(\frac{1}{qC^*}\right)^{1/q} \approx 1.$$

By virtue of (81) and (82) the minimax regret

$$\Delta\rho^* = L_0[1 - P_0(Y(X) < C^*)] = L_0[1-(1-h)^N]$$

and if $N \gg 1$, $q \gg 1$ then

$$\Delta\rho^* \approx L_0[1-\exp(-hN)] \approx L_0\left[1-\exp\left(\frac{1}{qC^*}\right)\right]$$

Assume that the components X_k and X_m are statistically independent for all $k \neq m$ which corresponds to the natural assumption of independent noises in different resolution elements. Then

$$\Lambda_i(x_1,\ldots,x_N) = \frac{dP_i(x_1,\ldots,x_N)}{dP_0(x_1,\ldots,x_N)} = \frac{dP_i(x_i)}{dP_0(x_i)} = \Lambda_i(x_i) \tag{84}$$

where $P_0(x_i)$ and $P_i(x_i)$ are the measures corresponding to a noise and to a mixture of a signal from target and a noise in the ith element, respectively ($\Lambda_i(x_i)$ is the likelihood ratio for this element). If X_i is the result of a quasi-optimal preprocessing of a mixture of a Gaussian signal with slow fluctuations and a white Gaussian noise, then X_i has exponential distributions with the densities (see, e.g., [3])

$$p_i(x_i) = \frac{1}{1+q_i}\exp\left(-\frac{x_i}{1+q_i}\right), \quad p_0(x_i) = p_k(x_i) = \exp(-x_i),\ k \neq i,\ x_i \geq 0,$$

where q_i is the signal-to-noise ratio in the ith resolution element.

In this special case the statistics $\Lambda_i(x_i)$ and $Y(x)$ have the form

$$\Lambda_i(x_i) = \frac{1}{1+q_i}\exp\left(\frac{q_i}{1+q_i}x_i\right),$$

$$Y(x) = \exp\left\{\max_{i\in\{1,\ldots,N\}}\left[\frac{q_i}{1+q_i}x_i + \ln(\alpha_i/(1+q_i))\right]\right\}. \tag{85}$$

Since $X_i,\ i = 1,\ldots,N$ are mutually independent, the distribution function $F_0(y) = P_0(Y(X) \leq y)$ of the statistic $Y(X)$ is expressed as follows

$$F_0(y) = \prod_{i=1}^{N} P_0(\alpha_i\Lambda_i(X_i) \leq y)$$

By (85) for $y \geq \alpha_i/(1+q_i)$

$$P_0(\alpha_i\Lambda_i(X_i) \leq y) = P_0\left(X_i \leq \frac{1+q_i}{q_i}\ln\frac{y(1+q_i)}{\alpha_i}\right) = 1 - \left[\frac{\alpha_i}{y(1+q_i)}\right]^{1+1/q_i}$$

and is equal to zero for $y < \alpha_i/(1+q_i)$. Hence

$$F_0(y) = \begin{cases} \prod_{i=1}^{N}\left[1-\left(\frac{\alpha_i}{1+q_i}\frac{1}{y}\right)^{1+1/q_i}\right], & y \geq \min_{i\in\{1,\ldots,N\}}\alpha_i/(1+q_i), \\ 0, & y < \min_{i\in\{1,\ldots,N\}}\alpha_i/(1+q_i). \end{cases} \tag{86}$$

Relations (81) and (82) give the desirable equation for γ and g^*.

It should be noted that in attempting to solve the above decision problem, the first step is to find the conditional distributions of the statistic $S(X)$ that represents rather difficult problem even in the simplest cases. The problem is essentially simplified if $L_1 = L$. Then $S(x) = Y(x)$ and it follows from (81) and (82) that the equation for γ and g^* takes the form (c.f. (67))

$$\int_{\{x:Y(x)<C^*\}} Y(x)dP_0(x) + \frac{L_0}{L}P_0(Y(X) < C^*)$$

$$+\gamma\left\{\int_{\{x:Y(x)=C^*\}} Y(x)dP_0(x) + \frac{L_0}{L}P_0(Y(X) = C^*)\right\} = \frac{L_0}{L}.$$

If P_i–distributions of $\Lambda_i(x)$ are continuous the MAR test becomes non-randomized ($\gamma = 1$) and we arrive at the following equation for $C^*(g^*)$

$$\int_0^{C^*} ydF_0(y) + F_0(C^*) = \frac{L_0}{L}, \tag{83}$$

where $F_1(y) = P_0(Y(X) \leq y)$ is the distribution function of the statistic $Y(X) = \max_{i\in\{1,\dots,N\}} \alpha_i\Lambda_i(X)$ under hypothesis H_0.

Example 12. Let $X = (X_1, \dots, X_N)$ be an N-component vector, where X_k represents an observation process in the kth resolution element of a radar system (a total number of such resolution element or channels is equal to N). The problem is to detect a target and to determine its location (i.e., to indicate the number of the element where a target is present) or to make the decision that a target is absent. The absence of a target is identified with the hypothesis H_0 and its presence in the ith resolution element — with the hypothesis H_i. Therefore L_0 is the loss due to a false alarm which is essentially greater than the loss L_1 due to target missing or the loss L due to an incorrect determination of its coordinates. Moreover, it would seem reasonable to suppose that an incorrect determination of the location of a target leads to its missing and hence $L_1 = L$. The probability of a target appearance π_0 is, of course, unknown but the conditional probabilities $\alpha_i = \Pr(H_i \mid \text{target is present})$ of the location of a target in the particular resolution element are known. Thus the optimal test has the form (77) where $S(x) = Y(x) = \max_{k\in\{1,\dots,N\}} \alpha_k\Lambda_k(x)$.

Additional notations:

$$a = L/L_1; \quad C^* = L_0/g^* L_1;$$

$$Y(x) = \max_{k\in\{1,\dots,N\}} \alpha_k \Lambda_k(x); \quad S(x) = aY(x) + (1-a)\bar{\Lambda}(x).$$

It follows from (75) and (76) that the MAR test can be represented in the form

$$\delta_0^*(x) = \begin{cases} 1, & S(x) < C^*, \\ \gamma, & S(x) = C^*, \\ 0, & S(x) > C^*; \end{cases} \tag{77}$$

$$\delta_j^*(x) = \begin{cases} 1, & S(x) > C^*, \ \alpha_j\Lambda_j(x) > \max_{k\in\{1,\dots,N\}\setminus j} \alpha_k\Lambda_k(x), \\ 1-\gamma, & S(x) = C^*, \ \alpha_j\Lambda_j(x) > \max_{k\in\{1,\dots,N\}\setminus j} \alpha_k\Lambda_k(x), \\ 0, & S(x) < C^*. \end{cases} \tag{78}$$

If $S(x) > C^*$ and maximum $Y(x)$ is attained for several hypotheses among $H_1, \dots, H_N$ we may accept any one, for instance, with minimal number. The same hypothesis is accepted with probability $1-\gamma$ if $S(x) = C^*$. Now it remains to specify equation (56) for the parameters g^* and γ.

Similarly (63) and (64) we have

$$\Delta\rho_1 = L_0 \int_{\mathcal{X}} [1 - \delta_0^*(x)] dP_0(x); \tag{79}$$

$$\Delta\rho_2 = (L_1 - L)\sum_{i=1}^N \alpha_i \int_{\mathcal{X}} [\delta_0^*(x) - \delta_0^0(x, \vec{p}_2)] dP_i(x) + L\sum_{i=1}^N \alpha_i \int_{\mathcal{X}} [\delta_i^0(x, \vec{p}_2) - \delta_i^*(x)] dP_i(x). \tag{80}$$

Combining (77) and (79), we get

$$\Delta\rho_1 = L\left\{1 - P_0(S(X) < C^*) - \gamma P_0(S(X) = C^*)\right\}. \tag{81}$$

Note that $\delta_0^0(x, \vec{p}_2) = 0$ for all x and $\delta_i^0(x, \vec{p}_2) = 1$ if $Y(x) = \alpha_i\Lambda_i(x)$, $\delta_i^0(x, \vec{p}_2) =$ 0 otherwise. Thus, using (77), (78), and (80), similarly (66) we obtain

$$\Delta\rho_2 = (L_1 - L)\sum_{i=1}^N \alpha_i \left[P_i(S(X) < C^*) + \gamma P_i(S(X) = C^*)\right] + L\left\{\int_{\{x:S(x)<C^*\}} Y(x) dP_0(x) + \gamma \int_{\{x:S(x)=C^*\}} Y(x) dP_0(x)\right\}. \tag{82}$$

Assume that there are $N+1$ hypotheses $H_i,\ i=0,1,\ldots,N$. Hypotheses $H_1,\ldots,H_N$ are "similar" in the sence that the losses associated with these hypotheses and corresponding decisions $d_j,\ j=1,\ldots,N$ do not depend on the particular hypothesis. In other words

$$L_{ij}=\begin{cases} L+l, & i\neq j,\\ l, & i=j, \quad i,j=1,\ldots,N.\end{cases}$$

On the contrary the loss $L_{0j}=L_0$ associated with the hypothesis H_0 when it is rejected are essencially greater then $L+l$. Let also $L_{i0}=L_1$ be the loss due to making the decision d_0 when H_i is true (accepting H_0 that is not true). Assume that the loss associated with any correct decision is equal to zero (i.e., $L_{ii}=l=0$ for all $i=0,1,\ldots,N$). There is no loss of generality in making this assumption.

Thus the loss function has the form

$$L_{ij}=\begin{cases} 0, & i=j, \quad i=0,1,\ldots,N,\\ L, & i\neq j, \quad i,j=1,\ldots,N,\\ L_1, & i\neq 0,\ j=0,\\ L_0, & i=0,\ j\neq 0,\end{cases} \tag{74}$$

where L is the loss associated with dismissing of hypotheses $H_1,\ldots,H_N$; L_1 is the loss when the hypothesis H_0 is accepted instead of any of $H_1,\ldots,H_N$ and L_0 is the loss due to rejecting H_0 when it is true. Let $L_0>L_1\geq L>0$.

The *a priori* probability $\pi_0=\Pr(H_0)$ is unknown, but conditional distribution of $H_1,\ldots,H_N$ under condition that one of them is definitely true is known (i.e., $\pi_i=(1-\pi_0)\alpha_i$ where $\alpha_i,\ i=1,\ldots,N$ are known, $\sum_{i=1}^N\alpha_i=1$). By $\Lambda_i(x)=dP_i(x)/dP_0(x)$ denote the likelihood ratio of hypotheses H_i and H_0, and let $\bar\Lambda(x)=\sum_{i=1}^N\alpha_i\Lambda_i(x)$, $g^*=(1-\mu^*)/\mu^*$ (μ^* is a least favorable probability of hypothesis H_0). Then the *a posteriori* probability that H_i is true is given by

$$P^*(H_i\mid x)=\begin{cases} 1/[1+g^*\bar\Lambda(x)], & i=0,\\ g^*\alpha_i\Lambda_i(x)/[1+g^*\bar\Lambda(x)], & i=1,\ldots,N.\end{cases} \tag{75}$$

By (74) the *a posteriori* risk is expressed as follows

$$r_j^*(x)=\begin{cases} L_1[1-P^*(H_0\mid x)], & j=0,\\ L_0P^*(H_0\mid x)+L[1-P^*(H_0\mid x)-P^*(H_j\mid x)], & j\neq 0.\end{cases} \tag{76}$$

where $X_{(1)} = \min\{X_1, \ldots, X_n\}$ is the first order statistic. Assume that $\theta_{i+1} - \theta_i > n^{-1} \ln(\alpha_i/\alpha_{i+1})$. Then

$$Y(x_{(1)}) = Y_i = \alpha_i \exp\{n(\theta_i - \theta_1)\} \text{ if } \theta_i \leq x_{(1)} < \theta_{i+1}, \; i = 1, \ldots, N, \quad (72)$$

where $\theta_{N+1} = \infty$.

It follows from (59), (60), and (72) that we have to take $C^* = Y_{k-1}$ for some k and consequently the optimal test is defined as

$$\delta_1^*(x_{(1)}) = \begin{cases} 0, & x_{(1)} \geq \theta_{k^*}, \\ 1, & x_{(1)} < \theta_{k^*-1}, \\ \gamma, & \theta_{k^*-1} \leq x_{(1)} < \theta_{k^*}, \end{cases}$$

$$\delta_j^*(x_{(1)}) = 0 \text{ for } j = 2, \ldots, k^* - 2;$$

$$\delta_{k^*-1}^*(x_{(1)}) = \begin{cases} 0, & x_{(1)} < \theta_{k^*-1}, \\ 1 - \gamma, & x_{(1)} \in [\theta_{k^*-1}, \theta_{k^*}); \end{cases}$$

$$\delta_j^*(x_{(1)}) = \begin{cases} 1, & x_{(1)} \in [\theta_j, \theta_{j+1}), \\ 0, & x_{(1)} \notin [\theta_j, \theta_{j+1}), \quad j = k^*, \ldots, N, \end{cases}$$

where the numbers k^* and γ are determined from the equation (see (67))

$$(Y_{k^*-2} + 1) P_1(X_{(1)} < \theta_{k^*-1}) + \gamma (Y_{k^*-1} + 1) P_1\{X_{(1)} \in [\theta_{k^*-1}, \theta_{k^*})\} = 1. \quad (73)$$

Here

$$P_1(X_{(1)} < y) = 1 - [1 - \exp\{-(y - \theta_1)\}]^n.$$

Let, for simplicity, $n = 1$ and $\theta_i = \theta_1 + i\Delta$, $\Delta > 0$. Then it follows from (73) that the number k^* is determined from the inequality

$$\alpha_{k^*-2} \left(e^{(k^*-2)\Delta} + 1 \right) e^{-(k^*-1)\Delta} < 1 < \alpha_{k^*-1} \left(e^{(k^*-1)\Delta} + 1 \right) e^{-k^*\Delta}$$

and

$$\gamma = \frac{1 - \alpha_{k^*-2} \left(e^{(k^*-2)\Delta} + 1 \right) e^{-(k^*-1)}}{\left(\alpha_{k^*-1} e^{(k^*-1)\Delta} + 1 \right) \left(e^{-k^*\Delta} - e^{-(k^*-1)\Delta} \right)}.$$

Above we have considered somewhat symmetric case when the loss due to making wrong decisions are greater than the loss due to making a correct decision by the same constant. Now consider the generalization of this problem for more complex loss function which occurs in various applications.

