Kybernetika

Andrey Novikov

Optimal sequential multiple hypothesis testing in presence of control variables

Kybernetika, Vol. 45 (2009), No. 3, 507--528

Persistent URL: http://dml.cz/dmlcz/140020

KYBERNETIKA — VOLUME 45 (2009), NUMBER 3, PAGES 507 – 528

OPTIMAL SEQUENTIAL MULTIPLE HYPOTHESIS TESTING IN PRESENCE OF CONTROL VARIABLES

Andrey Novikov

Suppose that at any stage of a statistical experiment a control variable X that affects the distribution of the observed data Y at this stage can be used. The distribution of Y depends on some unknown parameter θ, and we consider the problem of testing multiple hypotheses $H_1 : \theta = \theta_1$, $H_2 : \theta = \theta_2, \ldots$, $H_k : \theta = \theta_k$ allowing the data to be controlled by X, in the following sequential context. The experiment starts with assigning a value X_1 to the control variable and observing Y_1 as a response. After some analysis, another value X_2 for the control variable is chosen, and Y_2 as a response is observed, etc. It is supposed that the experiment eventually stops, and at that moment a final decision in favor of one of the hypotheses $H_1, \ldots$, H_k is to be taken. In this article, our aim is to characterize the structure of optimal sequential testing procedures based on data obtained from an experiment of this type in the case when the observations $Y_1, Y_2, \ldots, Y_n$ are independent, given controls $X_1, X_2, \ldots, X_n$, $n = 1, 2, \ldots$.

Keywords: sequential analysis, sequential hypothesis testing, multiple hypotheses, control variable, independent observations, optimal stopping, optimal control, optimal decision, optimal sequential testing procedure, Bayes, sequential probability ratio test

AMS Subject Classification: 62L10, 62L15, 60G40, 62C99, 93E20

1. INTRODUCTION. PROBLEM SET–UP

Let us suppose that at any stage of a statistical experiment a "control variable" X can be used, that affects the distribution of the observed data Y at this stage. "Statistical" means that the distribution of Y depends on some unknown parameter θ, and we have the usual goal of statistical analysis: to obtain some information about the true value of θ. In this work, we consider the problem of testing multiple hypotheses $H_1 : \theta = \theta_1$, $H_2 : \theta = \theta_2, \ldots$, $H_k : \theta = \theta_k$ allowing the data to be controlled by X, in the following "sequential" context.

The experiment starts with assigning a value X_1 to the control variable and observing Y_1 as a response. After some analysis, we choose another value X_2 for the control variable, and observe Y_2 as a response. Analyzing this, we choose X_3 for the third stage, get Y_3, and so on. In this way, we obtain a sequence $X_1, \ldots, X_n$, $Y_1, \ldots, Y_n$ of experimental data, $n = 1, 2, \ldots$. It is supposed that the experiment eventually stops, and at that moment a final decision in favor of one of $H_1, \ldots, H_k$ is to be taken.

In this article, our aim is to characterize the structure of optimal sequential procedures, based on this type of data, for testing the multiple hypotheses $H_1, \ldots, H_k$.

We follow [5] and [11] in our interpretation of "control variables". For example, in a regression experiment, with a dependent variable Y and an independent variable X, the variable X is a control variable in our sense, whenever the experimenter can vary its value before the next observation is taken. Another classical context for "control variables" in our sense is the experimental design, when one of some alternative treatments is assigned to every experimental unit before the experiment starts. The randomization, which is frequently used with both these type of "controlled" experiments, can be easily incorporated in our theory below as well.

There exist yet another concept of "control variables" introduced by Haggstrom [3], and largely used in [10] and many subsequent articles (see also [1] for results, closely related to [10], where "control variables" are not used). In the context of [10], a control variable, roughly speaking, is an integer variable whose value, at every stage of the experiment, is a prescription of a number of the additional observations to be taken at the next stage, if any. To some extent, it is related to our control variables as well, because it affects the distribution of subsequently observed data. It is very likely that our method will work for this type of "sequentially planned" experiments as well, but formally it does not fit our theory below, mainly because we do not allow that the cost of observations depend on X.

In this article, we follow very closely our article [6], where the case of $k = 2$ simple hypotheses was considered, and use a method based on the same ideas as in [7], where multiple hypothesis testing for experiments without control variables was studied.

For data vectors, let us write, briefly, $X^{(n)}$ instead of $(X_1, \ldots, X_n)$, $Y^{(n)}$ instead of $(Y_1, \ldots, Y_n)$, $n = 1, 2, \ldots$, etc. Let us define a (randomized) sequential hypothesis testing procedure as a triplet (χ, ψ, ϕ) of a *control policy* χ, a *stopping rule* ψ, and a *decision rule* ϕ, with

$$\chi = (\chi_1, \chi_2, \ldots, \chi_n, \ldots), \quad \psi = (\psi_1, \psi_2, \ldots, \psi_n, \ldots), \quad \phi = (\phi_1, \phi_2, \ldots, \phi_n, \ldots),$$

with the components described below.

The functions

$$\chi_n = \chi_n(x^{(n-1)}, y^{(n-1)}), \quad n = 1, 2, \ldots$$

are supposed to be measurable functions with values in the space of values of the control variable. The functions

$$\psi_n = \psi_n(x^{(n)}, y^{(n)}), \quad n = 1, 2, \ldots$$

are supposed to be some measurable functions with values in $[0, 1]$. Finally,

$$\phi_n = (\phi_{n1}, \phi_{n2}, \ldots, \phi_{nk}),$$

with

$$\phi_{ni} = \phi_{ni}(x^{(n)}, y^{(n)}), \quad i = 1, \ldots, k,$$

are supposed to be measurable non-negative functions such that

$$\sum_{i=1}^{k} \phi_{ni}(x^{(n)}, y^{(n)}) \equiv 1 \quad \text{for any } n = 1, 2 \ldots.$$

The interpretation of all these functions is as follows.

The experiments starts at stage $n = 1$ applying χ_1 to determine the initial control x_1. Using this control, the first data y_1 is observed.

At any stage $n \geq 1$: the value of $\psi_n(x^{(n)}, y^{(n)})$ is interpreted as the conditional probability *to stop and proceed to decision making*, given that we came to that stage and that the observations were $(y_1, y_2, \ldots, y_n)$ after the respective controls $(x_1, x_2, \ldots, x_n)$ have been applied. If there is no stop, the experiments continues to the next stage $(n + 1)$, defining first the new control value x_{n+1} by applying the control policy:

$$x_{n+1} = \chi_{n+1}(x_1, \ldots, x_n; y_1, \ldots, y_n)$$

and then taking an additional observation y_{n+1} using control x_{n+1}. Then the rule ψ_{n+1} is applied to $(x_1, \ldots, x_{n+1}; y_1, \ldots, y_{n+1})$ in the same way as as above, etc., until the experiment eventually stops.

It is supposed that when the experiment stops, a decision *to accept one and only one* of $H_1, \ldots, H_k$ is to be made. The function $\phi_{ni}(x^{(n)}, y^{(n)})$ is interpreted as the conditional probability *to accept* H_i, given that the experiment stops at stage n being $(y_1, \ldots, y_n)$ the data vector observed and $(x_1, \ldots, x_n)$ the respective controls applied.

The control policy χ generates, by the above process, a sequence of random variables $X_1, X_2, \ldots, X_n$, recursively by

$$X_{n+1} = \chi_{n+1}(X^{(n)}, Y^{(n)}).$$

The stopping rule ψ generates, by the above process, a random variable τ_ψ (*stopping time*) whose distribution is given by

$$P_\theta^\chi(\tau_\psi = n) = E_\theta^\chi(1 - \psi_1)(1 - \psi_2) \ldots (1 - \psi_{n-1})\psi_n. \tag{1}$$

Here, and throughout the paper, we interchangeably use ψ_n both for

$$\psi_n(x^{(n)}, y^{(n)})$$

and for

$$\psi_n(X^{(n)}, Y^{(n)}),$$

and so do we for any other function

$$F_n = F_n(x^{(n)}, y^{(n)}).$$

This does not cause any problem if we adopt the following agreement: when F_n is under probability or expectation sign, it is $F_n(X^{(n)}, Y^{(n)})$, otherwise it is $F_n(x^{(n)}, y^{(n)})$.

