Kybernetika

Andrey Novikov

Optimal sequential multiple hypothesis tests

Kybernetika, Vol. 45 (2009), No. 2, 309--330

Persistent URL: http://dml.cz/dmlcz/140077

KYBERNETIKA — VOLUME 45 (2009), NUMBER 2, PAGES 309 – 330

OPTIMAL SEQUENTIAL MULTIPLE HYPOTHESIS TESTS

Andrey Novikov

This work deals with a general problem of testing multiple hypotheses about the distribution of a discrete-time stochastic process. Both the Bayesian and the conditional settings are considered. The structure of optimal sequential tests is characterized.

Keywords: sequential analysis, hypothesis testing, multiple hypotheses, discrete-time stochastic process, dependent observations, optimal sequential test, Bayes sequential test

AMS Subject Classification: 62L10, 62L15, 60G40, 62C10

1. INTRODUCTION

Let $X_1, X_2, \dots, X_n, \dots$ be a discrete-time stochastic process, whose distribution depends on an unknown "parameter" θ. We consider the classical problem of testing multiple hypotheses $H_1 : \theta = \theta_1$, $H_2 : \theta = \theta_2$, $\dots$, $H_k : \theta = \theta_k$, $k \geq 2$.

The main goal of this article is to characterize the structure of optimal sequential tests in this problem.

Let us suppose that for any $n = 1, 2, \dots$, the vector $(X_1, X_2, \dots, X_n)$ has a probability "density" function

$$f_\theta^n(x_1, x_2, \dots, x_n) \tag{1}$$

(Radon–Nikodym derivative of its distribution) with respect to a product-measure

$$\mu^n = \underbrace{\mu \otimes \mu \otimes \cdots \otimes \mu}_{n \text{ times}},$$

for some σ-finite measure μ on the respective space.

We define a (randomized) sequential hypothesis test as a pair (ψ, ϕ) of a *stopping rule* ψ and a *decision rule* ϕ, with

$$\psi = (\psi_1, \psi_2, \dots, \psi_n, \dots),$$

and

$$\phi = (\phi_1, \phi_2, \dots, \phi_n, \dots).$$

The functions

$$\psi_n = \psi_n(x_1, x_2, \dots, x_n), \quad n = 1, 2, \dots,$$

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

$$\phi_n = \phi_n(x_1, x_2, \dots, x_n), \quad n = 1, 2, \dots$$

are supposed to be measurable vector-functions with k non-negative components $\phi_n^i = \phi_n^i(x_1, \dots, x_n)$:

$$\phi_n = (\phi_n^1, \dots, \phi_n^k),$$

such that $\sum_{i=1}^k \phi_n^i = 1$ for any $n = 1, 2, \dots$.

The interpretation of all these elements is as follows.

The value of $\psi_n(x_1, \dots, x_n)$ is interpreted as the conditional probability *to stop and proceed to decision making*, given that we came to stage n of the experiment and that the observations up to stage n were $(x_1, x_2, \dots, x_n)$. If there is no stop, the experiments continues to the next stage and an additional observation x_{n+1} is taken. Then the rule ψ_{n+1} is applied to $x_1, x_2, \dots, x_n, x_{n+1}$ 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 some of* $H_1, \dots, H_k$ is to be made. The function $\phi_n^i(x_1, \dots, x_n)$ is interpreted as the conditional probability *to accept* H_i, $i = 1, \dots, k$, given that the experiment stops at stage n being $(x_1, \dots, x_n)$ the data vector observed.

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

$$P_\theta(\tau_\psi = n) = \mathrm{E}_\theta(1 - \psi_1)(1 - \psi_2) \dots (1 - \psi_{n-1})\psi_n.$$

Here, and throughout the paper, we interchangeably use ψ_n both for $\psi_n(x_1, x_1, \dots, x_n)$ and for $\psi_n(X_1, X_1, \dots, X_n)$, and so do we for any other function of observations F_n. This does not cause any problem if we adopt the following agreement: when F_n is under probability or expectation sign, it is $F_n(X_1, \dots, X_n)$, otherwise it is $F_n(x_1, \dots, x_n)$.

For a sequential test (ψ, ϕ) let us define

$$\alpha_{ij}(\psi, \phi) = P_{\theta_i}(\,\text{accept}\, H_j) = \sum_{n=1}^{\infty} \mathrm{E}_{\theta_i}(1 - \psi_1) \dots (1 - \psi_{n-1})\psi_n \phi_n^j \tag{2}$$

and

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

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

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

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

Let θ be any fixed (and known) value of the parameter (we *do not* suppose, generally, that θ is one of θ_i, $i = 1, \dots k$).

In this article, we solve the two following problems:

Problem I. Minimize $N(\psi) = N(\theta;\psi)$ over all sequential tests (ψ,ϕ) subject to

$$\alpha_{ij}(\psi,\phi) \le \alpha_{ij}, \quad \text{for all } i = 1,\dots k, \text{ and for all } j \ne i, \tag{5}$$

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

Problem II. Minimize $N(\psi)$ over all sequential tests (ψ,ϕ) subject to

$$\beta_i(\psi,\phi) \le \beta_i, \quad \text{for all } i = 1,\dots k, \tag{6}$$

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

More general problems of minimizing an average cost of type

$$N(\psi) = \sum_{n=1}^{\infty} \mathrm{E}_\theta C_n (1-\psi_1)\dots(1-\psi_{n-1})\psi_n$$

with some cost function $C_n = C_n(X_1, X_2, \dots X_n)$ can be treated in essentially the same manner.

If $k = 2$ then Problems I and II are equivalent, because $\beta_1(\psi,\phi) = \alpha_{12}(\psi,\phi)$, and $\beta_2(\psi,\phi) = \alpha_{21}(\psi,\phi)$, by (2) and (3).

For independent and identically distributed (i.i.d.) observations and $k = 2$ the formulated problem, when $\theta \ne \theta_1$ and $\theta \ne \theta_2$, is known as the modified Kiefer–Weiss problem (see [10]), being the original Kiefer–Weiss problem minimizing $\sup_\theta N(\psi)$ under (5) (see [6]).

For the latter problem, taking into account the usual relations between Bayesian and minimax procedures, it seems to be reasonable to generalize our problem of minimizing $N(\theta;\psi)$ to that of minimizing

$$\int N(\theta;\psi)\, \mathrm{d}\pi(\theta),$$

with some "weight" measure π. From what follows it is easily seen that, under natural measurability conditions, our method works as well for this latter problem.

In Section 2, we reduce Problems I and II to an unconstrained minimization problem. The new objective function is the Lagrange-multiplier function $L(\psi;\phi)$.

In Section 3, we find

$$L(\psi) = \inf_\phi L(\psi,\phi),$$

where the infimum is taken over all decision rules.

In Section 4, we minimize $L(\psi)$ in the class of truncated stopping rules, i.e. such that $\psi_N \equiv 1$ for some $0 < N < \infty$.

In Section 5, we characterize the structure of optimal stopping rule ψ in the class of all stopping rules.

In Section 6, we apply the results obtained in Sections 2−5 to the solution of Problems I and II.