Example 10. Observations have exponential distribution with an unknown scale parameter $\theta > 0$:

$$P_\theta(X < x) = \begin{cases} 1 - \exp(-\theta x), & x \geq 0, \\ 0, & x < 0 \end{cases} \tag{71}$$

and $H_i : \theta = \theta_i,\ i = 1, \ldots, N$. Then

$$\Lambda_i(x) = v_i \exp\{(1 - v_i)x\},$$

$$Y(x) = \exp\{\max_{2 \leq i \leq N} [(1 - v_i)x + \ln \alpha_i v_i]\},$$

where $v_i = \theta_i/\theta_1$ and the threshold is determined from equation (68). The integrals in this equation are expressed in elementary functions and final transcendental equation can be solved numerically. An interesting peculiarity of the MAR test for the large values of the *a priori* probability β of the hypothesis H_1 (which is unknown) is that it suppresses some of the nearest hypotheses to H_1. For instance, if $N = 5$ and $\alpha_i = 0.25$ then $\delta_2^*(x) = 0$ for all $x \geq 0$. This property essentially distinguishes the MAR test from the popular maximum likelihood test which is optimal for a uniform *a priori* distribution of all hypotheses (i.e., for $\beta = 1/N$). However, for other β this test is not optimal and gives the risk

$$\rho(\beta) = \frac{1 - \beta}{4}[3(N - 2)(N - 1) + (N - 1)^{-1}] + \frac{\beta}{2}$$

which is essentially greater than the minimax risk if $\beta > 1/N$ or $< 1/N$.

Example 11. Let $\{P_\theta,\ |\theta| < \infty\}$ represents the family of exponential distributions with the shift θ, i.e. the density of P_θ has the form

$$p_\theta(x) = \begin{cases} \exp\{-(x - \theta)\}, & x \geq \theta, \\ 0, & x < \theta \end{cases}$$

and $H_i : \theta = \theta_i,\ i = 1, \ldots, N;\ \theta_i < \theta_{i+1}$. The likelihood ratios $\Lambda_i(x) = dP_i(x)/dP_1(x)$ for the vector $X = (X_1, \ldots, X_n)$ of i.i.d. observations

$$\Lambda_i(X) = \Lambda_i(X_{(1)}) = \begin{cases} \exp\{n(\theta_i - \theta_1)\}, & X_{(1)} \geq \theta_i, \\ 0, & X_{(1)} < \theta_i, \end{cases}$$

$$= L\left\{\int_{\{x:Y(x)<C^*\}} Y(x)dP_1(x) + \gamma \int_{\{x:Y(x)=C^*\}} Y(x)dP_1(x)\right\}. \qquad (66)$$

Finally, using (65) and (66), we arrive at the equation

$$\int_{\{x:Y(x)<C^*\}} Y(x)dP_1(x) + P_1(Y(X) < C^*)$$

$$+\gamma\left[\int_{\{x:Y(x)=C^*\}} Y(x)dP_1(x) + P_1(Y(X) = C^*)\right] = 1. \qquad (67)$$

If P_1–distributions of the likelihood ratios $\Lambda_i(X)$ are continuous then $\gamma = 1$ and equation (67) is simplified:

$$\int_0^{C^*} y dF_1(y) + F_1(C^*) = 1, \qquad (68)$$

where $F_1(y) = P_1(Y(X) \leq y)$ is the distribution function of the statistic $Y(X) = \max_{i\in\{2,\ldots,N\}} \alpha_i \Lambda_i(X)$.

The minimax risk regret $\Delta\rho^*$ is determined by expression (65) for the general case and equals

$$\Delta\rho^* = L[1 - F_1(C^*)] \qquad (69)$$

for the continuous case.

Now noting that relations (61) and (62) are valid for any test $D(x)$ and using Corollary 2, we find that the minimax test $D^m(x) = D_0(x, \lambda\vec{p}_1 + (1-\lambda)\vec{p}_2)$ has the same structure as $D^*(x)$ (see (59) and (60)) with a threshold C_m which satisfies the equation

$$\int_{\{x:Y(x)<C_m\}} Y(x)dP_1(x) + P_1(Y(X) < C_m)$$

$$+\gamma\left[\int_{\{x:Y(x)=C_m\}} Y(x)dP_1(x) + P_1(Y(X) = C_m)\right] = 1 + \frac{l_1 - \sum_{i=2}^N \alpha_i l_i}{L}. \qquad (70)$$

Comparing (67) and (70), it may be seen that $D^*(x)$ and $D^m(x)$ coincide for all cases when $l_i = l = const$, i.e. when the loss due to correct decisions are equal. In the last case the minimax risk $\rho^m = \Delta\rho^* + l$.

$$= \int_{\mathcal{X}} \sum_{i=2}^{N} \alpha_i [l_i \delta_i^*(x) + (l_i + L)(1 - \delta_i^*(x)) dP_i(x)$$

$$= \sum_{i=2}^{N} \alpha_i l_i + L \sum_{i=2}^{N} \alpha_i \int_{\mathcal{X}} [1 - \delta_i^*(x)] dP_i(x) \tag{62}$$

and hence

$$\Delta\rho_1 = L \int_{\mathcal{X}} [1 - \delta_1^*(x)] dP_1(x); \tag{63}$$

$$\Delta\rho_2 = L \sum_{i=2}^{N} \alpha_i \left\{ \int_{\mathcal{X}} \delta_i^0(x, \vec{p}_2) dP_i(x) - \int_{\mathcal{X}} \delta_i^*(x) dP_i(x) \right\}. \tag{64}$$

From (59) and (63) we get

$$\Delta\rho_1 = L \left\{ P_1(Y(X) > C^*) + (1 - \gamma) P_1(Y(X) = C^*) \right\}$$

$$= L \left\{ 1 - P_1(Y(X) < C^*) - \gamma P_1(Y(X) = C^*) \right\}. \tag{65}$$

Using (60) and evident fact that $\delta_i^0(x, \vec{p}_2) = 1$ if $Y(x) = Y_i(x)$, we obtain

$$\sum_{i=2}^{N} \alpha_i \int_{\mathcal{X}} \delta_i^0(x, \vec{p}_2) dP_i(x) = \sum_{i=2}^{N} \int_{\mathcal{X}} I(Y_i(x) = Y(x)) Y_i(x) dP_1(x)$$

$$= \sum_{i=2}^{N} \int_{\mathcal{X}} I(Y_i(x) = Y(x)) Y(x) dP_1(x) = \int_{\mathcal{X}} Y(x) dP_1(x);$$

$$\sum_{i=2}^{N} \alpha_i \int_{\mathcal{X}} \delta_i^*(x) dP_i(x) = \sum_{i=2}^{N} \int_{\mathcal{X}} I(Y_i(x) = Y(x), Y(x) > C^*) Y_i(x) dP_1(x)$$

$$+ \sum_{i=2}^{N} \gamma_i \int_{\mathcal{X}} I(Y_i(x) = Y(x), Y(x) = C^*) Y_i(x) dP_1(x)$$

$$= \int_{\{x: Y(x) > C^*\}} Y(x) dP_1(x) + (1 - \gamma) \int_{\{x: Y(x) = C^*\}} Y(x) dP_1(x)$$

$$= \int_{\{x: Y(x) \geq C^*\}} Y(x) dP_1(x) - \gamma \int_{\{x: Y(x) = C^*\}} Y(x) dP_1(x),$$

where $I(\cdot)$ is an indicator function.

These last relations together with (64) yield

$$\Delta\rho_2 = L \left\{ \int_{\mathcal{X}} Y(x) dP_1(x) - \int_{\{x: Y(x) \geq C^*\}} Y(x) dP_1(x) + \gamma \int_{\{x: Y(x) = C^*\}} Y(x) dP_1(x) \right\}$$

distribution $\Pi^*(\mu^*)$ and the *a posteriori* risk associated with the decision d_j are given by

$$P^*(H_i \mid x) = \begin{cases} \mu^*/\bar{\Lambda}(x), & i = 1, \\ (1-\mu^*)\alpha_i \Lambda_i(x)/\bar{\Lambda}(x), & i = 2, \dots, N \end{cases} \tag{57}$$

$$r_j^*(x) = \sum_{i=1}^{N} L_{ij} P^*(H_i \mid x) = \sum_{i=1}^{N} l_i P^*(H_i \mid x) + L[1 - P^*(H_j \mid x)]. \tag{58}$$

Since the optimal test minimizes *a posteriori* risk, it follows from (58) that it maximizes $P^*(H_i \mid x)$ and using (57) we obtain

$$\delta_1^*(x) = \begin{cases} 0, & Y(x) > C^*, \\ 1, & Y(x) < C^* \\ \gamma, & Y(x) = C^*, \end{cases} \tag{59}$$

$$\delta_j^*(x) = \begin{cases} 0, & Y(x) < C^* \text{ or } Y(x) > C^*,\ Y_j(x) < Y(x), \\ \gamma_j, & Y(x) = Y_j(x) \ge C^*, \quad j = 2, \dots, N, \end{cases} \tag{60}$$

where $Y_j(x) = \alpha_j \Lambda_j(x)$, $Y(x) = \max_{1 \le i \le N} \alpha_i \Lambda_i(x)$, $C^* = \mu^*/(1-\mu^*)$ and γ_j are arbitrary non-negative numbers such that $\sum_{j=2}^{N} \gamma_j = 1$ if $Y(x) > C^*$ and $\sum_{j=2}^{N} \gamma_j = 1 - \gamma$ if $Y(x) = C^*$.

Therefore the structure of the MAR test is determined but it remains to find the "threshold" C^* and the numbers γ, γ_j. To this end let us use equation (56). Clearly,

$$\rho(D_0(x, \vec{p}_1), \vec{p}_1) = l_1;$$

$$\rho(D_0(x, \vec{p}_2), \vec{p}_2) = \int_{\mathcal{X}} \sum_{i=2}^{N} \alpha_i [l_i \delta_i^0(x, \vec{p}_2) + \sum_{j \ne i} (l_i + L) \delta_j^0(x, \vec{p}_2)] dP_i(x)$$

$$= \int_{\mathcal{X}} \sum_{i=2}^{N} \alpha_i [l_i \delta_i^0(x, \vec{p}_2) + (l_i + L)(1 - \delta_i^0(x, \vec{p}_2))] dP_i(x)$$

$$= \sum_{i=2}^{N} \alpha_i l_i + L \sum_{i=2}^{N} \alpha_i \int_{\mathcal{X}} [1 - \delta_i^0(x, \vec{p}_2)] dP_i(x);$$

$$\rho(D^*(x), \vec{p}_1) = l_1 + L \int_{\mathcal{X}} [1 - \delta_1^*(x)] dP_1(x); \tag{61}$$

$$\rho(D^*(x), \vec{p}_2) = \int_{\mathcal{X}} \sum_{i=2}^{N} \alpha_i [l_i \delta_i^*(x) + \sum_{j \ne i} (l_i + L) \delta_j^*(x)] dP_i(x)$$

wrong is known. In other words $\pi_1 = \beta$ is unknown and $\pi_i = (1-\beta)\alpha_i$, $i = 2, \ldots, N$, where α_i are known, $\sum_2^N \alpha_i = 1$. Then the set $\mathcal{P}$ is a linear segment with vertices at the points $\vec{p}_1 = \{1, 0, \ldots, 0\}$, $\vec{p}_2 = \{0, \alpha_2, \ldots, \alpha_N\}$.

This case describes a variety of practical problems. The problem of detecting a target (signal) with an unknown probability of appearance by multichannel radar system can serve as an example. The fact of a target appearance represents a composite hypothesis: the number of a channel where the signal is located (i.e., the target location, target type, etc.) has to be established along with the fact of a target presence. These features (in particular, the probability of a target location after it appears) have known (conditional) distribution.

Another interesting example is the problem of classification and identification of an object (in particular, in images) when multiple past observations are available for the representatives of $N-1$ classes and hence consistent estimates of *a priori* probabilities of these classes can be constructed. But the Nth class is absolutely new, we know nothing about the probability of the presence of this class and *a priori* information is restricted by the fact of its possible existence.

Again for the sake of simplicity, consider the case when $M = N$. Let the loss function is defined as follows

$$L_{ij} = \begin{cases} l_i + L, & i \neq j, \\ l_i, & i = j, \quad i, j = 1, \ldots, N, \end{cases}$$

where $L > 0$, $l_i = L_{ii} \geq 0$. Thus the loss due to making a correct decision when H_i is true is equal to l_i and the loss due to incorrect decision is more than this loss by a positive constant L.

By Corollary 1 the MAR test is Baeys with respect to $\Pi^*(\mu^*) = \mu^* \vec{p}_1 + (1-\mu^*)\vec{p}_2$ (i.e., $D^*(x) = D_0(x, \mu^* \vec{p}_1 + (1-\mu^*)\vec{p}_2)$) and the constant $\mu^* \in (0,1)$ is determined by the equation

$$\Delta\rho_1 = \Delta\rho_2, \tag{56}$$

where $\Delta\rho_i = \Delta\rho(D^*(x), \vec{p}_i) = \rho(D^*(x), \vec{p}_i) - \rho(D_0(x, \vec{p}_i), \vec{p}_i)$.

