Optimal Sequential Tests for Two Simple Hypotheses

Andrey Novikov[1]

Department of Mathematics, Autonomous Metropolitan University - Iztapalapa, Mexico City, Mexico

Abstract: A general problem of testing two simple hypotheses about the distribution of a discrete-time stochastic process is considered. The main goal is to minimize an average sample number over all sequential tests whose error probabilities do not exceed some prescribed levels. As a criterion of minimization, the average sample number under a third hypothesis is used (modified Kiefer-Weiss problem).

For a class of sequential testing problems, the structure of optimal sequential tests is characterized. An application to the Kiefer-Weiss problem for discrete-time stochastic processes is proposed. As another application, the structure of Bayes sequential tests for two composite hypotheses, with a fixed cost per observation, is given. The results are also applied for finding optimal sequential tests for discrete-time Markov processes. In a particular case of testing two simple hypotheses about a location parameter of an autoregressive process of order 1, it is shown that the sequential probability ratio test has the Wald-Wolfowitz optimality property.

Keywords: Bayesian sequential testing; Composite hypotheses; Dependent observations; Discrete-time Markov process; Discrete-time stochastic process; Kiefer-Weiss problem; Optimal sequential test; Sequential analysis; Sequential hypothesis testing; Sequential probability ratio test; Two simple hypotheses;

Subject Classifications: 62L10; 62L15; 60G40; 62C10

1 INTRODUCTION. PROBLEM SETTING.

Let $X_1, X_2, \dots, X_n, \dots$ be a discrete-time stochastic process with a distribution P. We consider the problem of testing a simple hypothesis $H_0 : P = P_0$ versus a simple alternative $H_1 : P = P_1$. The main goal of this article is to characterize the structure of optimal sequential tests in this problem.

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

$$f_i^n = f_i^n(x_1, x_2, \dots, x_n) \tag{1.1}$$

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

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

[1]Address correspondence to Andrey Novikov, Departamento de Matemáticas, UAM-Iztapalapa, San Rafael Atlixco 186, col. Vicentina, C.P. 09340, México D.F., México; Fax: (52)5558044653; E-mail: an@xanum.uam.mx

with some σ-finite measure μ on the respective space, $i = 0, 1$.

Let us define a (randomized) sequential hypothesis test as a pair (ψ, ϕ) of a *stopping rule* ψ and a *decision rule* ϕ (see, for example, Wald (1950), Ferguson (1967), DeGroot (1970), Ghosh (1970), Ghosh, Mukhopadhyay and Sen (1997)), with the following interpretation.

It is supposed that

$$\psi = (\psi_1, \psi_2, \dots, \psi_n, \dots) \quad \text{and} \quad \phi = (\phi_1, \phi_2, \dots, \phi_n, \dots),$$

where the functions

$$\psi_n = \psi_n(x_1, x_2, \dots, x_n) \quad \text{and} \quad \phi_n = \phi_n(x_1, x_2, \dots, x_n)$$

are supposed to be measurable functions with values in $[0, 1]$, $n = 1, 2, \dots$.

The experiment starts with obtaining the value x_1 of the first observation of X_1 (stage $n = 1$). For any stage $n = 1, 2, \dots$, 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 the experiment came to stage n and that the observations of $(X_1, X_2, \dots, X_n)$ up to this stage 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} of 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 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_ψ (*stopping time*) whose distribution is given by

$$P(\tau_\psi = n) = E(1 - \psi_1)(1 - \psi_2) \dots (1 - \psi_{n-1})\psi_n. \tag{1.2}$$

In (1.2), we suppose that $\psi_n = \psi_n(X_1, X_2, \dots, X_n)$, unlike its previous definition as $\psi_n = \psi_n(x_1, x_2, \dots, x_n)$. We do this intentionally and systematically throughout the paper, applying, generally, for any $F_n = F_n(x_1, x_2, \dots, x_n)$ or $F_n = F_n(X_1, X_2, \dots, X_n)$, the following rule: if F_n is under the probability or expectation sign, then 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 *the type I error probability* as

$$\alpha(\psi, \phi) = P_0(\{\text{reject } H_0\} \cap \{\tau_\psi < \infty\}) = \sum_{n=1}^{\infty} E_0(1 - \psi_1) \dots (1 - \psi_{n-1})\psi_n\phi_n, \tag{1.3}$$

and *the type II error probability* as

$$\beta(\psi, \phi) = P_1(\{\text{accept } H_0\} \cap \{\tau_\psi < \infty\}) = \sum_{n=1}^{\infty} E_1(1-\psi_1) \dots (1-\psi_{n-1})\psi_n(1-\phi_n) \tag{1.4}$$

(here, and in what follows, $E_i(\cdot)$ stands for the expectation with respect to P_i, $i = 0, 1$).

Typically, one would like to keep the error probabilities below some specified levels:

$$\alpha(\psi, \phi) \le \alpha \quad \text{and} \quad \beta(\psi, \phi) \le \beta, \tag{1.5}$$

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

To characterize the number of observations until the final decision is taken, the *average sample number* is used:

$$\mathscr{N}(\psi) = E\tau_\psi. \tag{1.6}$$

If $P(\tau_\psi < \infty) = \sum_{n=1}^{\infty} P(\tau_\psi = n) = 1$, then

$$\mathscr{N}(\psi) = \sum_{n=1}^{\infty} nE(1-\psi_1)\dots(1-\psi_{n-1})\psi_n \tag{1.7}$$

(see (1.2)), otherwise $\mathscr{N}(\psi) = \infty$.

We suppose that the distribution of the process $X_1, X_2, \dots, X_n, \dots$ used for calculating (1.6) does not necessarily match one of the hypothesized distribution P_0 or P_1. Instead, we are using any (fixed) distribution P of the process $X_1, X_2, \dots, X_n, \dots$. Let $f^n = f^n(x_1, x_2, \dots, x_n)$ be the corresponding "density" of $(X_1, X_2, \dots, X_n)$, with respect to μ^n, when the process follows the distribution P, $n = 1, 2, \dots$.

Our main goal is minimizing $\mathscr{N}(\psi)$ over all sequential tests subject to (1.5). Let us refer to this problem as to hypothesis testing problem (P, P_0, P_1).

For independent and identically distributed (i.i.d.) observations, using $P = P_0$ or $P = P_1$ as a criterion of minimization in (1.6) is typical for the problem of sequential hypotheses testing (see Wald and Wolfowitz (1948), Wald (1950), Ferguson (1967), DeGroot (1970), Ghosh (1970), among many others). Minimizing $\mathscr{N}(\psi)$ for P different from both P_0 and P_1 is known as the modified Kiefer-Weiss problem, being the original Kiefer-Weiss problem minimizing $\sup_P \mathscr{N}(\psi)$ under (1.5) (see Kiefer and Weiss (1957), Weiss (1962), Lorden (1980), Schmitz (1993)).

In Section 2, we reduce the problem of minimizing $\mathscr{N}(\psi)$ under constraints (1.5) to a non-constrained minimization problem, with a Lagrange-multiplier function $L(\psi, \phi)$ as an objective function. Then we reduce further the problem of minimization, by finding

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

In Section 3, we solve the problem of minimization of $L(\psi)$, first in the class of truncated stopping rules, then, for a class of problems called truncatable, in the class of all stopping rules. In each case we obtain necessary and sufficient conditions of optimality.

In Section 4, we give some applications of our results to some general problems for discrete-time stochastic processes: to the Kiefer-Weiss problem, and to its modified version, to the Bayesian sequential testing of two composite hypotheses, and to optimal sequential tests for discrete-time Markov processes.

In Section 5, we lay down the proofs of the main results.

2 REDUCTION TO AN OPTIMAL STOPPING PROBLEM

To proceed with minimizing (1.6) over the tests subject to (1.5) let us define the following Lagrange-multiplier function:

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

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

The following theorem is nothing else than an application of the Lagrange multiplier method to the conditional problem above.

