International Journal of Pure and Applied Mathematics

Volume 45 No. 2 2008, 291-314

OPTIMAL SEQUENTIAL TESTS FOR TWO SIMPLE HYPOTHESES BASED ON INDEPENDENT OBSERVATIONS

Andrey Novikov

Department of Mathematics
UAM-Iztapalapa
San Rafael Atlixco 186, col. Vicentina, C.P. 09340, Mexico D.F., MEXICO
e-mail: an@xanum.uam.mx
url: http://mat.izt.uam.mx/profs/anovikov/en

Abstract: In this work, we consider a general problem of testing two simple hypotheses about the distribution of a discrete-time stochastic process with independent values. The structure of optimal sequential tests is characterized. As a criterion of optimization the average sample number under a third hypothesis is taken, which does not necessarily match one of the two hypotheses under consideration (a version of the modified Kiefer-Weiss problem).

In the particular case of independent and identically distributed observations, we describe the class of all sequential hypotheses tests which share the optimality property with the sequential probability ratio test (SPRT), as well as characterize the class of optimal sequential tests in the modified Kiefer-Weiss problem.

As an illustration of the general results, we characterize the structure of the optimal sequential tests for processes with independent values which are stationary beginning from some fixed time on.

AMS Subject Classification: 62L10, 62L15, 60G40
Key Words: sequential analysis, sequential hypothesis testing, two simple hypotheses, discrete-time stochastic process, independent observations, optimal sequential test, sequential probability ratio test

1. Introduction. Problem Set-Up

Let $X_1, X_2, \dots, X_n, \dots$ be a discrete-time stochastic process with independent

Received: April 11, 2008

values, whose distribution depends on an unknown "parameter" θ. We consider the classical problem of testing a simple hypothesis $H_0: \theta = \theta_0$ against a simple alternative $H_1: \theta = \theta_1$, $\theta_0 \neq \theta_1$.

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

Following Ferguson [5] 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

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

where the functions

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

and

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

are supposed to be some measurable functions with values in $[0, 1]$, which have the following meaning.

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 above, etc., until the experiment eventually stops.

It is supposed that when the experiment stops, a decision *to accept or to reject* H_0 is to be made. The function $\phi_n(x_1, \dots, x_n)$ is interpreted as the conditional probability *to reject* the null-hypothesis H_0, 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 (*stopping time*) whose distribution is given by

$$P_\theta(\tau = n) = E_\theta(1 - \psi_1)(1 - \psi_2) \dots (1 - \psi_{n-1})\psi_n.$$

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 we do 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 $F_n(X_1, \ldots, X_n)$, otherwise it is $F_n(x_1, \ldots, x_n)$.

For practical reasons, we are interested in tests satisfying

$$P_\theta(\tau < \infty) = \sum_{n=1}^{\infty} P_\theta(\tau = n) = 1 \tag{1.1}$$

for any θ involved in the problem, and, first of all, for $\theta = \theta_0$ and $\theta = \theta_1$.

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

$$\alpha(\psi, \phi) = P_{\theta_0}(\text{ reject } H_0) = \sum_{n=1}^{\infty} E_{\theta_0}(1 - \psi_1) \ldots (1 - \psi_{n-1})\psi_n \phi_n \tag{1.2}$$

and

$$\beta(\psi, \phi) = P_{\theta_1}(\text{ accept } H_0) = \sum_{n=1}^{\infty} E_{\theta_1}(1 - \psi_1) \ldots (1 - \psi_{n-1})\psi_n(1 - \phi_n). \tag{1.3}$$

The probability $\alpha(\psi, \phi)$ is called the *type I error probability*, and $\beta(\psi, \phi)$ is called the *type II error probability*. Normally, we would like to keep them below some specified levels:

$$\alpha(\psi, \phi) \leq \alpha \tag{1.4}$$

and

$$\beta(\psi, \phi) \leq \beta \tag{1.5}$$

with some $\alpha, \beta \in (0, 1)$.

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

$$N(\theta; \psi) = E_\theta \tau = \sum_{n=1}^{\infty} n E_\theta(1 - \psi_1) \ldots (1 - \psi_{n-1})\psi_n. \tag{1.6}$$

Our main goal is minimizing $N(\theta; \psi)$ over all sequential tests subject to (1.4) and (1.5).

For independent and identically distributed (i.i.d.) observations the problem of minimizing (1.6) under conditions (1.4) and (1.5), when $\theta \neq \theta_0$ and $\theta \neq \theta_1$, is known as the modified Kiefer-Weiss problem (see [15]), being the original Kiefer-Weiss problem minimizing $\sup_\theta N(\theta; \psi)$ under (1.4) and (1.5) (see [7]).

2. General Results. Existence and Uniqueness

In a rather standard way (see [10], see also an earlier work [1] for a slightly different sequential testing problem), the problem of minimization of (1.6) under constraints (1.4) and (1.5) is reduced to minimization of the Lagrange-multiplier function:

$$L(\psi, \phi) = N(\theta; \psi) + \lambda_0 \alpha(\psi, \phi) + \lambda_1 \beta(\psi, \phi) \tag{2.1}$$

where $\lambda_0 \geq 0$ and $\lambda_1 \geq 0$ are some constant multipliers.

The following theorem is an immediate consequence of the idea of the Lagrange multiplier method.

Let Δ be a class of tests.

Theorem 1. (see [11]) *Let there exist $\lambda_0 > 0$ and $\lambda_1 > 0$ and a test $(\psi^*, \phi^*) \in \Delta$ such that for any other test $(\psi, \phi) \in \Delta$*

$$L(\psi^*, \phi^*) \leq L(\psi, \phi) \tag{2.2}$$

holds and such that

$$\alpha(\psi^*, \phi^*) = \alpha \quad \textit{and} \quad \beta(\psi^*, \phi^*) = \beta. \tag{2.3}$$

Then for any test $(\psi, \phi) \in \Delta$ satisfying

$$\alpha(\psi, \delta) \leq \alpha \quad \textit{and} \quad \beta(\psi, \delta) \leq \beta \tag{2.4}$$

it holds

$$N(\theta; \psi^*) \leq N(\theta; \psi). \tag{2.5}$$

The inequality in (2.5) is strict if at least one of the equalities (2.4) is strict.

Because of Theorem 1, the problem is to find the structure of tests minimizing $L(\psi, \phi)$, over all sequential tests (ψ, ϕ).

And the first step to this, rather standard as well (see, e.g., [8], [5], [13], [3], [4], [10]), is to get a "universal" decision rule ϕ^* which minimizes $L(\psi, \phi)$ over all decision rules ϕ.

Let us suppose that any X_i has a probability "density" function

$$f_{\theta,i}(x)$$

(Radon-Nicodym derivative of its distribution) with respect to a σ-finite measure μ on the space of its "values", $i = 1, 2, 3 \ldots$

Because of the independence, for any $n = 1, 2, 3 \ldots$, the "vector" $(X_1, X_2,$

$\dots, X_n)$ of the first n observations has the "density" function

$$f_\theta^n(x_1, x_2, \dots, x_n) = \prod_{i=1}^n f_{\theta,i}(x_i)$$

with respect to the product-measure $\nu_n = \mu \otimes \mu \otimes \dots \otimes \mu$ of μ n times by itself.

Let I_A be the indicator function of the event A.