Denote $\Lambda_i(x) = \frac{dP_i(x)}{dP_1(x)}$ the likelihood ratio of hypotheses H_i and H_1 and let $\bar{\Lambda}(x) = \mu^* + (1-\mu^*)\sum_{k=2}^N \alpha_k \Lambda_k(x)$. After the element x has been drawn, the *a posteriori* probability that ith hypothesis is true under the *a priori* the results of Example 4, we see that $\rho(D_c^*(x), \pi_1) = L\theta_2^n/(\theta_1^n + \theta_2^n)$ for all $\pi_1 \in [0,1]$ and hence the relative efficiency of the tests

$$\mathcal{EF} = \frac{\rho(D_c^*(x), \pi_1) - \rho(D^*(x), \pi_1)}{\rho(D_c^*(x), \pi_1)} = \frac{1-\pi_1}{\pi_1} - \frac{\theta_2^n}{\theta_1^n}, \quad 0 < \pi_1 \leq \theta_1^n/(\theta_1^n + \theta_2^n).$$

It may be seen that efficiency increases when π_1 decreases and tends to infinity as π_1 tends to zero. For $\pi_1 > \theta_1^n/(\theta_1^n + \theta_2^n)$ the test (31) has advantage since the condition of optimality of the test (55) $\pi_1 \leq 0.5$ is not fulfilled. Note that $\theta_1^n/(\theta_1^n + \theta_2^n) > 0.5$, i.e. $D^*(x)$ has advantage in comparison with (31) even in less rigid conditions than simple ordering. This reflects the fact that the MAR test (55) is uniformly optimal in the conditions when $\pi_i \leq \pi_{i+1}(\theta_i/\theta_{i+1})^n$, $i = 1, \ldots, N-1$ $(\theta_i > \theta_{i+1})$.

The family of distributions for which the ordering of prior probabilities ensures the existence of a uniformly optimal test (which coincides with the MAR test) is of course a lot richer than in example considered. Really, let S be a sufficient statistic for the family of distributions $P_i(dx)$, $i = 1, \ldots, N$. Denote $F_i(ds)$ probability measures corresponding to S under H_i. Assume that there exists a collection of nonintersecting sets $\mathcal{S}_1, \ldots, \mathcal{S}_N$ such that for $s \in \mathcal{S}_k$, $k = 1, \ldots, N$

$$F_i(ds) = 0, \; i = k+1, \ldots, N; \quad F_k(ds) > \max_{1 \leq j < k} F_j(ds).$$

Then

$$\delta_j^*(s) = \begin{cases} 1, & s \in \mathcal{S}_j, \\ 0, & s \notin \mathcal{S}_j, \quad j = 1, \ldots, N, \end{cases}$$

$$\Delta\rho^* = \Delta\rho^m = 0$$

and this test is uniformly optimal (i.e. Bayes) for all Π belonging to the class $\{\Pi : \pi_1 \leq \pi_2 \leq \cdots \leq \pi_N\}$. The family of exponential distributions with unknown shift parameter (see Section 4.5, Example 11) can serve as another example.

4.5 Partial prior uncertainty: Case C3

Consider now Case C3 when the *a priori* probability of one of hypotheses (say H_1) is unknown and conditional prior distribution under condition that H_1 is

$(ny^{n-1}/\theta^n)dy,\ y \in [0,\theta]$ corresponding $X_{(n)}$ satisfy the following conditions: for $y \in (\theta_{k+1}, \theta_k],\ k = 1, \dots, N$

$$P^{(n)}_{\theta_i}(dy) = 0 \text{ if } i = k+1, \dots, N; \quad P^{(n)}_{\theta_k}(dy) > \max_{1 \le j \le k-1} P^{(n)}_{\theta_j}(dy).$$

Then obviously,

$$\delta^*_j(x_{(n)}) = \begin{cases} 1, & x_{(n)} \in (\theta_{j+1}, \theta_j], \\ 0, & x_{(n)} \notin (\theta_{j+1}, \theta_j], \quad j = 1, \dots, N \end{cases} \tag{55}$$

and the regret is equal to zero:

$$\Delta\rho_1 = \cdots = \Delta\rho_N = L\left[\int_{-\infty}^{\infty} dP^{(n)}_{\theta_N}(y) - \int_{-\infty}^{\infty} \delta^*_N(y) dP^{(n)}_{\theta_N}(y)\right]$$

$$= L\left[1 - \int_0^{\theta_N} \frac{ny^{n-1}}{\theta^n_N} dy\right] = 0.$$

This means that for all Π from the class of ordered *a priori* distributions the risk of the MAR test is equal to the Bayes risk, i.e. the MAR test gives the uniformly best solution. Thus simple ordering of the prior probabilities provides exhaustive *a priori* description for the model considered. The knowledge of the exact values of *a priori* probabilities only changes the value of the average risk but not the optimal test.

Note that the minimax test of course coincides with the MAR test and the average risk equals

$$\rho(D^*(x), \Pi) = \rho(D^m(x), \Pi) = L \sum_{i=1}^{N} \pi_i \int_{\mathcal{X}} [1 - \delta^*_i(x)] dP_i(x)$$

$$= L \sum_{i=1}^{N} \pi_i P_{\theta_i}(X_{(n)} \notin (\theta_{i+1}, \theta_i]) = L \sum_{i=1}^{N-1} \pi_i \left(\frac{\theta_{i+1}}{\theta_i}\right)^n.$$

In particular, if $N = 2$ then $\rho(D^*(x), \Pi) = L\pi_1(\theta_2/\theta_1)^n$, where $\pi_1 \le 0.5$. Let us compare this test with the MAR (minimax) test $D^*_c(x)$ in the case of complete prior uncertainty (see (31)) when we do not know in advance that the *a priori* probability of H_2 is more than or equal to the *a priori* probability of the first hypothesis (i.e., the condition $\pi_1 \le 0.5$ is not required). Using

$$= l\left[1 - F\left(\frac{a^* - \tilde{h} + c}{\sigma}\right) - F\left(\frac{a^* + \tilde{h} - c}{\sigma}\right)\right].$$

Obviously, if we take $a^* = 0$, then the left- and right-hand sides of this equation are equal to zero and hence the least favorable distribution is uniform. In particular, this result takes place for Gaussian observations with unknown mean (see Example 1).

4.4 Partial prior uncertainty: Case C2

Consider the problem for Case C2 when we know in advance that the *a priori* probabilities of hypotheses do not decrease, i.e. $0 < \pi_1 \le \pi_2 \le \cdots \le \pi_N < 1$. Then, as we established in Section 2, the set $\mathcal{P}$ represents a polyhedron with vertices at the points $\vec{p}_i = (N-i+1)^{-1}\{\underbrace{1,1,\ldots,1}_{N-i+1},\underbrace{0,0,\ldots,0}_{i-1}\}$, $i = 1,\ldots,N$.

For the sake of simplicity, we shall restrict ourselves to problems when the number of hypotheses is equal to that of decisions and the loss function is simple, i.e. $M = N$ and $L_{ij} = L(1-\delta_{ij})$. Then evidently, $\tilde{L}_{ij} = L(1-\delta_{ij})$ and

$$\Delta\rho_i = \Delta\rho(D^*(x),\vec{p}_i) = \frac{L}{N-i+1}\int_{\mathcal{X}}\sum_{k=i}^{N}\sum_{j\neq k}\delta_j^*(x)dP_k(x)$$

$$= \frac{L}{N-i+1}\left\{\int_{\mathcal{X}}\max_{i\le k\le N} dP_k(x) - \sum_{k=i}^{N}\int_{\mathcal{X}}\delta_k^*(x)dP_k(x)\right\}.$$

The set I^* is not necessarily identical with $\{1,\ldots,N\}$ and in order to find Π^* we have to use the complete set of equalities and inequalities (13) and (14). The only conclusion that can be made in the general case is that always $\mu_1^* \neq 0$.

As an illustration, consider the following example in which the MAR test is uniformly optimal.

Example 9. The model is the same as in Example 4, i.e. X_k, $k = 1,\ldots,n$ are i.i.d. in accordance with uniform distribution with unknown scale parameter θ; $H_i : \theta = \theta_i$; $\theta_i > \theta_{i+1}$, $i = 1,\ldots,N$ $(\theta_{N+1} = 0)$. The nth order statistic $X_{(n)} = \max(X_1,\ldots,X_n)$ is sufficient in this problem and has distribution (32). Therefore conditional probability measures $P_\theta^{(n)}(dy) =$

defined by system (13) where $i \in \{1,\dots,N\}$. It may be shown that

$$\psi_j^*(x) = \begin{cases} 1, & P^*(H_j \mid x) > \frac{l_j}{l_j+L}, \\ 0, & P^*(H_j \mid x) < \frac{l_j}{l_j+L}, \\ \gamma_j, & P^*(H_j \mid x) = \frac{l_j}{l_j+L}, \end{cases} \tag{53}$$

where $P^*(H_j \mid x) = \mu_j^* \exp\{Z_j(x)\}/\sum_{k=1}^N \mu_k^* \exp\{Z_k(x)\}$ is the *a posteriori* probability that H_j is true after the element x has been observed ($Z_j(x) = \log[dP_j(x)/dP_0(x)]$).

If the distributions of the log-likelihood ratios $Z_j(x)$ are continuous, then one can take $\gamma_j = 1$ and the test becomes non-randomized.

Thus it follows from (53) that the "refusing" decision d_0 ($\alpha_j = 0\ \forall j$) is never accepted, while all hypotheses are accepted ($\alpha_j = 1\ \forall j$) if $P^*(H_j \mid x) \geq l_j/(l_j + L)\ \forall j$.

For the case of two hypotheses the test becomes especially simple: $\alpha_1 = 1$, $\alpha_2 = 0$ if $Z(x) \geq h_2; \alpha_1 = 0$, $\alpha_2 = 1$ if $Z(x) \leq h_1$; $\alpha_1 = \alpha_2 = 1$ if $Z(x) \in (h_1, h_2)$, where $h_1 = \log(\mu_2^* l_1/\mu_1^* L)$, $h_2 = \log(\mu_2^* L/\mu_1^* l_2)$ and $Z(x) = \log[dP_1(x)/dP_2(x)]$.

By (20), (52), and (53) we have the following equation for $a^* = \log[(1-\mu_1^*)/\mu_1^*]$:

$$\begin{aligned} &LP_1(Z(X) \leq a^* - \tilde{h}_1) + l_2 P_1(Z(X) < a^* + \tilde{h}_2) \\ &= LP_2(Z(X) \geq a^* + \tilde{h}_2) + l_1 P_2(Z(X) > a^* - \tilde{h}_1), \end{aligned} \tag{54}$$

where $\tilde{h}_1 = \log(L/l_1)$, $\tilde{h}_2 = \log(L/l_2)$.

In the special case when $l_1 = l_2 = l$ and the distribution functions of $Z(X)$ have the form

$$P_1(Z(X) \leq y) = F\left(\frac{y-c}{\sigma}\right), \quad P_2(Z(X) \leq y) = F\left(\frac{y+c}{\sigma}\right),$$

where $F(y)$ is some continuous and symmetric relative to $y = 0$ distribution function ($|c| < \infty$, $0 < \sigma < \infty$), the values of $\tilde{h}_1$ and $\tilde{h}_2$ are equal to $\tilde{h} = \log(L/l)$ and we get from (54)

$$L\left[F\left(\frac{a^* - \tilde{h} - c}{\sigma}\right) + F\left(\frac{a^* + \tilde{h} + c}{\sigma}\right) - 1\right]$$

4.3.2 A problem with a possibility of making non-single-valued decisions

Suppose that the decision space consists of the values of the vector $\vec{\alpha} = (\alpha_1, \ldots, \alpha_N)$ with components $\alpha_i = \{0, 1\}$. Then the total number of decisions equals $M = 2^N$ where, as before, N is a number of possible hypotheses. The situation when $\alpha_i = 0$ and 1 means that the ith hypothesis is rejected and accepted, respectively. Thus we admit the possibility of rejecting all hypotheses (no decision making) and accepting any combination of hypotheses (in particular, only one hypothesis or all hypotheses), i.e., non-single-valued decisions are allowed.

This problem occurs, for example, in generalized classification of objects when the observed object which definitely belongs to one of the N possible classes at the same time can be related to several classes, to all classes or to none of the classes. For details, see Tartakovsky [4].

Let the loss function has the following "additive" form

$$L(H_i, \vec{\alpha}) = \begin{cases} L_0 + l_i & \text{if } \alpha_k = 0 \ \forall\, k, \\ (L + l_i)(1 - \alpha_i) + \sum_{k=1}^{N} l_k \alpha_k & \text{if at least one } \alpha_k = 1. \end{cases}$$

In other words l_k represents the cost of identification of the object with kth class, $L + l_i$ is the loss due to rejecting the true ith class and L_0 is the loss due to no decision making. Let $l_i < L_0 < L \ \forall i$. Then $\min_{\vec{\alpha}} L(H_i, \vec{\alpha}) = l_i$;

$$\tilde{L}_{ij} = \begin{cases} L_0 & \text{if all } \alpha_k = 0, \\ L(1 - \alpha_i) + \sum_{k=1, k \neq i}^{N} l_k \alpha_k & \text{if at least one } \alpha_k = 1, \end{cases} \tag{52}$$

$$\mathcal{E}_i(x) = L + L_0 \prod_{k=1}^{N} [1 - \psi_k(x)] + \sum_{k \neq i} l_k \psi_k(x) - L\psi_i(x),$$

where $\tilde{L}_{ij} = L(H_i, \vec{\alpha}) - \min_{\vec{\alpha}} L(H_i, \vec{\alpha})$ and $\psi_k(x) = \Pr(\alpha_k = 1 \mid x)$ is the probability of choosing $\alpha_k = 1$ (i.e., accept kth hypothesis) regardless of the choice of the other values of α_i. Clearly, the set of these probabilities completely defines the test. Comparing the last expression with (34), we see that $a_0 = c_i = L > 0$, $b_k = l_k \geq 0$, $k = 1, \ldots, N$; $b_0 = L_0 > 0$ and $b_k = 0$ for $k = N + 1, \ldots, 2^N$ and hence $\mu_i^* > 0$ for all $i = 1, \ldots, N$.