Theorem 2.1. *Let Δ be some class of sequential tests. Let exist $\lambda_0 > 0$ and $\lambda_1 > 0$ and a test $(\psi, \phi) \in \Delta$ such that $L(\psi, \phi; \lambda_0, \lambda_1) < \infty$, and such that for any other test $(\psi', \phi') \in \Delta$*

$$L(\psi, \phi; \lambda_0, \lambda_1) \leq L(\psi', \phi'; \lambda_0, \lambda_1) \tag{2.2}$$

holds and such that

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

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

$$\alpha(\psi', \phi') \leq \alpha \quad \text{and} \quad \beta(\psi', \phi') \leq \beta \tag{2.4}$$

it holds

$$\mathscr{N}(\psi) \leq \mathscr{N}(\psi'). \tag{2.5}$$

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

Proof. It is quite straightforward:

Let $(\psi', \phi') \in \Delta$ be any test satisfying (2.4). Because of (2.3) and (2.2),

$$\begin{aligned} \mathscr{N}(\psi) + \lambda_0 \alpha + \lambda_1 \beta &= \mathscr{N}(\psi) + \lambda_0 \alpha(\psi, \phi) + \lambda_1 \beta(\psi, \phi) && (2.6) \\ &\leq \mathscr{N}(\psi') + \lambda_0 \alpha(\psi', \phi') + \lambda_1 \beta(\psi', \phi') && (2.7) \\ &\leq \mathscr{N}(\psi') + \lambda_0 \alpha + \lambda_1 \beta, && (2.8) \end{aligned}$$

where to get the last inequality we used (2.4).

It follows from (2.6) – (2.8) that

$$\mathscr{N}(\psi) \leq \mathscr{N}(\psi').$$

The get the last statement of the theorem we note that if $\mathscr{N}(\psi) = \mathscr{N}(\psi')$ then there are equalities in (2.7) – (2.8) instead of the inequalities which is only possible if $\alpha(\psi', \phi') = \alpha$ and $\beta(\psi', \phi') = \beta$. □

Remark 2.1. *In fact, the method of Lagrange multipliers is used (sometimes implicitly) in various hypotheses testing problems, e.g., in Kiefer and Weiss (1957), Weiss (1962), Berk (1975), Lorden (1980), Castillo and García (1983), Müller-Funk et al. (1985), Schmitz (1993). In essence, the method of Lagrange multipliers is used in Lehmann (1959) in the proof of the fundamental lemma of Neyman-Pearson, where, to minimize $\beta(\psi, \phi)$ among all (non-sequential) tests such that $\alpha(\psi, \phi) \leq \alpha$, the minimum of $\beta(\psi, \phi) + \lambda \alpha(\psi, \phi)$ is found.*

In a way, the Bayesian approach in hypotheses testing can be considered as a variant of the Lagrange multipliers method as well. If, in particular, $P = \pi P_0 + (1-\pi) P_1$ in our definition (1.6) above, with some π, $0 < \pi < 1$, then the Lagrange-multiplier function (2.1) is nothing else than the Bayesian risk (see Wald and Wolfowitz (1948), Wald (1950), Lehmann (1959), Ferguson (1967), DeGroot (1970), Zacks (1971), Shiryayev (1978), Cochlar and Vrana (1978), Liu and Blostein (1992), Ghosh, Mukhopadhyay and Sen (1997), among many others), and the classical proofs of the optimality of the sequential probability ratio test (SPRT) for i.i.d. observations are using arguments of type of Theorem 2.1 (see, e.g., Wald and Wolfowitz (1948), Lehmann (1959), Ferguson (1967)).

Because of Theorem 2.1, from now on, our main focus will be on the unrestricted minimization of $L(\psi, \phi) = L(\psi, \phi; \lambda_0, \lambda_1)$, over all sequential tests, for any $\lambda_0 > 0$ and $\lambda_1 > 0$.

For any stopping rule $\psi = (\psi_1, \psi_2, \dots)$ let us denote

$$s_n^\psi = (1-\psi_1)\dots(1-\psi_{n-1})\psi_n \quad \text{and} \quad c_n^\psi = (1-\psi_1)\dots(1-\psi_{n-1}), \tag{2.9}$$

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

Let I_A be the indicator function of the event A.

The following theorem, in a rather standard way, lets us find optimal decision rules for any given stopping rule ψ.

Theorem 2.2. *For any $\lambda_0 \geq 0$ and $\lambda_1 \geq 0$ and for any sequential test (ψ, ϕ)*

$$\lambda_0 \alpha(\psi, \phi) + \lambda_1 \beta(\psi, \phi) \geq \sum_{n=1}^{\infty} \int s_n^\psi \min\{\lambda_0 f_0^n, \lambda_1 f_1^n\} d\mu^n \tag{2.10}$$

with an equality if and only if

$$I_{\{\lambda_0 f_0^n < \lambda_1 f_1^n\}} \leq \phi_n \leq I_{\{\lambda_0 f_0^n \leq \lambda_1 f_1^n\}} \tag{2.11}$$

μ^n-almost everywhere on $S_n^\psi = \big\{(x_1, \dots, x_n) : s_n^\psi(x_1, \dots, x_n) > 0\big\}$ for any $n = 1, 2, \dots$.

The proof of Theorem 2.2 can be found in Appendix.

Let us denote

$$L(\psi) = L(\psi; \lambda_0, \lambda_1) = \inf_\phi L(\psi, \phi; \lambda_0, \lambda_1).$$

Corollary 2.1. *If $P(\tau_\psi < \infty) = 1$, then*

$$L(\psi) = \sum_{n=1}^{\infty} \int s_n^\psi (n f^n + l_n) d\mu^n, \tag{2.12}$$

where, by definition,

$$l_n = \min\{\lambda_0 f_0^n, \lambda_1 f_1^n\}. \tag{2.13}$$

If $P(\tau_\psi < \infty) < 1$ then $L(\psi) = \infty$.

Proof. This follows from Theorem 2.2 by (2.1), in view of (1.7). $\square$

Remark 2.2. *Theorem 2.2 establishes the form of optimal decision rules, which turn out to be Bayesian. It is essentially Theorem 5.2.1 in Ghosh, Mukhopadhyay and Sen (1997) applied to the natural 0-1 loss function (see also Cochlar and Vrana (1978)).*

By Theorem 2.2, the problem of minimization of $L(\psi, \phi; \lambda_0, \lambda_1)$ is reduced now to the problem of minimization of $L(\psi; \lambda_0, \lambda_1)$, that is, to an optimal stopping problem. Indeed, if there is a ψ such that

$$L(\psi; \lambda_0, \lambda_1) = \inf_{\psi'} L(\psi'; \lambda_0, \lambda_1),$$

then, adding to ψ any decision rule ϕ satisfying (2.11), by Theorem 2.2 we have that for any sequential test (ψ', ϕ'):

$$L(\psi, \phi; \lambda_0, \lambda_1) = L(\psi; \lambda_0, \lambda_1) \leq L(\psi'; \lambda_0, \lambda_1) \leq L(\psi', \phi'; \lambda_0, \lambda_1). \tag{2.14}$$

In particular, in this way we obtain tests (ψ, ϕ) satisfying (2.2), which is crucial for solving the original conditional problem (see Theorem 2.1).

3 OPTIMAL STOPPING RULES

In this section, we characterize the structure of stopping rules minimizing $L(\psi)$, first in the class of truncated stopping rules, then in the class of all stopping rules, supposing that the hypothesis testing problem is truncatable.

3.1 Optimal Truncated Stopping Rules

Here we solve the problem of minimization of $L(\psi)$ in the class of truncated stopping rules, that is, in the class $\mathscr{D}^N$ of

$$\psi = (\psi_1, \psi_2, \dots, \psi_{N-1}, 1, \dots). \tag{3.1}$$

The following lemma takes over a large part of work of doing this.