The universal decision rule is given by the following

Theorem 2. (see [11]) *For any $\lambda_0 \geq 0$ and $\lambda_1 \geq 0$ and for any sequential test (ψ, ϕ)*

$$L(\psi, \phi) \geq L(\psi, \phi^*) \tag{2.6}$$

$$= \sum_{n=1}^\infty \int (1-\psi_1)\dots(1-\psi_{n-1})\psi_n \left[n f_\theta^n + \min\{\lambda_0 f_{\theta_0}^n, \lambda_1 f_{\theta_1}^n\} \right] d\nu_n \tag{2.7}$$

with

$$\phi^* = (\phi_1^*, \phi_2^*, \dots, \phi_n^*, \dots), \tag{2.8}$$

where

$$\phi_n^* = \phi_n^*(x_1, \dots, x_n) = I_{\left\{ \lambda_0 f_{\theta_0}^n(x_1,\dots,x_n) \leq \lambda_1 f_{\theta_1}^n(x_1,\dots,x_n) \right\}}. \tag{2.9}$$

Now, the problem of minimizing $L(\psi, \phi)$ is reduced to that of minimizing $L(\psi, \phi^*)$: indeed, if there is (for some class of stopping rules Δ) a stopping rule $\psi^* \in \Delta$ such that $L(\psi^*, \phi^*) \leq L(\psi, \phi^*)$ for any other stopping rule $\psi \in \Delta$, then for any sequential test (ψ, ϕ) with $\psi \in \Delta$ we have

$$L(\psi, \phi) \geq L(\psi, \phi^*) \geq L(\psi^*, \phi^*),$$

where the first inequality is due to Theorem 2. Thus, our problem is reduced to an optimal stopping problem, namely, to that of minimizing, over all stopping rules $\psi \in \Delta$,

$$L(\psi) \equiv L(\psi, \phi^*) = \sum_{n=1}^\infty \int (1-\psi_1)\dots(1-\psi_{n-1})\psi_n \left[n f_\theta^n + l_n \right] d\nu_n, \tag{2.10}$$

where, by definition,

$$l_n \equiv \min\{\lambda_0 f_{\theta_0}^n, \lambda_1 f_{\theta_1}^n\}.$$

Let us consider first the class Δ_N of all *truncated* stopping rules, i.e. those with $\psi_N \equiv 1$.

The following theorem is an immediate consequence of Theorem 3 in [11].

Theorem 3. *Let $\psi \in \Delta_N$ be any (truncated) stopping rule. Then*

$$L(\psi) \geq 1 + \int V_1^N(x) d\mu(x) , \tag{2.11}$$

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

$$V_k^N = \min\{l_k, f_\theta^k + \int V_{k+1}^N d\mu(x_{k+1})\}. \tag{2.12}$$

The lower bound in (2.11) is achieved if

$$\begin{aligned} \psi_{N-1} &= I_{\{l_{N-1} \leq f_\theta^{N-1} + \int V_N^N d\mu(x_N)\}} , \\ \psi_{N-2} &= I_{\{l_{N-2} \leq f_\theta^{N-2} + \int V_{N-1}^N d\mu(x_{N-1})\}} , \\ &\dots \\ \psi_1 &= I_{\{l_1 \leq f_\theta^1 + \int V_2^N d\mu(x_2)\}} . \end{aligned} \tag{2.13}$$

Remark 1. Necessary and sufficient conditions of optimality (see Remark 5 in [11]):

If the lower bound (2.14) is achieved by a test $\psi \in \Delta_N$, then for $r = 1, 2, \dots N-1$

$$\psi_r = I_{\{l_r < f_\theta^r + \int V_{r+1}^N d\mu(x_{r+1})\}} + \gamma_r I_{\{l_r = f_\theta^r + \int V_{r+1}^N d\mu(x_{r+1})\}}$$

ν_r-almost anywhere on

$$\{(x_1, x_2, \dots, x_r) : \ (1-\psi_1)(1-\psi_2)\dots(1-\psi_{r-1}) > 0\},$$

where $\gamma_r = \gamma_r(x_1, \dots, x_r)$ are some measurable functions, $0 \leq \gamma_r \leq 1$, for any $r = 1, 2, \dots, N-1$.

On the other hand, there is an equality in (2.11) for any stopping rule of this type.

To treat the case of "purely sequential" tests, we need some additional work to be done. First of all, let us define for any stopping rule ψ

$$\begin{aligned} L_N(\psi) = \sum_{n=1}^{N-1} \int (1-\psi_1)\dots(1-\psi_{n-1})\psi_n(n f_\theta^n + l_n) d\nu_n \\ + \int (1-\psi_1)\dots(1-\psi_{N-1}) \left(N f_\theta^N + l_N\right) d\nu_N . \end{aligned} \tag{2.14}$$

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

From the proof of Lemma 3 [11] it is easy to see that if $N(\theta; \psi) < \infty$ then

$L_N(\psi) \to L(\psi)$, as $N \to \infty$, if

$$\int (1-\psi_1)(1-\psi_2)\dots(1-\psi_{N-1}) l_N d\nu_N \to 0, \text{ as } N \to \infty. \tag{2.15}$$

This gives place to the following definition.

Let us call the problem of hypotheses testing *truncatable* if (2.15) is satisfied for any stopping rule ψ with $N(\theta;\psi) < \infty$.

A sufficient condition for that is obviously

$$\int l_N d\nu_N \to 0, \text{ as } N \to \infty. \tag{2.16}$$

Lemma 1. *For any case of i.i.d. observations, i.e.*

$$f_{\theta_0,j}(x) = f_{\theta_0}(x) \text{ and } f_{\theta_1,j}(x) = f_{\theta_1}(x)$$

for any $j = 1, 2, \dots,$ *the problem of testing* $H_0 : \theta = \theta_0$ *vs.* $H_1 : \theta = \theta_1$ *is truncatable if*

$$\mu\{x : f_{\theta_0}(x) \neq f_{\theta_1}(x)\} > 0. \tag{2.17}$$

Proof. Let us start with noting that

$$\int l_N d\nu_N = E_{\theta_0} \min\{\lambda_0, \lambda_1 Z_N\}, \tag{2.18}$$

where

$$Z_N = Z_N(X_1, \dots, X_N) = \prod_{i=1}^N \frac{f_{\theta_1}(X_i)}{f_{\theta_0}(X_i)}.$$

We show first that $Z_N \to 0$ as $N \to \infty$ in P_{θ_0}-probability. Indeed, for any $\epsilon > 0$

$$P_{\theta_0}(Z_N > \epsilon) = P_{\theta_0}((Z_N)^{1/2} > \epsilon^{1/2}) \leq \epsilon^{-1/2} E_{\theta_0} \left(\prod_{i=1}^N \frac{f_{\theta_1}(X_i)}{f_{\theta_0}(X_i)} \right)^{1/2} \to 0$$

$$= \epsilon^{-1/2} \left(E_{\theta_0} \left(\frac{f_{\theta_1}(X_k)}{f_{\theta_0}(X_k)} \right)^{1/2} \right)^N \to 0$$

as $N \to \infty$, because, by (2.17),

$$E_{\theta_0} \left(\frac{f_{\theta_1}(X)}{f_{\theta_0}(X)} \right)^{1/2} < 1.$$

Now, it suffices to note that

$$\xi_N = \min\{\lambda_0, \lambda_1 Z_N\} \to 0, \text{ as } N \to \infty,$$

in P_{θ_0}-probability as well, and

$$0 \le \xi_N \le \lambda_0,$$

so

$$E_{\theta_0}\xi_N \to 0, \text{ as } N \to \infty,$$

which is equivalent to (2.16). □

Remark 2. For testing $H_0 : \theta = \theta_0$ vs. $H_1 : \theta = \theta_1$ condition (2.17) is always supposed because otherwise H_0 and H_1 state the same distribution, and the statistical problem is meaningless.

Remark 3. Similarly to Lemma 1, any testing problem for which

$$f_{\theta_0,j}(x) = f_{\theta_0}(x) \text{ and } f_{\theta_1,j}(x) = f_{\theta_1}(x)$$

for any $j \ge k$, with some natural $k > 1$, is truncatable if (2.17) is fulfilled.