By Theorem 1 and Corollary 1 the optimal test $\{\psi_k^*(x)\}$ is completely

$$= 2L_0 \int_{-\infty}^{\infty} [F_0(y+h+a^*) - F_0(y+h-a^*)] f_1(y) dy, \tag{51}$$

which, obviously, has the solution $a^* = 0$. In other words the least favorable distribution is uniform ($\mu_1^* = 1/2$) for any continuous distributions $F_0(y)$ and $F_1(y)$.

Let, for instance, the probability densities

$$g_1(t) = \frac{1}{\sqrt{2\pi}} \exp\{-(t-\theta)^2/2\}, \quad g_0(t) \frac{1}{\sqrt{2\pi}} \exp\{-t^2/2\}, \ \ \theta > 0.$$

Then $Z(x) = \theta(x_1 - x_2)$, $a^* = 0$ and the regret

$$\Delta\rho^* = (L - L_0)\Phi\left(-\frac{\log[(L-L_0)/L] + \theta^2}{\sqrt{2}\theta}\right) + L_0 \Phi\left(\frac{\log[(L-L_0)/L] - \theta^2}{\sqrt{2}\theta}\right).$$

Example 8. Let the model be the same as in Example 1. Then

$$P_1(Z(X) \le y) = \Phi\left(\frac{y - q/2}{\sqrt{q}}\right), \ \ P_2(Z(X) \ge y) = 1 - \Phi\left(\frac{y + q/2}{\sqrt{q}}\right)$$

and it follows from (50) that a^* satisfies the following equation

$$L\left[\Phi\left(\frac{a^* - h - q/2}{\sqrt{q}}\right) + \Phi\left(\frac{a^* + h + q/2}{\sqrt{q}}\right) - 1\right]$$

$$= L_0\left[\Phi\left(\frac{a^* - h - q/2}{\sqrt{q}}\right) + \Phi\left(\frac{a^* + h + q/2}{\sqrt{q}}\right)\right]$$

$$-L_0\left[\Phi\left(\frac{a^* + h - q/2}{\sqrt{q}}\right) + \Phi\left(\frac{a^* - h + q/2}{\sqrt{q}}\right)\right].$$

If $a^* = 0$, then the left- and right-hand sides are equal to zero and hence the least favorable distribution is again uniform.

Note that this result remains true not only for Gaussian distribution but for any continuous distribution $F(y)$ which is symmetric relative to zero and

$$P_1(Z(X) < y) = F\left(\frac{y - c}{\sigma}\right), \quad P_2(Z(X) < y) = F\left(\frac{y + c}{\sigma}\right)$$

($|c| < \infty$, $0 < \sigma < \infty$).

where $Z(x) = \log \frac{dP_1(x)}{dP_2(x)}$ and hence

$$\delta_0^*(x) = \begin{cases} 1, & h_1 < Z(x) < h_2, \\ 0, & Z(x) \notin [h_1, h_2], \\ 1-\gamma_2, & Z(x) = h_1, \\ 1-\gamma_1, & Z(x) = h_2, \end{cases}$$

$$\delta_1^*(x) = \begin{cases} 1, & Z(x) > h_2, \\ 0, & Z(x) < h_2, \\ \gamma_1, & Z(x) = h_2, \end{cases} \quad \delta_2^*(x) = \begin{cases} 1, & Z(x) < h_1, \\ 0, & Z(x) > h_1, \\ \gamma_2, & Z(x) = h_1, \end{cases} \tag{49}$$

where $\gamma_i \in [0,1]$, $h_1 = \log\left[\frac{\mu_2^*}{\mu_1^*}\frac{L_0}{L-L_0}\right]$, $h_2 = \log\left[\frac{\mu_2^*}{\mu_1^*}\frac{L-L_0}{L_0}\right]$.

If $L_0 < L \le 2L_0$ then $h_1 \ge h_2$ and the decision d_0 is never accepted.

It follows from (48) and (49) that the thresholds h_1, h_2 and the constants γ_1, γ_2 satisfy the equation

$$\begin{gathered} L\{P_1(Z(X) < h_1) + \gamma_2 P_1(Z(X) = h_1)\} + L_0\{P_1(h_1 < Z(X) < h_2) \\ +(1-\gamma_2)P_1(Z(X) = h_1) + (1-\gamma_1)P_1(Z(X) = h_2)\} \\ = L\{P_2(Z(X) > h_2) + \gamma_1 P_2(Z(X) = h_2)\} + L_0\{P_2(h_1 < Z(X) < h_2) \\ +(1-\gamma_2)P_2(Z(X) = h_1) + (1-\gamma_1)P_2(Z(X) = h_2)\} \end{gathered}$$

which has, generally speaking, nonunique solution.

If P_1–distribution of the log-likelihood ratio $Z(X)$ is continuous, the test is non-randomized ($\gamma_1 = \gamma_2 = 1$) and this equation is simplified

$$\begin{aligned} &LP_1\{Z(X) \le a^* - h\} + L_0 P_1\{Z(X) \in (a^* - h, a^* + h)\} \\ &= LP_2\{Z(X) \ge a^* + h\} + L_0 P_2\{Z(X) \in (a^* - h, a^* + h)\}, \end{aligned} \tag{50}$$

where $a^* = \log[(1-\mu_1^*)/\mu_1^*]$ and $h = \log[(L-L_0)/L_0]$.

Example 7. Consider the extension of Example 2 for the case with additional decision d_0 (no one of H_1 and H_2 is accepted). Using (28) and (50), we obtain the following equation for a^*

$$L \int_{-\infty}^{\infty} [F_0(y + h + a^*) - F_0(y + h - a^*)] f_1(y) dy$$

the decision d_j, respectively ($P_0(\cdot)$ is some basic measure). By (46)

$$r_j^*(x) = \begin{cases} L_0 + \sum_{i=1}^N l_i P^*(H_i \mid x), & j = 0, \\ L[1 - P^*(H_j \mid x)] + \sum_{i=1}^N l_i P^*(H_i \mid x), & j = 1, \ldots, N. \end{cases}$$

Since the MAR test is Bayes ralative to μ^*, it minimizes *a posteriori* risk and we have

$$\delta_0^*(x) = \begin{cases} 1, & \max_k P^*(H_k \mid x) < 1 - L_0/L, \\ 0, & \max_k P^*(H_k \mid x) > 1 - L_0/L, \\ \gamma, & \max_k P^*(H_k \mid x) = 1 - L_0/L, \end{cases}$$

$$\delta_j^*(x) = \begin{cases} 1, & P^*(H_j \mid x) > 1 - L_0/L,\ P^*(H_j \mid x) > P^*(H_k \mid x)\ \forall k \neq j, \\ 0, & \max_i P^*(H_i \mid x) < 1 - L_0/L, \\ \tilde{\gamma}_j, & P^*(H_j \mid x) = \max_{i \neq j} P^*(H_i \mid x) > 1 - L_0/L, \\ \gamma_j, & P^*(H_j \mid x) = \max_i P^*(H_i \mid x) = 1 - L_0/L, \quad j = 1, \ldots, N, \end{cases}$$

where $\tilde{\gamma}_j$ and γ_j are non-negative constants such that $\sum_{j=1}^N \tilde{\gamma}_j = 1$, $\sum_{j=1}^N \gamma_j = 1 - \gamma$, $\gamma \in [0, 1]$. Note that if $1 - L_0/L < 1/N$, i.e. $L_0 \geq (N-1)L/N$, then the decision d_0 is never accepted.

Using (20) and (47), we get the system of equations for μ_i^*, $\tilde{\gamma}_i$ and γ_i:

$$\Delta\rho^* = \Delta\rho_i = L \int_{\mathcal{X}} (1 - \delta_i^*(x)) dP_i(x)$$

$$- (L - L_0) \int_{\mathcal{X}} \delta_0^*(x)) dP_i(x), \quad i = 1, \ldots, N. \tag{48}$$

Similarly the parameters of the minimax test which has the same form are determined from the following equations

$$\rho^m = R_i(D^m(x)) = l_i + L \int_{\mathcal{X}} (1 - \delta_i^m(x)) dP_i(x)$$

$$-(L - L_0) \int_{\mathcal{X}} \delta_0^m(x)) dP_i(x), \quad i = 1, \ldots, N.$$

Consequently again the structures of the MAR and minimax tests coincide if the loss due to making correct decisions are equal ($l_i = l = const$).

Let us specify these tests for the case of two hypotheses. Obviously,

$$P^*(H_1 \mid x) = 1 - P^*(H_2 \mid x) = \frac{(\mu_1^*/\mu_2^*) \exp\{Z(x)\}}{1 + (\mu_1^*/\mu_2^*) \exp\{Z(x)\}},$$

4.3 Many hypotheses, $M > N$

We shall consider now two problems in which the decision space richer than the space of hypotheses.

4.3.1 Hypotheses testing with a possibility of a no decision making

Assume that along with the decisions $d_1, \ldots, d_N$ the additional decision d_0 can be made which means that no one of the hypotheses is accepted. In this case $M = N+1,\ D(x) = \{\delta_0(x), \delta_1(x), \ldots, \delta_N(x)\}$ and $\sum_{i=1}^N \delta_i(x) = 1 - \delta_0(x) < 1$, where $1 - \delta_0(x)$ is a probability that at least one of $H_1, \ldots, H_N$ is accepted. Consider the following loss function

$$L_{ij} = (L + l_i)(1 - \delta_{ij}) + l_i \delta_{ij}, \quad L_{i0} = L_0 + l_i, \quad i, j = 1, \ldots, N, \tag{46}$$

where $L > \max(L_0, l_i),\ \delta_{ii} = 1,\ \delta_{ij} = 0,\ i \neq j$ (a Kronecker symbol). Therefore the cost of accepting true ith hypothesis equals l_i, the accepting a wrong decision increases this loss by $L > l_i$ and when we do not make any terminal decision this loss increases by $L_0 < L$. Obviously,

$$\tilde{L}_{i0} = L_0, \quad \tilde{L}_{ij} = \begin{cases} L, & i \neq j, \\ 0, & i = j, \quad i, j = 1, \ldots, N \end{cases} \tag{47}$$

and

$$\mathcal{E}_i(x) = L_0 \delta_0(x) + L \sum_{k \neq i} \delta_k(x) = L_0 + (L - L_0) \sum_{k \neq i} \delta_k(x) - L_0 \delta_i(x).$$

Hence condition (34) is fulfilled with $c_i = a_0 = L_0 > 0,\ b_k = L - L_0 > 0,\ \psi_k(x) = \delta_k(x)$ and all μ_i^* are positive. By Theorem 1 the MAR test is Bayes and the constants μ_i^* are determined by conditions (20).

Let

$$P^*(H_i \mid X = x) = P^*(H_i \mid x) = \frac{\mu_i^* \frac{dP_i(x)}{dP_0(x)}}{\sum_{k=1}^N \mu_k^* \frac{dP_k(x)}{dP_0(x)}}$$

and $r_j^*(x) = \sum_{i=1}^N L_{ij} P^*(H_i \mid x)$ denote *a posteriori* probability of the hypothesis H_i under prior $\mu^* = (\mu_1^*, \ldots, \mu_N^*)$ and *a posteriori* risk corresponding to

Similarly for the general case

$$\delta_j^m(x) = \delta_j^*(x) = \begin{cases} \gamma_{jk}, & x_{(n)} \in (\theta_{k+1}, \theta_k],\ k = j, \ldots, N, \\ 0, & x_{(n)} > \theta_j, \quad j = 1, \ldots, N, \end{cases} \tag{44}$$

where $\theta_{N+1} = 0$. By (36) and (44)

$$R_i(D^*(x)) = L\left\{P_i(X_{(n)} > \theta_i) + \sum_{k=i}^{N}(1-\gamma_{ik})P_i(X_{(n)} \in (\theta_{k+1}, \theta_k])\right\}$$

$$= L\sum_{k=i}^{N}(1-\gamma_{ik})\frac{\theta_k^n - \theta_{k+1}^n}{\theta_i^n}, \quad i = 1, \ldots, N. \tag{45}$$

Using (45) and (35), we find that

$$\gamma_{11} = 1, \quad \gamma_{NN} = \theta_1^n / \sum_{k=1}^{N} \theta_k^n$$

and the other values $\gamma_{ik} \in [0,1],\ k = 1, \ldots, j;\ j = 2, \ldots N$ satisfy the system of equations

$$\gamma_{1N} + \gamma_{2N} + \cdots + \gamma_{NN} = 1,$$
$$\gamma_{1N-1} + \gamma_{2N-1} + \cdots + \gamma_{N-1N-1} = 1,$$
$$\vdots$$
$$\gamma_{12} + \gamma_{22} = 1;$$
$$\gamma_{NN} = \sum_{k=i}^{N}(\theta_k^n - \theta_{k+1}^n)\gamma_{ik}/\theta_i^n, \quad i = 1, \ldots, N-1.$$

Evidently, the test (44) is nonunique for $N \geq 3$ (but unique for $N = 2$, see Example 4). It follows from Lemma 1 that any appropriate values of γ_{ik} can be chosen and this does not reduce the effectiveness of the test.

The minimax regret and average risk are equal to

$$\Delta\rho^* = \rho(D^*(x)) = \rho(D^m(x)) = L(1 - \gamma_{NN}) = L\sum_{k=2}^{N}\theta_k^n / \sum_{k=1}^{N}\theta_k^n.$$

Note that similar results can be obtained not only for a uniform distribution but for any distribution for which the likelihood ratios are piecewise constant on some sets $\mathcal{X}_k$. The sets $\mathcal{X}_k$ may be finite or infinite (see, e.g., Example 11 below concerning the family of shifted exponential distributions).

where $X_{(n)} = \max(X_1, \dots, X_n)$.