2. REDUCTION TO NON–CONSTRAINED MINIMIZATION

In this section, the 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 Problem I

To proceed with minimizing $N(\psi)$ over the sequential tests subject to (5), let us define the following Lagrange-multiplier function:

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

where $\lambda_{ij}\geq 0$ are some constant multipliers. Recall that $N(\psi)=\mathrm{E}_\theta\tau_\psi$, where θ is the fixed value of parameter for which the average sample number (4) is to be minimized. Generally, we *do not* suppose that θ is one of θ_i, $i=1,\dots,k$.

Let Δ be a class of tests.

The following theorem is a direct application of the Lagrange multipliers method.

Theorem 1. Let exist $\lambda_{ij}>0$, $i=1,\dots,k$, $j=1,\dots,k$, $j\neq i$, and a test $(\psi^*,\phi^*)\in\Delta$ such that for all sequential tests $(\psi,\phi)\in\Delta$

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

holds and such that

$$\alpha_{ij}(\psi^*,\phi^*)=\alpha_{ij}\quad\text{for all}\quad i=1,\dots,k,\ \text{and for all } j\neq i. \tag{9}$$

Then for all $(\psi,\phi)\in\Delta$ such that

$$\alpha_{ij}(\psi,\phi)\leq\alpha_{ij}\quad\text{for all}\quad i=1,\dots,k,\ \text{and for all } j\neq i, \tag{10}$$

it holds

$$N(\psi^*)\leq N(\psi). \tag{11}$$

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

P r o o f. Let $(\psi,\phi)\in\Delta$ be any sequential test satisfying (10). Because of (8)

$$\begin{aligned} L(\psi^*,\phi^*) &= N(\psi^*)+\sum_{j\neq i}\lambda_{ij}\alpha_{ij}(\psi^*,\phi^*)\\ &\leq L(\psi,\phi)=N(\psi)+\sum_{j\neq i}\lambda_{ij}\alpha_{ij}(\psi,\phi)\leq N(\psi)+\sum_{j\neq i}\lambda_{ij}\alpha_{ij}, \end{aligned} \tag{12}$$

where to get the last inequality we used (5).

So,

$$N(\psi^*)+\sum_{j\neq i}\lambda_{ij}\alpha_{ij}(\psi^*,\phi^*)\leq N(\psi)+\sum_{j\neq i}\lambda_{ij}\alpha_{ij},$$

and taking into account conditions (9) we get from this that

$$N(\psi^*)\leq N(\psi).$$

To get the last statement of the theorem we note that if $N(\psi^*)=N(\psi)$ then there are equalities in (12) instead of inequalities which is only possible if $\alpha_{ij}(\psi,\phi)=\alpha_{ij}$ for any $i,j=1,\dots k$, $j\neq i$. □

Remark 1. The author owes the idea of the use of the Lagrange-multiplier method in sequential hypotheses testing to Berk [1]. Essentially, the method of Lagrange multipliers is implicitly used in the monograph of Lehmann [7] in the proof of the fundamental lemma of Neyman–Pearson. In a way, the Bayesian approach in hypotheses testing can be considered as a variant of the Lagrange-multiplier method as well.

Remark 2. All our results below can be adapted to the Bayesian context by choosing appropriate Lagrange multipliers and using

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

instead of $N(\theta;\psi)$ in $L(\psi,\phi)$ above. From this point of view, we extend and complement the results of Cochlar [2] about the existence of Bayesian sequential tests.

More generally, all our results are applicable as well for minimization of

$$\int N(\theta;\psi)\,\mathrm{d}\pi(\theta),$$

where π is any probability measure (see Remarks 6 and 11 below).

2.2. Reduction to non-constrained minimization in Problem II

Very much like in the preceding section, define

$$L(\psi,\phi) = N(\psi) + \sum_{i=1}^{k} \lambda_i \beta_i(\psi,\phi), \tag{13}$$

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

In a very similar manner to Theorem 1, we have

Theorem 2. Let exist $\lambda_i > 0$, $i = 1,\dots,k$, and a sequential test $(\psi^*,\phi^*) \in \Delta$ such that for all $(\psi,\phi) \in \Delta$

$$L(\psi^*,\phi^*) \leq L(\psi,\phi) \tag{14}$$

holds and such that

$$\beta_i(\psi^*,\phi^*) = \beta_i \quad \text{for all} \quad i = 1,\dots k. \tag{15}$$

Then for all sequential tests $(\psi,\phi) \in \Delta$ such that

$$\beta_i(\psi,\phi) \leq \beta_i \quad \text{for all} \quad i = 1,\dots k, \tag{16}$$

it holds

$$N(\psi^*) \leq N(\psi). \tag{17}$$

The inequality in (17) is strict if at least one of the equalities (16) is strict.

3. OPTIMAL DECISION RULES

Due to Theorems 1 and 2, Problem I is reduced to minimizing (7) and Problem II is reduced to minimizing (13). But (13) 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, from now on, we will only solve the problem of minimizing $L(\psi, \phi)$ defined by (7).

In particular, in this section we find

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

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

Let I_A be the indicator function of the event A.

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

$$L(\psi, \phi) \geq N(\psi) + \sum_{n=1}^{\infty} \int (1 - \psi_1) \dots (1 - \psi_{n-1}) \psi_n l_n \, \mathrm{d}\mu^n, \tag{18}$$

where

$$l_n = \min_{1 \leq j \leq k} \sum_{i \neq j} \lambda_{ij} f^n_{\theta_i}. \tag{19}$$

Supposing that $N(\psi)$ is finite, the right-hand side of (18) is attained if and only if

$$\phi^j_n \leq I_{\left\{\sum_{i \neq j} \lambda_{ij} f^n_{\theta_i} = l_n\right\}} \tag{20}$$

for all $j = 1, \dots k$, μ^n-almost anywhere on

$$S^{\psi}_n = \{(x_1, \dots, x_n) : s_n(x_1, \dots, x_n) > 0\},$$

where $s^{\psi}_n(x_1, \dots, x_n) = s^{\psi}_n = (1 - \psi_1) \dots (1 - \psi_{n-1}) \psi_n$, for all $n = 1, 2, \dots$.

Proof. Inequality (18) is equivalent to

$$\sum_{1 \leq i, j \leq k;\, j \neq i} \lambda_{ij} \alpha_{ij}(\psi, \phi) \geq \sum_{n=1}^{\infty} \int (1 - \psi_1) \dots (1 - \psi_{n-1}) \psi_n l_n \, \mathrm{d}\mu^n. \tag{21}$$

We prove it by finding a lower bound for the left-hand side of (21) and proving that this lower bound is attained if ϕ satisfies (20).