If

$$f_{\theta_0,j}(x) = f_{\theta_0,j+T}(x) \text{ and } f_{\theta_1,j}(x) = f_{\theta_1,j+T}(x)$$

for any $j \ge 1$, with some natural $T > 1$ (the periodical case mentioned in [9]), then similar arguments show that the testing problem is truncatable if

$$\mu\{x : f_{\theta_0,j}(x) \ne f_{\theta_1,j}(x)\} > 0, \tag{2.19}$$

at least for some $j = 1, \dots T - 1$.

Remark 4. Even if a problem is not truncatable, Theorem 3 still can be used to find optimal sequential tests in Δ_N for any $N \ge 1$.

For any $r = 1, 2, \dots$ let us define

$$V_r = V_r(x_1, \dots, x_r) = \lim_{N\to\infty} V_r^N(x_1, \dots, x_r)$$

(which always exists because by Lemma 4 [11]

$$V_r^N(x_1, \dots, x_r) \ge V_r^{N+1}(x_1, \dots, x_r) \tag{2.20}$$

for any $N \ge 1$ and $r = 1, \dots N$).

Lemma 2. *For any truncatable testing problem*

$$\inf_\psi L(\psi) = 1 + \int V_1(x)d\mu(x)\,, \tag{2.21}$$

where the infimum is taken over all stopping rules ψ.

Proof. Let us denote

$$U = \inf_\psi L(\psi), \quad U_N = 1 + \int V_1^N(x)d\mu(x).$$

By Theorem 3, for any $N = 1, 2, \ldots$

$$U_N = \inf_{\psi \in \Delta_N} L(\psi).$$

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

$$\lim_{N \to \infty} U_N \geq U. \tag{2.22}$$

Let us show first that there is an equality in (2.22).

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

$$U_N \geq U + 3\epsilon \tag{2.23}$$

for all sufficiently large N.

On the other hand, by the definition of U there exists a ψ such that $U \leq L(\psi) \leq U + \epsilon$.

Because, for a truncated testing problem, $L_N(\psi) \to L(\psi)$, as $N \to \infty$, by Lemma 3 [11], we have that

$$L_N(\psi) \leq U + 2\epsilon \tag{2.24}$$

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

$$U_N \leq U + 2\epsilon$$

for all sufficiently large N, which contradicts (2.23).

Now, (2.21) follows from the Lebesgue Monotone Convergence Theorem, because

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

in view of (2.20). □

For any truncatable testing problem the following theorem is an immediate consequence of Theorem 4 and Theorem 5 [11] and Lemma 2.

Theorem 4. *If there exists an optimal stopping rule ψ^*:*

$$\inf_{\psi} L(\psi) = L(\psi^*), \tag{2.25}$$

then

$$\psi_r^* = I_{\{l_r < f_\theta^r + \int V_{r+1} d\mu(x_{r+1})\}} + \gamma_r I_{\{l_r = f_\theta^r + \int V_{r+1} d\mu(x_{r+1})\}} \tag{2.26}$$

ν_r-almost anywhere on $\{(1 - \psi_1^*) \ldots (1 - \psi_{r-1}^*) > 0\}$ with some measurable function $\gamma_r = \gamma_r(x_1, \ldots, x_r)$ such that $0 \leq \gamma_r \leq 1$, for any $r = 1, 2, 3, \ldots$.

On the other hand, any ψ^* defined by (2.26) is optimal in the sense of (2.25) *if*

$$\sum_{n=1}^{\infty} E_\theta(1-\psi_1^*)\dots(1-\psi_{n-1}^*)\psi_n^* = 1. \tag{2.27}$$

If (2.27) is not fulfilled for any choice of γ_r, $r=1,2,\dots$, then there is no optimal sequential test for the testing problem.

Remark 5. A simple sufficient condition for (2.27), and hence for the optimality, is that

$$P_\theta\left(\min\left\{\lambda_0 Z_N^0, \lambda_1 Z_N^1\right\} > 1\right) \to 0, \tag{2.28}$$

as $N\to\infty$, where

$$Z_j = \prod_{i=1}^{N} \frac{f_{\theta_j,i}(X_i)}{f_{\theta,i}(X_i)}, \quad j=0,1.$$

Indeed, in view of (2.26)

$$E_\theta(1-\psi_1^*)(1-\psi_2^*)\dots(1-\psi_N^*) \le E_\theta(1-\psi_N^*) \le P_\theta(l_N > f_\theta^N$$
$$+\int V_{N+1}d\mu(x_{N+1})) \le P_\theta(l_N > f_\theta^N) = P_\theta\left(\min\left\{\lambda_0 Z_N^0, \lambda_1 Z_N^1\right\} > 1\right) \to 0,$$

as $N\to\infty$, by (2.28).

Thus,

$$E_\theta(1-\psi_1^*)(1-\psi_2^*)\dots(1-\psi_N^*) \to 0, \quad \text{as} \quad N\to\infty\,, \tag{2.29}$$

which is equivalent to (2.27).

3. Structure of Optimal Tests for Independent Observations

In a large class of testing problems for independent observations the optimal tests of Theorems 3 and 4 take a much simpler form. In this section we will see that they are based on "probability ratios" for the distributions involved.

Let us suppose that for any $i=1,2,\dots$

$$\{x : f_{\theta,i}(x)=0\} \subset \{x : f_{\theta_0,i}(x)=0\} \cup \{x : f_{\theta_1,i}(x)=0\}. \tag{3.1}$$

In some sense, this means that the distribution corresponding to θ "lies between" those corresponding to θ_0 and θ_1.

For example, if

$$f_{\theta,i}(x) = \frac{1}{\theta} \text{ for } x\in[0,\theta]$$

(uniform distribution on $[0,\theta]$) then (3.1) is fulfilled if $\theta \in [\theta_0, \theta_1]$ and not fulfilled otherwise.

Supposing that (3.1) is satisfied, the Lagrange-multiplier function $L(\psi)$ (2.2) can be expressed in terms of the probability ratios

$$\begin{aligned} z_n^0 &= z_n^0(x_1, \ldots, x_n) = \prod_{i=1}^n \frac{f_{\theta_0,i}(x_i)}{f_{\theta,i}(x_i)} \quad \text{and} \\ z_n^1 &= z_n^1(x_1, \ldots, x_n) = \prod_{i=1}^n \frac{f_{\theta_1,i}(x_i)}{f_{\theta,i}(x_i)}. \end{aligned} \tag{3.2}$$

Indeed,

$$L(\psi) = \sum_{n=1}^{\infty} \int (1-\psi_1)\ldots(1-\psi_{n-1})\psi_n \left[n f_\theta^n + \min\{\lambda_0 f_{\theta_0}^n, \lambda_1 f_{\theta_1}^n\}\right] I_{\{f_\theta^n \neq 0\}} d\nu_n$$
$$+ \sum_{n=1}^{\infty} \int (1-\psi_1)\ldots(1-\psi_{n-1})\psi_n \min\{\lambda_0 f_{\theta_0}^n, \lambda_1 f_{\theta_1}^n\} I_{\{f_\theta^n = 0\}} d\nu_n.$$

But $\min\{\lambda_0 f_{\theta_0}^n, \lambda_1 f_{\theta_1}^n\} = 0$ on $\{f_\theta^n = 0\}$ because of (3.1), so the second term is null. Thus,

$$L(\psi) = \sum_{n=1}^{\infty} E_\theta (1-\psi_1)\ldots(1-\psi_{n-1})\psi_n \left[n + \min\{\lambda_0 z_n^0, \lambda_1 z_n^1\}\right] \tag{3.3}$$

(cf. [10]). Here, and in what follows, we assume that any expectation corresponding to a "density" function $f(x)$

$$Eg(X) = \int g(x) f(x) d\mu(x)$$

is understood as

$$Eg(X) = \int g(x) f(x) I_{\{f(x)\neq 0\}} d\mu(x),$$

so we do not need to care about the definition of $g(x)$ on $\{f(x) = 0\}$.