Lemma 3.1. *Let $k \geq 1$ be any natural number, and let $v_{k+1} = v_{k+1}(x_1, x_2, \dots, x_{k+1})$ be any non-negative measurable function. Then*

$$\sum_{n=1}^{k} \int s_n^\psi (nf^n + l_n) d\mu^n + \int c_{k+1}^\psi \left((k+1) f^{k+1} + v_{k+1} \right) d\mu^{k+1}$$

$$\geq \sum_{n=1}^{k-1} \int s_n^\psi (nf^n + l_n) d\mu^n + \int c_k^\psi \left(kf^k + v_k \right) d\mu^k, \tag{3.2}$$

where

$$v_k = \min\{l_k, f^k + R_k\}, \tag{3.3}$$

being

$$R_k = R_k(x_1, \dots, x_k) = \int v_{k+1}(x_1, \dots, x_{k+1}) d\mu(x_{k+1}).$$

There is an equality in (3.2) if and only if

$$I_{\{l_k < f^k + R_k\}} \leq \psi_k \leq I_{\{l_k \leq f^k + R_k\}} \tag{3.4}$$

μ^k-almost everywhere on $C_k^\psi = \{(x_1, \dots, x_k) : c_k^\psi (x_1, \dots, x_{k-1}) > 0\}$.

We lay down the proof of Lemma 3.1 in Appendix.

Applying Lemma 3.1 to the Lagrange-multiplier function

$$L_N(\psi) = \sum_{n=1}^{N-1} \int s_n^\psi (nf^n + l_n) d\mu^n + \int c_N^\psi \left(Nf^N + l_N \right) d\mu^r \tag{3.5}$$

of any truncated stopping rule $\psi \in \mathscr{D}^N$, consecutively for $k = N-1, N-2, \dots, r$, we easily have

Theorem 3.1. *Let $\psi \in \mathscr{D}^N$ be any (truncated) stopping rule. Then for any $1 \leq r \leq N-1$ the following inequalities hold true*

$$L_N(\psi) \geq \sum_{n=1}^{r} \int s_n^\psi (nf^n + l_n) d\mu^n + \int c_{r+1}^\psi \left((r+1) f^{r+1} + V_{r+1}^N \right) d\mu^{r+1} \tag{3.6}$$

$$\geq \sum_{n=1}^{r-1} \int s_n^\psi (nf^n + l_n) d\mu^n + \int c_r^\psi \left(rf^r + V_r^N \right) d\mu^r, \tag{3.7}$$

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

$$V_k^N = \min\{l_k, f^k + R_k^N\}, \tag{3.8}$$

being

$$R_k^N = R_k^N(x_1, \dots, x_k) = \int V_{k+1}^N(x_1, \dots, x_{k+1}) d\mu(x_{k+1}). \tag{3.9}$$

The lower bound in (3.7) is attained if and only if

$$I_{\{l_k < f^k + R_k^N\}} \leq \psi_k \leq I_{\{l_k \leq f^k + R_k^N\}} \tag{3.10}$$

μ^k*-almost everywhere on* C_k^ψ*, for any* $k = r, r+1, \dots, N-1$.

In particular, for $r = 1$ we have

Corollary 3.1. *For any* $\psi \in \mathscr{D}^N$

$$L(\psi) \geq 1 + R_0^N, \tag{3.11}$$

where

$$R_0^N = \int V_1^N(x_1) d\mu(x_1).$$

There is an equality in (3.11) if and only if ψ_k *satisfy (3.10)* μ^k*-almost everywhere on* C_k^ψ*, for any* $k = 1, 2, \dots, N-1$.

Remark 3.1. *For the Bayesian problem (when* $P = \pi P_0 + (1-\pi) P_1$ *with* $\pi \in (0,1)$*), an optimal truncated (non-randomized) stopping rule can also be obtained from Theorem 5.2.2 in Ghosh, Mukhopadhyay and Sen (1997).*

Remark 3.2. *Despite that any* ψ *such that* $L(\psi) = 1 + R_0^N$*, by Theorem 3.1, is optimal in the class* $\mathscr{D}^N$ *of all truncated rules, it only makes practical sense if*

$$\min\{\lambda_0, \lambda_1\} > 1 + R_0^N.$$

The reason is that $l_0 = \min\{\lambda_0, \lambda_1\}$ *can be considered as "the* $L(\psi)$*" function for a trivial sequential test* (ψ_0, ϕ_0) *which, without taking any observations, makes the decision* $\phi_0 = I_{\{\lambda_0 \leq \lambda_1\}}$*. In this case there are no observations (*$\mathscr{N}(\psi_0) = 0$*) and it is easily seen that*

$$L(\psi_0, \phi_0) = \lambda_0 \alpha(\psi_0, \phi_0) + \lambda_1 \beta(\psi_0, \phi_0) = \min\{\lambda_0, \lambda_1\} = l_0.$$

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* $\mathscr{D}^N$*.*

3.2 Optimal Non-Truncated Stopping Rules

In this section we characterize the structure of general sequential tests minimizing $L(\psi) = L(\psi; \lambda_0, \lambda_1)$ (see (2.1)).

Let us define for any stopping rule ψ, and for any natural $N \geq 1$,

$$L_N(\psi) = L_N(\psi; \lambda_0, \lambda_1) = L(\psi^N; \lambda_0, \lambda_1),$$

where $\psi^N = (\psi_1, \psi_2, \dots, \psi_{N-1}, 1, \dots)$ is the rule ψ truncated at N.

By (3.5),

$$L_N(\psi) = \sum_{n=1}^{N-1} \int s_n^\psi (nf^n + l_n) d\mu^n + \int c_N^\psi \left(Nf^N + l_N \right) d\mu^N. \tag{3.12}$$

Because ψ^N is truncated, the results of the preceding section apply, in particular, the inequalities of Theorem 3.1.

The idea of what follows is to make $N \to \infty$, to obtain some lower bounds for $L(\psi)$ from (3.6) - (3.7). To do this, we need some conditions which guarantee that $L_N(\psi) \to L(\psi)$ as $N \to \infty$.

Let us call a hypothesis testing problem (P, P_0, P_1) *truncatable* if for any $\lambda_0 > 0$ and $\lambda_1 > 0$ it holds

$$\lim_{N\to\infty} L_N(\psi; \lambda_0, \lambda_1) = L(\psi; \lambda_0, \lambda_1) \tag{3.13}$$

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

The following lemma characterizes the property of being a problem truncatable.

Lemma 3.2. *The hypothesis testing problem (P, P_0, P_1) is truncatable if and only if for any stopping rule ψ, such that $E\tau_\psi < \infty$, for any $\lambda_0 > 0$, $\lambda_1 > 0$ it holds*

$$\lim_{n\to\infty} \int c_n^\psi \min\{\lambda_0 f_0^n, \lambda_1 f_1^n\} d\mu^n = 0. \tag{3.14}$$

The proof of Lemma 3.2 can be found in Appendix.

Corollary 3.2. *The hypothesis testing problem (P, P_0, P_1) is truncatable if any of the following three conditions is satisfied:*

i) for any stopping rule ψ, from $E\tau_\psi < \infty$ it follows that

$$P_0(\tau_\psi < \infty) = 1 \quad \textit{or} \quad P_1(\tau_\psi < \infty) = 1, \tag{3.15}$$

ii) for some $\lambda_0 > 0$ and $\lambda_1 > 0$

$$\lim_{n\to\infty} \int \min\{\lambda_0 f_0^n, \lambda_1 f_1^n\} d\mu^n = 0.$$

iii)

$$\frac{f_1^n(X_1, \ldots, X_n)}{f_0^n(X_1, \ldots, X_n)} \to 0, \quad \textit{as} \quad n \to \infty,$$

in P_0-probability.

Because of its technical character, we lay down the proof of Corollary 3.2 in Appendix.