For a sequential testing procedure (χ, ψ, ϕ) let us define

$$\alpha_{ij}(\chi, \psi, \phi) = P_{\theta_i}(\,\text{accept}\, H_j) = \sum_{n=1}^{\infty} E_{\theta_i}^{\chi}(1-\psi_1)\ldots(1-\psi_{n-1})\psi_n\phi_{nj} \tag{2}$$

and

$$\beta_i(\chi, \psi, \phi) = P_{\theta_i}(\,\text{accept any}\, H_j \,\text{different from}\, H_i) = \sum_{j\neq i} \alpha_{ij}(\chi, \psi, \phi). \tag{3}$$

The probabilities $\alpha_{ij}(\chi, \psi, \phi)$ for $j \neq i$ can be considered "individual" error probabilities and $\beta_i(\chi, \psi, \phi)$ "gross" error probability, under hypothesis H_i, $i = 1, 2, \ldots, k$, of the sequential testing procedure (χ, ψ, ϕ).

Another important characteristic of a sequential testing procedure is the *average sample number*:

$$N(\theta; \chi, \psi) = E_{\theta}^{\chi}\tau_{\psi} = \begin{cases} \displaystyle\sum_{n=1}^{\infty} nP_{\theta}^{\chi}(\tau_{\psi} = n), \text{ if } P_{\theta}^{\chi}(\tau_{\psi} < \infty) = 1, \\ \infty \quad \text{otherwise.} \end{cases} \tag{4}$$

In this article, we solve the two following problems:

Problem I. Minimize $N(\chi, \psi) = N(\theta_1; \chi, \psi)$ over all sequential testing procedures (χ, ψ, ϕ) subject to

$$\alpha_{ij}(\chi, \psi, \phi) \leq \alpha_{ij}, \quad \text{for any } i = 1, \ldots k, \text{ and for any } j \neq i, \tag{5}$$

where $\alpha_{ij} \in (0, 1)$ (with $i, j = 1, \ldots k$, $j \neq i$) are some constants.

Problem II. Minimize $N(\chi, \psi) = N(\theta_1; \chi, \psi)$ over all sequential testing procedures (χ, ψ, ϕ) subject to

$$\beta_i(\chi, \psi, \phi) \leq \beta_i, \quad \text{for any } i = 1, \ldots k, \tag{6}$$

with some constants $\beta_i \in (0, 1)$, $i = 1, \ldots, k$.

In Section 2, we reduce the problem of minimizing $N(\chi, \psi)$ under constraints (5) (or (6)) to an unconstrained minimization problem. The new objective function is the Lagrange-multiplier function $L(\chi, \psi, \phi)$.

Then, finding

$$L(\psi, \phi) = \inf_{\phi} L(\chi, \psi, \phi)$$

we reduce the problem further to a problem of finding optimal control policy and stopping rule.

In Section 3, we solve the problem of minimization of $L(\chi, \psi)$ in a class of control-and-stopping strategies.

In Section 4, the likelihood ratio structure for optimal strategy is given.

In Section 5, we apply the results obtained in Sections 2 – 4 to the solution of Problems I and II.

The final Section 6 contains some additional results, examples and discussion.

2. REDUCTION TO A PROBLEM OF OPTIMAL CONTROL AND STOPPING

In this section, Problems I and II will be reduced to unconstrained optimization problems using the idea of the Lagrange multipliers method.

2.1. Reduction to non-constrained minimization in Problems I and II

The following two theorems are practically Theorem 1 and Theorem 2 in [7]. They reduce Problem I and Problem II to respective unconstrained minimization problems, using the idea of the Lagrage multipliers method.

For Problem I, let us define $L(\chi, \psi, \phi)$ as

$$L(\chi, \psi, \phi) = N(\chi, \psi) + \sum_{1 \le i,j \le k;\, i \ne j} \lambda_{ij} \alpha_{ij}(\chi, \psi, \phi) \tag{7}$$

where $\lambda_{ij} \ge 0$ are some constant multipliers.

Let Δ be a class of sequential testing procedures.

Theorem 1. Let exist $\lambda_{ij} > 0$, $i = 1, \ldots, k$, $j = 1, \ldots, k$, $j \ne i$, and a testing procedure $(\chi^*, \psi^*, \phi^*) \in \Delta$ such that for any other testing procedure $(\chi, \psi, \phi) \in \Delta$

$$L(\chi^*, \psi^*, \phi^*) \le L(\chi, \psi, \phi) \tag{8}$$

holds (with $L(\chi, \psi, \phi)$ defined by (7)), and such that

$$\alpha_{ij}(\chi^*, \psi^*, \phi^*) = \alpha_{ij} \quad \text{for any } i = 1, \ldots k, \text{ and for any } j \ne i. \tag{9}$$

Then for any testing procedure $(\chi, \psi, \phi) \in \Delta$ such that

$$\alpha_{ij}(\chi, \psi, \delta) \le \alpha_{ij} \quad \text{for any } i = 1, \ldots k, \text{ and for any } j \ne i, \tag{10}$$

it holds

$$N(\chi^*, \psi^*) \le N(\chi, \psi). \tag{11}$$

The inequality in (11) is strict if at least one of the equalities (10) is strict.

For Problem II, let now $L(\chi, \psi, \phi)$ be defined as

$$L(\chi, \psi, \phi) = N(\chi, \psi) + \sum_{i=1}^{k} \lambda_i \beta_i(\chi, \psi, \phi), \tag{12}$$

where $\lambda_i \ge 0$ are the Lagrange multipliers.

Theorem 2. Let exist $\lambda_i > 0$, $i = 1, \ldots, k$, and a testing procedure $(\chi^*, \psi^*, \phi^*) \in \Delta$ such that for any other testing procedure $(\chi, \psi, \phi) \in \Delta$

$$L(\chi^*, \psi^*, \phi^*) \le L(\chi, \psi, \phi) \tag{13}$$

holds (with $L(\chi, \psi, \phi)$ defined by (12)), and such that

$$\beta_i(\chi^*, \psi^*, \phi^*) = \beta_i \quad \text{for any } i = 1, \ldots k. \tag{14}$$

Then for any testing procedure $(\chi, \psi, \phi) \in \Delta$ such that

$$\beta_i(\chi, \psi, \delta) \leq \beta_i \quad \text{for any } i = 1, \dots k, \tag{15}$$

it holds

$$N(\chi^*, \psi^*) \leq N(\chi, \psi). \tag{16}$$

The inequality in (16) is strict if at least one of the equalities (15) is strict.

2.2. Optimal decision rules

Due to Theorems 1 and 2, Problem I is reduced to minimizing (7) and Problem II is reduced to minimizing (12). But (12) is a particular case of (7), namely, when $\lambda_{ij} = \lambda_i$ for any $j = 1, \dots, k$, $j \neq i$ (see (2) and (3)). Because of that, we will only solve the problem of minimizing $L(\chi, \psi, \phi)$ defined by (7).

In particular, in this section we find

$$\inf_{\phi} L(\chi, \psi, \phi),$$

and the corresponding decision rule ϕ, at which this infimum is attained.

Let I_A be the indicator function of the event A.

From this time on, we suppose that for any $n = 1, 2, \dots$, the random variable Y, when a control x is applied, has a probability "density" function

$$f_\theta(y|x) \tag{17}$$

(Radon–Nikodym derivative of its distribution) with respect to a σ-finite measure μ on the respective space. We are supposing as well that, at any stage $n \geq 1$, given control values $x_1, x_2, \dots x_n$ applied, the observations $Y_1, Y_2, \dots, Y_n$ are independent, i. e. their joint probability density function, conditionally on given controls $x_1, x_2, \dots x_n$, can be calculated as

$$f_\theta^n(x_1, \dots, x_n; y_1, \dots, y_n) = \prod_{i=1}^n f_\theta(y_i|x_i), \tag{18}$$

with respect to the product-measure $\mu^n = \mu \otimes \dots \otimes \mu$ of μ n times by itself. It is easy to see that any expectation, which uses a control policy χ, can be expressed as

$$E_\theta^\chi g(Y^{(n)}) = \int g(y^{(n)}) f_\theta^{n,\chi}(y^{(n)}) \, \mathrm{d}\mu^n(y^{(n)}),$$

where

$$f_\theta^{n,\chi}(y^{(n)}) = \prod_{i=1}^n f_\theta(y_i|x_i)$$

with

$$x_i = \chi_i(x^{(i-1)}, y^{(i-1)}) \tag{19}$$

for any $i = 1, 2, \dots$.