To do this, we will use the following simple

Lemma 1. Let $\phi_1, \dots, \phi_k$ and $F_1, \dots F_k$ be some measurable non-negative functions on a measurable space with a measure μ, such that

$$\sum_{i=1}^{k} \phi_i(x) \equiv 1,$$

and such that

$$\int \min_{1 \leq i \leq k} F_i(x) \, \mathrm{d}\mu(x) < \infty.$$

Then
$$\int \left(\sum_{i=1}^{k} \phi_i(x) F_i(x) \right) \mathrm{d}\mu(x) \geq \int \min_{1 \leq i \leq k} F_i(x) \, \mathrm{d}\mu(x) \tag{22}$$
with an equality in (22) if and only if
$$\phi_i \leq I_{\{F_i = \min\limits_{1 \leq j \leq k} F_j\}} \quad \text{for any} \quad i = 1, 2, \ldots, k, \tag{23}$$
μ-almost anywhere.

P r o o f. To prove (22) it suffices to show that
$$\int \left(\sum_{i=1}^{k} \phi_i(x) F_i(x) \right) \mathrm{d}\mu(x) - \int \min_{1 \leq i \leq k} F_i(x) \, \mathrm{d}\mu(x) \geq 0, \tag{24}$$
because the second integral is finite by the conditions of the Lemma.

But (24) is equivalent to
$$\int \sum_{i=1}^{k} \phi_i(x) (F_i(x) - \min_{1 \leq j \leq k} F_j(x)) \, \mathrm{d}\mu(x) \geq 0, \tag{25}$$
being this trivial because the function under the integral sign is non-negative.

Because of this, there is an equality in (25) if and only if
$$\sum_{i=1}^{k} \phi_i(x) (F_i(x) - \min_{1 \leq j \leq k} F_j(x)) = 0$$
μ-almost anywhere, which is only possible if (23) holds true. □

Starting with the proof of (21), let us give to the left-hand side of it the form
$$\sum_{1 \leq i,j \leq k;\, j \neq i} \lambda_{ij} \alpha_{ij}(\psi, \phi)$$
$$= \sum_{n=1}^{\infty} \int (1 - \psi_1) \ldots (1 - \psi_{n-1}) \psi_n \sum_{j=1}^{k} \left(\sum_{1 \leq i \leq k;\, i \neq j} \lambda_{ij} f_{\theta_i}^n \right) \phi_n^j \, \mathrm{d}\mu^n \tag{26}$$
(see (2)).

Applying Lemma 1 to each summand in (26) we immediately have:
$$\sum_{1 \leq i,j \leq k;\, j \neq i} \lambda_{ij} \alpha_{ij}(\psi, \phi) \geq \sum_{n=1}^{\infty} \int (1 - \psi_1) \ldots (1 - \psi_{n-1}) \psi_n l_n \, \mathrm{d}\mu^n \tag{27}$$
with an equality if and only if
$$\phi_n^j \leq I_{\{\sum_{i \neq j} \lambda_{ij} f_{\theta_i}^n = l_n\}}$$
for all $1 \leq j \leq k$, μ^n-almost anywhere on S_n^{ψ}, for all $n = 1, 2, \ldots$. □

Remark 3. It is easy to see, using (4) and (27), that

$$L(\psi)=\inf_{\phi}L(\psi,\phi)=\sum_{n=1}^{\infty}\int(1-\psi_1)\dots(1-\psi_{n-1})\psi_n\,(nf_\theta^n+l_n)\,\mathrm{d}\mu^n \tag{28}$$

if $P_\theta(\tau_\psi<\infty)=1$ and $L(\psi)=\infty$ otherwise.

Remark 4. In the Bayesian context of Remark 2, the "if"-part of Theorem 3 can also be derived from Theorem 5.2.1 [5].

4. TRUNCATED STOPPING RULES

Our next goal is to find a stopping rule ψ minimizing the value of $L(\psi)$ in (28).

In this section, we solve, as an intermediate step, the problem of minimization of $L(\psi)$ in the class of truncated stopping rules, that is, in the class Δ^N of

$$\psi=(\psi_1,\psi_2,\dots,\psi_{N-1},1,\dots). \tag{29}$$

For any $\psi\in\Delta^N$ let us define

$$L_N(\psi)=\sum_{n=1}^{N}\int(1-\psi_1)\dots(1-\psi_{n-1})\psi_n\,(nf_\theta^n+l_n)\,\mathrm{d}\mu^n \tag{30}$$

(see (28)).

The following lemma takes over a large part of work of minimizing $L_N(\psi)$ over $\psi\in\Delta^N$.

Lemma 2. Let $r\geq 2$ be any natural number, and let $v_r=v_r(x_1,x_2,\dots,x_r)$ be any measurable function such that $\int v_r\,\mathrm{d}\mu^r<\infty$. Then

$$\begin{aligned}\sum_{n=1}^{r-1}\int(1-\psi_1)\dots(1-\psi_{n-1})\psi_n(nf_\theta^n+l_n)\,\mathrm{d}\mu^n\\ +\int(1-\psi_1)\dots(1-\psi_{r-1})\,(rf_\theta^r+v_r)\,\mathrm{d}\mu^r\\ \geq\sum_{n=1}^{r-2}\int(1-\psi_1)\dots(1-\psi_{n-1})\psi_n(nf_\theta^n+l_n)\,\mathrm{d}\mu^n\\ +\int(1-\psi_1)\dots(1-\psi_{r-2})\big((r-1)f_\theta^{r-1}+v_{r-1}\big)\,\mathrm{d}\mu^{r-1},\end{aligned} \tag{31}$$

where

$$v_{r-1}=\min\Big\{l_{r-1},f_\theta^{r-1}+\int v_r(x_1,\dots,x_r)\,\mathrm{d}\mu(x_r)\Big\}. \tag{32}$$

There is an equality in (31) if and only if

$$I_{\{l_{r-1}<f_\theta^{r-1}+\int v_r(x_1,\dots,x_r)\,\mathrm{d}\mu(x_r)\}}\leq\psi_{r-1}\leq I_{\{l_{r-1}\leq f_\theta^{r-1}+\int v_r(x_1,\dots,x_r)\,\mathrm{d}\mu(x_r)\}}, \tag{33}$$

μ^{r-1}-almost anywhere on C_{r-1}^{ψ}, where, by definition,

$$C_n^{\psi} = \{(x_1, \dots, x_n) : (1 - \psi_1(x_1)) \dots (1 - \psi_{n-1}(x_1, \dots, x_{n-1})) > 0\}$$

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

P r o o f. Let us start with the following simple consequence of Lemma 1.

Lemma 3. Let χ, ϕ, F_1, F_2 be some measurable functions on a measurable space with a measure μ, such that

$$0 \le \chi(x) \le 1, \quad 0 \le \phi(x) \le 1, \quad F_1(x) \ge 0, \quad F_2(x) \ge 0,$$

and

$$\int \min\{F_1(x), F_2(x)\} \, \mathrm{d}\mu(x) < \infty.$$

Then

$$\int \chi(x)(\phi(x)F_1(x) + (1 - \phi(x))F_2(x)) \, \mathrm{d}\mu(x) \ge \int \chi(x) \min\{F_1(x), F_2(x)\} \, \mathrm{d}\mu(x) \quad (34)$$

with an equality if and only if

$$I_{\{F_1(x) < F_2(x)\}} \le \phi(x) \le I_{\{F_1(x) \le F_2(x)\}} \quad (35)$$

μ-almost anywhere on $\{x : \chi(x) > 0\}$.