For example, there is no problem with the definition of z_n^0 and z_n^1 on the right-hand side of (3.3) when $f_\theta^n = 0$.

In view of (3.3), it seems very natural that in this case the optimal stopping rules of Theorems 3 and 4 are functions of z_n^0 and z_n^1 only. Let us prove this and find the structure of the corresponding optimal tests in terms of z_n^0 and z_n^1.

Lemma 3. *Under condition (3.1), for any $N \geq 1$ and $r = 1, 2, \ldots, N$*

$$V_r^N = f_\theta^r \rho_r^N (z_r^0, z_r^1) I_{\{f_\theta^r > 0\}}, \tag{3.4}$$

where the functions $\rho_r^N = \rho_r^N(z_0, z_1)$, are defined for any $z_0, z_1 \geq 0$ for any

$N \geq 1$, $r = 1, \ldots, N$ in the following way:

$$\rho_N^N(z_0, z_1) = g(z_0, z_1)$$

with $g(z_0, z_1) \equiv \min\{\lambda_0 z_0, \lambda_1 z_1\}$, and recursively for any $r = N, N-2, \ldots, 2$

$$\rho_{r-1}^N(z_0, z_1) = \min\left\{g(z_0, z_1), 1 + E_\theta \rho_r^N\left(z_0 \frac{f_{\theta_0,r}(X_r)}{f_{\theta,r}(X_r)}, z_1 \frac{f_{\theta_1,r}(X_r)}{f_{\theta,r}(X_r)}\right)\right\}. \quad (3.5)$$

Proof. By induction over $r = N, \ldots, 1$. For $r = N$ we have

$V_N^N = \min\{\lambda_0 f_{\theta_0}^N, \lambda_1 f_{\theta_1}^N\} = f_\theta^N \min\{\lambda_0 z_N^0, \lambda_1 z_N^1\} I_{\{f_\theta^N > 0\}} = f_\theta^N g(z_N^0, z_N^1) I_{\{f_\theta^N > 0\}}$ by virtue of (3.1).

Let us suppose that (3.4) is satisfied for some $r = 2, \ldots, N$.

Then

$$V_{r-1}^N = f_\theta^{r-1} \min\{g(z_{r-1}^0, z_{r-1}^1), 1 + \int f_{\theta,r}(x_r) \rho_r^N(z_r^0, z_r^1) d\mu(x_r)\}$$

$$= f_\theta^{r-1} \min\{g(z_{r-1}^0, z_{r-1}^1), 1$$

$$+ \int f_{\theta,r}(x_r) \rho_r^N\left(z_{r-1}^0 \frac{f_{\theta_0,r}(x_r)}{f_{\theta,r}(x_r)}, z_{r-1}^1 \frac{f_{\theta_1,r}(x_r)}{f_{\theta,r}(x_r)}\right) d\mu(x_r)\} = f_\theta^{r-1} \rho_{r-1}^N(z_{r-1}^0, z_{r-1}^1)$$

if $f_\theta^{r-1} > 0$.

If $f_\theta^{r-1} = 0$ then by (3.1) again we have that $\min\{\lambda_0 f_{\theta_0}^{r-1}, \lambda_1 f_{\theta_1}^{r-1}\} = 0$, so, by (2.12), $V_{r-1}^N = 0$.

Thus, (3.4) is satisfied for $r - 1$ as well. □

Lemma 4. *For any $N \geq 1$, $r = 1, 2, \ldots, N$*

$$\rho_r^N(z_0, z_1) \geq \rho_r^{N+1}(z_0, z_1) \quad (3.6)$$

for any $z_0 \geq 0$, $z_1 \geq 0$.

Proof. By induction over $r = N, N-1, \ldots, 1$ again.

For $r = N$

$$\rho_N^N(z_0, z_1) = g(z_0, z_1) \geq \rho_N^{N+1}(z_0, z_1)$$

by (3.5).

Let us suppose (3.6) for some $r = N, N-1, \ldots, 2$. Then

$$\rho_{r-1}^N(z_0, z_1) = \min\left\{g(z_0, z_1), 1 + E_\theta \rho_r^N\left(z_0 \frac{f_{\theta_0,r}(X_r)}{f_{\theta,r}(X_r)}, z_1 \frac{f_{\theta_1,r}(X_r)}{f_{\theta,r}(X_r)}\right)\right\}$$

$$\geq \min\left\{g(z_0, z_1), 1 + E_\theta \rho_r^{N+1}\left(z_0 \frac{f_{\theta_0,r}(X_r)}{f_{\theta,r}(X_r)}, z_1 \frac{f_{\theta_1,r}(X_r)}{f_{\theta,r}(X_r)}\right)\right\} = \rho_{r-1}^{N+1}(z_0, z_1). \ \square$$

By Lemma 4, for any $r = 1, 2, \dots$ there exists

$$\rho_r(z_0, z_1) = \lim_{N\to\infty} \rho_r^N(z_0, z_1).$$

Passing to the limit, as $N \to \infty$, in (3.5), we have

$$\rho_{r-1}(z_0, z_1) = \min\left\{g(z_0, z_1), 1 + E_\theta \rho_r\left(z_0 \frac{f_{\theta_0,r}(X_r)}{f_{\theta,r}(X_r)}, z_1 \frac{f_{\theta_1,r}(X_r)}{f_{\theta,r}(X_r)}\right)\right\} \tag{3.7}$$

for any $r = 2, 3, \dots$.

Let

$$R_r(z_0, z_1) = 1 + E_\theta \rho_r\left(z_0 \frac{f_{\theta_0,r}(X_r)}{f_{\theta,r}(X_r)}, z_1 \frac{f_{\theta_1,r}(X_r)}{f_{\theta,r}(X_r)}\right).$$

Any optimal test in is now of the form

$$\psi_r^* = I_{\{g(z_r^0, z_r^1) < R_{r+1}(z_r^0, z_r^1)\}} + \gamma_r I_{\{g(z_r^0, z_r^1) = R_{r+1}(z_r^0, z_r^1)\}}, \tag{3.8}$$

ν_r-almost anywhere on $\{(1-\psi_1)\dots(1-\psi_{r-1}) > 0\} \cap \{f_\theta^r > 0\}$, with some measurable function $\gamma_r = \gamma_r(x_1, \dots, x_r)$, $0 \le \gamma_r \le 1$, for any $r = 1, 2, 3, \dots$ (Theorem 4).

For the case of independent *and identically distributed* observations considered in [10] (which is always truncatable by Lemma 1) we now have a broader class of optimal tests: the optimal tests of type (3) [10] correspond to a particular case of (3.8), when $\gamma_r \equiv 1$ o $\gamma_r \equiv 0$ for any $r \ge 1$. This extension is because of the randomization of stopping rules we admit in this article. This extension is irrelevant for the problem of minimization of $L(\psi)$ as such, but may be important for the original problem of minimizing the average sample number (1.6) under restrictions (1.4) and (1.5) (see Theorem 1), because it broadens the class of tests to seek for complying with (2.3), quite like the randomization of decision rules in the Neyman-Pearson problem is important for finding tests with a given α-level (see, for example, [8]).

In particular, any test (3.8) is optimal in the sense if Theorem 1, i.e. for any sequential test (ψ, ϕ) such that

$$\alpha(\psi, \phi) \le \alpha(\psi^*, \phi^*) \quad \text{and} \quad \beta(\psi, \phi) \le \beta(\psi^*, \phi^*) \tag{3.9}$$

(with the decision rule ϕ^* defined in Theorem 2) it holds

$$N(\theta; \psi^*) \le N(\theta, \psi), \tag{3.10}$$

and (3.10) is strict if one of the inequalities in (3.9) is strict.