Let $L_{ij} = L,\ i \neq j;\ \ L_{ii} = 0$. At first consider the case $N = 3$. Using (37) and (43), we obtain

$$\delta_1^m(x) = \delta_1^*(x) = \begin{cases} 1 - \gamma_{33} - \gamma_{23}, & x_{(n)} \le \theta_3, \\ 1 - \gamma_{22}, & \theta_3 < x_{(n)} \le \theta_2, \\ 1, & x_{(n)} > \theta_2, \end{cases}$$

$$\delta_2^m(x) = \delta_2^*(x) = \begin{cases} \gamma_{23}, & x_{(n)} \le \theta_3, \\ \gamma_{22}, & \theta_3 < x_{(n)} \le \theta_2, \\ 0, & x_{(n)} > \theta_2, \end{cases}$$

$$\delta_3^m(x) = \delta_3^*(x) = \begin{cases} \gamma_{33}, & x_{(n)} \le \theta_3, \\ 0, & x_{(n)} > \theta_3. \end{cases}$$

Now taking into account (32) and (36), we get

$$R_1(D^*(x)) = L\left[\gamma_{22}(\theta_2^n - \theta_3^n) + (\gamma_{23} + \gamma_{33})\theta_3^n\right]/\theta_1^n,$$

$$R_2(D^*(x)) = L\left[(1 - \gamma_{22}(\theta_2^n - \theta_3^n) + (1 - \gamma_{23})\theta_1^n\right]/\theta_2^n,$$

$$R_3(D^*(x)) = L(1 - \gamma_{33}).$$

Thus we have three equations (see (35))

$$R_1(\gamma_{22}, \gamma_{33}, \gamma_{23}) = R_2(\gamma_{22}, \gamma_{23}), \quad R_1(\gamma_{22}, \gamma_{33}, \gamma_{23}) = R_3(\gamma_{33}),$$

$$R_2(\gamma_{22}, \gamma_{23}) = R_3(\gamma_{33}).$$

Solving this system, we obtain

$$\gamma_{33} = \frac{\theta_1^n}{\theta_1^n + \theta_2^n + \theta_3^n}; \quad \gamma_{22} = \alpha;$$

$$\gamma_{23} = \left[\frac{\theta_2^n \theta_3^n}{\theta_1^n + \theta_2^n + \theta_3^n} - \alpha(\theta_2^n - \theta_1^n)\right]/\theta_1^n,$$

where $\alpha \in [0, 1]$ is arbitrary constant such that the right-hand side of the last equality belongs to the interval $[0, 1]$. The minimax regret coincides with the average risk and with the conditional risk:

$$\Delta\rho^* = \rho(D^*(x)) = \rho(D^m(x)) = R_i(D^m(x)) = L\frac{\theta_2^n + \theta_3^n}{\theta_1^n + \theta_2^n + \theta_3^n}.$$

Obviously, $D^*(x)$ is also a minimax test and

$$\Delta\rho^* = \rho^m = 1 - \int_{-\infty}^{\infty} [1 - F_1(y)]^{N-1} f_0(y) dy. \tag{42}$$

It should be remarked that the problem considered occurs in radar applications when it is known exactly that the signal presents in one of the channels but it is unknown in what channel. It is necessary to indicate the channel where the signal is located. The statistical characteristics of the noises in the channels are described by probability density $g_0(\cdot)$ and the mixture of the signal and noise – by the density $g_1(\cdot)$ (regardless of the true channel). If the signal and noise are Gaussian processes and specific preprocessing is carried out (see, e.g., [3]) then the distributions are exponential with densities

$$g_0(y) = \exp(-y), \quad g_1(y) = \frac{1}{1+q} \exp\left(-\frac{y}{1+q}\right), \quad y \geq 0,$$

where q characterizes signal-to-noise ratio. Consequently we have

$$Z(X_i) = \frac{q}{1+q} X_i - \log(1+q),$$

$$F_0(y) = 1 - \exp\left\{-\frac{1+q}{q}[y + \log(1+q)]\right\},$$

$$F_1(y) = 1 - \exp\left\{-\frac{1}{q}[y + \log(1+q)]\right\}, \quad y \geq -\log(1+q)$$

and expression (42) yields

$$\Delta\rho^* = \rho^m = \frac{N-1}{N+q}.$$

Example 6. Assume, as in Example 4, that X_k, $k = 1, \ldots, n$ are independent and identically distrideded random variables with common uniform distribution with a scale parameter θ and $X = (X_1, \ldots, X_n)$. The values of parameter θ are assumed to be unknown and hypotheses have the form $H_i : \theta = \theta_i$, $i = 1, \ldots, N$, $N > 2$, $\theta_1 > \theta_2 > \ldots > \theta_N$. Analogously (30) for $j < k$

$$Z_j(X) - Z_k(X) = \begin{cases} n \log(\theta_k/\theta_j), & X_{(n)} < \theta_k, \\ \infty, & X_{(n)} \geq \theta_k, \end{cases} \tag{43}$$

are independent and the density of the joint distribution under hypothesis H_i has the form

$$p_i(x) = g_1(x_i) \prod_{k \neq i} g_0(x_k), \quad i = 1, \ldots, N. \tag{41}$$

As the basic density for the log-likelihood ratio it is convenient to take the density $g_0(\cdot)$ (i.e. $Z_i(x) = \log[p_i(x)/\prod_{k=1}^N g_0(x_k)]$). Then

$$Z_i(x) = Z(x_i) = \log \frac{g_1(x_i)}{g_0(x_i)}.$$

Assume that P_i-distributions of $Z(X_j)$ are continuous and do not depend on i, j (symmetric case), i.e.

$$P_i(Z(X_j) < y) = \begin{cases} F_0(y), & i \neq j, \\ F_1(y), & i = j, \quad i, j = 1, \ldots, N, \end{cases}$$

where $F_0(y) \neq F_1(y)$ are continuous distribution functions. Then, obviously,

$$P_i\left(\bigcap_{k \neq i} \{Z_i(x) - Z_k(x) > h_{ki}\}\right) = \int_{-\infty}^{\infty} \prod_{k \neq i} [1 - F_1(y + h_{ki})] f_0(y) dy, \; i = 1, \ldots, N$$

and using (40) we obtain the system of equations for the "thresholds" $h_{ki} = \log(\mu_k^*/\mu_i^*)$ of the MAR test

$$\int_{-\infty}^{\infty} \prod_{k \neq i} [1 - F_1(y + h_{ki})] f_0(y) dy = \int_{-\infty}^{\infty} \prod_{k=1}^{N-1} [1 - F_1(y + h_{ki})] f_0(y) dy, \; i = 1, \ldots, N-1,$$

where $f_0(y) = dF_0(y)/dy$.

The solution of this system is $h_{ki} = 0$ for all $k \neq i$ and $i = 1, \ldots, N$ and hence the least favorable distribution is uniform, $\mu_i^* = 1/N, \; i = 1, \ldots, N$. It could be expected because the problem is completely symmetric. Thus the MAR test represents the maximum likelihood test:

$$\delta_j^*(x) = 1 \text{ if } Z_j(x_j) = \max_{k \in \{1, \ldots, N\}} Z_k(x_k).$$

(i.e., decide on d_j where j is the smallest integer for which the likelihood ratio $\Lambda_j(x)$ attains maximum).

where $\Lambda_i(x) = dP_i(x)/dP_0(x),\ i = 1,\dots,N$ are the likelihood ratios for hypotheses H_i with respect to some convenient basic measure P_0 and γ_{jk} are some non-negative numbers such that $\sum_{j=1}^N \sum_{k\neq j} \gamma_{jk} = 1$.

Denote $Z_i(x) = \log \Lambda_i(x)$ the log-likelihood ratio, $h_{kj} = \log(\mu_k^*/\mu_j^*)$. After simple manipulations we get

$$\delta_j^*(x) = \begin{cases} 0, & Z_j(x) - Z_k(x) < h_{kj} \text{ at least for one } k \neq j, \\ 1, & Z_j(x) - Z_k(x) > h_{kj}\ \forall k \neq j, \\ \gamma_{jk}, & Z_j(x) - Z_k(x) = h_{kj},\ \ k \neq j, \quad j = 1,\dots,N, \end{cases} \tag{37}$$

In the case when P_i-distributions of $Z_j(x)$ are continuous we can take $\gamma_{jk} = 1$ and the test $D^*(x)$ becomes non-randomized:

$$\delta_j^*(x) = \begin{cases} 0, & Z_j(x) - Z_k(x) < h_{kj} \text{ at least for one } k \neq j, \\ 1, & Z_j(x) - Z_k(x) \geq h_{kj}\ \forall k \neq j, \quad j = 1,\dots,N. \end{cases} \tag{38}$$

Let

$$A_{ik}(h_{ki}) = \{X : Z_i(X) - Z_k(X) > h_{ki}\},\ D_{ik}(h_{ki}) = \{X : Z_i(X) - Z_k(X) = h_{ki}\}.$$

Using (35) and (37), we obtain the following system of equations for h_{kj} and γ_{jk}

$$P_i\left\{\bigcap_{k\neq i} A_{ik}(h_{ki})\right\} + \sum_{k\neq i} \gamma_{ik} P_i\left\{D_{ik}(h_{ki})\right\} =$$

$$P_N\left\{\bigcap_{k\neq N} A_{Nk}(h_{kN})\right\} + \sum_{k=1}^{N-1} \gamma_{Nk} P_N\left\{D_{Nk}(h_{kN})\right\},\quad i = 1,\dots,N-1. \tag{39}$$

For the continuous case this system is simplified

$$P_i\left(\bigcap_{k\neq i}\{Z_i(X) - Z_k(X) > h_{ki}\}\right) = P_N\left(\bigcap_{k\neq N}\{Z_i(X) - Z_N(X) > h_{kN}\}\right). \tag{40}$$

Example 5. Consider the generalization of Example 2 for arbitrary $N > 2$. One can observe the N-component vector $X = (X_1,\dots,X_N)$, where X_i is a random variable which is identified with the observation process in the ith channel of a multichannel (N-channel) system. The components X_i and X_j

Assume that when the ith hypothesis is true the loss due to making a correct decision $L_{ii} = l_i$ and due to a wrong decision $L_{ij} = L + l_i,\ i \neq j,\ L > 0$. Then $\tilde{L}_{ij} = L(1 - \delta_{ij})$ and

$$\mathcal{E}_i(x) = \sum_{j \neq i} L\delta_j(x) = L(1 - \delta_i(x)).$$

Hence condition (34) holds with $b_k = 0,\ c_i = a_0 = L$ and $\psi_i(x) = \delta_i(x)$ and the optimal test $D^*(x)$ can be determined by the system (see (20))

$$\int_{\mathcal{X}} [1 - \delta_i^*(x)] dP_i(x) = \int_{\mathcal{X}} [1 - \delta_N^*(x)] dP_N(x), \quad i = 1, \ldots, N-1. \qquad (35)$$

The minimax regret is then the probability of making a wrong decision times L, i.e.

$$\Delta\rho^* = L \int_{\mathcal{X}} [1 - \delta_i^*(x)] dP_i(x), \quad i = 1, \ldots, N.$$

Similarly, for the minimax test $D^m(x)$ we have

$$R_i(D^m(x)) = l_i + L \int_{\mathcal{X}} [1 - \delta_i^m(x)] dP_i(x) \qquad (36)$$

and consequently $D^m(x)$ satisfies the equations

$$l_i + L \int_{\mathcal{X}} [1 - \delta_i^m(x)] dP_i(x) = l_N + L \int_{\mathcal{X}} [1 - \delta_N^m(x)] dP_N(x), \quad i = 1, \ldots, N-1.$$

It follows from this that if $l_i = l,\ i = 1, \ldots, N$ then $D^m(x) = D^*(x)$ and

$$\rho^m = \Delta\rho^* + l = l + L \int_{\mathcal{X}} [1 - \delta_i^*(x)] dP_i(x), \quad i = 1, \ldots, N.$$

Let us determine the general structure of these tests for the particular case when $l_i = 0,\ L = 1$ ("0–1" or so-called simple loss function). By Theorem 1 the test $D^*(x) = D_0(x, \Pi^*)$ is Bayes relative to $\Pi^* = (\mu_1^*, \ldots, \mu_N^*)$ and hence has the form (2) where $\pi_i = \mu_i^*, \gamma_j = \gamma_j^*$. For the case considered this gives

$$\delta_j^*(x) = \begin{cases} 1, & \mu_j^* \Lambda_j(x) = \max_{k \in \{1,\ldots,N\}} \mu_k^* \Lambda_k(x) > \mu_l^* \Lambda_l(x) \ \ \forall l \neq j, \\ 0, & \mu_j^* \Lambda_j(x) < \max_{k \in \{1,\ldots,N\}} \mu_k^* \Lambda_k(x), \\ \gamma_{jl}, & \mu_j^* \Lambda_j(x) = \mu_l^* \Lambda_l(x) = \max_{k \in \{1,\ldots,N\}} \mu_k^* \Lambda_k(x), \ \ l \neq j, \end{cases}$$

The minimax test has the same form as δ_1^* but

$$\gamma^m = \frac{L_{11} - L_{22} + \tilde{L}_{12}\beta^n}{\tilde{L}_{21} + \tilde{L}_{12}\beta^n}.$$

Of course as for the general case, the MAR test completely coincides with the minimax test ($\gamma = \gamma^m$) if $L_{11} = L_{22}$.

Using (24), (31)–(33), we get the minimax regret

$$\Delta\rho^* = \tilde{L}_{21}\gamma = \tilde{L}_{21}\frac{\tilde{L}_{12}\beta^n}{\tilde{L}_{21} + \tilde{L}_{12}\beta^n}.$$

For the symmetric case ($L_{11} = L_{22} = 0,\ L_{21} = L_{12} = L$), we have

$$\Delta\rho^* = \rho^m = L\beta^n/(1 + \beta^n).$$

4.2 Many hypotheses, $M = N > 2$

We continue to consider the case where the number of decisions is equal to the number of hypotheses and hence again the decision d_j is interpreted as the decision to accept the hypothesis that P_j is the true distribution. But now N is assumed to be an arbitrary number more than 2. In this case a least favorable distribution could be concentrated, in principle, not on the whole set (some of μ_i^* can be equal to zero) and the set I^*, generally speaking, does not coincide with $\{1,\ldots,N\}$. However, it is rather easy to prove that sufficient condition for $I^* = \{1,\ldots,N\}$ (i.e., $\mu_i^* > 0$ for all $i = 1,\ldots,N$) are the following one.