Similarly, for any function $F_n = F_n(x^{(n)}, y^{(n)})$ let us define

$$F_n^\chi(y^{(n)}) = F_n(x^{(n)}, y^{(n)})$$

where $x_1, \dots, x_n$ are defined by (19).

As a first step of minimization of $L(\chi, \psi, \phi)$, let us prove the following

Theorem 3. For any $\lambda_{ij} \geq 0$, $i = 1, \ldots, k$, $j \neq i$, and for any sequential testing procedure (χ, ψ, ϕ)

$$L(\chi, \psi, \phi) \geq N(\chi, \psi) + \sum_{n=1}^{\infty} \int (1 - \psi_1^\chi) \ldots (1 - \psi_{n-1}^\chi) \psi_n^\chi l_n^\chi \, d\mu^n, \tag{20}$$

where

$$l_n = \min_{1 \leq j \leq k} \sum_{i \neq j} \lambda_{ij} f_{\theta_i}^n. \tag{21}$$

The right-hand side of (20) is attained if

$$\phi_{nj} \leq I_{\left\{\sum_{i \neq j} \lambda_{ij} f_{\theta_i}^n = l_n\right\}} \tag{22}$$

for any $n = 1, 2, \ldots$ and for any $j = 1, \ldots k$.

Proof. Let us suppose that $N(\chi, \psi) < \infty$, otherwise (20) is trivial. Then let us prove an equivalent to (20) inequality:

$$\sum_{1 \leq i,j \leq k;\, j \neq i} \lambda_{ij} \alpha_{ij}(\chi, \psi, \phi) \geq \sum_{n=1}^{\infty} \int (1 - \psi_1^\chi) \ldots (1 - \psi_{n-1}^\chi) \psi_n^\chi l_n^\chi \, d\mu^n. \tag{23}$$

The left-hand side of it can be represented as

$$\begin{aligned} &\sum_{1 \leq i,j \leq k;\, j \neq i} \lambda_{ij} \alpha_{ij}(\chi, \psi, \phi) \\ &= \sum_{n=1}^{\infty} \int (1 - \psi_1^\chi) \ldots (1 - \psi_{n-1}^\chi) \psi_n^\chi \sum_{j=1}^{k} \left(\sum_{1 \leq i \leq k;\, i \neq j} \lambda_{ij} f_{\theta_i}^{n,\chi} \right) \phi_{nj}^\chi \, d\mu^n \end{aligned} \tag{24}$$

(see (2)).

Applying Lemma 1 [7] to each summand on the right-hand side of (24) we immediately have:

$$\sum_{1 \leq i,j \leq k;\, j \neq i} \lambda_{ij} \alpha_{ij}(\chi, \psi, \phi) \geq \sum_{n=1}^{\infty} \int (1 - \psi_1^\chi) \ldots (1 - \psi_{n-1}^\chi) \psi_n^\chi l_n^\chi \, d\mu^n \tag{25}$$

with an equality if

$$\phi_{nj} \leq I_{\{\sum_{i \neq j} \lambda_{ij} f_{\theta_i}^n = l_n\}}$$

for any $n = 1, 2, \ldots$ and for any $1 \leq j \leq k$. □

Remark 1. It is easy to see, using (4) and (25), that

$$L(\chi, \psi) = \inf_{\phi} L(\chi, \psi, \phi) = \sum_{n=1}^{\infty} \int (1 - \psi_1^\chi) \ldots (1 - \psi_{n-1}^\chi) \psi_n^\chi \left(n f_{\theta_1}^{n,\chi} + l_n^\chi \right) d\mu^n, \tag{26}$$

with l_n defined by (21), if $P_{\theta_1}^\chi(\tau_\psi < \infty) = 1$, and $L(\chi, \psi) = \infty$ otherwise.

Problem I is reduced now to the problem of finding strategies (χ, ψ) which minimize $L(\chi, \psi)$. Indeed, if there is a (χ^*, ψ^*) such that

$$L(\chi^*, \psi^*) = \inf_{(\chi,\psi)} L(\chi, \psi),$$

then for any ϕ^* satisfying

$$\phi^*_{nj} \leq I_{\left\{\sum_{i\neq j} \lambda_{ij} f^n_{\theta_i} = l_n\right\}}$$

(see (22)), by Theorem 3 for any (χ, ψ, ϕ)

$$L(\chi^*, \psi^*, \phi^*) = L(\chi^*, \psi^*) \leq L(\chi, \psi) = L(\chi, \psi, \phi^*),$$

thus, the conditions of Theorem 1 are fulfilled with $\alpha_{ij} = \alpha_{ij}(\chi^*, \psi^*, \phi^*)$ for $i, j = 1, \ldots, k$, $i \neq j$.

Because of this, in what follows we solve the problem of minimizing $L(\chi, \psi)$.

Let us denote, for the rest of this article,

$$s^\psi_n = (1 - \psi_1) \ldots (1 - \psi_{n-1})\psi_n \quad \text{and} \quad c^\psi_n = (1 - \psi_1) \ldots (1 - \psi_{n-1})$$

for any $n = 1, 2, \ldots$ (being $s^\psi_1 \equiv \psi_1$ and $c^\psi_1 \equiv 1$). Respectively,

$$s^{\psi,\chi}_n = (1 - \psi^\chi_1) \ldots (1 - \psi^\chi_{n-1})\psi^\chi_n \quad \text{and} \quad c^{\psi,\chi}_n = (1 - \psi^\chi_1) \ldots (1 - \psi^\chi_{n-1})$$

for any $n = 1, 2, \ldots$ (being $s^{\psi,\chi}_1 \equiv \psi^\chi_1$ and $c^{\psi,\chi}_1 \equiv 1$ as well).

Let also

$$C^{\psi,\chi}_n = \{y^{(n)} : (1 - \psi^\chi_1(y^{(1)})) \ldots (1 - \psi^\chi_{n-1}(y^{(n-1)})) > 0\},$$

for any $n \geq 2$, and let $C^{\psi,\chi}_1$ be the space of all $y^{(1)}$, and finally let

$$\bar{C}^{\psi,\chi}_n = \{y^{(n)} : (1 - \psi^\chi_1(y^{(1)})) \ldots (1 - \psi^\chi_n(y^{(n)})) > 0\},$$

for any $n \geq 1$.

3. OPTIMAL CONTROL AND STOPPING

In this section, the problem of finding strategies (χ, ψ) minimizing $L(\chi, \psi)$ (see (26)) will be solved.

3.1. Truncated stopping rules

In this section, we solve, as an intermediate step, the problem of minimization of $L(\chi, \psi)$ over all strategies (χ, ψ) with truncated stopping rules, i.e. such ψ that

$$\psi = (\psi_1, \psi_2, \ldots, \psi_{N-1}, 1, \ldots). \tag{27}$$

Let Δ^N be the class of stopping rules ψ of type (27), where N is any integer, $N \geq 2$.

The following Theorem can be proved in the same way as Theorem 4.2 in [6].

Theorem 4. Let $\psi \in \Delta^N$ be any (truncated) stopping rule, and χ any control policy. Then for any $1 \le r \le N-1$ the following inequalities hold true

$$L(\chi,\psi) \ge \sum_{n=1}^{r} \int s_n^{\psi,\chi}(n f_{\theta_1}^{n,\chi} + l_n^{\chi})\,\mathrm{d}\mu^n + \int c_{r+1}^{\psi,\chi}\left((r+1) f_{\theta_1}^{r+1,\chi} + V_{r+1}^{N,\chi}\right)\mathrm{d}\mu^{r+1} \tag{28}$$

$$\ge \sum_{n=1}^{r-1} \int s_n^{\psi,\chi}(n f_{\theta_1}^{n,\chi} + l_n^{\chi})\,\mathrm{d}\mu^n + \int c_r^{\psi,\chi}\left(r f_{\theta_1}^{r,\chi} + V_r^{N,\chi}\right)\mathrm{d}\mu^r, \tag{29}$$

where $V_N^N \equiv l_N$, and recursively for $n = N, N-1, \dots 2$

$$V_{n-1}^N = \min\{l_{n-1}, f_{\theta_1}^{n-1} + R_{n-1}^N\}, \tag{30}$$

with

$$R_{n-1}^N = R_{n-1}^N(x^{(n-1)}; y^{(n-1)}) = \min_{x_n} \int V_n^N(x_1,\dots,x_n; y_1,\dots,y_n)\,\mathrm{d}\mu(y_n). \tag{31}$$

The lower bound in (29) is attained if and only if

$$I_{\{l_n^{\chi} < f_{\theta_1}^{n,\chi} + R_n^{N,\chi}\}} \le \psi_n^{\chi} \le I_{\{l_n^{\chi} \le f_{\theta_1}^{n,\chi} + R_n^{N,\chi}\}} \tag{32}$$

μ^n-almost everywhere on $C_n^{\psi,\chi}$ and

$$R_n^{N,\chi}(y^{(n)}) = \int V_{n+1}^{N,\chi}\,\mathrm{d}\mu(y_{n+1}) \tag{33}$$

μ^n-almost everywhere on $\bar{C}_n^{\psi,\chi}$, for any $n = r, \dots, N-1$.