P r o o f. Defining $\phi_1(x) \equiv \phi(x)$ and $\phi_2(x) \equiv 1 - \phi(x)$, from Lemma 1 we immediately obtain (34), with an equality if and only if

$$\phi_1(x) = \phi(x) \le I_{\{\chi(x)(F_1(x) - \min\{F_1(x), F_2(x)\}) = 0\}} \quad (36)$$

and

$$\phi_2(x) = 1 - \phi(x) \le I_{\{\chi(x)(F_2(x) - \min\{F_1(x), F_2(x)\}) = 0\}} \quad (37)$$

μ-almost anywhere. Expressing $\phi(x)$ from (36) and (37) we have that there is an equality in (34) if and only if

$$I_{\{\chi(x)(F_2(x) - \min\{F_1(x), F_2(x)\}) > 0\}} \le \phi(x) \le I_{\{\chi(x)(F_1(x) - \min\{F_1(x), F_2(x)\}) = 0\}}$$

μ-almost anywhere, which is equivalent to

$$I_{\{F_1(x) < F_2(x)\}} \le \phi(x) \le I_{\{F_1(x) \le F_2(x)\}} \ \mu - \text{almost anywhere on } \{\chi(x) > 0\}. \qquad \square$$

To start with the proof of Lemma 2 let us note that for proving (31) it is sufficient to show that

$$\int (1 - \psi_1) \dots (1 - \psi_{r-2}) \psi_{r-1} ((r-1) f_\theta^{r-1} + l_{r-1}) \, \mathrm{d}\mu^{r-1}$$
$$+ \int (1 - \psi_1) \dots (1 - \psi_{r-1}) \, (r f_\theta^r + v_r) \, \mathrm{d}\mu^r$$

$$\geq \int (1-\psi_1)\dots(1-\psi_{r-2})\left((r-1)f_\theta^{r-1}+v_{r-1}\right)\mathrm{d}\mu^{r-1}. \tag{38}$$

By Fubini's theorem the left-hand side of (38) is equal to

$$\int (1-\psi_1)\dots(1-\psi_{r-2})\psi_{r-1}((r-1)f_\theta^{r-1}+l_{r-1})\,\mathrm{d}\mu^{r-1}$$
$$+\int (1-\psi_1)\dots(1-\psi_{r-1})\left(\int (rf_\theta^r+v_r)\,\mathrm{d}\mu(x_r)\right)\mathrm{d}\mu^{r-1}$$
$$=\int (1-\psi_1)\dots(1-\psi_{r-2})[\psi_{r-1}((r-1)f_\theta^{r-1}+l_{r-1})$$
$$+(1-\psi_{r-1})\int (rf_\theta^r+v_r)\,\mathrm{d}\mu(x_r)]\,\mathrm{d}\mu^{r-1}. \tag{39}$$

Because $f_\theta^r(x_1,\dots,x_r)$ is a joint density function of $(X_1,\dots,X_r)$, we have

$$\int f_\theta^r(x_1,\dots,x_r)\,\mathrm{d}\mu(x_r)=f_\theta^{r-1}(x_1,\dots,x_{r-1}),$$

so that the right-hand side of (39) transforms to

$$\int (1-\psi_1)\dots(1-\psi_{r-2})[(r-1)f_\theta^{r-1}$$
$$+\psi_{r-1}l_{r-1}+(1-\psi_{r-1})(f_\theta^{r-1}+\int v_r\,\mathrm{d}\mu(x_r))]\,\mathrm{d}\mu^{r-1}. \tag{40}$$

Applying Lemma 3 with

$$\chi=(1-\psi_1)\dots(1-\psi_{r-2}),\quad \phi=\psi_{r-1},$$
$$F_1=l_{r-1},\quad F_2=f_\theta^{r-1}+\int v_r\,\mathrm{d}\mu_r,$$

we see that (40) is greater or equal than

$$\int (1-\psi_1)\dots(1-\psi_{r-2})\left[(r-1)f_\theta^{r-1}+\min\left\{l_{r-1},f_\theta^{r-1}+\int v_r\,\mathrm{d}\mu(x_r)\right\}\right]\mathrm{d}\mu^{r-1}$$
$$=\int (1-\psi_1)\dots(1-\psi_{r-2})[(r-1)f_\theta^{r-1}+v_{r-1}]\,\mathrm{d}\mu^{r-1}, \tag{41}$$

by the definition of v_{r-1} in (32).

Moreover, by the same Lemma 3, (40) is equal to (41) if and only if (33) is satisfied μ^{r-1}-almost anywhere on C_{r-1}^ψ. □

Let now $\psi\in\Delta^N$ be any truncated stopping rule.

By (30) we have

$$L_N(\psi)=\sum_{n=1}^{N-1}\int (1-\psi_1)\dots(1-\psi_{n-1})\psi_n(nf_\theta^n+l_n)\,\mathrm{d}\mu^n$$

$$+\int (1-\psi_1)\dots(1-\psi_{N-1})\left(Nf_\theta^N+l_N\right)\,\mathrm{d}\mu^r. \tag{42}$$

Let $V_N^N \equiv l_N$. Applying Lemma 2 with $r=N$ and $v_N=V_N^N$ we have

$$L_N(\psi)\geq \sum_{n=1}^{N-2}\int (1-\psi_1)\dots(1-\psi_{n-1})\psi_n(nf_\theta^n+l_n)\,\mathrm{d}\mu^n$$

$$+\int (1-\psi_1)\dots(1-\psi_{N-2})\left((N-1)f_\theta^{N-1}+V_{N-1}^N\right)\mathrm{d}\mu^{N-1}, \tag{43}$$

where $V_{N-1}^N=\min\{l_{N-1},f_\theta^{N-1}+\int V_N^N\,\mathrm{d}\mu(x_N)\}$. Also by Lemma 2, the inequality in (43) is in fact an equality if

$$\psi_{N-1}=I_{\{l_{N-1}\leq f_\theta^{N-1}+\int V_N^N\,\mathrm{d}\mu(x_N)\}}. \tag{44}$$

Applying Lemma 2 to the right-hand side of (43) again we see that

$$L_N(\psi)\geq \sum_{n=1}^{N-3}\int (1-\psi_1)\dots(1-\psi_{n-1})\psi_n(nf_\theta^n+l_n)\,\mathrm{d}\mu^n$$

$$+\int (1-\psi_1)\dots(1-\psi_{N-3})\left((N-2)f_\theta^{N-2}+V_{N-2}^N\right)\mathrm{d}\mu^{N-2}, \tag{45}$$

where $V_{N-2}^N=\min\{l_{N-2},f_\theta^{N-2}+\int V_{N-1}^N\,\mathrm{d}\mu(x_{N-1})\}$. There is an equality in (45) if (44) holds and

$$\psi_{N-2}=I_{\{l_{N-2}\leq f_\theta^{N-2}+\int V_{N-1}^N\,\mathrm{d}\mu(x_{N-1})\}}, \tag{46}$$

etc.