Denote $\mathcal{E}_i(x) = \sum_{j=1}^M \delta_j(x)\tilde{L}_{ij}$ (here at first we suppose that $M \geq N$). Let $\psi_k(x) \in [0,1],\ k = 1,\ldots,M$ be some functions and $a_0 > 0,\ b_k \geq 0,\ c_i > 0$ be some constants. If $\mathcal{E}_i(x)$ can be represented in the form

$$\mathcal{E}_i(x) = a_0 + \sum_{k=1,k\neq i}^{M} b_k\psi_k(x) - c_i\psi_i(x), \quad i = 1,\ldots,N \tag{34}$$

then the set $I^* = \{1,\ldots,N\}$, all values $\mu_i^* > 0$ and can be determined by the system of equations (13).

For large value of θ_1/v we have the following approximate formula for the maximum regret of the minimax test

$$\Delta\rho^m = \frac{(L_{21}-L_{11})(L_{12}-L_{11})}{L_{12}+L_{21}-2L_{11}-\Delta}$$

and relative efficiency of the MAR and minimax tests is expressed as

$$\mathcal{EF} = \frac{\Delta\rho^m - \Delta\rho^*}{\Delta\rho^m} = \frac{\Delta}{L_{21}-L_{11}}.$$

Since usually $L_{21} \gg L_{22}$, $\mathcal{EF} \ll 1$ and the advantage of the MAR test is nonessential.

Example 4. Consider the case when the condition of continuity of the log-likelihood ratio is not satisfied. Let $X = (X_1, \dots, X_n)$ be a sample from the uniform distribution with the density $p_\theta(x_i) = \theta^{-1}$, $x_i \in [0, \theta]$ ($p_\theta(x_i) = 0$, $x_i \notin [0, \theta]$), where θ takes values θ_1 and θ_2 ($\theta_1 > \theta_2$) under H_1 and H_2, respectively. Clearly,

$$Z(X) = \begin{cases} n \log \beta, & X_{(n)} < \theta_2, \\ \infty, & X_{(n)} \ge \theta_2, \end{cases} \tag{30}$$

where $\beta = \theta_2/\theta_1$ and $X_{(n)} = \max(X_1, \dots, X_n)$.

By (23) and (30) the MAR test can be represented in the following form

$$\delta_1^*(x_{(n)}) = \begin{cases} 1, & x_{(n)} \ge \theta_2, \\ \gamma, & x_{(n)} < \theta_2. \end{cases} \tag{31}$$

Obviously,

$$P_\theta(X_{(n)} < y) = \begin{cases} (y/\theta)^n, & y \in [0, \theta], \\ 0, & y \notin [0, \theta]. \end{cases} \tag{32}$$

Now using (24), (31) and (32), we obtain

$$\gamma = \frac{\tilde{L}_{12}\beta^n}{\tilde{L}_{21} + \tilde{L}_{12}\beta^n}, \tag{33}$$

where $\tilde{L}_{12} = L_{12} - L_{11}$, $\tilde{L}_{21} = L_{21} - L_{22}$ and the MAR test is completely determined.

The log-likelihood ratio $Z(x) = (\theta_2 - \theta_1)x + \log(\theta_1/\theta_2)$ has continuous distributions ($i = 1, 2$)

$$P_i(Z(X) < y) = \begin{cases} 1 - \exp\left\{-\frac{\theta_i}{\theta_2 - \theta_1}[y + \log(\theta_2/\theta_1)]\right\}, & y \geq -\log(\theta_2/\theta_1), \\ 0, & y < -\log(\theta_1/\theta_2). \end{cases}$$

Additional notations:

$$v = \theta_2 - \theta_1, \;\; l^* = (L_{21} - L_{22})/(L_{12} - L_{11}),$$

$$l = (L_{12} - L_{11})/(L_{12} - L_{22}), \;\; l_m = (L_{21} - L_{22})/(L_{12} - L_{22}),$$
$$C = (1 + v/\theta_1)\exp(h), \;\; C_m = (1 + v/\theta_2)\exp(h_m).$$

Since the distributions of the log-likelihood ratio are continuous, we can put $\gamma = \gamma_m = 0$ in (23), (24) and (25). After simple manipulations we arrive at the following equations for the constants C and C_m of the MAR and minimax tests, respectively,

$$C^{1+\theta_1/v} - C - l^* = 0,$$

$$C_m^{1+\theta_1/v} - C_m - l_m = 0.$$

For large values of the parameter θ_1/v ($\theta_1/v \gg \max\{1, \log(l + l_m)\}$) we get from this

$$C \approx (1 + l^*)^{v/(v+\theta_1)}, \quad C_m \approx (1 + l_m)^{v/(v+\theta_1)},$$
$$\Delta\rho^* = (L_{21} - L_{22})P_2(Z(X) \geq h) = (L_{21} - L_{22})C^{-(1+\theta_1/v)}$$
$$\approx (L_{21} - L_{22})(1 + l^*)^{-1} = \frac{(L_{21} - L_{22})(L_{12} - L_{11})}{L_{12} + L_{21} - 2L_{11} - \Delta};$$
$$\rho^m = (L_{21} - L_{22})P_2(Z(X) \geq h_m) + L_{22} = (L_{21} - L_{22})(1 - C_m^{-(1+\theta_1/v)}) + L_{11}$$
$$\approx L_{11} + \frac{(L_{21} - L_{11})(L_{12} - L_{11})}{L_{12} + L_{21} - 2L_{11} - \Delta},$$

where $\Delta = L_{22} - L_{11}$.

Obviously,

$$\Delta\rho^m = \max_{\Pi}(\rho^m - \rho_0(\Pi)) = \rho^m - \min_{\Pi}\rho_0(\Pi) = \rho^m - \min(L_{11}, L_{22})$$

$$= (L_{12} - L_{11})(1 - C_m^{-(1+\theta_1/v)}).$$

If $L_{11} = L_{22}$ then the MAR test coincides with the minimax one. If, furthermore, $L_{12} = L_{21} = L$ then equation (29) gives $h = 0$ and $\mu_1^* = \lambda_1^m = 1/2$ (uniform distribution). In this case the minimax regret and average risk equal

$$\Delta\rho^* = \rho_m = R_i(D^m(x)) = L\int_{-\infty}^{\infty}[1 - F_0(y)]f_1(y)dy, \quad i = 1, 2.$$

Let, for instance, the distributions of X_k be exponential with densities

$$g_0(x_k) = \theta_0 \exp(-\theta_0 x_k), \quad g_1(x_k) = \theta_1 \exp(-\theta_1 x_k), \quad x_k \geq 0$$

(and 0 for $x_k < 0$), where $0 < \theta_1 < \theta_0$. Then

$$\tilde{Z}(x_k) = (\theta_0 - \theta_1)x_k - \log(\theta_0/\theta_1),$$

$$F_i(y) = 1 - \exp\left\{-\frac{\theta_i}{\theta_0 - \theta_1}[y + \log(\theta_0/\theta_1)]\right\}, \quad y \geq -\log(\theta_0/\theta_1), \; i = 0, 1$$

and we get

$$\Delta\rho^* = \rho_m = L\theta_1/(\theta_0 + \theta_1).$$

For the case of Gaussian variables when $X_k \sim \mathcal{N}(\theta, 1)$ under H_k and $X_k \sim \mathcal{N}(0, 1)$ under H_j, $j \neq k$, $k = 1, 2$, we have

$$\tilde{Z}(x_k) = \theta x_k - \theta^2/2, \quad Z(x) = \theta(x_1 - x_2),$$

$$\Delta\rho^* = \rho_m = L\Phi(-\theta/2).$$

Example 3. Consider asymmetric case when $L_{11} < L_{22}$, $L_{12} > L_{21}$ and the conditional distributions of the observations X are exponential

$$P_i(X < x) = 1 - \exp(-\theta_i x), \quad x \geq 0, \quad i = 1, 2,$$

where $0 < \theta_1 < \theta_2$. This problem occurs in radar systems when one wants to detect a Gaussian signal with slow fluctuations in the presence of a white Gaussian noise (see, e.g., [3], [4]) and in image processing when one wants to detect the anomaly in the radio image. The decision d_2 in the situation when H_1 is true is identified with a false alarm and $L_{12} \gg L_{21}$.

Example 2. Let $X = (X_1, X_2)$ be a two component process which is observed in a "two-channel" system, i.e. X_i is the observation process in the ith channel ($i = 1, 2$). The components X_1 and X_2 are statistically independent and have conditional joint densities (under hypotheses H_1 and H_2, respectively) of the form

$$p_1(x_1, x_2) = g_1(x_1)g_0(x_2), \quad p_2(x_1, x_2) = g_0(x_1)g_1(x_2),$$

where $g_0(y)$ and $g_1(y)$ are some probability densities. Thus the log-likelihood ratio

$$Z(x) = \log \frac{p_1(x)}{p_2(x)} = \tilde{Z}(x_1) - \tilde{Z}(x_2),$$

where $\tilde{Z}(y) = \log[g_1(y)/g_0(y)]$.

Suppose that the P_i-distributions of $\tilde{Z}(X_k)$ are continuous and have the form

$$P_i(\tilde{Z}(X_k) < y) = F_0(y),\ k \neq i; \quad P_i(\tilde{Z}(X_i) < y) = F_1(y),\ i, k = 1, 2, \quad (28)$$

where $F_0(y)$ and $F_1(y)$ are some distribution functions ($F_0 \neq F_1$). Conditions (28) describe specific symmetric case and hold, for instance, in the problem of signal detection in a multi-channel system (two-channel in our case) when statistical characteristics of observations in channels do not depend on the number of the channel where the signal is located (see, e.g., [3], [4]).

In view of the continuity of the distributions, $\gamma = 0$ in (23) and equation (24) takes the form

$$(L_{12} - L_{11})P_1(\tilde{Z}(X_1) - \tilde{Z}(X_2) < h) = (L_{21} - L_{22})P_2(\tilde{Z}(X_1) - \tilde{Z}(X_2) \geq h).$$

Now using (28), we obtain

$$P_2(\tilde{Z}(X_1) - \tilde{Z}(X_2) \geq h) = \int_{-\infty}^{\infty} [1 - F_0(y + h)] f_1(y) dy,$$

$$P_1(\tilde{Z}(X_1) - \tilde{Z}(X_2) < h) = \int_{-\infty}^{\infty} [1 - F_0(y - h)] f_1(y) dy,$$

where $f_1(y) = {F_1}'(y)$ is the density of $F_1(y)$.

Thus we can rewrite the equation for the constant h in the form

$$\tilde{L}_{12} \int_{-\infty}^{\infty} [1 - F_0(y - h)] f_1(y) dy = \tilde{L}_{21} \int_{-\infty}^{\infty} [1 - F_0(y + h)] f_1(y) dy. \quad (29)$$

Obviously, if $h = 0$ then the least favorable distribution is uniform ($\mu_1^* = 1 - \mu_2^* = \lambda_1^m = 1 - \lambda_2^m = 1/2$). In particular this is fulfilled if

$$P_1(Z(X) < h) = F\left(\frac{h-c}{\sigma}\right), \quad P_2(Z(X) < h) = F\left(\frac{h+c}{\sigma}\right),$$

where $\sigma > 0$, $|c| < \infty$ and the distribution function $F(y)$ is continuous and symmetric with respect to $y = 0$.

Note also that in all cases when P_1–distribution of $Z(X)$ is continuous we can put $\gamma = 0$ in (23)–(27) and consider only non-randomized tests.

Example 1. Let $X \sim \mathcal{N}(\theta_i, 1)$ be a Gaussian variable with unit variance and the mean θ_i under hypothesis H_i, $i = 1, 2$. Then

$$Z(x) = (\theta_1 - \theta_2)x - \frac{1}{2}(\theta_1^2 - \theta_2^2), \quad E_1 Z(X) = -E_2 Z(X) = \frac{q}{2},$$

$$D_2 Z(X) = D_1 Z(X) = q,$$

where $q = (\theta_1 - \theta_2)^2$ and hence

$$P_1(Z(X) < h) = \Phi\left(\frac{h - q/2}{\sqrt{q}}\right), \quad P_2(Z(X) \geq h) = 1 - \Phi\left(\frac{h + q/2}{\sqrt{q}}\right).$$

Here and in the sequel E_i and D_i denote operators of conditional expectation and variance, respectively, and

$$\Phi(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} \exp(-t^2/2)dt$$

is a standard normal distribution function. Therefore for the symmetric case ($L_{11} = L_{22} = 0$, $L_{12} = L_{21} = L$) using (27) ($\gamma = 0$), we obtain

$$\Phi\left(\frac{h + q/2}{\sqrt{q}}\right) + \Phi\left(\frac{h - q/2}{\sqrt{q}}\right) = 1,$$

i.e. $h = 0$ and the least favorable distribution is uniform. The regret and the minimax risk equal

$$\Delta\rho^* = \rho^m = L\Phi(-\sqrt{q}/2).$$

where

$$Z(x) = \log \frac{dP_1(x)}{dP_2(x)}, \quad h = \log \frac{(L_{21} - L_{22})\mu_2^*}{(L_{12} - L_{11})\mu_1^*}, \quad \gamma \in [0,1].$$

Since the loss due to making wrong decisions are always more than due to correct ones, $\tilde{L}_{11} = \tilde{L}_{22} = 0$, $\tilde{L}_{ij} = L_{ij} - L_{ii}$, $j \neq i$, $i = 1, 2$ and we have

$$\Delta\rho_1 = (L_{12} - L_{11}) \int_{\mathcal{X}} (1 - \delta_1^*(x)) dP_1(x), \quad \Delta\rho_2 = (L_{21} - L_{22}) \int_{\mathcal{X}} \delta_1^*(x) dP_2(x),$$

Thus the MAR test is completely defined by two parameters h and γ and we can write the equation (13) ($I^* = \{1,2\}$) with respect to h and γ (instead of $\mu_1^* = 1 - \mu_2^*$):

$$\Delta\rho^* = (L_{21} - L_{22}) \left[P_2(Z(X) > h) + \gamma P_2(Z(X) = h)\right]$$

$$= (L_{12} - L_{11}) \left[P_1(Z(X) < h) + (1 - \gamma) P_1(Z(X) = h)\right]. \qquad (24)$$