Remark 3.3. *In fact, conditions* ii) *and* iii) *of Corollary 3.2 are equivalent. It is easily seen from its proof.*

Corollary 3.3. *Any Bayesian hypotheses testing problem $((\pi P_0 + (1 - \pi)P_1, P_0, P_1)$, $\pi \in (0, 1))$ is truncatable.*

Proof. In this case $f^n = \pi f_0^n + (1 - \pi) f_1^n$ with some $\pi \in (0, 1)$. Thus,

$$\mathcal{N}(\psi) = \pi E_0 \tau_\psi + (1 - \pi) E_1 \tau_\psi < \infty$$

implies that (3.15) is satisfied. By Corollary 3.2, it follows that the problem is truncatable. □

Corollary 3.4. *Any of the hypotheses testing problems* (P_0, P_0, P_1) *and* (P_1, P_0, P_1) *is truncatable.*

Proof. Indeed, if, for example, $E_0 \tau_\psi < \infty$, then (3.15) is trivial, so, by Corollary 3.2, the problem (P_0, P_0, P_1) is truncatable. □

The second fact we need for passing to the limit in (3.6) – (3.7) is about the behaviour of the functions V_k^N defined by (3.8).

Lemma 3.3. *For any* $k \geq 1$ *and for any* $N \geq k$

$$V_k^N \geq V_k^{N+1}. \tag{3.16}$$

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

Let $k = N$. Then by (3.8)

$$V_N^{N+1} = \min\{l_N, f^N + \int V_{N+1}^{N+1} d\mu(x_{N+1})\} \leq l_N = V_N^N.$$

If we suppose that (3.16) is satisfied for some k, $N \geq k > 1$, then

$$\begin{aligned} V_{k-1}^N &= \min\{l_{k-1}, f^{k-1} + \int V_k^N d\mu(x_k)\} \\ &\geq \min\{l_{k-1}, f^{k-1} + \int V_k^{N+1} d\mu(x_k)\} = V_{k-1}^{N+1}. \end{aligned}$$

Thus, (3.16) is satisfied for $k-1$ as well, which completes the induction. □

It follows from Lemma 3.3 that for any fixed $k \geq 1$ the sequence V_k^N is non-increasing. So, there exists

$$V_k = \lim_{N \to \infty} V_k^N. \tag{3.17}$$

Passing to the limit, as $N \to \infty$, in (3.6) – (3.7), we get

Lemma 3.4. *If the hypothesis testing problem* (P, P_0, P_1) *is truncatable, then for any stopping rule* ψ *and for any* $r \geq 1$ *the following inequalities hold*

$$L(\psi) \geq \sum_{n=1}^{r} \int s_n^\psi (n f^n + l_n) d\mu^n + \int c_{r+1}^\psi \left((r+1) f^{r+1} + V_{r+1}\right) d\mu^{r+1} \tag{3.18}$$

$$\geq \sum_{n=1}^{r-1} \int s_n^\psi (n f^n + l_n) d\mu^n + \int c_r^\psi \left(r f^r + V_r\right) d\mu^r, \tag{3.19}$$

where

$$V_k = \min\{l_k, f^k + R_k\}, \tag{3.20}$$

with

$$V_k = \lim_{N \to \infty} V_k^N, \quad k = 1, 2, 3, \ldots$$

and

$$R_k = R_k(x_1, x_2, \ldots, x_k) = \int V_{k+1}(x_1, \ldots, x_{k+1}) d\mu(x_{k+1}), \quad k = 0, 1, 2, \ldots \tag{3.21}$$

In particular, for any stopping rule ψ

$$L(\psi) \geq 1 + R_0. \tag{3.22}$$

Proof. The left-hand side of (3.6) tends to the left-hand side of (3.18), as $N \to \infty$, because the problem is truncatable. The right-hand side of (3.6) tends to the right-hand side of (3.18), as $N \to \infty$, by the monotone convergence theorem, in view of Lemma 3.3. The right-hand side of (3.7) tends to the right-hand side of (3.19) by the same reason. At last, passing to the limit, as $N \to \infty$, on the both sides of (3.8) gives us (3.20). □

The following lemma shows that, for a truncatable testing problem, the lower bound in (3.22) is, in fact, the infimum value of the left-hand side of (3.22).

Lemma 3.5. *If the hypothesis testing problem* (P, P_0, P_1) *is truncatable, then*

$$\inf_{\psi} L(\psi) = 1 + R_0. \tag{3.23}$$

The proof of Lemma 3.5 is laid down in Appendix.

Remark 3.4. *Again (see Remark 3.1), for Bayesian problems Lemma 3.5 can be obtained from Theorem 5.2.3 in Ghosh, Mukhopadhyay and Sen (1997), applying it to the natural 0-1 loss function.*

At last, the following theorem characterizes all optimal stopping rules in the hypothesis testing problem (P, P_1, P_2), if this is truncatable.

Theorem 3.2. *Let the testing problem* (P, P_0, P_1) *be truncatable. Then*

$$L(\psi) = \inf_{\psi'} L(\psi') \tag{3.24}$$

if and only if

$$I_{\{l_k < f^k + R_k\}} \leq \psi_k \leq I_{\{l_k \leq f^k + R_k\}} \tag{3.25}$$

μ^k*-almost everywhere on* C_k^{ψ}*, for any* $k = 1, 2, \ldots$.

The proof of Theorem 3.2 can be found in Appendix.

Remark 3.5. *Once again (see Remark 3.2), the optimal stopping rule* ψ *from Theorem 3.2 only makes sense if* $l_0 > 1 + \int V_1 d\mu(x_1)$*, because otherwise the trivial rule which does not take any observations, gives a lesser value* (l_0) *than* $L(\psi)$*.*

Remark 3.6. *In a particular case of Bayesian hypothesis testing, when* $P = \pi P_0 + (1-\pi)P_1$ *with some* $\pi \in (0,1)$*, Theorem 3.2 gives all optimal stopping rules for the Bayesian problem considered in Cochlar and Vrana (1978). To prove the existence of the optimal stopping rule, Cochlar and Vrana (1978) used the general theory of optimal stopping (see, for example, Chow et al. (1971) or Shiryayev (1978)). Our direct treatment of a (more general) stopping problem here gives a more specific structure of optimal stopping rules in the Bayesian problem considered in Cochlar and Vrana (1978).*

In the case of independent observations, Liu and Blostein (1992) give the structure of a Bayesian stopping rule for the problem of testing a simple hypothesis against a simple alternative. Their treatment of the problem is based on a direct approach of Ferguson (1967). Our Theorem 3.2, based on the same principles (see also Ghosh, Mukhopadhyay and Sen (1997)), allows to characterize the structure of all Bayesian stopping rules in this problem (see Novikov (2008a) for details). The solution of the *(conditional) problem of minimizing the average sample number under restrictions on the type I and type II error probabilities, was not so successful in Liu and Blostein (1992) (see Schmitz (1995)). Using our approach, we were able to characterize the structure of optimal sequential tests in the conditional problem as well (see Novikov (2008a)).*

If the problem (P, P_0, P_1) is not truncatable, we still can use the results obtained above, restricting the class of stopping rules to the class $\mathscr{F}$ of *truncatable* stopping rules ψ, i.e. such that $\lim_{N\to\infty} L_N(\psi) = L(\psi)$ whenever $P(\tau_\psi < \infty) = 1$. It is easy to see that for this class

$$\inf_{\psi\in\mathscr{F}} L(\psi) = 1 + R_0$$

still holds (see the proof of Lemma 3.5).

So a variant of Theorem 3.2 can be formulated as follows.

Theorem 3.3. *If there is a $\psi \in \mathscr{F}$ such that*

$$L(\psi) = \inf_{\psi'\in\mathscr{F}} L(\psi') \tag{3.26}$$

then

$$I_{\{l_k<f^k+R_k\}} \leq \psi_k \leq I_{\{l_k\leq f^k+R_k\}} \tag{3.27}$$

μ^k-almost everywhere on C_k^ψ, for any $k = 1, 2, \dots$.

On the other hand, if a stopping rule ψ satisfies (3.27) μ^k-almost everywhere on C_k^ψ, for any $k = 1, 2, \dots$, and $\psi \in \mathscr{F}$, then for this ψ (3.26) holds.