Remark 2. It is supposed in Theorem 4, and in what follows in this article, that all the functions R_{n-1}^N defined by (31) are well-defined and measurable, for any $n = 2, \dots, N$, and for any $N = 1, 2, \dots$.

The following Corollary characterizes optimal strategies with truncated stopping rules. It immediately follows from Theorem 4 applied for $r = 1$.

Corollary 1. For any truncated stopping rule $\psi \in \Delta^N$, and for any control rule χ

$$L(\chi,\psi) \ge 1 + R_0^N, \tag{34}$$

where

$$R_0^N = \min_{x_1} \int V_1^N(x_1; y_1)\,\mathrm{d}\mu(y_1). \tag{35}$$

The lower bound in (34) is attained if and only if (32) is satisfied μ^n-almost everywhere on $C_n^{\psi,\chi}$ and (33) is satisfied μ^n-almost everywhere on $\bar{C}_n^{\psi,\chi}$, for any $n = 1, 2, \dots, N-1$ and, additionally,

$$R_0^N = \int V_1^N(\chi_1; y_1)\,\mathrm{d}\mu(y_1). \tag{36}$$

Remark 3. It is obvious that the testing procedure attaining the lower bound in (34) is optimal among all truncated testing procedures with $\psi \in \Delta^N$. But it only makes practical sense if

$$l_0 = \min_{1\le j\le k} \sum_{i\neq j} \lambda_{ij} > 1 + R_0^N.$$

The reason is that l_0 can be considered as "the $L(\chi,\psi)$" function for a trivial sequential testing procedure (χ_0, ψ_0, ϕ_0) which, without taking any observations, applies any decision rule ϕ_0 such that $\phi_{0j} \le I_{\{\sum_{i\neq j}\lambda_{ij}=l_0\}}$ for any $j = 1, \ldots, k$. In this case there are no observations ($N(\theta;\psi_0) = 0$), χ_0 is nothing, and it is easily seen that

$$L(\chi_0, \psi_0, \phi_0) = \sum_{j=1}^{k} \sum_{i\neq j} \lambda_{ij}\phi_{0j} = l_0.$$

Thus, the inequality

$$l_0 \le 1 + R_0^N$$

means that the trivial testing procedure (χ_0, ψ_0, ϕ_0) is not worse than the best testing procedure with ψ from Δ^N.

Because of this, we may think that

$$V_0^N = \min\{l_0, 1 + R_0^N\}$$

is the minimum value of $L(\chi,\psi)$ when taking no observations is permitted. It is obvious that this is a particular case of (30) with $n = 1$, if we define $f_\theta^0 \equiv 1$.

3.2. General stopping rules

In this section we characterize the structure of general sequential testing procedures minimizing $L(\chi,\psi)$.

Let us define for any stopping rule ψ and any control policy χ

$$L_N(\chi,\psi) = \sum_{n=1}^{N-1} \int s_n^{\psi,\chi}(n f_{\theta_1}^{n,\chi} + l_n^\chi)\,\mathrm{d}\mu^n + \int c_N^{\psi,\chi}\left(N f_{\theta_1}^{N,\chi} + l_N^\chi\right)\mathrm{d}\mu^N. \tag{37}$$

This is the Lagrange-multiplier function corresponding to ψ truncated at N, i. e. the rule with the components $\psi^N = (\psi_1, \psi_2, \ldots, \psi_{N-1}, 1, \ldots)$, $L_N(\chi,\psi) = L(\chi,\psi^N)$.

Since ψ^N is truncated, the results of the preceding section apply, in particular, the inequalities of Theorem 4.

The idea of what follows is to make $N \to \infty$, to obtain some lower bounds for $L(\chi,\psi)$ from (28) - (29). Obviously, we need that $L_N(\chi,\psi) \to L(\chi,\psi)$ as $N \to \infty$. A manner to guarantee this is using the following definition.

Let us denote by $\mathscr{F}$ the set of all strategies (χ,ψ) such that

$$\lim_{n\to\infty} E_{\theta_i}^\chi (1-\psi_1)\ldots(1-\psi_n) = 0 \quad \text{for any} \quad i = 1, 2, \ldots, k. \tag{38}$$

It is easy to see that (38) is equivalent to

$$P_{\theta_i}^\chi(\tau_\psi < \infty) = 1 \quad \text{for any} \quad i = 1, 2, \ldots, k$$

(see (1)).

Lemma 1. For any strategy $(\chi,\psi)\in\mathscr{F}$

$$\lim_{N\to\infty} L_N(\chi,\psi)=L(\chi,\psi).$$

Proof. Practically coincides with that of Lemma 5.1 in [6] (with $f_{\theta_1}^n$ instead of $f_{\theta_0}^n$), except that in order to show the convergence

$$\int c_N^{\psi,\chi} l_N^{\chi}\,\mathrm{d}\mu^N\to 0,\quad N\to\infty,$$

we use the following estimate:

$$\int c_N^{\psi,\chi} l_N^{\chi}\,\mathrm{d}\mu^N\le \max_{i\neq j}\lambda_{ij}\sum_{i=1}^{k}\int c_N^{\psi,\chi} f_{\theta_i}^{N,\chi}\,\mathrm{d}\mu^N=\max_{i\neq j}\lambda_{ij}\sum_{i=1}^{k}E_{\theta_i}^{\chi}c_N^{\psi}\to 0 \tag{39}$$

as $N\to\infty$, because of (38). □

The second fact we need is about the behaviour of the functions V_r^N which participate in the inequalities of Theorem 4, as $N\to\infty$.

Lemma 2. For any $n\ge 1$ and for any $N\ge n$

$$V_n^N\ge V_n^{N+1}. \tag{40}$$

Proof. Completely analogous to the proof of Lemma 5.2 [6] (with $f_{\theta_1}^n$ instead of $f_{\theta_0}^n$). □

It follows from Lemma 2 that for any fixed $n\ge 1$ the sequence V_n^N is non-increasing. So, there exists

$$V_n=\lim_{N\to\infty}V_n^N. \tag{41}$$

Now, passing to the limit, as $N\to\infty$, in (28) and (29) with $\psi=\psi^N$, we have the following Theorem. The left-hand side of (28) tends to $L(\chi,\psi)$ by Lemma 1. Passing to the limit on the right hand side of (28) and in (29) is possible by Lebesgue's monotone convergence theorem, by virtue of Lemma 2.