The minimax test has the same form as $\delta_1^*(x)$ but, generally speaking, with $\gamma^m \neq \gamma$ and $h^m \neq h$. These constants satisfy the equation

$$\rho^m = R_1(D^m(x)) = R_2(D^m(x))$$

$$= (L_{21} - L_{22}) \left[P_2(Z(X) > h^m) + \gamma^m P_2(Z(X) = h^m)\right] + L_{22}$$

$$= (L_{12} - L_{11}) \left[P_1(Z(X) < h^m) + (1 - \gamma^m) P_1(Z(X) = h^m)\right] + L_{11}. \qquad (25)$$

Comparing (24) with (25), it may be seen that $\delta_1^*(x)$ and $\delta_1^m(x)$ coincide if $L_{11} = L_{22}$, i.e. when the losses due to correct decisions are equal. Moreover, if $L_{11} = L_{22} = 0$ then

$$\Delta\rho^* = \rho^m = R_i(D^m(x)) = L_{21} \left[P_2(Z(X) > h) + \gamma P_2(Z(X) = h)\right]$$

$$= L_{12} \left[P_1(Z(X) < h) + (1 - \gamma) P_1(Z(X) = h)\right], \ i = 1, 2. \qquad (26)$$

If besides $L_{21} = L_{12} = L$ (symmetric case), equation (24) for h and γ (h^m and γ^m) takes the form

$$P_2(Z(X) > h) + P_1(Z(X) > h)$$

$$+ \gamma \left[P_2(Z(X) = h) + P_1(Z(X) = h)\right] = 1. \qquad (27)$$

$\sum_{i=1}^N \mu_i^* = 1$. Then $D^*(x) = D_0(x, \Pi^*)$ and it follows from Corollary 1 that μ_i^* are determined by the system of N equations

$$\Delta\rho^* = \Delta\rho_i = \int_{\mathcal{X}} \sum_{j=1}^{M} \delta_j^*(x) \tilde{L}_{ij} dP_i(x), \quad i = 1, \dots, N, \tag{20}$$

where

$$\tilde{L}_{ij} = L_{ij} - \min_{k \in \{1,\dots,M\}} L_{ik}. \tag{21}$$

Similarly under assumption $\lambda_i^m > 0,\ i = 1, \dots, N$ we get from (16) the system of equations for a least favorable distribution Π^m of the minimax test

$$\rho^m = \rho_i = \int_{\mathcal{X}} \sum_{j=1}^{M} \delta_j^m(x) L_{ij} dP_i(x), \quad i = 1, \dots, N. \tag{22}$$

Obviously, ρ_i coincides with the conditional risk $R_i(D^m(x))$ that reflects a well-known Wald's result [2].

It would be interesting to obtain the conditions under which the set $I^* = \{1, \dots, N\}$ ($I^m = \{1, \dots, N\}$). But we do not discuss this problem here, because it would take us too far astray. We shall consider now the cases when this assumption is obviously fulfilled. However, for a simple sufficient condition, see Section 4.2, relation (34).

4.1 Two hypotheses, $M = N = 2$

The special case where the number of decisions is equal to that of hypotheses is of particular interest. In this case we may interpret d_j as the decision to accept the hypothesis H_j that P_j is the true probability measure.

If $N = M = 2$ then, by Theorem 5.1 of Wald [2], a Bayes solution relative to any *a priori* distribution $\Pi = (\pi_1, \pi_2)$ represents a likelihood ratio test and hence the MAR test has the form

$$\delta_1^*(x) = 1 - \delta_2^*(x) = \begin{cases} 1, & Z(x) > h, \\ \gamma, & Z(x) = h, \\ 0, & Z(x) < h, \end{cases} \tag{23}$$

$$\rho(D^m(x),\Pi) < \rho_m(\mathcal{P}) \quad \textit{for all} \quad \Pi \in B_{\mathcal{P}} \backslash B^m. \qquad (17)$$

If the set $\mathcal{P}$ is a convex polyhedron with vertices $\vec{p}_i$, $i = 1, \ldots, k$, $k \leq N$ then similarly (13) and (14) we have

$$\rho^m(\mathcal{P}) = \rho_i = \max_{l \in \{1,\ldots,k\}} \rho_l \quad \text{for all} \quad i \in I^m, \qquad (18)$$

$$\rho_i < \rho^m(\mathcal{P}) \quad \text{for all} \quad i \in \{1,\ldots,k\} \backslash I^m, \qquad (19)$$

where $\rho_i = \rho(D^m(x), \vec{p}_i)$, $\Pi^m = \sum_{i \in I^m} \vec{p}_i \lambda_i^m$, $\sum_{i \in I^m} \lambda_i^m = 1$, $\lambda_i^m > 0$ for $i \in I^m$.

It should be noted that while randomization does not reduce the Bayes risk under complete *a priori* information, the system of equalities (10), (11) and (16), (17) (respectively, (13), (14) and (18), (19)) are not always solvable for the class of non-randomized tests. Consequently, generally speaking, in order to find the MAR and minimax tests we have to consider randomized tests as well.

In spite of Theorems 1 and 2 completely define the structures of the optimal tests (in general case) it is not still so easy to carry out this program. Further specialization of $D^*(x)$ and $D^m(x)$ requires specialization of a set $\mathcal{P}$, a set of decisions $\{d_1, \ldots, d_M\}$ and the loss function and sometimes even a distribution of observations. A number of important specific and at the same time relatively general problems will be considered in the next sections. Note also that it is interesting to obtain more or less general conditions under which the minimax and MAR tests coincide. However, it seems that this problem is not less complicated than to find the structure of optimal tests and we shall find the answer only for some cases, in particular for the case of complete prior uncertainty.

4 Complete Prior Uncertainty

Consider the case C1 in the list of Section 2. Suppose that a least favorable distribution is concentrated on the whole set $B_{\mathcal{P}} = \{\vec{p}_1, \ldots, \vec{p}_N\}$, $\vec{p}_i = (\delta_{i1}, \ldots, \delta_{iN})$, i.e. $\Pi^* = (\mu_1^*, \ldots, \mu_N^*)$, where $\mu_i^* > 0$ for all $i = 1, \ldots, N$ and which is fulfilled for any $D(x)$ and the proof is complete.

For a particular case when $\mathcal{P}$ is a polyhedron with vertices $\vec{p}_i$, $i = 1, \dots, k$ the boundary $B_{\mathcal{P}} = \{\vec{p}_1, \dots, \vec{p}_k\}$, the set B^* represents some subset of $\{\vec{p}_1, \dots, \vec{p}_k\}$ with indices $i \in I^* \subseteq \{1, \dots, k\}$ and the problem becomes especially simple.

Corollary 1 *Let the set $\mathcal{P}$ be a convex polyhedron with vertices $\vec{p}_i$, $i = 1, \dots, k$, $k \leq N$ and $\Delta\rho^*(\mathcal{P}) = \min_{D(x)} \max_{i \in \{1,\dots,k\}} \Delta\rho(D(x), \vec{p}_i)$. Then $D^*(x) = D_0(x, \Pi^*)$,*

$$\Delta\rho^*(\mathcal{P}) = \Delta\rho_i = \max_{l \in \{1,\dots,k\}} \Delta\rho_l \quad \textit{for all} \quad i \in I^*, \tag{13}$$

$$\Delta\rho_i < \max_{l \in \{1,\dots,k\}} \Delta\rho_l \quad \textit{for all} \quad i \in \{1, \dots, k\} \backslash I^*, \tag{14}$$

where $\Delta\rho_i = \Delta\rho(D^*(x), \vec{p}_i)$, $\Pi^* = \sum_{i \in I^*} \vec{p}_i \mu_i^*$, $\sum_{i \in I^*} \mu_i^* = 1$, $\mu_i^* > 0$ for $i \in I^*$.

Thus it turns out that the MAR test is a Bayes test relative to some specific (least favorable) *a priori* distribution Π^* which is determined by equalities and inequalities (10) and (11) (or (13) and (14) for polyhedron). This could be expected and reflects Wald's theorem on completeness of the class of Bayes decision rules [2].

Let us now reformulate the results for a minimax test $D^m(x)$ which is obtained by applying the minimax principle to the average risk:

$$\rho^m(\mathcal{P}) = \rho(D^m(x), \mathcal{P}) = \inf_{D(x)} \sup_{\Pi \in \mathcal{P}} \rho(D(x), \Pi). \tag{15}$$

The same kind of argument as in proof of Theorem 1 gives the following result.

Theorem 2 *The minimax test $D^m(x)$ is the Bayes test with respect to the a priori distribution $\Pi^m = \int_{B^m} \Pi \, d\lambda(\Pi)$, where the measure $\lambda(\Pi)$, $\lambda(B^m) = 1$ and the set B^m are chosen such that*

$$\rho^m(\mathcal{P}) = \rho(D^m(x), \Pi) = \max_{\Pi \in B_{\mathcal{P}}} \rho(D^m(x), \Pi) \quad \textit{for all} \quad \Pi \in B^m, \tag{16}$$

3 The Structure of the Optimal Tests

Let

$$\Delta\rho^*(\mathcal{P}) = \inf_{D(x)} \sup_{\Pi\in\mathcal{P}} \Delta\rho(D(x),\Pi) \tag{9}$$

denotes the minimax regret and $\mu(\Pi)$ be some measure on $B^* \subseteq B_{\mathcal{P}}$ such that $\mu(B^*) = 1$ and $d\mu(\Pi) > 0$ for $\Pi \in B^*$.

The following theorem determines the general structure of the MAR test.

Theorem 1 *The MAR test $D^*(x)$ is the Bayes test with respect to the a priori distribution $\Pi^* = \int_{B^*} \Pi \, d\mu(\Pi)$, where $\mu(\Pi)$ and B^* are chosen such that the following equality and inequality are satisfied simultaneously*

$$\Delta\rho^*(\mathcal{P}) = \Delta\rho(D^*(x),\Pi) = \max_{\Pi\in\mathcal{P}} \Delta\rho(D^*(x),\Pi) \quad \textit{for all} \quad \Pi \in B^*, \tag{10}$$

$$\Delta\rho(D^*(x),\Pi) < \Delta\rho^*(\mathcal{P}) \quad \textit{for all} \quad \Pi \in B_{\mathcal{P}}\backslash B^*. \tag{11}$$

Proof. By (7) for any $D(x)$

$$\max_{\Pi\in\mathcal{P}} \Delta\rho(D(x),\Pi) = \max_{\Pi\in B_{\mathcal{P}}} \Delta\rho(D(x),\Pi) \geq \max_{\Pi\in B^*} \Delta\rho(D(x),\Pi)$$

$$= \int_{B^*} \max_{\Pi\in B^*} \Delta\rho(D(x),\Pi) d\mu(\Pi) \geq \int_{B^*} \Delta\rho(D(x),\Pi) d\mu(\Pi). \tag{12}$$

Clearly,

$$\int_{B^*} \Delta\rho(D,\Pi) d\mu(\Pi) = \rho(D,\Pi^*) - \int_{B^*} \rho(D_0,\Pi) d\mu(\Pi)$$

$$\geq \rho(D^*,\Pi^*) - \int_{B^*} \rho(D_0,\Pi) d\mu(\Pi) = \int_{B^*} \Delta\rho(D^*,\Pi) d\mu(\Pi).$$

However, by the condition of the theorem the last value is equal to

$$\max_{\Pi\in B^*} \Delta\rho(D^*(x),\Pi) = \max_{\Pi\in B_{\mathcal{P}}} \Delta\rho(D^*(x),\Pi)$$

that together with (12) gives the inequality

$$\max_{\Pi\in\mathcal{P}} \Delta\rho(D(x),\Pi) \geq \max_{\Pi\in\mathcal{P}} \Delta\rho(D^*(x),\Pi)$$

In what follows we shall establish the structure of the optimal test $D^*(x)$ for arbitrary prior constraints and specialize this structure for the particular constraints of the type C1–C3 and for various particular examples. Note that for some types of loss function, e.g. for zero–one loss function, and complete *a priori* uncertainty (case C1 above) the structures of the tests $D^*(x)$ and $D^m(x)$ coincide and furthermore $\Delta\rho(D^*(x)) = R_i(D^*(x)) = R_i(D^m(x)) = const$ for all $i = 1, \dots, N$, i.e. the conditional risk gives simultaneously the value of the average risk regret. In the sequel we shall call the test $D^m(x)$ minimax test and the test $D^*(x)$ minimax average regret (MAR) test.

The next proposition is useful for practical calculations and simplifies the search for the MAR test.

Lemma 1 *1. The regret $\Delta\rho(D(x), \Pi)$ is a concave function of Π for any fixed $D(x)$.*

2. The functional $\sup_{\Pi\in\mathcal{P}} \Delta\rho(D(x), \Pi)$ is concave.

The proof of this lemma is very simple and might be omitted.

As a consequence of the first part $\max_{\Pi\in\mathcal{P}} \Delta\rho(D(x), \Pi)$ is attained only on the boundary of the set $\mathcal{P}$. More precisely, let $B_{\mathcal{P}}$ be the boundary of the set $\mathcal{P}$, i.e. a subset of $\mathcal{P}$ such that for any $\Pi_1, \Pi_2 \in B_{\mathcal{P}}$ and any $\lambda \in (0, 1)$ the point $(1 - \lambda)\Pi_1 + \lambda\Pi_2 \notin B_{\mathcal{P}}$. Then we have

$$\max_{\Pi\in\mathcal{P}} \Delta\rho(D(x), \Pi) = \max_{\Pi\in B_{\mathcal{P}}} \Delta\rho(D(x), \Pi) \tag{7}$$

and, in particular, if the set $\mathcal{P}$ is a convex polyhedron with vertices at the points $\vec{p}_i$, $i = 1, \dots, k$ then clearly $B_{\mathcal{P}} = \{\vec{p}_1, \dots, \vec{p}_k\}$ and

$$\max_{\Pi\in\mathcal{P}} \Delta\rho(D(x), \Pi) = \max_{i\in\{1,\dots,k\}} \Delta\rho(D(x), \vec{p}_i). \tag{8}$$

This result essentially simplifies the problem of minimizing the regret for the most interesting cases C1, C2, and C3 (see above).