Thus, if there exists an optimal stopping rule in $\mathscr{F}$, then it can be found by (3.27) for $k = 1, 2, \dots$. And if a stopping rule ψ satisfying (3.27), μ^k-almost everywhere on C_k^ψ, for $k = 1, 2, \dots$, fails to belong to $\mathscr{F}$, then there is no optimal stopping rule in $\mathscr{F}$. It is worth noting that even in this case, we can find a sequence of truncated stopping rules $\psi_N \in \mathscr{D}^N$, by Corollary 3.1, such that $L(\psi_N) = 1 + R_0^N \to 1 + R_0 = \inf_{\psi'\in\mathscr{F}} L(\psi')$, as $N \to \infty$.

Remark 3.7. *The class $\mathscr{F}$ defined above is the largest class of stopping rules for which our method works. For practical purposes, however, a more restricted class of stopping rules may seem statistically more attractive, even when the problem is truncatable. For example, let $\mathscr{F}_1$ be defined as the class of all stopping rules such that $P_0(\tau_\psi < \infty) = 1$* and *$P_1(\tau_\psi < \infty) = 1$. Then, obviously, with $\mathscr{F}_1$ instead of $\mathscr{F}$, Theorem 3.3 still holds. However, this makes harder to comply with the requirement that $\psi \in \mathscr{F}_1$, in order that the optimal stopping rule exist. For example, for i.i.d. observations, in Hawix and Schmitz (1998) we find an example of a (non-randomized) test with a stopping rule ψ satisfying (3.27) for any $k = 1, 2, \dots$, but such that $\psi \notin \mathscr{F}_1$, meaning that there is no optimal stopping rule in $\mathscr{F}_1$.*

4 APPLICATIONS

4.1 Modified Kiefer-Weiss Problem

In this section, we provide a solution, to the problem of minimizing $E\tau_\psi$ in the class of all sequential tests (ψ, ϕ) with error probabilities not exceeding some given levels (see Introduction).

Combining Theorems 2.1, 2.2 and 3.2 we immediately have the following

Theorem 4.1. *Let the hypothesis testing problem (P, P_0, P_1) be truncatable. Let $\lambda_0 > 0$ and $\lambda_1 > 0$ be any numbers and ψ such that for any $n = 1, 2, \ldots$*

$$I_{\{l_n < f^n + R_n\}} \leq \psi_n \leq I_{\{l_n \leq f^n + R_n\}} \tag{4.1}$$

μ^n-almost everywhere on C_n^ψ, where l_n is defined in (2.13), and R_n is defined in (3.21), and any ϕ such that

$$I_{\{\lambda_0 f_0^n < \lambda_1 f_1^n\}} \leq \phi_n \leq I_{\{\lambda_0 f_0^n \leq \lambda_1 f_1^n\}} \tag{4.2}$$

μ^n-almost everywhere on S_n^ψ, $n = 1, 2, \ldots$

Then for any (ψ', ϕ') such that

$$\alpha(\psi', \phi') \leq \alpha(\psi, \phi) \quad \textit{and} \quad \beta(\psi', \phi') \leq \beta(\psi, \phi) \tag{4.3}$$

it holds

$$E\tau_\psi \leq E\tau_{\psi'}. \tag{4.4}$$

The inequality in (4.4) is strict if at least one of the inequalities in (4.3) is strict.

If there are equalities in all of the inequalities in (4.3) and (4.4) then ψ' satisfies (4.1) (with ψ_n' instead of ψ_n) μ^n-almost everywhere on $C_n^{\psi'}$ for any $n = 1, 2, \ldots$, and ϕ' satisfies (4.2) (with ϕ_n' instead of ϕ_n) μ^n-almost everywhere on $S_n^{\psi'}$ for any $n = 1, 2, \ldots$.

4.2 Kiefer-Weiss Problem

In this section, we generalize the problem of Kiefer and Weiss (1957) to general discrete-time stochastic processes and propose a method for solving it, based on our results above.

Let the distributions of the stochastic process we observe, $X_1, X_2, \ldots, X_n, \ldots$ be defined by a parametric family of joint density functions $\{f_\theta^n(x_1, \ldots, x_n), \theta \in \Theta, n \geq 1\}$. We suppose that $f_\theta^n(x_1, \ldots, x_n)$ is measurable with respect to $(\theta, x_1, \ldots, x_n)$ for any $n \geq 1$.

The Kiefer-Weiss problem is to minimize

$$\sup_{\theta \in \Theta} E_\theta \tau_\psi$$

in the class of all sequential tests (ψ, ϕ) such that

$$\alpha(\psi, \phi) = \sum_{n=1}^{\infty} E_{\theta_0} s_n^\psi \phi_n \leq \alpha \quad \text{and} \quad \beta(\psi, \phi) = \sum_{n=1}^{\infty} E_{\theta_1} s_n^\psi (1 - \phi_n) \leq \beta,$$

where $\alpha, \beta \in [0, 1)$ are some constants.

Applying the well-known arguments relating Bayesian and minimax methods, we get the following

Theorem 4.2. *Let $\lambda_0 > 0$, $\lambda_1 > 0$ be any constants, and let there exist a probability measure π such that*

$$\inf_{\psi', \phi'} \left(\int_\Theta E_\theta \tau_{\psi'} d\pi(\theta) + \lambda_0 \alpha(\psi', \phi') + \lambda_1 \beta(\psi', \phi') \right) \tag{4.5}$$

is attained at some test (ψ, ϕ) for which

$$\sup_{\theta \in \Theta} E_\theta \tau_\psi = \int_\Theta E_\theta \tau_\psi d\pi(\theta).$$

Then for any (ψ', ϕ') such that

$$\alpha(\psi', \phi') \leq \alpha(\psi, \phi) \quad \textit{and} \quad \beta(\psi', \phi') \leq \beta(\psi, \phi) \tag{4.6}$$

$$\sup_{\theta \in \Theta} E_\theta \tau_\psi \leq \sup_{\theta \in \Theta} E_\theta \tau_{\psi'}. \tag{4.7}$$

If there is an equality in (4.7), then (ψ', ϕ') attains the infimum (4.5) as well, and, additionally,

$$\sup_{\theta \in \Theta} E_\theta \tau_{\psi'} = \int_\Theta E_\theta \tau_{\psi'} d\pi(\theta).$$

Proof. It is very similar to that of Theorem 2.1 and is omitted here. □

Remark 4.1. *We can apply Theorem 3.2 when seeking for tests attaining (4.5). Indeed, if we define $P(\cdot) = \int P_\theta(\cdot) d\pi(\theta)$, then Theorem 3.2, applied to the hypothesis testing problem $(P, P_{\theta_0}, P_{\theta_1})$, gives all the tests attaining (4.5), provided that the problem is truncatable.*

Remark 4.2. *In the particular case of one-point distribution π in Theorem 4.2, we can apply Theorem 3.2 to the testing problem $(P_\theta, P_{\theta_0}, P_{\theta_1})$ directly, for finding tests that attain (4.5). In this case, the objective function to be minimized,*

$$E_\theta \tau_\psi + \lambda_0 \alpha(\psi, \phi) + \lambda_1 \beta(\psi, \phi), \tag{4.8}$$

is nothing else than the Lagrange-multiplier function for the modified Kiefer-Weiss problem. For i.i.d. observations, essentially this was used by Weiss (1962) for solving the original Kiefer-Weiss problem in some particular cases. He finds a test (ψ, ϕ) minimizing (4.8), with some specific $\theta \in \Theta$, and such that

$$\sup_{\theta' \in \Theta} E_{\theta'} \tau_\psi = E_\theta \tau_\psi.$$

Thus, the examples of Weiss (1962) may serve as examples of application of Theorem 4.2, where for his construction $\pi(\theta) = 1$ should be taken.

Remark 4.3. *By Corollary 3.2, both $(P, P_{\theta_0}, P_{\theta_1})$ in Remark 4.1 and $(P_\theta, P_{\theta_0}, P_{\theta_1})$ in Remark 4.2 are truncatable if*

$$\int \min\{\lambda_0 f_{\theta_0}^n, \lambda_1 f_{\theta_1}^n\} d\mu^n = E_{\theta_0} \min\left\{\lambda_0, \lambda_1 \frac{f_{\theta_1}^n}{f_{\theta_0}^n}\right\} \to 0, \quad \textit{as} \quad n \to \infty \tag{4.9}$$

for some $\lambda_0 > 0$ and $\lambda_1 > 0$ (see condition ii) of the Corollary). By Theorem 2.2, the left-hand side of (4.9) is the minimum value of the weighted sum of error probabilities, when the (non-sequential) tests are based on the first n observations. Because of this, it should be expected that it approaches zero, as $n \to \infty$, for any reasonable statistical hypothesis testing problem. In particular, for i.i.d. observations this is always the case if

$$\mu\{x : f_{\theta_0}^1(x) \neq f_{\theta_1}^1(x)\} > 0$$

(see Novikov (2008a)).