Theorem 5. Let $(\chi,\psi)\in\mathscr{F}$ be any control-stopping strategy. Then for any $r\ge 1$ the following inequalities hold

$$L(\chi,\psi)\ge\sum_{n=1}^{r}\int s_n^{\psi,\chi}(nf_{\theta_1}^{n,\chi}+l_n^{\chi})\,\mathrm{d}\mu^n+\int c_{r+1}^{\psi,\chi}\left((r+1)f_{\theta_1}^{r+1,\chi}+V_{r+1}^{\chi}\right)\mathrm{d}\mu^{r+1} \tag{42}$$

$$\ge\sum_{n=1}^{r-1}\int s_n^{\psi,\chi}(nf_{\theta_1}^{n,\chi}+l_n^{\chi})\,\mathrm{d}\mu^n+\int c_n^{\psi,\chi}\left(rf_{\theta_1}^{r,\chi}+V_r^{\chi}\right)\mathrm{d}\mu^r, \tag{43}$$

where

$$V_r=\min\{l_r,f_{\theta_1}^r+R_r\}, \tag{44}$$

being

$$R_r = R_r(x^{(r)}, y^{(r)}) = \min_{x_{r+1}} \int V_{r+1}(x^{(r+1)}, y^{(r+1)}) \, \mathrm{d}\mu(y_{r+1}). \tag{45}$$

In particular, for $r = 1$, the following lower bound holds true:

$$L(\chi, \psi) \geq 1 + \int V_1^{\chi} \, \mathrm{d}\mu(y_1) \geq 1 + R_0, \tag{46}$$

where, by definition,

$$R_0 = \min_{x_1} \int V_1(x_1, y_1) \, \mathrm{d}\mu(y_1).$$

Exactly as in [6] (see Lemma 5.4 [6]) it can be proved that the right-hand side of (46) coincides with

$$\inf_{(\chi,\psi)\in\mathscr{F}} L(\chi, \psi).$$

In fact, this is true for any $\mathscr{F}$ such that $(\chi, \psi) \in \mathscr{F}$ implies $L_N(\chi, \psi) \to L(\chi, \psi)$ as $N \to \infty$.

The following theorem characterizes the structure of optimal strategies.

Theorem 6. If there is a strategy $(\chi, \psi) \in \mathscr{F}$ such that

$$L(\chi, \psi) = \inf_{(\chi',\psi')\in\mathscr{F}} L(\chi', \psi'), \tag{47}$$

then

$$I_{\{l_r^{\chi} < f_{\theta_1}^{r,\chi} + R_r^{\chi}\}} \leq \psi_r^{\chi} \leq I_{\{l_r^{\chi} \leq f_{\theta_1}^{r,\chi} + R_r^{\chi}\}} \tag{48}$$

μ^r-almost everywhere on $C_r^{\psi,\chi}$, and

$$\int V_{r+1}^{\chi}(y^{(r+1)}) \, \mathrm{d}\mu(y_{r+1}) = R_r^{\chi} \tag{49}$$

μ^r-almost everywhere on $\bar{C}_r^{\psi,\chi}$, for any $r = 1, 2\ldots$, where χ_1 is defined in such a way that

$$\int V_1^{\chi} \, \mathrm{d}\mu(y_1) = R_0. \tag{50}$$

On the other hand, if a strategy (ψ, χ) satisfies (48) μ^r-almost everywhere on $C_r^{\psi,\chi}$, and satisfies (49) μ^r-almost everywhere on $\bar{C}_r^{\psi,\chi}$, for any $r = 1, 2\ldots$, where χ_1 is such that (50) is fulfilled, and $(\psi, \chi) \in \mathscr{F}$, then (47) holds.

P r o o f. Almost literally coincides with the proof of Theorem 5.5 [6] (substituting $f_{\theta_0}^n$ by $f_{\theta_1}^n$), with the omission of the proof that $(\psi, \chi) \in \mathscr{F}$ in the "if"-part (see (76) and (77) in [6]), because now it is a condition of Theorem 6. □

Remark 4. Theorem 6 treats the optimality among strategies which take at least one observation. If we allow to take no observations, there is a possibility that the trivial testing procedure (see Remark 3) gives a better result. It is easy to see that this happens if and only if

$$l_0 < 1 + R_0.$$

4. LIKELIHOOD RATIO STRUCTURE OF OPTIMAL STRATEGY

In this section, we will give to the optimal strategy in Theorem 6 an equivalent form related to the likelihood ratio process, supposing that all the distributions given by f_{θ_i} are absolutely continuous with respect to that given by f_{θ_1}. More precisely, we will suppose that for any x

$$\{y : f_{\theta_1}(y|x) = 0\} \subset \bigcap_{i>1} \{y : f_{\theta_i}(y|x) = 0\}. \tag{51}$$

Let us start with defining the likelihood ratios:

$$Z_n^r = Z_n^r(x^{(n)}, y^{(n)}) = \prod_{i=1}^{n} \frac{f_{\theta_r}(y_i|x_i)}{f_{\theta_1}(y_i|x_i)}, \quad r > 1,$$

and let $Z_n = (Z_n^2, \dots, Z_n^k)$.

Let us introduce then the following sequence of functions $\rho_r = \rho_r(z)$, $r = 0, 1, \dots$, where $z = (z_2, \dots z_k)$.

Let

$$\rho_0(z) = g(z) \equiv \min_j \sum_{i \neq j} \lambda_{ij} z_i, \tag{52}$$

where, by definition, $z_1 \equiv 1$. Let for $r = 1, 2, 3, \dots$, recursively,

$$\rho_r(z) = \min\left\{g(z), 1+\min_x \int f_{\theta_1}(y|x)\rho_{r-1}\left(z_2\frac{f_{\theta_2}(y|x)}{f_{\theta_1}(y|x)}, \dots, z_k\frac{f_{\theta_k}(y|x)}{f_{\theta_1}(y|x)}\right) \mathrm{d}\mu(y)\right\} \tag{53}$$

(we are supposing that all ρ_r, $r = 0, 1, 2, \dots$ are well-defined and measurable functions of z). It is easy to see that (see (30), (31))

$$V_N^N = f_{\theta_1}^N \rho_0(Z_N),$$

and for $r = N-1, N-2, \dots, 1$

$$V_r^N = f_{\theta_1}^r \rho_{N-r}(Z_r). \tag{54}$$

It is not difficult to see (very much like in Lemma 2) that

$$\rho_r(z) \geq \rho_{r+1}(z)$$

for any $r = 0, 1, 2, \dots$, so there exists

$$\rho(z) = \lim_{n\to\infty} \rho_n(z). \tag{55}$$

Using arguments similar to those used for obtaining Theorem 5, it can be shown, starting from (53), that

$$\rho(z) = \min\left\{g(z), 1 + R(z)\right\}, \tag{56}$$

where

$$R(z) = \min_x \int f_{\theta_1}(y|x)\rho\left(z_2\frac{f_{\theta_2}(y|x)}{f_{\theta_1}(y|x)}, \dots, z_k\frac{f_{\theta_k}(y|x)}{f_{\theta_1}(y|x)}\right) \mathrm{d}\mu(y). \tag{57}$$

Let us pass now to the limit, as $N \to \infty$, in (54). We see that

$$V_k = f_{\theta_1}^k \rho(Z_k).$$

Using this expression in Theorem 6 we get

Theorem 7. If there exists a strategy $(\chi, \psi) \in \mathscr{F}$ such that

$$L(\chi, \psi) = \inf_{(\chi', \psi') \in \mathscr{F}} L(\chi', \psi'), \tag{58}$$

then

$$I_{\{g(Z_r^\chi) < 1 + R(Z_r^\chi)\}} \leq \psi_r^\chi \leq I_{\{g(Z_r^\chi) \leq 1 + R(Z_r^\chi)\}} \tag{59}$$

$P_{\theta_1}^\chi$-almost sure on

$$\{y^{(r)} :\ (1 - \psi_1^\chi(y^{(1)})) \dots (1 - \psi_{r-1}^\chi(y^{(r-1)})) > 0\}, \tag{60}$$

and

$$\int f_{\theta_1}(y|\chi_{r+1}) \rho \left(Z_r^{2,\chi} \frac{f_{\theta_2}(y|\chi_{r+1})}{f_{\theta_1}(y|\chi_{r+1})}, \dots, Z_r^{k,\chi} \frac{f_{\theta_k}(y|\chi_{r+1})}{f_{\theta_1}(y|\chi_{r+1})} \right) \mathrm{d}\mu(y) = R(Z_r^\chi) \tag{61}$$

$P_{\theta_1}^\chi$-almost sure on

$$\{y^{(r)} :\ (1 - \psi_1^\chi(y^{(1)})) \dots (1 - \psi_r^\chi(y^{(r)})) > 0\}, \tag{62}$$

where χ_1 is defined in such a way that

$$\int f_{\theta_1}(y|\chi_1) \rho \left(\frac{f_{\theta_2}(y|\chi_1)}{f_{\theta_1}(y|\chi_1)}, \dots, \frac{f_{\theta_k}(y|\chi_1)}{f_{\theta_1}(y|\chi_1)} \right) \mathrm{d}\mu(y) = R(1). \tag{63}$$

On the other hand, if (χ, ψ) satisfies (59) $P_{\theta_1}^\chi$-almost sure on (60) and satisfies (61) $P_{\theta_1}^\chi$-almost sure on (62), for any $r = 1, 2, \dots$, where χ_1 satisfies (63), and $(\chi, \psi) \in \mathscr{F}$, then (χ, ψ) satisfies (58).