Repeating the applications of Lemma 2, we finally get

$$L_N(\psi)\geq \int \left(f_\theta^1+V_1^N\right)\mathrm{d}\mu^1=1+\int V_1^N\,\mathrm{d}\mu(x_1), \tag{47}$$

and a series of conditions on ψ, starting from (44), (46), etc., under which $L(\psi)$ is equal to the right-hand side of (47). Because Lemma 2 also gives necessary and sufficient conditions for attaining the equality, we also have necessary conditions for attaining the lower bound in (47).

In this way, formally, we have the following

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

$$L_N(\psi)\geq \sum_{n=1}^{r}\int (1-\psi_1)\dots(1-\psi_{n-1})\psi_n(nf_\theta^n+l_n)\,\mathrm{d}\mu^n$$
$$+\int (1-\psi_1)\dots(1-\psi_r)\left((r+1)f_\theta^{r+1}+V_{r+1}^N\right)\mathrm{d}\mu^{r+1} \tag{48}$$

$$\geq \sum_{n=1}^{r-1} \int (1-\psi_1)\dots(1-\psi_{n-1})\psi_n(nf_\theta^n + l_n)\,\mathrm{d}\mu^n$$
$$+ \int (1-\psi_1)\dots(1-\psi_{r-1})\left(rf_\theta^r + V_r^N\right)\mathrm{d}\mu^r, \tag{49}$$

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

$$V_m^N = \min\{l_m, f_\theta^m + R_m^N\}, \tag{50}$$

with $$R_m^N = R_m^N(x_1, \dots, x_m) = \int V_{m+1}^N(x_1, x_2, \dots, x_{m+1})\,\mathrm{d}\mu(x_{m+1}). \tag{51}$$

The lower bound in (49) is attained if and only if for any $m = r, \dots, N-1$

$$I_{\{l_m < f_\theta^m + R_m\}} \leq \quad \psi_m \quad \leq I_{\{l_m \leq f_\theta^m + R_m\}} \quad \mu^m\text{-almost anywhere on} \quad C_m^\psi. \tag{52}$$

In particular, conditions (52) with $m = 1, 2, \dots, N-1$ are necessary and sufficient for being $\psi = (\psi_1, \dots, \psi_{N-1}, 1, \dots)$ an optimal truncated rule in Δ^N. The minimum value of $L(\psi)$, over $\psi \in \Delta^N$, is equal to

$$1 + \int V_1^N\,\mathrm{d}\mu(x_1) = 1 + R_0^N.$$

Remark 5. Despite that any $\psi = (\psi_1, \dots, \psi_{N-1}, 1, \dots)$ satisfying (52) for $m = 1, \dots, N-1$ is optimal among all truncated tests in Δ^N, it only makes practical sense if

$$l_0 > 1 + R_0^N$$

where l_0 defined as

$$l_0 \equiv \min_{1 \leq j \leq k} \sum_{1 \leq i \leq k, i \neq j} \lambda_{ij}.$$

The reason is that l_0 can be considered as "the $L(\psi)$" function for a trivial sequential test (ψ_0, ϕ_0) which, without taking any observations, makes a decision according to any $\phi_0 = (\phi_0^1, \dots, \phi_0^k)$ such that

$$\phi_0^j \leq I_{\{\sum_{i \neq j} \lambda_{ij} = l_0\}}, \quad 1 \leq j \leq k.$$

In this case, there are no observations ($N(\psi_0) = 0$) and it is easily seen that

$$L(\psi_0, \phi_0) = \sum_{1 \leq i,j \leq k, i \neq j} \lambda_{ij}\alpha_{ij}(\psi_0, \phi_0) = l_0.$$

Thus, the inequality

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

means that the trivial test (ψ_0, ϕ_0) is not worse than the best truncated test in Δ^N.

Because of that, we consider V_0^N defined by (50) for $m = 0$, where, by definition, $f_\theta^0 = 1$, as the minimum value of $L(\psi)$ in Δ^N, in the case it is allowed not to take any observations.

Remark 6. It is not difficult to see from the proof of Lemma 2, that the problem of the optimal testing when the cost of the experiment is defined as

$$\int N(\psi)\,\mathrm{d}\pi(\theta), \tag{53}$$

with some measure π (see Remark 2), under suitable measurability conditions, can receive essentially the same treatment. The corresponding optimal stopping rule in Δ^N will be defined by

$$\psi_r = I_{\{l_r \leq \int f_\theta^r d\pi(\theta) + \int V_{r+1}^N\,\mathrm{d}\mu_{r+1}\}} \tag{54}$$

for $r = 1, 2, \ldots, N-1$, with V_r^N defined recursively as

$$V_{r-1}^N = \min\left\{l_{r-1}, \int f_\theta^{r-1}\mathrm{d}\pi(\theta) + \int V_r^N\,\mathrm{d}\mu(x_r)\right\}, \tag{55}$$

starting from $r = N$, in which case $V_N^N \equiv l_N$.

In the Bayesian context of Remark 2 the optimality of (54)–(55) with $\lambda_{ij} = \pi_i L_{ij}$, where L_{ij} are some non-negative losses, $i \neq j$, can be derived also from Theorem 5.2.2 [5]. Our Theorem 4 gives, additionally to that, a necessary condition of optimality, providing the structure of *all* Bayesian truncated tests. Essentially, they are randomizations of (54):

$$I_{\{l_r < \int f_\theta^r \mathrm{d}\pi(\theta) + \int V_{r+1}^N\,\mathrm{d}\mu_{r+1}\}} \leq \psi_r \leq I_{\{l_r \leq \int f_\theta^r \mathrm{d}\pi(\theta) + \int V_{r+1}^N \mathrm{d}\mu_{r+1}\}},$$

for $r = 1, 2, \ldots, N-1$.

In purely Bayesian context, such conditions may be irrelevant, because any Bayesian test gives the same (minimum) value of the Bayesian risk. Nevertheless, for our (conditional) Problems I and II, it may be important to have a broader class of optimal tests, for easier compliance with (9) in Theorem 1 (or with (15) in Theorem 2), just like the randomization of decision rule is important for finding tests with a given α-level in the Neyman–Pearson problem (see, for example, [7]).

5. GENERAL STOPPING RULES

In this section we characterize the structure of general stopping rules minimizing $L(\psi)$.

Let us define for any stopping rule ψ

$$L_N(\psi) = \sum_{n=1}^{N-1} \int (1-\psi_1)\ldots(1-\psi_{n-1})\psi_n(nf_\theta^n + l_n)\,\mathrm{d}\mu^n + \int (1-\psi_1)\ldots(1-\psi_{N-1})\left(Nf_\theta^N + l_N\right)\mathrm{d}\mu^N. \tag{56}$$

(cf. (42)). This is the Lagrange-multiplier function for ψ truncated at N, i.e. the rule with the components $\psi^N = (\psi_1, \psi_2, \ldots, \psi_{N-1}, 1, \ldots)$, $L_N(\psi) = L(\psi^N)$.

Because 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 $L(\psi)$ from (48)–(49).

To be able to do this, we need some "approximation properties" for $L(\psi)$, to guarantee that $L_N(\psi) \to L(\psi)$, as $N \to \infty$, at least for stopping rules ψ for which $P_\theta(\tau_\psi < \infty) = 1$.