4.3 Bayes Sequential Testing of Two Composite Hypotheses

In this section, we apply our main Theorem 3.2 to Bayesian testing of two composite hypotheses.

As in preceding section, let us suppose that we have a parametric family of joint "density" functions $\{f_\theta(x_1,\dots,x_n), \theta \in \Theta, n \geq 1\}$ where it is supposed that any $f_\theta(x_1,\dots,x_n)$ is measurable with respect to $(\theta, x_1,\dots,x_n)$, $n = 1, 2, \dots$. We consider in this section the problem of sequential testing $H_0 : \theta \in \Theta_0$ vs. $H_1 : \theta \in \Theta_1$, where Θ_0 and Θ_1 are some subsets of Θ such that $\Theta_0 \cap \Theta_1 = \emptyset$.

Let us suppose that there exists a prior distribution π on Θ.

For any sequential test (ψ, ϕ) the Bayes risk is defined as

$$w(\psi,\phi) = c\int_\Theta E_\theta \tau_\psi d\pi(\theta) + a\int_{\Theta_0} \alpha_\theta(\psi,\phi) d\pi(\theta) + b\int_{\Theta_1} \beta_\theta(\psi,\phi) d\pi(\theta). \tag{4.10}$$

where

$$\alpha_\theta(\psi,\phi) = \sum_{n=1}^\infty E_\theta s_n^\psi \phi_n \quad \text{and} \quad \beta_\theta(\psi,\phi) = \sum_{n=1}^\infty E_\theta s_n^\psi (1-\phi_n)$$

for $\theta \in \Theta$, being $c > 0$ some unitary observation cost, and $a \geq 0$ and $b \geq 0$ some losses due to incorrect decisions.

We call a test (ψ, ϕ) Bayesian (for a given π) if its Bayesian risk (4.10) is a minimum among all sequential tests. In this section we show how to find Bayesian sequential tests for any prior distribution π.

Using the Fubini theorem, it is easy to see that

$$\int_{\Theta_0} \alpha_\theta(\psi,\phi) d\pi(\theta) = \sum_{n=1}^\infty \int s_n^\psi \phi_n \int_{\Theta_0} f_\theta^n d\pi(\theta) d\mu^n = \pi_0 \sum_{n=1}^\infty \int s_n^\psi \phi_n g_0^n d\mu^n, \tag{4.11}$$

where $g_0^n \equiv \int_{\Theta_0} f_\theta^n d\pi(\theta)/\pi_0$, $\pi_0 = \pi(\Theta_0)$, and

$$\int_{\Theta_1} \beta_\theta(\psi,\phi) d\pi(\theta) = \sum_{n=1}^\infty \int s_n^\psi (1-\phi_n) \int_{\Theta_1} f_\theta^n d\pi(\theta) d\mu^n = \pi_1 \sum_{n=1}^\infty \int s_n^\psi (1-\phi_n) g_1^n d\mu^n, \tag{4.12}$$

where $g_1^n \equiv \int_{\Theta_1} f_\theta^n d\pi(\theta)/\pi_1$, $\pi_1 = \pi(\Theta_1)$, and that

$$\int_\Theta E_\theta \tau_\psi d\pi(\theta) = \pi_0 \sum_{n=1}^\infty n \int s_n^\psi g_0^n d\mu^n + \pi_1 \sum_{n=1}^\infty n \int s_n^\psi g_1^n d\mu^n$$

Therefore, the Bayes risk (4.10) is equivalent to

$$w(\psi,\phi) = \sum_{n=1}^\infty \int s_n^\psi \left(cng^n + \phi_n a\pi_0 g_0^n + (1-\phi_n) b\pi_1 g_1^n\right) d\mu^n, \tag{4.13}$$

where $g^n = \pi_0 g_0^n + \pi_1 g_1^n = \int_\Theta f_\theta^n d\pi(\theta)$. Thus, $w(\psi,\phi)/c$ has the same structure as $L(\psi,\phi)$ in (2.1), so we can use all the theory of Section 2 - Section 3 for finding tests (ψ,ϕ) minimizing $w(\psi,\phi)$, i.e. Bayesian sequential tests.

To formulate the main theorem (Theorem 3.2) in this Bayesian context, we need to re-define the key elements for the structure of optimal tests.

Let

$$l_N = \min\{\pi_0 a g_0^N, \pi_1 b g_1^N\}, \quad N = 1, 2, \dots. \tag{4.14}$$

Define

$$V_N^N = l_N \tag{4.15}$$

and, recursively,

$$V_k^N = \min\{l_k, c g^k + \int V_{k+1}^N d\mu(x_{k+1})\} \tag{4.16}$$

for any $k = N-1, N-2, \dots, 1$.

Let further

$$V_k = \lim_{N\to\infty} V_k^N \tag{4.17}$$

for any $k = 1, 2, \dots$, and

$$R_k = R_k(x_1, \dots, x_k) = \int V_{k+1}(x_1, \dots, x_{k+1}) d\mu(x_{k+1}), \tag{4.18}$$

for $k = 0, 1, \dots$. Theorem 3.2, with the help of Theorem 2.2, transforms now to

Theorem 4.3. *Let l_k be defined by (4.14) and R_k by (4.18) by means of (4.17) and (4.15) and (4.16). Then*

$$w(\psi, \phi) = \inf_{\psi', \phi'} w(\psi', \phi') \tag{4.19}$$

if and only if

$$I_{\{l_k < c g^k + R_k\}} \le \psi_k \le I_{\{l_k \le c g^k + R_k\}} \tag{4.20}$$

μ^k-almost everywhere on C_k^ψ, for any $k = 1, 2, \dots$, and

$$I_{\{\pi_0 a g_0^k < \pi_1 b g_1^k\}} \le \phi_k \le I_{\{\pi_0 a g_0^k \le \pi_1 b g_1^k\}} \tag{4.21}$$

μ^k-almost everywhere on S_k^ψ, for any $k = 1, 2, \dots$.

Remark 4.4. *Because, by definition, any test (ψ, ϕ) satisfying (4.19) is Bayesian, (4.20) and (4.21) give all Bayesian sequential tests for two composite hypotheses $H_0 : \theta \in \Theta_0$ and $H_0 : \theta \in \Theta_1$. In particular, for two simple hypotheses ($\Theta_0 = \{\theta_0\}$ and $\Theta_1 = \{\theta_1\}$), Theorem 4.3 characterizes the structure of all Bayesian tests for the problem considered in Cochlar and Vrana (1978). For independent, but not necessarily identically distributed observations, it gives the structure of all Bayesian sequential tests for the problem considered in Liu and Blostein (1992) (see also Section 2 in Cochlar and Vrana (1978); more related results can be found in Novikov (2008a)).*

For composite hypotheses, it gives the structure of all Bayesian tests for the problem considered in Section 9.4 of Zacks (1971) in the particular case of $k = 2$ hypotheses, a linear cost function ($K(x_1, \dots, x_n) \equiv n$), and constant losses due to incorrect decisions. In Novikov (2009), there is a generalization of our present results to $k \ge 2$ hypotheses. In particular, the structure of Bayesian sequential multiple hypothesis tests can be easily obtained from Theorem 6 in Novikov (2009), just like we did it for the case of two hypotheses in this section. More general loss functions, in both problems, can be treated using a more general approach of Novikov (2008b).

There is a vast literature on asymptotic (as $c \to 0$) shapes of Bayesian sequential tests for two composite hypotheses when sampling from an exponential family, starting from Schwarz (1962) (see an exhaustive list of references in Lai (1997), see also Lai (2001)).