5. APPLICATION TO THE CONDITIONAL PROBLEMS

In this section, we apply the results obtained in the preceding sections to minimizing the average sample size $N(\chi, \psi) = E_{\theta_1}^\chi \tau_\psi$ over all sequential testing procedures with error probabilities not exceeding some prescribed levels (see Problems I and II in Section 1).

Combining Theorems 1, 3 and 6, we immediately have the following solution to Problem I.

Theorem 8. Let $(\chi, \psi) \in \mathscr{F}$ satisfy the conditions of Theorem 6 with $\lambda_{ij} > 0$, $i, j = 1, \dots, k$, $i \neq j$ (recall that l_n, V_n, and R_n are functions of λ_{ij}), and let ϕ be any decision rule satisfying (22).

Then for any sequential testing procedure $(\chi', \psi', \phi') \in \mathscr{F}$ such that

$$\alpha_{ij}(\chi', \psi', \phi') \leq \alpha_{ij}(\chi, \psi, \phi) \quad \text{for any} \quad i, j = 1, \dots, k,\ i \neq j, \tag{64}$$

it holds

$$N(\chi', \psi') \geq N(\chi, \psi). \tag{65}$$

The inequality in (65) is strict if at least one of the inequalities in (64) is strict.

If there are equalities in all of the inequalities in (64) and (65), then (χ', ψ') satisfies the condition of Theorem 6 as well (with χ' instead of χ and ψ' instead of ψ).

P r o o f. The only thing to be proved is the last assertion.

Let us suppose that

$$\alpha_{ij}(\chi', \psi', \phi') = \alpha_{ij}(\chi, \psi, \phi), \quad \text{for any} \quad i, j = 1, \dots, k,\ i \neq j,$$

and

$$N(\chi', \psi') = N(\chi, \psi).$$

Then, obviously,

$$L(\chi, \psi, \phi) = L(\chi, \psi) = L(\chi', \psi', \phi') \geq L(\chi', \psi') \tag{66}$$

(see (7)) and Remark 1.

By Theorem 6, there can not be strict inequality in the last inequality in (66), so $L(\chi, \psi) = L(\chi', \psi')$. From Theorem 6 it follows now that (χ', ψ') satisfies the conditions of Theorem 6 as well. □

Analogously, combining Theorems 2, 3 and 6, we also have the following solution to Problem II.

Theorem 9. Let $(\chi, \psi) \in \mathscr{F}$ satisfy the conditions of Theorem 6 with $\lambda_{ij} = \lambda_i > 0$ for any $i = 1, \dots k$ and for any $j = 1, \dots, k$, and let ϕ be any decision rule such that

$$\phi_{nj} \leq I_{\left\{\sum_{i \neq j} \lambda_i f_{\theta_i}^n = \min_j \sum_{i \neq j} \lambda_i f_{\theta_i}^n\right\}}$$

for any $j = 1, \dots, k$ and for any $n = 1, 2, \dots$.

Then for any sequential testing procedure $(\chi', \psi', \phi') \in \mathscr{F}$ such that

$$\beta_i(\chi', \psi', \phi') \leq \beta_i(\chi, \psi, \phi) \quad \text{for any} \quad i = 1, \dots, k, \tag{67}$$

it holds

$$N(\chi', \psi') \geq N(\chi, \psi). \tag{68}$$

The inequality in (68) is strict if at least one of the inequalities in (67) is strict.

If there are equalities in all of the inequalities in (67) and (68), then (χ', ψ') satisfies the conditions of Theorem 6 with $\lambda_{ij} = \lambda_i$, $i, j = 1, \dots, k$, $i \neq j$, as well (with χ' instead of χ and ψ' instead of ψ).

6. ADDITIONAL RESULTS, EXAMPLES AND DISCUSSION

6.1. Some general remarks

Remark 5. The class $\mathscr{F}$ defined by (38) can be extended in such a way that Theorem 6 remains valid. It can be defined as the class of all the strategies (χ, ψ) for which

$$\lim_{n \to \infty} E_{\theta_i}^{\chi} (1 - \psi_1) \dots (1 - \psi_n) = 0 \tag{69}$$

for at least $k - 1$ different values of θ_i. To see this it is sufficient to notice that for any strategy in this extended class

$$L_N(\chi, \psi) \to L(\chi, \psi), \quad \text{as} \quad N \to \infty,$$

because (see the proof of Lemma 1)

$$\int c_N^{\psi,\chi} l_N^\chi \,\mathrm{d}\mu^N \le \sum_{1\le i\le k, i\ne j} \lambda_{ij} \int c_N^{\psi,\chi} f_{\theta_i}^{N,\chi} \,\mathrm{d}\mu^N = \sum_{1\le i\le k, i\ne j} \lambda_{ij} E_{\theta_i}^\chi c_N^\psi \to 0, \quad N\to\infty,$$

if j corresponds to θ_j for which (69) does not hold.

Obviously, Theorem 6 remains valid with this extension of $\mathscr{F}$.

Moreover, in the same way, Theorem 6 remains valid if $\mathscr{F}$ is defined as the class of all strategies (χ,ψ) for which

$$L_N(\chi,\psi) \to Ł(\chi,\psi), \quad N\to\infty.$$

But the statistical meaning of this class is not clear, so we prefer for $\mathscr{F}$ one of the definitions above.

Remark 6. In the same way as in the preceding sections, a more general problem than just minimizing $N(\theta_1;\chi,\psi)$ can be treated (see (4) and Problems I and II thereafter).

Namely, we can minimize any convex combination of the average sample numbers, or

$$\sum_{i=1}^k c_i N(\theta_i;\chi,\psi),$$

where $c_i\ge 0$, $i=1,\dots,k$, are arbitrary but fixed constants. More exactly, if we modify the definition of the functions V_r^N in (30) to

$$V_{r-1}^N = \min\left\{ l_{r-1}, \sum_{i=1}^k c_i f_{\theta_i}^{r-1} + R_{r-1}^N \right\}, \tag{70}$$

for $r=N,\dots,2$, being, as before,

$$V_r = \lim_{N\to\infty} V_r^N,$$

and, respectively, change (48) in Theorem 6 to

$$I_{\{l_r^\chi < \sum_{i=1}^k c_i f_{\theta_i}^{r,\chi} + R_r^\chi\}} \le \psi_r^\chi \le I_{\{l_r^\chi \le \sum_{i=1}^k c_i f_{\theta_i}^{r,\chi} + R_r^\chi\}} \tag{71}$$

then Theorem 6 remains valid. Theorems 4, 7, 8 and 9 can be modified respectively.

6.2. An example

In this Section we show how our results can be applied to a concrete statistical model.

Let us suppose that any stage of our experiment is a regression experiment with a normal response. More specifically, we are supposing that the distribution of Y, given a value of the control variable X, is normal with mean value θX and a know variance σ^2, say $\sigma^2=1$.

Thus,
$$f_\theta(y|x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(y-\theta x)^2}{2}\right\}, \quad -\infty < y < \infty \tag{72}$$

For simplicity, let us take $k = 2$ simple hypotheses, for example, $H_1 : \theta = 1$ and $H_2 : \theta = 2$, and suppose that the control variable takes only two values, say, $x = 1$ and $x = 2$.

Condition (51) is fulfilled in an obvious way.

Let $\lambda_{12} > 0$ and $\lambda_{21} > 0$ be two arbitrary constants. We start with defining

$$\rho_0(z) = g(z) \equiv \min\{\lambda_{12}, \lambda_{21} z\},$$

(see (52)).

Next, we calculate
$$\frac{f_2(y|x)}{f_1(y|x)} = \exp\{xy - 3x^2/2\},$$
and
$$\rho_{n+1}(z) = \min\{g(z), 1 + \min_{x=1,2} \int_{-\infty}^{\infty} \rho_n(z \exp\{xy - 3x^2/2\}) \frac{\exp\{-(y-x)^2/2\}}{\sqrt{2\pi}}\, \mathrm{d}y,$$

for $n = 0, 1, 2, \ldots$ (see (53)).