More than that, if $\alpha(\psi, \phi) = \alpha(\psi^*, \phi^*)$, $\beta(\psi, \phi) = \beta(\psi^*, \phi^*)$ and $N(\theta; \psi) = N(\theta, \psi^*)$ then by Theorem 4 the stopping rule ψ has to have the form (3.8) as well (probably, with some other γ_r, $r = 1, 2, \dots$).

There is a detailed study of properties of the non-randomized optimal tests in [10] for the i.i.d. case.

Our focus in what follows will be on the "non-stationary" observations case (see [9]), i.e. the case when $f_{\theta,r}$ vary "in time". Because the approach based on Bayesian formulation (see [9]) was not very successful (see [12]), we would like to propose another one, based on the present theory.

We will focus our attention to the minimization of $N(\theta_0;\psi)$, i.e. minimization of the average sample size under *one* of the two competing hypotheses. Because the hypotheses are interchangeable, this will obviously give us a way to minimize $N(\theta_1;\psi)$ as well. This not necessarily gives us *the same* test, but occasionally it *does* (e.g. for i.i.d. observations). If it is *not* the same test, we will have *two* optimal tests, and at least we always have the choice, which one to prefer. As an alternative, we can minimize a weighted average sample size, in form of

$$\pi N(\theta_0;\psi)+(1-\pi)N(\theta_1;\psi),$$

which can be solved using essentially the same technique.

The existence of the optimal sequential test from Bayesian point of view is proved in [4] for a very broad class of discrete-time stochastic processes, on the basis of the general theory of optimal stopping (see, for example, [3]). A more direct approach in [9] for independent observations, based on the technique of [5], gives much more explicit results on the structure of the sequential Bayesian tests. Our general results [11], based on the same principles as the treatment of sequential problems in [5], allow to unify all existing results. In what follows we show how it works for the case of independent observations.

We will use the form (3.8) of the optimal sequential test in this particular case of $\theta=\theta_0$.

Because in this case $z_r^0\equiv 1$ (see (3.2)) let us express all the elements of (3.8) in terms of

$$z_r\equiv z_r^1=\prod_{i=1}^r\frac{f_{\theta_1,i}}{f_{\theta_0,i}}$$

(see (3.2)), simply omitting the first argument in all the functions involved. In particular, we have:

$$g(z)=\min\{\lambda_0,\lambda_1 z\},\quad \rho_N^N(z)=g(z), \tag{3.11}$$

$$\rho_{r-1}^N(z)=\min\left\{g(z),1+E_{\theta_0}\rho_r^N\left(z\frac{f_{\theta_1,r}(X_r)}{f_{\theta_0,r}(X_r)}\right)\right\}, \tag{3.12}$$

$$\rho_r(z)=\lim_{N\to\infty}\rho_r^N(z), \tag{3.13}$$

$$\rho_{r-1}(z) = \min\left\{g(z), 1 + E_{\theta_0}\rho_r\left(z\frac{f_{\theta_1,r}(X_r)}{f_{\theta_0,r}(X_r)}\right)\right\}, \tag{3.14}$$

$$R_r(z) = E_{\theta_0}\rho_r\left(z\frac{f_{\theta_1,r}(X_r)}{f_{\theta_0,r}(X_r)}\right). \tag{3.15}$$

Finally, by Theorem 4 we have that any optimal sequential test minimizing $L(\psi)$ has the form (see (3.8)):

$$\psi_r^* = I_{\{g(z_r)<R_{r+1}(z_r)\}} + \gamma_r I_{\{g(z_r)=R_{r+1}(z_r)\}}, \tag{3.16}$$

ν_r-almost anywhere on $\{(1-\psi_1)\dots(1-\psi_{r-1})>0\}\cap\{f_{\theta_0}^r>0\}$, with some measurable function $\gamma_r=\gamma_r(x_1,\dots,x_r)$, $0\le\gamma_r\le 1$, for any $r=1,2,3,\dots$.

A more specific structure of the optimal stopping rule can be obtained from the properties of the functions involved (i.e., $g(z), \rho_r(z), R_r(z)$).

Lemma 5. *The functions $\rho_r(z)$ defined in (3.11)-(3.15) have the following properties:*

i) they are concave and continuous on $[0,\infty)$,

ii) they are non-decreasing on $[0,\infty)$,

iii) $\rho_r^N(0)=\rho_r(0)=0$ *and* $\rho_r^N(\infty)=\rho_r(\infty)=\lambda_0$,

Proof. Let us start with the following simple lemma which will be useful at different stages of the proof.

Lemma 6. *Let ϕ be any concave function on* $[0,\infty)$. *Then*

$$E_{\theta_0}\phi\left(z\frac{f_{\theta_1,r}(X_r)}{f_{\theta_0,r}(X_r)}\right)$$

is a concave function of z.

Proof. This is a straightforward consequence of the concavity of ϕ. ☐

i) Let us prove by induction over $r=N,N-1,\dots,1$ that ρ_r^N are concave.

For $r=N$ $\rho_r^N(z)=g(z)$ is concave as a minimum of two concave functions (in this particular case, lineal ones).

Suppose that $\rho_r^N(z)$ is concave. Then, observing (3.12), we see that ρ_{r-1}^N is a minimum of two concave functions (the second one is concave by Lemma 6). Thus, ρ_{r-1}^N is concave as well.

The function $\rho_r(z)=\lim_{N\to\infty}\rho_r^N(z)$ is also concave as a limit of concave functions. Because $\rho_r^N(z)\ge 0$ and $\rho_r(z)\ge 0$ are concave, they are continuous in $(0,\infty)$ (see, for example, Section 3.18 of [6]).

In addition, $\rho_r^N(z) \le g(z)$ (by (3.12)), and $\rho_r(z) \le g(z)$ (by (3.14)), and $g(z) \to 0$ as $z \to 0$, then the continuity at $z = 0$ will follow from iii).

ii) By induction as well.

For $r = N$ it is obvious. If $\rho_r^N(z)$ is non-decreasing then by (3.12) ρ_{r-1}^N is non-decreasing as well. Thus, ρ_r is non-decreasing as a limit of non-decreasing functions.

iii) Starting from $\rho_N^N(0) = g(0) = 0$, by a straightforward induction we obtain that $\rho_r^N(0) = 0$ for any $r \le N$. Thus, $\rho_r(0) = \lim_{N\to\infty} \rho_r^N(0) = 0$.

By a similar induction, using the Lebesgue Monotone Convergence Theorem, we have $\rho_r^N(\infty) = \lambda_0$.

By ii), the limit $\lim_{z\to\infty} \rho_r(z) = \lambda_r$ exists for any $r = 1, 2, \ldots$. Passing to the limit, as $z \to \infty$, in (3.15) we see that $\lambda_{r-1} = \min\{\lambda_0, 1 + \lambda_r\}$, for any $r = 2, 3, \ldots$.

If for some $r \ge 2$ $\lambda_r = \lambda_0$ then, obviously, $\lambda_i = \lambda_0$ for any $i \le r$.

Let us suppose that there is a finite number k such that $\lambda_i = 1 + \lambda_{i+1} < \lambda_0$ for any $i \ge k$. This easily conducts to a contradiction because in such a case for any $m > k$ we would have

$$\lambda_k = 1 + \lambda_{k+1} = 2 + \lambda_{k+2} = \cdots = (m - k) + \lambda_m < \lambda_0,$$

which can not be true if $m - k > \lambda_0$ because $\lambda_m \ge 0$. Thus, $\lambda_r = \lambda_0$ for any $r \ge 1$.