4.4 Discrete-Time Markov Process

In this section, we apply Theorem 4.1 for optimal sequential testing of two simple hypotheses about the distribution of a Markov process.

Let us suppose that, under the respective hypothesis H_j, $j = 0, 1$, the conditional "density" of X_i, given $X_{i-1} = x_{i-1}$, is

$$f_{j,i}(x_i|x_{i-1}), \quad i = 2, 3, \dots$$

and let $f_j(x_1)$ be the initial "density" (all the densities being with respect to some σ-finite measure μ). In this case, the joint density function of $(X_1, \dots, X_k)$, under H_j, is

$$f_j^k = f_j^r(x_1, \dots, x_r) = f_j(x_1) \prod_{i=2}^{k} f_{j,i}(x_i|x_{i-1}), \quad k = 1, 2, \dots, \quad j = 0, 1. \tag{4.22}$$

Let us define for any $k = 1, 2, \dots$ the likelihood ratio as

$$z_k = \begin{cases} f_1^k / f_0^k, & \text{if} \quad f_0^k > 0, \\ \infty \quad \text{if} \quad f_0^k = 0 \quad \text{and} \quad f_1^k > 0 \\ 0 \quad \text{if} \quad f_0^k = 0 \quad \text{and} \quad f_1^k = 0. \end{cases}$$

In what follows we construct optimal sequential tests, which minimize $E_0 \tau_\psi$ under restrictions $\alpha(\psi, \phi) \le \alpha$ and $\beta(\psi, \phi) \le \beta$ (see Section 4.1).

Let us introduce the following family of functions $\rho_k^n = \rho_k^n(z, x)$, $k = 0, 1, \dots,$ $n = 0, 1, \dots$. Let

$$\rho_k^0(z, x) = g(z) \equiv \min\{\lambda_0, \lambda_1 z\}, \tag{4.23}$$

for any $k = 0, 1, 2, \dots$, and, recursively, for $n = 1, 2, \dots$,

$$\rho_{k-1}^n(z, x) = \min\left\{ g(z), 1 + \int \rho_k^{n-1}\left(z \frac{f_{1,k}(x_k|x)}{f_{0,k}(x_k|x)}, x_k \right) f_{0,k}(x_k|x) d\mu(x_k) \right\} \tag{4.24}$$

for $k \ge 2$, and

$$\rho_0^n(z, x) = \rho_0^n(z) = \min\left\{ g(z), 1 + \int \rho_1^{n-1}\left(z \frac{f_1(x_1)}{f_0(x_1)}, x_1 \right) f_0(x_1) d\mu(x_1) \right\}. \tag{4.25}$$

It is not difficult to see, by induction, that

$$V_N^N(x_1, \dots, x_N) = \rho_N^0(z_N, x_N) f_0^N(x_1, \dots, x_N),$$

and

$$V_k^N(x_1, \dots, x_k) = \rho_k^{N-k}(z_k, x_k) f_0^k(x_1, \dots, x_k)$$

for $k = N-1, N-2, \dots, 1$, (see (3.8)), and that

$$R_{k-1}^N = \int \rho_k^{N-k}\left(z_{k-1} \frac{f_{1,k}(x_k|x_{k-1})}{f_{0,k}(x_k|x_{k-1})}, x_k \right) f_{0,k}(x_k|x_{k-1}) d\mu(x_k) f_0^{k-1} \tag{4.26}$$

(see (3.9)).

It is easy to see (very much like in Lemma 3.3) that for any fixed $k = 0, 1, \dots$

$$\rho_k^N(z, x) \ge \rho_k^{N+1}(z, x), \quad N = 0, 1, 2, \dots$$

Thus, there exists

$$\rho_k(z,x) = \lim_{N\to\infty} \rho_k^N(z,x), \tag{4.27}$$

and, passing to the limit in (4.24), as $n \to \infty$, we have for $k \geq 2$:

$$\rho_{k-1}(z,x) = \min\left\{g(z), 1 + \int \rho_k\left(z\frac{f_{1,k}(x_k|x)}{f_{0,k}(x_k|x)}, x_k\right) f_{0,k}(x_k|x)d\mu(x_k)\right\}. \tag{4.28}$$

Analogously, from (4.25) we have

$$\rho_0(z) = \min\left\{g(z), 1 + \int \rho_1\left(z\frac{f_1(x_1)}{f_0(x_1)}, x_1\right) f_0(x_1)d\mu(x_1)\right\}. \tag{4.29}$$

Let the integral on the right-hand side of (4.28) be denoted as $r_{k-1}(z,x)$, so (4.28) is equivalent to

$$\rho_{k-1}(z,x) = \min\{g(z), 1 + r_{k-1}(z,x)\}, \tag{4.30}$$

where

$$r_{k-1}(z,x) = \int \rho_k\left(z\frac{f_{1,k}(x_k|x)}{f_{0,k}(x_k|x)}, x_k\right) f_{0,k}(x_k|x)d\mu(x_k), \quad k \geq 2. \tag{4.31}$$

Analogously, (4.29) is equivalent to

$$\rho_0(z) = \min\{g(z), 1 + r_0(z)\}, \tag{4.32}$$

where

$$r_0(z) = \int \rho_1\left(z\frac{f_1(x_1)}{f_0(x_1)}, x_1\right) f_0(x_1)d\mu(x_1). \tag{4.33}$$

Finally, passing to the limit, as $N \to \infty$, in (4.26), we have for $k \geq 2$

$$R_{k-1} = r_{k-1}(z_{k-1}, x_{k-1})f_0^{k-1}, \tag{4.34}$$

and

$$R_0 = r_0(1).$$

Using these expressions in Theorem 4.1, we immediately have

Theorem 4.4. *Let $\lambda_0 > 0$ and $\lambda_1 > 0$ be any numbers and let ψ be any stopping time such that for any $n = 1, 2, \ldots$*

$$I_{\{g(z_n)<1+r_n(z_n,x_n)\}} \leq \psi_n \leq I_{\{g(z_n)\leq 1+r_n(z_n,x_n)\}} \tag{4.35}$$

μ^n-almost everywhere on $C_n^\psi \cap \{f_0^n > 0\}$, where g is defined by (4.23), and r_n is defined by (4.31), (4.27), and (4.24). And let ϕ be any decision rule such that

$$I_{\{\lambda_0 f_0^n < \lambda_1 f_1^n\}} \leq \phi_n \leq I_{\{\lambda_0 f_0^n \leq \lambda_1 f_1^n\}} \tag{4.36}$$

μ^n-almost everywhere on S_n^ψ, $n = 1, 2, \ldots$

Then for any (ψ', ϕ') such that

$$\alpha(\psi', \phi') \leq \alpha(\psi, \phi) \quad \text{and} \quad \beta(\psi', \phi') \leq \beta(\psi, \phi) \tag{4.37}$$

it holds

$$E_0\tau_\psi \leq E_0\tau_{\psi'}. \tag{4.38}$$

The inequality in (4.38) is strict if at least one of the inequalities in (4.37) is strict.

If there are equalities in all of the inequalities in (4.37) and (4.38) then ψ' satisfies (4.35) (with ψ_n' instead of ψ_n) μ^n-almost everywhere on $C_n^{\psi'} \cap \{f_0^n > 0\}$ for any $n = 1, 2, \ldots$, and ϕ' satisfies (4.36) (with ϕ_n' instead of ϕ_n) μ^n-almost everywhere on $S_n^{\psi'}$ for any $n = 1, 2, \ldots$.

Remark 4.5. *It is not difficult to see that all $\rho_k^n(z,x)$, and $\rho_k(z,x)$, for any x fixed, are concave, continuous non-decreasing functions of z, for any $k = 0,1,2,\dots$, and for any $n = 0,1,\dots$. Because of this, the inequality $g(z_n) \le 1 + r_n(z_n, x_n)$ defining an optimal way of stopping at any stage n, is equivalent to*

$$z_n \notin (A_n(x_n), B_n(x_n)),$$

with some $0 < A_n(x) \le B_n(x) < \infty$, $n = 1,2,\dots$.