Let $\rho(z) = \lim_{n\to\infty} \rho_n(z)$, and

$$R(z) = \min_{x=1,2} \int_{-\infty}^{\infty} \rho(z \exp\{xy - 3x^2/2\}) \frac{\exp\{-(y-x)^2/2\}}{\sqrt{2\pi}}\, \mathrm{d}y.$$

Now, by Theorem 7, an optimal strategy will be defined on the basis of the likelihood ratio process

$$Z_n = \exp\left\{\sum_{i=1}^{n} (X_i Y_i - 3X_i^2/2)\right\},$$

being the optimal stopping time $\tau = \min\{n : g(Z_n) \le 1 + R(Z_n)\}$, whereas at each stage $n = 1, 2, \ldots$ the next control value $X_{n+1} = x$ ($x = 1$ or $x = 2$) is defined in such a way that

$$R(Z_n) = \int_{-\infty}^{\infty} \rho(Z_n \exp\{xy - 3x^2/2\}) \frac{\exp\{-(y-x)^2/2\}}{\sqrt{2\pi}}\, \mathrm{d}y,$$

starting from X_1 defined as x ($x = 1$ or $x = 2$) for which

$$R(1) = \int_{-\infty}^{\infty} \rho(\exp\{xy - 3x^2/2\}) \frac{\exp\{-(y-x)^2/2\}}{\sqrt{2\pi}}\, \mathrm{d}y.$$

When the test terminates at some stage $\tau = n$, we should reject H_1, if $\lambda_{21} Z_n \ge \lambda_{12}$, and accept H_1 otherwise (see Theorem 4).

One can vary the error probability levels of this test by changing the values of λ_{12} and λ_{21}.

6.3. Bayesian testing of multiple hypotheses

In this section we characterize the structure of Bayesian multiple hypothesis tests.

Let $\pi_i > 0$, $i = 1, \ldots, k$ be prior probabilities of H_i, $i = 1, \ldots, k$, respectively, $\sum_{i=1}^k \pi_i = 1$, and let $w_{ij} \geq 0$, $i, j = 1, \ldots, k$, be some losses due to incorrect decisions (we assume that $w_{ii} = 0$ for any $i = 1, \ldots, k$). Then, for any sequential testing procedure (χ, ψ, ϕ), we define the Bayes risk as

$$r(\chi, \psi, \phi) = \sum_{i=1}^k \pi_i \left(cE_{\theta_i}^\chi \tau_\psi + \sum_{j=1}^k w_{ij} \alpha_{ij}(\chi, \psi, \phi) \right), \tag{73}$$

where $c > 0$ is some unitary observation cost (cf. Section 9.4 of [13], see also Chapter 5 of [2] for a more general sequential Bayesian decision theory, both monographs treating non-controlled experiments). Let us call *Bayesian* any testing procedure (χ, ψ, ϕ) minimizing (73).

In this section, we show that the Bayesian testing procedures always exist, and characterize the structure of both truncated and non-truncated Bayesian testing procedures for the controlled experiments.

To formulate our results, we use the notation of Sections 1 – 5, but we have to re-define some elements have been defined therein.

First of all, it is easy to see from Theorem 3 that the optimal decision rule ϕ has the following form. Let

$$l_n = \min_{1 \leq j \leq k} \sum_{i=1}^k \pi_i w_{ij} f_{\theta_i}^n. \tag{74}$$

(cf. (21)). Then the decision rule ϕ is optimal ($\inf_{\phi'} r(\chi, \psi, \phi') = r(\chi, \psi, \phi)$ for any χ and ψ) if

$$\phi_{nj} \leq I_{\{\sum_{i=1}^k \pi_i w_{ij} f_{\theta_i}^n = l_n\}} \tag{75}$$

for any $j = 1, \ldots, k$ and for any $n = 1, 2, \ldots$ (see Theorem 3).

Let Π be the prior distribution defined by π_i, $i = 1, \ldots, k$, and let, by definition,

$$f_\Pi^n = \sum_{i=1}^k \pi_i f_{\theta_i}^n$$

for any $n = 1, 2, \ldots$.

For any $N = 1, 2, \ldots$ let us define

$$V_N^N = l_N, \tag{76}$$

and for any $n = N-1, N-2, \ldots, 1$, recursively,

$$V_n^N = \min\{l_n, cf_\Pi^n + R_n\}, \tag{77}$$

where

$$R_n^N = R_n^N(x^{(n)}, y^{(n)}) = \min_{x_{n+1}} \int V_{n+1}^N(x_1, \ldots, x_{n+1}; y_1, \ldots, y_{n+1})\, \mathrm{d}\mu(y_{n+1}). \tag{78}$$

Let also

$$R_0^N = \min_{x_1} \int V_1^N(x_1; y_1)\, \mathrm{d}\mu(y_1). \tag{79}$$

The following Theorem characterizes Bayesian procedures with truncated stopping rules and can be proved in exactly the same way as Corollary 1.

Theorem 10. Let χ be any control policy, $\psi \in \Delta^N$ be any (truncated) stopping rule and ϕ any decision rule satisfying (75) for any $j = 1, \ldots, k$ and for any $n = 1, 2, \ldots$. Then

$$r(\chi, \psi, \phi) \geq c + R_0^N. \tag{80}$$

There is an equality in (80) if and only if

$$I_{\{l_n^\chi < c f_\Pi^{n,\chi} + R_n^{N,\chi}\}} \leq \psi_n^\chi \leq I_{\{l_n^\chi \leq c f_\Pi^{n,\chi} + R_n^{N,\chi}\}} \tag{81}$$

μ^n-almost everywhere on $C_n^{\psi,\chi}$ and

$$R_n^{N,\chi}(y^{(n)}) = \int V_{n+1}^{N,\chi}(y^{(n+1)})\, \mathrm{d}\mu(y_{n+1}) \tag{82}$$

μ^n-almost everywhere on $\bar{C}_n^{\psi,\chi}$, for any $n = 1, \ldots, N-1$, and, additionally,

$$R_0^N = \int V_1^N(\chi_1; y_1)\, \mathrm{d}\mu(y_1). \tag{83}$$

Let now $V_n = \lim_{N\to\infty} V_n^N$, $n = 1, 2, \ldots$ Respectively, $R_n = \lim_{N\to\infty} R_n^N$, $n = 0, 1, 2, \ldots$.

Theorem 11. Let χ be any control policy, ψ any stopping rule, and ϕ any decision rule satisfying (75) for any $j = 1, \ldots, k$ and for any $n = 1, 2, \ldots$. Then

$$r(\chi, \psi, \phi) \geq c + R_0. \tag{84}$$

There is an equality in (84) if and only if

$$I_{\{}l_n^\chi < c f_\Pi^{n,\chi} + R_n^\chi\} \leq \psi_n^\chi \leq I_{\{l_n^\chi \leq c f_\Pi^{n,\chi} + R_n^\chi\}} \tag{85}$$

μ^n-almost everywhere on $C_n^{\psi,\chi}$ and

$$R_n^\chi(y^{(n)}) = \int V_{n+1}^\chi(y^{(n+1)})\, \mathrm{d}\mu(y_{n+1}) \tag{86}$$

μ^n-almost everywhere on $\bar{C}_n^{\psi,\chi}$, for any $n = 1, 2 \ldots$, and, additionally,

$$R_0 = \int V_1(\chi_1; y_1)\, \mathrm{d}\mu(y_1). \tag{87}$$

P r o o f. First of all we need to prove that (84) holds for *any* strategy (χ, ψ). Obviously, it suffices to prove this only for such (χ, ψ) that $r(\chi, \psi, \phi) < \infty$. But this latter fact implies, in particular, that $\sum_{i=1}^{k} \pi_i E_{\theta_i}^{\chi} \tau_{\psi} < \infty$ (see (73)). Because $\pi_i > 0$ for any $i = 1, \ldots k$, it follows that (χ, ψ) satisfies (38), so

$$r(\chi, \psi^N, \phi) \to r(\chi, \psi, \phi), \quad N \to \infty,$$

where ψ^N, by definition, is $(\psi_1, \psi_2, \ldots, \psi_{N-1}, 1, \ldots)$ (see the proof of Lemma 1).