As follows from the second part of Lemma, a local minimum of the functional $\sup_{\Pi\in\mathcal{P}} \Delta\rho(D(x), \Pi)$ coincides with a global one and hence if we found some test minimizing this functional further search can be terminated. Any of several available tests may be accepted – only the principle of convenience is important.

Since a uniformly optimal solution of the problem which minimizes the conditional risk

$$R_i(D(x)) = \int_{\mathcal{X}} \sum_{j=1}^{M} L_{ij}\delta_j(x) dP_i(x) \quad (4)$$

for all $i = 1, \ldots, N$ does not exist (and consequently a Bayes solution depends on an unknown *a priori* distribution) it is necessary to consider alternative methods. Wald [2] proposed to minimize the maximum value of the risk (4) and proved that this solution was a Bayes solution with respect to a least favorable *a priori* distribution. Moreover, he proved that the minimax test $D^m(x)$ gives constant conditional risk for all i for which least favorable distribution is positive. In particular,

$$R_1(D^m(x)) = R_2(D^m(x)) = \ldots = R_N(D^m(x)) \quad (5)$$

if this distribution is positive for all $i = 1, \ldots, N$.

However, the minimax solution in its traditional form has at least two drawbacks. Firstly, it does not take into account any *a priori* information about the distribution of the hypotheses, i.e. only the case C1 in the list above fits with this particular form. Secondly, it does not take into account a deviation of a minimax average risk from a Bayes risk (3). Hence the regret $\Delta\rho(D^m(x), D_0(x), \Pi) = \rho(D^m(x), \Pi) - \rho(D_0(x), \Pi)$ can be large for Π which differs from a least favorable distribution and we cannot control this regret. In this sense the traditional minimax method is too cautious.

These drawbacks motivate us to consider another, in our opinion, more reasonable minimax approach which makes it possible to control the excess of a risk over Bayes risk. More precisely, we propose to use the minimax principle not to conditional risk (4) but to the regret $\Delta\rho(D(x), \Pi) = \rho(D(x), \Pi) - \rho(D_0(x), \Pi)$, i.e. the optimal test $D^*(x)$ is determined from the condition

$$\Delta\rho(D^*(x), \mathcal{P}) = \inf_{D(x)} \sup_{\Pi \in \mathcal{P}} \Delta\rho(D(x), \Pi), \quad (6)$$

where $\Delta\rho(D(x), \Pi) \geq 0$ and $\mathcal{P}$ is a given set that describes our prior knowledge. Therefore we adopt as our optimal solution the one which will give the minimum of the maximal regret $\Delta\rho(D(x), \Pi)$ over all possible *a priori* distributions from the set $\mathcal{P}$.

This technique was proposed in [1] for the general decision making problem with nuisance parameters as the basis of an adaptive (empirical) Bayes approach.

The most restrictive assumption among those we used above is the knowledge of an *a priori* distribution. In practice almost always this distribution is unknown and a Bayes test cannot be used.

Consider the problem when Π is unknown and prior information is restricted by assumption that Π belongs to some set of distributions $\mathcal{P}$ which contains more than a single point. Some typical examples which frequently occur in various applications are as follows.

C1. Complete prior uncertainty: π_i are arbitrary nonnegative values, $\sum_{i=1}^N \pi_i = 1$. In this case $\mathcal{P}$ represents a convex polyhedron with vertices at $\vec{p}_i = (p_{i1}, \ldots, p_{iN})$, where $p_{ik} = \delta_{ik}$ (δ_{ik} is a Kronecker symbol, $\delta_{ik} = 0, i \neq k$, $\delta_{ii} = 1$).

C2. π_i, $i = 1, \ldots, N$ represent an ordered sequence: $\pi_{i-1} \leq \pi_i$, $\pi_0 = 0$. Again the set $\mathcal{P}$ is a convex polyhedron with vertices at the points

$$\vec{p}_1 = \frac{1}{N}\{1, 1, \ldots, 1\},\ \vec{p}_2 = \frac{1}{N-1}\{1, 1, \ldots, 1, 0\}, \ldots, \vec{p}_N = \{1, 0, \ldots, 0\}.$$

C3. Partial prior uncertainty when probabilities $\pi_1, \ldots, \pi_r$ are completely unknown and remaining probabilities π_i, $i = r+1, \ldots, N$ are proportional to known values α_i, $\sum_{r+1}^N \alpha_i = 1$. In other words $\pi_i = (1-\beta)\alpha_i$, $i = r+1, \ldots, N$, where $\beta = \sum_1^r \pi_i$ is unknown and α_i are known. This means that the conditional prior distribution under condition that one of the hypotheses $H_{r+1}, \ldots, H_N$ is true is known ($\Pr(H_i \mid \text{either } H_{r+1}, \ldots, H_N \text{ is true}) = \alpha_i$). Obviously, $\mathcal{P}$ is a polyhedron with vertices at the points

$$\vec{p}_1 = \{1, 0, \ldots, 0\},\ \vec{p}_2 = \{0, 1, 0, \ldots, 0\}, \ldots,$$

$$\vec{p}_r = \{\underbrace{0, \ldots, 0}_{r-1}, 1, \underbrace{0, \ldots, 0}_{N-r}\},\ \vec{p}_{r+1} = \{0, \ldots, 0, \alpha_{r+1}, \ldots, \alpha_N\}.$$

In the important particular case when $r = 1$, $\pi_1 = \beta$ (unknown) and $\pi_i = (1-\beta)\alpha_i$, $i = 2, \ldots, N$ the set $\mathcal{P}$ is a linear segment, $\vec{p}_1 = \{1, 0, \ldots, 0\}$, $\vec{p}_2 = \{0, \alpha_2, \ldots, \alpha_N\}$.

C4. Π is known with the accuracy to a small deviation from the known value Π_0 which does not exceed some constant β_0: $\Pi = \Pi_0 + \beta\vec{e}$, where $\vec{e}$ is a unit vector and $0 \leq \beta \leq \beta_0$.

made additionally to "ordinary" decision d_j which means that the hypothesis H_j $(j = 1, \ldots, N)$ is accepted (for details, see Section 4.3).

As usual a vector function $D(x) = (\delta_1(x), \ldots, \delta_M(x))$ with M components $\delta_j(x)$ satisfying the conditions

$$\delta_j(x) \geq 0, \quad \sum_{j=1}^{M} \delta_j(x) = 1$$

is called a (randomized) statistical test (decision function, rule, procedure, algorithm) of a hypothesis, where $\delta_j(x)$ is the probability that we shall make the decision d_j when the element $X = x$ is observed. Thus $D(x)$ is a probability measure on the set of decisions $\{d_1, \ldots, d_M\}$. If for any x the probability measure $D(x)$ assigns the probability 1 to a single element of $\{d_1, \ldots, d_M\}$ (i.e., $\delta_j(x)$ takes only values 0 and 1), then the test is called non-randomized (deterministic) and may be identified with a measurable function $d(x)$ taking values in the set $\{d_1, \ldots, d_M\}$.

The consequences of the adoption of the decision d_j when the hypothesis H_i is true is evaluated by the loss $L(H_i, d_j) = L_{ij} \in [0, \infty),\ i = 1, \ldots, N; j = 1, \ldots, M$.

Any *a priori* distribution can be represented by a vector $\Pi = (\pi_1, \ldots, \pi_N)$, where $\pi_i = \Pr(H_i) \geq 0$ denotes the *a priori* probability that H_i is true, $\sum_{i=1}^{N} \pi_i = 1$. If the *a priori* distribution Π is known then the degree of preference given to the various tests can be evaluated by an average risk

$$\rho(D(x), \Pi) = \int_{\mathcal{X}} \sum_{j=1}^{M} \delta_j(x) \sum_{i=1}^{N} L_{ij} \pi_i dP_i(x). \tag{1}$$

The optimal test $D_0(x) = (\delta_1^0(x), \ldots, \delta_M^0(x))$ minimizing $\rho(D(x), \Pi)$ is called a Bayes test and has the form

$$\delta_j^0(x) = \begin{cases} 0, & \sum_{i=1}^{N} L_{ij} \pi_i dP_i(x) > \min_{k \in \{1, \ldots, M\}} \sum_{i=1}^{N} L_{ik} \pi_i dP_i(x), \\ \gamma_j, & \sum_{i=1}^{N} L_{ij} \pi_i dP_i(x) = \min_{k \in \{1, \ldots, M\}} \sum_{i=1}^{N} L_{ik} \pi_i dP_i(x), \end{cases} \tag{2}$$

where $\gamma_j,\ j = 1, \ldots, M$ are arbitrary non-negative values, $\sum_{j=1}^{M} \gamma_j = 1$. Obviously, we can take $\gamma_n = 1$ for arbitrary n among the others for which the second equality in (2) is satisfied (say, the minimum value). Thus a Bayes rule is non-randomized and its average risk is given by

$$\rho_0 = \rho(D_0, \Pi) = \int_{\mathcal{X}} \min_{j \in \{1, \ldots, M\}} \sum_{i=1}^{N} L_{ij} \pi_i dP_i(x). \tag{3}$$

for obtaining the best adaptive Bayes decision making rules which use estimates of unknown parameters of distributions of observed data. The results show that usually this approach gives quite attractive solution.

In the present paper we use the above mentioned "regret" ideas in the problem that includes testing several simple hypotheses and unknown (completely or partially) *a priori* distribution of hypotheses. As could be expected the "minimax regret" approach similarly to the ordinary minimax approach gives a Bayes solution with respect to some specific, in some sense least favorable distribution. The proofs of the major general results are quite simple and our main goal was to illustrate these results for the variety of particular cases which are of great importance for applications.

2 Notations, Formulation of the Problem and Auxiliary Results

A variety of practical problems involving detection, classification, identification and pattern recognition can be formulated as the following multiple decision problem. Let $(P, \mathcal{F}, \Omega)$ be a stochastic space with standard assumptions and one can observe a random vector X with values in some measurable space $\{\mathcal{X}, \mathcal{U}\}$ (in other words X is a measurable mapping of $\{\Omega, \mathcal{F}\}$ in $\{\mathcal{X}, \mathcal{U}\}$). We shall suppose that $P = \{P_i, i = 1, \ldots, N\}$, where P_i are some completely known probability measures. The true probability distribution of the observed data X is not known and the problem consists in testing finitely many ($N < \infty$) statistical hypotheses H_i, $i = 1, \ldots, N$, where the hypothesis H_i means that the true probability measure is P_i.

Since the space $\mathcal{X}$ is arbitrary the element X can represent a scalar or vector process which is observed in both discrete and continuous time. However, we shall restrict ourselves to cases when the observation time is fixed in advance, i.e. only non-sequential procedures will be considered.

The problem is formalized as making one of the M decisions d_j, $j = 1, \ldots, M$, where $N \leq M < \infty$. It should be mentioned that the value of M, generally speaking, is not equal to N. In other words the space of decisions can be richer than the space of hypotheses. For instance, the decision that no one of P_i is true or that several H_i are accepted simultaneously can be

from various fields such as detection, identification, classification and pattern recognition.

1 Introduction

A number of important applied problems such as detection, identification, classification and pattern recognition can be reduced to a testing several statistical hypotheses. The case of complete prior information when a Bayes solution can be used is exotic because an *a priori* distribution of hypotheses and conditional probability distributions of observed data are almost always unknown at least partially. Thus we usually deal with prior uncertainty when a Bayes test cannot be applied successfully. The type of prior uncertainty depends on particular circumstances.

There are many approaches to the solution of a decision making problem under limited prior knowledge [1], including a minimax one. The minimax approach relative to conditional risk was introduced and studied by Wald [2]. The main advantage of this approach is that it uses only the main data of the problem that we really have and does not use any additional, frequently subjective information. At the same time its peculiarity and sometimes drawback is the constancy of the conditional risk for all hypotheses and hence the average risk of the minimax test is constant for any prior distribution of hypotheses. Firstly, it means that only the case of complete uncertainty relative to an *a priori* distribution is involved into consideration and if we have some additional information about the character of this distribution this information is not used. The second drawback is that the traditional minimax approach does not use any information about characteristics of the Bayes rule which is optimal under complete prior knowledge. As a result it may lead to undesirable excess of the minimax average risk over the average risk of another reasonable rule (in particular, over Bayes risk) for almost all prior distributions except several ones. At the same time it is quite clear that for most cases the control of the deviation (regret) of the minimax average risk from the average risk of the Bayes solution is desirable.

In contrast to "purely" minimax approach Repin and Tartakovsky [1] proposed to use the principle of minimax to the average risk regret in the problems involving nuisance parameters and successfully applied this method

On the Minimax Approach to Overcoming Prior Uncertainty and Application to Pattern Recognition Problems

Alexander Tartakovsky
Department of Mathematics
University of California at Los Angeles
California, USA

Key words: minimax tests; minimax average regret tests; least favorable distributions; prior uncertainty.

Abstract

We consider the problem of testing many statistical hypotheses with incomplete information on *a priori* distribution of hypotheses. A minimax deviation (regret) of the average risk from the Bayes risk for a known *a priori* distribution serves as the optimality criterion. In contrast to traditional minimax method (with respect to a conditional risk) our approach makes it possible to control the excess of the risk over the risk of the optimal rule under complete prior knowledge. It appears that the optimal rule is Bayesian relative to some specific prior distribution. Generally speaking the structure of this "least favorable" distribution depends on the type of prior uncertainty, loss function and the distribution of observations. We present the relations which enable us to find a least favorable distribution and parameters of the minimax rule and minimax regret. Several types of prior uncertainty are considered: from complete uncertainty to partial uncertainty with various types of restrictions. The general results are illustrated by examples