To finalize the proof just note that the concavity of $R_r(z)$ is due to Lemma 6, the continuity at $z = 0$ follows again from the Lebesgue monotone convergence theorem, as well as the fact that $R_r(0) = 0$ and $R_r(\infty) = \lambda_0$. $\square$

The above properties allow to give to the inequality

$$g(z_r) < 1 + R_{r+1}(z_r)$$

defining the optimal stopping rule in (3.16) a simple equivalent form:

$$z_r \notin [A_r, B_r].$$

Let us define for any $r = 1, 2 \ldots$

$$A_r = \sup\{z : 0 \le z \le \lambda_0/\lambda_1,\ 1 + R_{r+1}(z) \ge g(z)\},$$

$$B_r = \inf\{z : \lambda_0/\lambda_1 \le z,\ 1 + R_{r+1}(z) \ge g(z)\}.$$

Because $1 + R_{r+1}(0) = 1 > g(0) = 0$ and $1 + R_{r+1}(\infty) = 1 + \lambda_0 > g(\infty) = \lambda_0$ (by Lemma 5) it follows that $A_r > 0$ y $B_r < \infty$.

On the other hand, if $A_r = B_r = \lambda_0/\lambda_1$ then $1 + R_{r+1}(z) \ge g(z)$ for any $z \ge 0$, and $\{1 + R_{r+1}(z) \ge g(z)\}$ is non-empty if and only if $1 + R_{r+1}(\lambda_0/\lambda_1) =$

$g(\lambda_0/\lambda_1) = \lambda_0$.

Otherwise $A_r < \lambda_0/\lambda_1 < B_r$. In this case, we have the following

Lemma 7. *$A_r < z < B_r$ if and only if $g(z) > 1 + R_{r+1}(z)$.*

Proof. By Lemma 5 all the functions involved are continuous, so $g(A_r) = 1 + R_{r+1}(A_r)$ and $g(B_r) = 1 + R_{r+1}(B_r)$. By definition, we have further $1 + R_{r+1}(z) < g(z)$ for $z \in (A_r, \lambda_0/\lambda_1]$. For the same reason $1 + R_{r+1}(z) < g(z)$ for $z \in [\lambda_0/\lambda_1, B_r)$. If now $z \in [0, A_r)$ then there exists $1 \geq \gamma > 0$ such that $z = (1-\gamma)A_r$. As $R_{r+1}(0) = 0$, by the concavity of $R_{r+1}(z)$ we have

$$R_{r+1}((1-\gamma)A_r) \geq (1-\gamma)R_{r+1}(A_r) = (1-\gamma)(g(A_r) - 1).$$

Therefore,

$$\begin{aligned} 1 + R_{r+1}(z) &\geq 1 + (1-\gamma)(g(A_r) - 1) = \gamma + (1-\gamma)g(A_r) \\ &> (1-\gamma)g(A_r) = g((1-\gamma)A_r) = g(z), \end{aligned}$$

which implies $1 + R_{r+1}(z) > g(z)$.

It can be proved analogously that if $z \in (B, \infty)$ then $1 + R_{r+1}(z) > g(z)$ as well. □

Any optimal test (3.16) is now of the form

$$\psi_r^* = \begin{cases} I_{\{z_r < A_r \text{ or } z_r > B_r\}} + \gamma_r I_{\{z_r = A_r \text{ or } z_r = B_r\}} & \text{if } 1 + R_{r+1}(\lambda_0/\lambda_1) \leq \lambda_0, \\ 1 & \text{otherwise}, \end{cases} \tag{3.17}$$

ν_r-almost anywhere on $\{(1-\psi_1)\dots(1-\psi_{r-1}) > 0\} \cap \{f_{\theta_0}^r > 0\}$, with some measurable function $\gamma_r = \gamma_r(x_1, \dots, x_r)$, $0 \leq \gamma_r \leq 1$, where A_r and B_r are two roots (possibly coinciding) of the equation

$$1 + R_{r+1}(z) = g(z)\,, \tag{3.18}$$

which are uniquely defined by Lemma 7, and $A_r \leq \lambda_0/\lambda_1 \leq B_r$, for any $r = 1, 2, 3, \dots$.

4. Applications

In this section, we obtain a complete solution to the problem of optimal sequential hypotheses testing for a particular case of the above framework, when the process is stationary from some time k on, $k \geq 1$.

4.1. Finitely Non-Stationary Process

Let us suppose that $f_{\theta,i}(x) = f_\theta(x)$ for $i \geq k$ with some natural $k \geq 1$, for both $\theta = \theta_0$ and $\theta = \theta_1$. By Remark 3, the problem of testing $H_0 : \ \theta = \theta_0$ vs. $H_1 : \ \theta = \theta_1$ is truncatable.

We will suppose that

$$\mu\{x : f_{\theta_0}(x) \neq f_{\theta_1}(x)\} > 0, \tag{4.1}$$

otherwise it is easy to see that the optimal sequential test is truncated at $k-1$.

It follows from (3.11) and (3.12) that

$$\rho_r^N(z) = \rho_{r+1}^{N+1}(z) \tag{4.2}$$

for any $r = k-1, k, \ldots, N$. Passing in (4.2) to the limit, as $N \to \infty$, we have that

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

for any $r = k-1, k, \ldots$. Let us denote $\rho(z) \equiv \rho_{k-1}(z)$. Applying (3.14) with $r = k$ we immediately have

$$\rho(z) = \min\{g(z), 1 + E_{\theta_0}\rho\left(z\frac{f_{\theta_1}(X_k)}{f_{\theta_0}(X_k)}\right)\}$$

for any $z \geq 0$.

Applying (3.14) for $r = k-1, k-2, \ldots, 2$ successively again, we obtain

$$\rho_{k-2}(z) = \min\{g(z), 1 + E_{\theta_0}\rho\left(z\frac{f_{\theta_1,k-1}(X_{k-1})}{f_{\theta_0,k-1}(X_{k-1})}\right)\},$$

$$\rho_{k-3}(z) = \min\{g(z), 1 + E_{\theta_0}\rho_{k-2}\left(z\frac{f_{\theta_1,k-2}(X_{k-2})}{f_{\theta_0,k-2}(X_{k-2})}\right)\}, \ldots$$

$$\rho_1(z) = \min\{g(z), 1 + E_{\theta_0}\rho_2\left(z\frac{f_{\theta_1,2}(X_2)}{f_{\theta_0,2}(X_2)}\right)\}.$$

In particular, the bounds A_r and B_r of the continuation region (see (3.18)), are defined for $r = 1, 2, \ldots, k-2$ from

$$g(z) = 1 + E_{\theta_0}\rho_{r+1}\left(z\frac{f_{\theta_1,r+1}(X_{r+1})}{f_{\theta_0,r+1}(X_{r+1})}\right), \tag{4.3}$$

and $A_r \equiv A$ and $B_r \equiv B$ for $r = k-1, k, \ldots$, where A and B are defined from

$$g(z) = 1 + E_{\theta_0}\rho\left(z\frac{f_{\theta_1}(X_k)}{f_{\theta_0}(X_k)}\right). \tag{4.4}$$

By Remark 5, any test with bounds defined by (4.3)–(4.4) is optimal: in this case (2.28) is fulfilled because $Z_N^0 \equiv 1$ (recall that $\theta = \theta_0$), and $Z_N^1 \to 0$ in

P_{θ_0}-probability, as $N \to \infty$ (see the proof of Lemma 1).

Combining Theorem 4 with Theorems 1 and 2 and with the form of optimal sequential test (3.17) we have

Theorem 5. *For any $\lambda_0 > 0$ and $\lambda_1 > 0$ any test (ψ^*, ϕ^*) based on stopping time (3.17) with A_r, B_r defined from (4.3) and A, B defined from (4.4), with ϕ^* defined in Theorem 2, is optimal in the following sense: for any sequential test (ψ, ϕ) such that*

$$\alpha(\psi, \phi) \leq \alpha(\psi^*, \phi^*) \quad \text{and} \quad \beta(\psi, \phi) \leq \beta(\psi^*, \phi^*) \tag{4.5}$$

it holds

$$N(\theta_0; \psi^*) \leq N(\theta_0; \psi). \tag{4.6}$$

The inequality in (4.6) is strict if at least one of the inequalities in (4.5) is strict. If there are equalities in all the inequalities in (4.5) and (4.6) then ψ has form (3.17), with possibly another choice of γ_r, $r = 1, 2, \ldots$.