The rest of the proof of the "only if"-part is completely analogous to the corresponding part of the proof of Theorem 6 (or Theorem 5.5 [6]).

To prove the "if"-part, first it can be shown, analogously to the proof of Theorem 5.5 [6], that

$$\sum_{n=1}^{r} \int s_n^{\psi,\chi} \left(cnf_{\Pi}^{n,\chi} + l_n^{\chi}\right) \mathrm{d}\mu^n + \int c_{r+1}^{\psi,\chi} \left(c(r+1) f_{\Pi}^{r+1,\chi} + V_{r+1}^{\chi}\right) \mathrm{d}\mu^{r+1} = c + R_0, \quad (88)$$

for any $r = 0, 1, 2, \ldots$, if (ψ, χ) satisfies (85) – (87).

Because $c > 0$, we have from (88), in particular, that

$$\sum_{i=1}^{k} \pi_i P_{\theta_i}^{\chi}(\tau_{\psi} \geq r + 1) = \int c_{r+1}^{\psi,\chi} f_{\Pi}^{r+1,\chi} \mathrm{d}\mu^{r+1} \leq \frac{c + R_0}{c(r + 1)} \to 0 \quad \text{as} \quad r \to \infty.$$

Because $\pi_i > 0$ for all $i = 1, \ldots, k$, this implies that for (χ, ψ) (38) is fulfilled. It follows from (88) now that

$$\lim_{r\to\infty} \sum_{n=1}^{r} \int s_n^{\psi,\chi}(cnf_{\Pi}^{n,\chi} + l_n^{\chi}) \,\mathrm{d}\mu^n = r(\chi, \psi, \phi) \leq c + R_0.$$

Along with (84) this gives that $r(\chi, \psi, \phi) = c + R_0$, i. e. there is an equality in (84). □

6.4. Experiments without control

In this section we draw consequences for statistical experiments without control.

Let us suppose that the density of Y given X does not depend on X: $f_{\theta}(y|x) \equiv f_{\theta}(y)$ for any y and for any θ, meaning that there is no way to control the flow of the experiment, and the observations $Y_1, Y_2, \ldots$ are independent and identically distributed (i.i.d.) random "variables" with probability "density" function $f_{\theta}(y)$. We can incorporate this particular case in the above scheme of controlled experiments thinking that there is some (fictitious) unique value of control variable at each stage of the experiment, thus, being any control policy trivial.

Because of this, any (sequential) testing procedure has in effect only two components in this case: a stopping rule ψ and a decision rule ϕ. So we use the notation of Section 6.3, simply omitting any mention of the control policy. For example, for any testing procedure (ψ, ϕ) the Bayesian risk (73) is now:

$$r(\psi, \phi) = \sum_{i=1}^{k} \pi_i \left(cE_{\theta_i} \tau_{\psi} + \sum_{j=1}^{k} w_{ij} \alpha_{ij}(\psi, \phi) \right). \quad (89)$$

Respectively, $f_\theta^n = f_\theta^n(y^{(n)}) = \prod_{i=1}^n f_\theta(y_i)$ in (74) now, and the functions V_n^N, R_n^N, V_n, R_n, etc. of the preceding section are all functions of $y^{(n)}$ only.

Theorem 11 of Section 6.3 now transforms to

Theorem 12. Let ψ be any stopping rule and ϕ any decision rule satisfying (75) for any $j = 1, \dots, k$ and for any $n = 1, 2, \dots$. Then

$$r(\psi, \phi) \geq c + R_0. \tag{90}$$

There is an equality in (90) if and only if

$$I_{\{l_n < cf_\Pi^n + R_n\}} \leq \psi_n^\chi \leq I_{\{l_n \leq cf_\Pi^n + R_n\}} \tag{91}$$

μ^n-almost everywhere on C_n^ψ for any $n = 1, 2, \dots$, where

$$R_n = R_n(y_1, \dots, y_n) = \int V_{n+1}(y_1, \dots, y_{n+1})\, \mathrm{d}\mu(y_{n+1}),$$

being, for any $n = 1, 2, \dots$, $V_n(y^{(n)}) = \lim_{N\to\infty} V_n^N(y^{(n)})$, where $V_N^N \equiv l_N$, and

$$V_n^N(y^{(n)}) = \min\left\{ l_n(y^{(n)}), cf_\Pi^n(y^{(n)}) + \int V_{n+1}^N(y^{(n+1)})\, \mathrm{d}\mu(y_{n+1}) \right\}$$

for any $n = N-1, \dots, 1$, $N = 1, 2, \dots$

In particular, this Theorem gives all solutions to the problem of Bayesian testing of multiple simple hypotheses for independent and identically distributed observations when the cost of observations is linear (see Section 9.4 of [13] and suppose that $K(X_1, \dots, X_n) \equiv n$ therein). A more general case of composite hypotheses (as stated in Section 9.4 [13]), can be treated, using essentially the same method as in the present article, with the help of the results of [8], even when the observations are dependent.

In the particular case of two hypotheses ($k = 2$) a Bayesian test of Theorem 12 given by

$$\psi_n^\chi = I_{\{l_n \leq cf_\Pi^n + R_n\}}, \quad n = 1, 2, \dots,$$

has the form of the Sequential Probability Ratio Test (SPRT, see [12]), being all other Bayesian tests (91) randomizations at its boundaries (see [9] for closely related results).

ACKNOWLEDGEMENTS

The author thanks the anonymous referees for reading the manuscript very carefully, and for their valuable comments and suggestions.

The author greatly appreciates the support of the Autonomous Metropolitan University, Mexico City, Mexico, where this work was done, and the support of the National System of Investigators (SNI) of CONACyT, Mexico.

This work is also partially supported by Mexico's CONACyT Grant no. CB-2005-C01-49854-F.

(Received April 18, 2008.)

REFERENCES

[1] N. Cressie and P. B. Morgan: The VRPT: A sequential testing procedure dominating the SPRT. Econometric Theory *9* (1993), 431–450.

[2] M. Ghosh, N. Mukhopadhyay, and P. K. Sen: Sequential Estimation. John Wiley, New York – Chichester – Weinheim – Brisbane – Singapore – Toronto 1997.

[3] G. W. Haggstrom: Optimal stopping and experimental design. Ann. Math. Statist. *37* (1966), 7–29.

[4] G. Lorden: Structure of sequential tests minimizing an expected sample size. Z. Wahrsch. verw. Geb. *51* (1980), 291–302.

[5] M. B. Malyutov: Lower bounds for the mean length of a sequentially planned experiment. Soviet Math. (Iz. VUZ) *27* (1983), 11, 21–47.

[6] A. Novikov: Optimal sequential testing of two simple hypotheses in presence of control variables. Internat. Math. Forum *3* (2008), 41, 2025–2048. Preprint arXiv:0812.1395v1 [math.ST] (http://arxiv.org/abs/0812.1395)

[7] A. Novikov: Optimal sequential multiple hypothesis tests. Kybernetika *45* (2009), 2, 309–330.

[8] A. Novikov: Optimal sequential procedures with Bayes decision rules. Preprint arXiv:0812.0159v1 [math.ST](http://arxiv.org/abs/0812.0159)

[9] A. Novikov: Optimal sequential tests for two simple hypotheses based on independent observations. Internat. J. Pure Appl. Math. *45* (2008), 2, 291–314.

[10] N. Schmitz: Optimal Sequentially Planned Decision Procedures. (Lecture Notes in Statistics 79.) Springer-Verlag, New York 1993.

[11] I. N. Volodin: Guaranteed statistical inference procedures (determination of the optimal sample size). J. Math. Sci. *44* (1989), 5, 568–600.

[12] A. Wald and J. Wolfowitz: Optimum character of the sequential probability ratio test. Ann. Math. Statist. *19* (1948), 326–339.

[13] S. Zacks: The Theory of Statistical Inference. John Wiley, New York – London – Sydney – Toronto 1971.

Andrey Novikov, Departamento de Matemáticas, Universidad Autónoma Metropolitana – Unidad Iztapalapa, San Rafael Atlixco 186, col. Vicentina, C.P. 09340, México D.F. México.
e-mail: an@xanum.uam.mx
http://mat.izt.uam.mx/profs/anovikov/en