Lemma 4. Suppose that ψ is a stopping rule such that $P_\theta(\tau_\psi < \infty) = 1$.

(i) If $L(\psi) < \infty$ and

$$\int (1-\psi_1)\dots(1-\psi_{n-1}) l_n \, \mathrm{d}\mu^n \to 0, \quad \text{as} \quad n \to \infty, \tag{57}$$

then

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

(ii) If $L(\psi) = \infty$ then $L_N(\psi) \to \infty$.

P r o o f. Let $L(\psi) < \infty$. Let us calculate the difference between $L(\psi)$ and $L_N(\psi)$ in order to show that it goes to zero as $N \to \infty$. By (56)

$$\begin{aligned} L(\psi) - L_N(\psi) = \sum_{n=1}^{\infty} \int (1-\psi_1)\dots(1-\psi_{n-1})\psi_n (n f_\theta^n + l_n) \, \mathrm{d}\mu^n \\ - \sum_{n=1}^{N-1} \int (1-\psi_1)\dots(1-\psi_{n-1})\psi_n (n f_\theta^n + l_n) \, \mathrm{d}\mu^n \\ - \int (1-\psi_1)\dots(1-\psi_{N-1}) \left(N f_\theta^N + l_N\right) \mathrm{d}\mu^N \\ = \sum_{n=N}^{\infty} \int (1-\psi_1)\dots(1-\psi_{n-1})\psi_n (n f_\theta^n + l_n) \, \mathrm{d}\mu^n \\ - \int (1-\psi_1)\dots(1-\psi_{N-1}) \left(N f_\theta^N + l_N\right) \mathrm{d}\mu^N. \end{aligned} \tag{58}$$

The first summand converges to zero, as $N \to \infty$, being the tail of a convergent series (this is because $L(\psi) < \infty$).

We have further

$$\int (1-\psi_1)\dots(1-\psi_{N-1}) l_N \, \mathrm{d}\mu^N \to 0$$

as $N \to \infty$, because of (57).

It remains to show that

$$\int (1-\psi_1)\dots(1-\psi_{N-1}) N f_\theta^N \, \mathrm{d}\mu^N = N P_\theta(\tau_\psi \geq N) \to 0 \text{ as } N \to \infty. \tag{59}$$

But this is again due to the fact that $L(\psi) < \infty$ which implies that

$$\mathrm{E}_\theta \tau_\psi = \sum_{n=1}^{\infty} n P_\theta(\tau_\psi = n) < \infty.$$

Because this series is convergent, $\sum_{n=N}^{\infty} nP_\theta(\tau_\psi = n) \to 0$. Thus, using the Chebyshev inequality we have

$$NP_\theta(\tau_\psi \geq N) \leq \mathrm{E}_\theta \tau_\psi I_{\{\tau_\psi \geq N\}} = \sum_{n=N}^{\infty} nP_\theta(\tau_\psi = n) \to 0$$

as $N \to \infty$, which completes the proof of (59).

Let now $L(\psi) = \infty$.

This means that

$$\sum_{n=1}^{\infty} \int (1-\psi_1)\dots(1-\psi_{n-1})\psi_n(nf_\theta^n + l_n)\,\mathrm{d}\mu^n = \infty$$

which immediately implies by (56) that

$$L_N(\psi) \geq \sum_{n=1}^{N-1} \int (1-\psi_1)\dots(1-\psi_{n-1})\psi_n(nf_\theta^n + l_n)\,\mathrm{d}\mu^n \to \infty. \qquad \square$$

Lemma 4 gives place to the following definition.

Let us say that our testing problem is *truncatable* if (57) holds for any ψ with $\mathrm{E}_\theta \tau_\psi < \infty$.

From Lemma 4 it immediately follows

Corollary 1. For any truncatable problem

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

for any stopping rule ψ such that $P_\theta(\tau_\psi < \infty) = 1$.

Remark 7. It is obvious from (57) that a testing problem is truncatable, in particular, if

$$\int l_n\,\mathrm{d}\mu^n \to 0, \quad \text{as} \quad n \to \infty. \tag{60}$$

Let us denote by $\alpha_{ij}(n,\phi)$ the error probability of a test corresponding to a fixed number n of observations, when the decision rule ϕ is applied. From Theorem 3 it follows that the left-hand side of (60) is the minimum weighted error sum:

$$\int l_n\,\mathrm{d}\mu^n = \inf_\phi \sum_{1\leq i,j\leq k, i\neq j} \lambda_{ij}\alpha_{ij}(n,\phi),$$

where the infimum is taken over all decision rules ϕ.

Thus, (60) requires a very natural behaviour of a statistical testing problem, namely that the minimum weighted error sum, over all fixed-sample size tests, tend to zero, as the sample size n tends to infinity.

Remark 8. Any Bayesian problem (with $N(\psi) = \sum_{i=1}^{k} N(\theta_i;\psi)\pi_i$ in (7), where $\pi_i > 0$, $i = 1, \ldots, k$) is truncatable. Indeed, if $N(\psi) < \infty$ then $\mathrm{E}_{\theta_i}\tau_\psi = N(\theta_i;\psi) < \infty$ for all $i = 1, \ldots, k$. Because of this,

$$\int (1-\psi_1)\ldots(1-\psi_{n-1}) l_n \,\mathrm{d}\mu^n \le \int (1-\psi_1)\ldots(1-\psi_{n-1}) \left(\sum_{i=2}^{k} \lambda_{i1} f_{\theta_i}^n\right) \mathrm{d}\mu^n$$
$$\le \sum_{i=2}^{k} \lambda_{i1} P_{\theta_i}(\tau_\psi \ge n) \to 0, \quad \text{as} \quad n \to \infty,$$

thus, (57) is fulfilled.

Our main results below will refer to truncatable testing problems.

To go on with the plan of passing to the limit, as $N \to \infty$, in the inequalities of Theorem 4, let us turn now to the behaviour of V_r^N, as $N \to \infty$.

Lemma 5. For any $r \ge 1$ and for any $N \ge r$

$$V_r^N \ge V_r^{N+1}. \tag{61}$$

P r o o f. By induction over $r = N, N-1, \ldots, 1$.

Let $r = N$. Then by (50)

$$V_N^{N+1} = \min\left\{l_N, f_\theta^N + \int V_{N+1}^{N+1} \mathrm{d}\mu(x_{N+1})\right\} \le l_N = V_N^N.$$

If we suppose that (61) is satisfied for some r, $N \ge r > 1$, then

$$V_{r-1}^N = \min\left\{l_{r-1}, f_\theta^{r-1} + \int V_r^N \,\mathrm{d}\mu(x_r)\right\}$$
$$\ge \min\left\{l_{r-1}, f_\theta^{r-1} + \int V_r^{N+1} \,\mathrm{d}\mu(x_r)\right\} = V_{r-1}^{N+1}.$$

Thus, (61) is satisfied for $r-1$ as well, which completes the induction. □

It follows from Lemma 5 that for any fixed $r \ge 1$ the sequence V_r^N is non-increasing. So, there exists

$$V_r = \lim_{N\to\infty} V_r^N. \tag{62}$$

Now, everything is prepared for passing to the limit, as $N \to \infty$, in (48) and (49) with $\psi = \psi^N$. If $P_\theta(\tau_\psi < \infty) = 1$, then the left-hand side of (48) by Lemma 4 tends to $L(\psi)$, whereas passing to the limit in the other two parts under the integral sign is justified by the Lebesgue monotone convergence theorem, in view of Lemma 5. For the same reason, passing to the limit as $N \to \infty$ is possible in (50) (see (51)).