If all the observations are identically distributed, except, possibly, the first one ($k = 2$ in the above case), it is immediate that the bounds of the continuation region are constant ($A_r \equiv A$ and $B_r \equiv B$ for any $r = 1, 2, \ldots$), see (4.4). Hence, the optimal test in Theorem 5 acts as the usual sequential probability ratio test (SPRT) for i.i.d. observations. Because the SPRT minimizes not only $N(\theta_0; \psi)$ but also $N(\theta_1; \psi)$, we will pay some more attention to this case in the following subsection, and prove that, for $k = 2$, any test of Theorem 5 minimizes both $N(\theta_0; \psi)$ and $N(\theta_1; \psi)$ under conditions (4.5), exactly like an SPRT does.

4.2. Stationary Process with Non-Stationary Initial Distribution

In this subsection we treat the case $f_{\theta,i}(x) = f_\theta(x)$ for $i \geq 2$ when $\theta = \theta_0$ and $\theta = \theta_1$, supposing that $f_{\theta,1}(x)$ does not necessarily coincide with $f_\theta(x)$. If it does, for both $\theta = \theta_0$ and $\theta = \theta_1$, we have the i.i.d. observations as a particular case of this model.

The aim of this subsection is to prove

Theorem 6. *For any $\lambda_0 > 0$ and $\lambda_1 > 0$ any test (ψ^*, ϕ^*) with*

$$\psi_r^* = I_{\{z_r < A \text{ or } z_r > B\}} + \gamma_r I_{\{z_r = A \text{ or } z_r = B\}}, \tag{4.7}$$

ν_r-almost anywhere on $\{(1 - \psi_1^*) \ldots (1 - \psi_{r-1}^*) > 0\} \cap \left(\{f_{\theta_0}^r > 0\} \cup \{f_{\theta_1}^r > 0\}\right)$, with some measurable function $\gamma_r = \gamma_r(x_1, \ldots, x_r)$, $0 \leq \gamma_r \leq 1$, $r \geq 1$, where

A and B are two roots of the equation

$$1 + R(z) = g(z), \tag{4.8}$$

with ϕ^* defined in Theorem 2, is optimal in the following sense: for any sequential test (ψ, ϕ) such that

$$\alpha(\psi, \phi) \le \alpha(\psi^*, \phi^*) \quad \text{and} \quad \beta(\psi, \phi) \le \beta(\psi^*, \phi^*) \tag{4.9}$$

it holds

$$N(\theta_0; \psi^*) \le N(\theta_0; \psi) \quad \text{and} \quad N(\theta_1; \psi^*) \le N(\theta_1; \psi). \tag{4.10}$$

The inequalities in (4.10) are strict if at least one of the inequalities in (4.9) is strict. If there are equalities in the inequalities in (4.9) and in one of inequalities (4.10) then ψ has form (4.7), with possibly another choice of γ_r, $r = 1, 2, \ldots$.

Proof. To prove the theorem, it is sufficient to show that for *any* sequential test of type (4.7) with *any* $0 < A < B < \infty$ there exist $\lambda_0 > 0$ and $\lambda_1 > 0$ such that A and B are roots of (4.8) (recall that $R(z)$, $g(z)$ are functions of λ_0 and λ_1, as well as of the densities f_{θ_0} and f_{θ_1}).

First of all, let us recall that we derive the optimal tests of type (4.7) from minimizing

$$N(\theta_0; \psi) + \lambda_0 \alpha(\psi, \phi) + \lambda_1 \beta(\psi, \phi) \tag{4.11}$$

with any λ_0 and λ_1. It is obvious that we would obtain the same results from minimizing, instead of (4.11),

$$cN(\theta_0; \psi) + \lambda \alpha(\psi, \phi) + \beta(\psi, \phi) \tag{4.12}$$

with any $c > 0$ and $\lambda > 0$. Essentially, (4.12) is a re-parametrization of (4.11). All the results above are easily reformulated in terms of c and λ instead of λ_0 and λ_1. In particular, we have

$$\rho(z; c, \lambda) = \min\{g(z; \lambda), c + R(z; c, \lambda)\},$$

where

$$g(z; \lambda) = \min\{\lambda, z\}, \quad \text{and} \quad R(z) = E_{\theta_0} \rho\left(z \frac{f_{\theta_1}(X_2)}{f_{\theta_0}(X_2)}; c, \lambda\right),$$

being $\rho(z; c, \lambda) = \lim_{n\to\infty} \rho_n(z; c, \lambda)$, where, in turn, $\rho_n(z)$ are defined starting from

$$\rho_0(z; c, \lambda) = g(z; \lambda),$$

and for $n = 1, 2, \ldots$ defined recursively:

$$\rho_n(z; c, \lambda) = \min\{g(z; \lambda), c + E_{\theta_0} \rho_{n-1}\left(z \frac{f_{\theta_1}(X_2)}{f_{\theta_0}(X_2)}; c, \lambda\right)\}.$$

As well, (4.8) transforms to

$$c + R(z; c, \lambda) = g(z; \lambda). \tag{4.13}$$

In exactly the same way as in Lemma 5 it can be shown that $R(z; c, \lambda)$ is concave and non-decreasing in c and λ, when other variables are held fixed. From this immediately follows, in particular, that it is continuous in $c \in (0, \infty)$. Let us show that $R(z; c, \lambda)$ is continuous at $c = 0$ for any fixed z and λ as well.

By Lemma 2 we have that

$$c + R(z; c, \lambda) = \inf_{\psi}\{cN(\theta_0, \psi) + \lambda\alpha(\psi, \phi^*) + z\beta(\psi, \phi^*)\} \tag{4.14}$$

for an "auxiliary" testing problem $H_0 : \theta = \theta_0$ vs. $H_1 : \theta = \theta_1$ when all the observations are supposed identically distributed, i.e. $f_{\theta,1}(x) = f_\theta(x)$, for $\theta = \theta_0$ and $\theta = \theta_1$, $c > 0$.

Let us take a test based on a fixed sample size: $\psi_i \equiv 0$, $i < n$,$\psi_n \equiv 1$ and recall that by Theorem 2 $\phi_n^* = I_{\{zf_{\theta_1}^n - \lambda f_{\theta_0}^n \geq 0\}}$. Then the right-hand side of (4.14) does not exceed

$$\begin{aligned} &cn + \lambda P_{\theta_0}(zf_{\theta_1}^n - \lambda f_{\theta_0}^n \geq 0) + zP_{\theta_1}(zf_{\theta_1}^n - \lambda f_{\theta_0}^n < 0) \\ &= cn + \lambda P_{\theta_0}(f_{\theta_1}^n / f_{\theta_0}^n \geq \lambda / z) + zP_{\theta_1}(f_{\theta_0}^n / f_{\theta_1}^n > z/\lambda). \end{aligned} \tag{4.15}$$

In the proof of Lemma 1 it is shown that the second summand in (4.15) tends to 0 as $n \to \infty$. In the same way it can be shown that so does the third summand on (4.15) as well. Thus, for any $\epsilon > 0$ (4.15) does not exceed

$$nc + 2\epsilon \tag{4.16}$$

for n large enough, say $n \geq N$. If now $c < \epsilon/N$ then (4.16) does not exceed 3ϵ. This proves that the right-hand side of (4.14) tends to 0, as $c \to 0$. Formally speaking, the left-hand side of (4.14) is not defined for $c = 0$. But in view of the fact that the right-hand side of (4.14) tends to 0 as $c \to 0$, it is convenient to put $R(z; 0, \lambda) = 0$ for any $\lambda \geq 0$ and for any $z \geq 0$, which makes $R(z; c, \lambda)$ continuous in $c \in [0, \infty)$.

Lemma 8. *For any fixed z $R(z; c, \lambda)$ is continuous, as a function of (c, λ), at any point (c_0, λ_0) with $c_0 \geq 0$, $\lambda_0 > 0$.*

Proof. Let us fix any two numbers $c_0 \geq 0$ and $\lambda_0 > 0$. Then for any λ_1, λ_2 such that $\lambda_1 < \lambda_0 < \lambda_2$ we have

$$R(z; c; \lambda_1) \leq R(z; c, \lambda) \leq R(z; c; \lambda_2)$$

if $\lambda \in [\lambda_1, \lambda_2]$. Now,

$$\lim_{c \to c_0} R(z; c; \lambda_1) \leq \liminf_{c \to c_0, \lambda \to \lambda_0} R(z; c, \lambda) \leq \limsup_{c \to c_0, \lambda \to \lambda_0} R(z; c, \lambda) \leq \lim_{c \to c_0} R(z; c; \lambda_2),$$

or

$$R(z;c_0;\lambda_1) \le \liminf_{c\to c_0,\lambda\to\lambda_0} R(z;c,\lambda) \le \limsup_{c\to c_0,\lambda\to\lambda_0} R(z;c,\lambda) \le R(z;c_0;\lambda_2). \quad (4.17)$$

Now, letting $\lambda_1 \to \lambda_0$ and $\lambda_2 \to \lambda_0$ in (4.17) and noting that

$$R(z;c_0;\lambda) \to R(z;c_0;\lambda_0), \text{ as } \lambda \to \lambda_0$$

we obtain that

$$\lim_{c\to c_0,\lambda\to\lambda_0} R(z;c,\lambda) = R(z;c_0,\lambda_0)\,. \qquad \square$$

Let now A and B be any two numbers, $0 < A < B < \infty$. For any λ, $\lambda \in [A,B]$, let us define $c = c(\lambda)$ as a solution of the equation

$$c + R(A;c,\lambda) = A. \quad (4.18)$$

We know that $c = c(\lambda)$ exists and is unique, because the left-hand side of (4.18) is a continuous and strictly increasing function of c. Moreover, $c(\lambda)$ is a continuous function of λ, as an implicit function (4.18) defined by a function which is continuous in two variables (Lemma 8). In addition, $c(\lambda) \ge 0$, because $R(A;c,\lambda) \le \min\{A,\lambda\} = A$ for $\lambda \ge A$.

Let us define now

$$G(\lambda) = \lambda - R(B;c(\lambda),\lambda) - c(\lambda),$$

that is continuous function of λ as a composition of two continuous functions.

Let us show that

$$G(A) < 0, \quad \text{and} \quad G(B) > 0. \quad (4.19)$$

Indeed,

$$G(A) = A - R(B;c(A),A) - c(A) < A - R(A;c(A),A) - c(A) = 0$$

(by (4.18)).

Let us show now that

$$G(B) = B - R(B;c(B),B) - c(B) > 0. \quad (4.20)$$

Taking into account that, by (4.18),

$$c(B) + R(A;c(B),B) = A,$$

we see that (4.20) is equivalent to

$$B - R(B;c(B),B) > A - R(A;c(B),B).$$

But this is due to the fact that $z - R(z;c(B),B)$ is a strictly increasing, at least in a neighbourhood of $z = B$: indeed, it is an obviously convex non-negative function, so there is a $z_0 \ge 0$ such that it is strictly increasing for $z \ge z_0$. Here z_0 cannot be greater or equal than B because otherwise $R(z;c(B),B)$ would coincide with $g(z, B)$ for any $z \geq 0$, which is only possible if $f_{\theta_0}(x) = f_{\theta_1}(x)$ μ-almost anywhere.

Thus, (4.19) is proved, so there exists $\lambda \in (A, B)$ such that $G(\lambda) = 0$, that is, we found $\lambda \in (A, B)$ and $c = c(\lambda)$ such that

$$c + R(A; c, \lambda) = A \quad \text{and} \quad c + R(B; c, \lambda) = \lambda,$$

which is equivalent to (4.13). □

Remark 6. In a particular case of i.i.d. observations, we have here an alternative, more direct, proof of the optimality of the SPRT (the SPRT corresponds to γ_r which coincides with $I_{\{z_r=A\}}$ or with $I_{\{z_r=B\}}$ for any $r = 1, 2, \ldots$ in (4.7)).

The uniqueness result of Theorem 6 is an extension, to randomized SPRT's, of the uniqueness results of Wijsman [16] (see also Remark 6 in Section 7 of [10]).

Remark 7. For i.i.d. observations, only the case $A \leq 1 \leq B$ has practical sense, because otherwise there is a (trivial) test which, without taking any observations, achieves a lesser value of the Lagrange multiplier function $L(\psi)$ (see also [14], [2], [10], among many others). For the present case, this is not true any more: the sequential test in Theorem 6 is only meaningful from the practical point of view if

$$\min\{\lambda_0, \lambda_1\} \geq 1 + E_{\theta_0} \rho \left(\frac{f_{\theta_1,1}(X_1)}{f_{\theta_0,1}(X_1)} \right),$$

or $A_1 \leq 1 \leq B_1$, in terms of Theorem 5.

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 supported by Mexico's CONACyT Grant No. CB-2005-C01-49854-F.

References

[1] R.H. Berk, Locally most powerful sequentialtests, *Annals of Statistics*, **3** (1975), 373-381.

[2] D.L. Burkholder, R.A. Wijsman, Optimumproperties and admissibility of sequential tests, *Ann. Math. Statist.*, **34** (1963), 1 – 17.

[3] Y.S. Chow, H. Robbins, D. Siegmund, *Great Expectations: The Theory of Optimal Stopping*, Boston, Houghton Mifflin Company (1971).

[4] J. Cochlar, I. Vrana, On the optimum sequential test oftwo hypotheses for statistically dependent observations, *Kybernetika*, **14** (1978), 57-69.

[5] T.S. Ferguson, *Mathematical Statistics: A Decision Theoretic Approach*, New York, Academic Press (1967).

[6] G.H. Hardy, J.E. Littlewood, G. Polya, *Inequalities*, Cambridge University Press (1934).

[7] J. Kiefer, L. Weiss, Some properties of generalized sequential probability ratio tests, *Ann. Math.Statis.*, **28** (1957), 57-75.

[8] E.L. Lehmann, *Testing Statistical Hypotheses*, New York, John Wiley and Sons, Inc., London, Chapman and Hall, Ltd. (1959).

[9] Y. Liu, D. Blostein, Optimality of the sequential probability ratio test for nonstationary observations, *IEEE Transactions on Information Theory*, **28**, No. 1 (1992), 177-182.

[10] G. Lorden, Structure of sequential tests minimizing an expected sample size, *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **51** (1980), 291-302.

[11] A. Novikov, Optimal sequential tests for twosimple hypotheses, *Sequential Analysis*, To Appear.

[12] N. Schmitz, Comment on: "Optimality of the sequential probability ratio testfor nonstationary observations" by Y. Liu and S.D. Blostein, *IEEE Trans. Inform. Theory*, **41**, No. 3 (1995), 861.

[13] A.N. Shiryayev, *Optimal Stopping Rules*, New York, Springer-Verlag (1978).

[14] A. Wald, J. Wolfowitz, Optimum character of the sequential probability ratio test, *Ann. Math. Statistics*, **19** (1948), 326-339.

[15] L. Weiss, On sequential tests which minimize the maximum expected samplesize, *Journal of American Statistical Association*, **57** (1962), 551-566.

[16] R.A. Wijsman, Existence, uniqueness, and monotonicity of sequential probability ratio tests, *Ann. Math. Statist.*, **34** (1963), 1541-1548.