In this way, for a truncatable testing problem we get the following

Theorem 5. Let ψ be any stopping rule. Then for any $r \ge 1$ the following inequalities hold

$$L(\psi) \ge \sum_{n=1}^{r} \int (1-\psi_1)\ldots(1-\psi_{n-1})\psi_n (n f_\theta^n + l_n) \,\mathrm{d}\mu^n$$

$$+\int(1-\psi_1)\dots(1-\psi_r)\left((r+1)f_\theta^{r+1}+V_{r+1}\right)\mathrm{d}\mu^{r+1} \tag{63}$$

$$\geq\sum_{n=1}^{r-1}\int(1-\psi_1)\dots(1-\psi_{n-1})\psi_n(nf_\theta^n+l_n)\,\mathrm{d}\mu^n$$

$$+\int(1-\psi_1)\dots(1-\psi_{r-1})\,(rf_\theta^r+V_r)\,\mathrm{d}\mu^r, \tag{64}$$

where
$$V_m=\min\{l_m,f_\theta^m+R_m\} \tag{65}$$

with
$$R_m=R_m(x_1,\dots,x_m)=\int V_{m+1}(x_1,\dots,x_{m+1})\,\mathrm{d}\mu(x_{m+1}) \tag{66}$$

for any $m\geq 1$.

In particular, the following lower bound holds true:

$$L(\psi)\geq 1+\int V_1\,\mathrm{d}\mu(x_1)=1+R_0. \tag{67}$$

In comparison to Theorem 4, Theorem 5 is lacking a very essential element: the structure of the test achieving the lower bound on the right-hand side of (67). In case this test exists, by virtue of (67) it has to be optimal.

First of all, let us show that if the optimal test exists, it reaches the lower bound on the right-hand side of (67). More exactly, we prove

Lemma 6. For any truncatable testing problem

$$\inf_\psi L(\psi)=1+R_0. \tag{68}$$

P r o o f. Let us denote

$$U=\inf_\psi L(\psi),\quad U_N=1+R_0^N,$$

where R_0^N is defined in Theorem 4.

By Theorem 4, for any $N=1,2,\dots$

$$U_N=\inf_{\psi\in\Delta^N}L(\psi).$$

Obviously, $U_N\geq U$ for any $N=1,2,\dots$, so

$$\lim_{N\to\infty}U_N\geq U. \tag{69}$$

Let us show first that in fact there is an equality in (69).

Suppose the contrary, i.e. that $\lim_{N\to\infty}U_N=U+4\varepsilon$, with some $\varepsilon>0$. We immediately have from this that

$$U_N\geq U+3\varepsilon \tag{70}$$

for all sufficiently large N.

On the other hand, by the definition of U there exists a ψ such that $U \le L(\psi) \le U + \varepsilon$. Because, by Lemma 4, $L_N(\psi) \to L(\psi)$, as $N \to \infty$, we have that

$$L_N(\psi) \le U + 2\varepsilon \tag{71}$$

for all sufficiently large N as well. Because, by definition, $L_N(\psi) \ge U_N$, we have that

$$U_N \le U + 2\varepsilon$$

for all sufficiently large N, which contradicts (70).

Thus,

$$\lim_{N\to\infty} U_N = U.$$

Now, to get (68) we note that, by the Lebesgue's monotone convergence theorem,

$$U = \lim_{N\to\infty} U_N = 1 + \lim_{N\to\infty} \int V_1^N(x)\,\mathrm{d}\mu(x) = 1 + \int V_1(x)\,\mathrm{d}\mu(x) = 1 + R_0,$$

thus, $U = 1 + R_0$. □

Remark 9. For the Bayesian context (see Remark 2), Lemma 6 can be derived from Theorem 5.2.3 [5] if (60) is supposed (see also Section 7.2 of [4] or Section 9.4 of [11]).

The following theorem gives the structure of the optimal stopping rule for a truncatable testing problem.

Theorem 6.

$$L(\psi) = \inf_{\psi'} L(\psi'), \tag{72}$$

if and only if

$$I_{\{l_m < f_\theta^m + R_m\}} \le \psi_m \le I_{\{l_m \le f_\theta^m + R_m\}} \quad \mu^m\text{-almost anywhere on} \quad C_m^\psi \tag{73}$$

for all $m = 1, 2 \ldots$.

P r o o f. Let ψ be any stopping rule. By Theorem 5 for any fixed $r \ge 1$ the following inequalities hold:

$$\begin{aligned} L(\psi) \ \ge\ & \sum_{n=1}^{r} \int (1-\psi_1)\ldots(1-\psi_{n-1})\psi_n(nf_\theta^n + l_n)\,\mathrm{d}\mu^n \\ & + \int (1-\psi_1)\ldots(1-\psi_r)\left((r+1)f_\theta^{r+1} + V_{r+1}\right)\mathrm{d}\mu^{r+1} \qquad (74) \\ \ge\ & \sum_{n=1}^{r-1} \int (1-\psi_1)\ldots(1-\psi_{n-1})\psi_n(nf_\theta^n + l_n)\,\mathrm{d}\mu^n \\ & + \int (1-\psi_1)\ldots(1-\psi_{r-1})\,(rf_\theta^r + V_r)\,\mathrm{d}\mu^r \qquad (75) \\ & \ge \ldots \\ \ge\ & \int \psi_1(f_\theta^1 + l_1)\,\mathrm{d}\mu^1 + \int (1-\psi_1)\left(2f_\theta^2 + V_2\right)\mathrm{d}\mu^2 \qquad (76) \\ \ge\ & 1 + \int V_1\,\mathrm{d}\mu(x_1) = 1 + R_0. \qquad (77) \end{aligned}$$

Let us suppose that $L(\psi) = 1 + R_0$. Then, by Lemma 6, there are equalities in all the inequalities (74)–(77). Applying the "only if"-part of Lemma 2 and using (65) and (66), successively, starting from the last inequality (77), we get that (73) has to be satisfied for any $m = 1, 2, \dots$. The first part of the Theorem is proved.

Let now ψ be any test satisfying (73).

Applying the "if"-part of Lemma 2 and using (65) and (66) again, we see that all the inequalities in (75)–(77) are in fact equalities for

$$\psi^r = (\psi_1, \psi_2, \dots, \psi_r, 1, \dots).$$

In particular, this means that there exists

$$\lim_{r\to\infty} \Bigg[\sum_{n=1}^{r} \int (1-\psi_1)\dots(1-\psi_{n-1})\psi_n (n f_\theta^n + l_n)\, \mathrm{d}\mu^n$$

$$+ \int (1-\psi_1)\dots(1-\psi_r)\left((r+1) f_\theta^{r+1} + V_{r+1}\right) \mathrm{d}\mu^{r+1} \Bigg] = 1 + R_0. \tag{78}$$

It follows from (78) that

$$\begin{aligned} &\limsup_{r\to\infty} \textstyle\int (1-\psi_1)\dots(1-\psi_r)(r+1) f_\theta^{r+1}\, \mathrm{d}\mu^{r+1} \\ &\quad = \limsup_{r\to\infty} (r+1) P_\theta(\tau_\psi \ge r+1) \le 1 + R_0, \end{aligned}$$

which implies that $\lim_{r\to\infty} P_\theta(\tau_\psi \ge r+1) = 0$. Thus, $P_\theta(\tau_\psi < \infty) = 1$.

From (78), it follows as well that

$$\lim_{r\to\infty} \sum_{n=1}^{r} \int [(1-\psi_1)\dots(1-\psi_{n-1})\psi_n (n f_\theta^n + l_n)]\, \mathrm{d}\mu^n \le 1 + R_0. \tag{79}$$

But the left-hand side of (79) is $L(\psi)$ (because $P_\theta(\tau_\psi < \infty) = 1$) and hence

$$L(\psi) \le 1 + R_0. \tag{80}$$

On the other hand, by virtue of Theorem 5,

$$L(\psi) \ge 1 + R_0,$$

which proves, together with (80), that $L(\psi) = 1 + R_0$. □

Remark 10. Once again (see Remark 5), the optimal stopping rule ψ from Theorem 6 only makes practical sense if $l_0 > 1 + R_0$.

Remark 11. From the results of this section it is not difficult to see that the same method works as well for minimizing

$$\int N(\psi)\, \mathrm{d}\pi(\theta)$$

(see Remark 6).

Repeating the steps which led us to Theorem 6 we get that the corresponding optimal stopping rule has the form

$$\psi_r = I_{\left\{l_r \le \int f_\theta^r \,\mathrm{d}\pi(\theta) + \int V_{r+1}\,\mathrm{d}\mu(x_{r+1})\right\}},\ r = 1, 2, 3, \dots, \tag{81}$$

with

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

being V_r^N defined for $r = N-1, N-2, \dots, 1$ recursively by

$$V_r^N = \min\left\{l_r, \int f_\theta^r \,\mathrm{d}\pi(\theta) + \int V_{r+1}^N \,\mathrm{d}\mu(x_{r+1})\right\}$$

starting from $V_N^N \equiv l_N$.

In a particular case of Remark 2 and

$$\int N(\theta;\psi)\,\mathrm{d}\pi(\theta) = \sum_{i=1}^k \pi_i N(\theta_i;\psi)$$

being $\lambda_{ij} = L_{ij}\pi_i$, this gives an optimal stopping rule for the Bayesian problem considered in [2].

In particular, for $k = 2$, this gives an optimal stopping rule for the Bayesian problem considered in [3].

6. APPLICATIONS TO THE CONDITIONAL PROBLEMS

In this section, we apply the results obtained in the preceding sections to minimizing the average sample size $N(\psi) = \mathrm{E}_\theta \tau_\psi$ over all sequential testing procedures with error probabilities not exceeding some prescribed levels (see Problems I and II in Section 1). Recall that we are supposing that our problems are truncatable (see Section 5).

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

Theorem 7. Let ψ satisfy (73) for all $m = 1, 2, \dots$, with any $\lambda_{ij} > 0$, $i, j = 1, \dots, k$, $i \ne j$, (recall that R_m and l_m in (73) are functions of λ_{ij}), and let ϕ be any decision rule satisfying (20).

Then for all sequential testing procedure (ψ', ϕ') such that

$$\alpha_{ij}(\psi', \phi') \le \alpha_{ij}(\psi, \phi) \quad \text{for all} \quad i, j = 1, \dots, k,\ i \ne j, \tag{82}$$

it holds

$$N(\psi') \ge N(\psi). \tag{83}$$

The inequality in (83) is strict if at least one of the inequalities in (82) is strict.

If there are equalities in all of the inequalities in (82) and (83), then ψ' satisfies (73) for all $m = 1, 2, \dots$ as well (with ψ' instead of ψ).

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

Let us suppose that

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

and

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

Then, obviously,

$$L(\psi',\phi') = L(\psi,\phi) = L(\psi) \geq L(\psi') \tag{84}$$

(see (7) and Remark 3).

By Theorem 6, there can not be strict inequality in (84), so $L(\psi)=L(\psi')$. From Theorem 6 it follows now that ψ' satisfies (73) as well. □

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

Theorem 8. Let ψ satisfy (73) for all $m=1,2,\dots,$ with $\lambda_{ij}=\lambda_i$ for all $j=1,\dots,k$, where $\lambda_i>0$, $i=1,\dots k$ are any numbers, and let ϕ be any decision rule such that

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

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

Then for any sequential test (ψ',ϕ') such that

$$\beta_i(\psi',\phi') \leq \beta_i(\psi,\phi) \quad \text{for any} \quad i=1,\dots,k, \tag{85}$$

it holds

$$N(\psi') \geq N(\psi). \tag{86}$$

The inequality in (86) is strict if at least one of the inequalities in (85) is strict.

If there are equalities in all of the inequalities in (85) and (86), then ψ' satisfies (73) for all $m=1,2,\dots$ as well (with ψ' instead of ψ).

Remark 12. There are examples of applications of Theorem 7 (or 8), in the case of two simple hypotheses based on independent observations, in [9].

A numerical example related to the modified Kiefer–Weiss problem for independent and identically distributed observations can be found in [8]. Obviously, our Theorem 7 provides, for this particular case, randomized versions of the optimal sequential test studied in [8] (see also [9]).

ACKNOWLEDGEMENTS

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.

The author thanks the anonymous referees for reading the article carefully and for their valuable suggestions and comments.

(Received November 2, 2007.)

REFERENCES

[1] R. H. Berk: Locally most powerful sequential tests. Ann. Statist. *3* (1975), 373–381.

[2] J. Cochlar: The optimum sequential test of a finite number of hypotheses for statistically dependent observations. Kybernetika *16* (1980), 36–47.

[3] J. Cochlar and I. Vrana: On the optimum sequential test of two hypotheses for statistically dependent observations. Kybernetika *14* (1978), 57–69.

[4] T. S. Ferguson: Mathematical Statistics: A Decision Theoretic Approach. Academic Press, New York 1967.

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

[6] J. Kiefer and L. Weiss: Some properties of generalized sequential probability ratio tests. Ann. Math. Statist. *28* (1957), 57–75.

[7] E. L. Lehmann: Testing Statistical Hypotheses. Wiley, New York; Chapman & Hall, London 1959.

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

[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] L. Weiss: On sequential tests which minimize the maximum expected sample size. J. Amer. Statist. Assoc. *57* (1962), 551–566.

[11] Sh. Zacks: The Theory of Statistical Inference. 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