If the Markov process is homogeneous, i.e.

$$f_{0,i}(x_i|x_{i-1}) = f_0(x_i|x_{i-1}) \quad \text{and} \quad f_{1,i}(x_i|x_{i-1}) = f_1(x_i|x_{i-1})$$

for any $i = 2,3,\dots$, then, obviously, $\rho_k(z,x) = \rho_1(z,x)$ for any $k = 2,3,\dots$, and therefore $A_n(x) = A(x)$ and $B_n(x) = B(x)$ for any $n = 1,2,\dots$.

It should be noted that in the case of homogeneous Markov process with a finite number of values, Schmitz (1968) sketched a proof of the fact that for any prior distribution there exists a Bayes sequential test with a stopping time of type (in our terms)

$$\tau = \min\{n : z_n \notin (A(x_n), B(x_n))\}$$

with $A(x) > 0$ and $B(x) < \infty$.

Remark 4.6. *Wald (1945) proposed the use of the sequential probability ratio test for any discrete-time stochastic process, and through decades this has been a popular topic in sequential analysis of stochastic processes, in particular, Markov chains (see, for example, Schmitz and Süselbeck (1983), where other references can be found). From Theorem 4.4 we see that, generally speaking, SPRTs are not optimal for discrete-time Markov processes, unless $r_n(z,x)$ in Theorem 4.4 does not depend on both x and n. In the following section we give a class of Markov processes where this is the case.*

4.5 An Example: Optimal Sequential Test for Location Parameter of AR(1) Process.

In this section, we apply the results of the preceding section to a specific model of an autoregressive process of order 1 (AR(1) process).

Let $X_1, X_2, \dots$ be a Markov process defined in the following way:

$$X_{n+1} - \theta = a(X_n - \theta) + \epsilon_n, \tag{4.39}$$

for $n = 1,2,\dots$, where $\epsilon_1, \epsilon_2, \dots$ are i.i.d. random variables with a probability density function $f(x)$ (with respect to the Lebesgue measure λ on the real line). We also suppose that X_1 is independent of $(\epsilon_1, \epsilon_2, \dots)$. Let us apply the results above for constructing optimal sequential tests for testing $H_0 : \theta = \theta_0$ vs. $H_1 : \theta = \theta_1$, $\theta_0 \ne \theta_1$, supposing that a in (4.39) is a fixed known constant.

For this process,

$$f_{j,i}(x_i|x_{i-1}) = f(x_i - ax_{i-1} - \theta_j(1-a)), \tag{4.40}$$

$i = 2,3,\dots$, $j = 0,1$. Let $f_0(x) = h(x - \theta_0)$ and $f_1(x) = h(x - \theta_1)$ be initial densities of X_1, with respect to λ, under H_0 and H_1, respectively.

The joint density functions (4.22) are now

$$f_j^n(x_1, \dots, x_n) = h(x_1 - \theta_j) \prod_{i=2}^{n} f(x_i - ax_{i-1} - \theta_j(1-a)), \quad j = 0, 1. \tag{4.41}$$

We use Theorem 4.4 for construction of tests which minimize $E_0 \tau_\psi$ in the class of all sequential tests such that $\alpha(\psi, \phi) \le \alpha$ and $\beta(\psi, \phi) \le \beta$, with some constant α and β.

We start with defining

$$\rho_0(z) = g(z)$$

(see 4.23). To apply (4.24), let us calculate the integral on the right-hand side of it:

$$\int \rho^{n-1} \left(z \frac{f(x_i - ax - \theta_1(1-a))}{f(x_i - ax - \theta_0(1-a))} \right) f(x_i - ax - \theta_0(1-a)) d\lambda(x_i)$$

$$= \int \rho^{n-1} \left(z \frac{f(y - (\theta_1 - \theta_0)(1-a))}{f(y)} \right) f(y) d\lambda(y)$$

because of invariance of the Lebesgue measure. We see that this integral does not depend on x, thus, (4.24) converts to

$$\rho^n(z) = \rho^n(z, x) = \min \left\{ g(z), 1 + \int \rho^{n-1} \left(z \frac{f(y - (\theta_1 - \theta_0)(1-a))}{f(y)} \right) f(y) d\lambda(y) \right\} \tag{4.42}$$

and is to be applied for any $n = 1, 2, \dots$.

With

$$\rho(z) = \lim_{n \to \infty} \rho^n(z) \tag{4.43}$$

(see (4.27)), we have from (4.42)

$$\rho(z) = \min\{g(z), 1 + r(z)\},$$

where

$$r(z) = r_k(z, x) = \int \rho \left(z \frac{f(y - (\theta_1 - \theta_0)(1-a))}{f(y)} \right) f(y) d\lambda(y) \tag{4.44}$$

(see (4.31)).

Theorem 4.4, applied to this special case, gives now all the optimal sequential tests in the following form.

Theorem 4.5. *Let $\lambda_0 > 0$ and $\lambda_1 > 0$ be any numbers and let ψ be any stopping time such that for any $n = 1, 2, \dots$*

$$I_{\{g(z_n) < 1 + r(z_n)\}} \le \psi_n \le I_{\{g(z_n) \le 1 + r(z_n)\}} \tag{4.45}$$

μ^n-almost everywhere on $C_n^\psi \cap \{f_0^n > 0\}$, where $g(z)$ is defined by (4.23), and $r(z)$ is defined by (4.44), through (4.43), and (4.42). And let ϕ be any decision rule such that

$$I_{\{\lambda_0 f_0^n < \lambda_1 f_1^n\}} \le \phi_n \le I_{\{\lambda_0 f_0^n \le \lambda_1 f_1^n\}} \tag{4.46}$$

μ^n-almost everywhere on S_n^ψ, $n = 1, 2, \dots$

Then for any (ψ', ϕ') such that

$$\alpha(\psi', \phi') \le \alpha(\psi, \phi) \quad \textit{and} \quad \beta(\psi', \phi') \le \beta(\psi, \phi) \tag{4.47}$$

it holds

$$E_0\tau_\psi \le E_0\tau_{\psi'}. \tag{4.48}$$

The inequality in (4.48) is strict if at least one of the inequalities in (4.47) is strict.

If there are equalities in all of the inequalities in (4.47) and (4.48) then ψ' satisfies (4.45) (with ψ'_n instead of ψ_n) μ^n-almost everywhere on $C_n^{\psi'} \cap \{f_0^n > 0\}$ for any $n = 1, 2, \ldots$, and ϕ' satisfies (4.46) (with ϕ'_n instead of ϕ_n) μ^n-almost everywhere on $S_n^{\psi'}$ for any $n = 1, 2, \ldots$.

It is easy to see that the optimal tests in Theorem 4.5 are randomized SPRTs, i.e. that the inequality $g(z) \le 1 + r(z)$, defining a way of optimal stopping in (4.45), is equivalent to $z \notin (A, B)$, with some $0 < A \le B < \infty$ (see Remark 4.5). In fact, the optimal tests in Theorem 4.5 coincide with those of a sequential testing problem for independent observations.

Indeed, let $Y_1 = X_1$, $Y_2 = X_2 - aX_1$, $Y_3 = X_3 - aX_2$, $\ldots$. Then, by (4.39), $Y_n = \theta(1-a) + \epsilon_{n-1}$, $n = 2, 3, \ldots$, with densities

$$f_{\theta,n}(x_n) = f(x_n - \theta(1-a)),$$

$n \ge 2$, and

$$f_{\theta,1}(x_1) = h(x_1 - \theta).$$

and Y_1 is independent of $Y_2, Y_3, \ldots$.

Because, obviously, the likelihood ratio z_n based on $X_1, \ldots, X_n$ coincides with the likelihood ratio based on $Y_1, \ldots, Y_n$ as defined in Novikov (2008a), it is easy to see that any optimal test in Theorem 4.5 coincides with a respective optimal test in Theorem 6 in Novikov (2008a). Thus, in this particular case of AR(1) process we have the following Theorem.

Theorem 4.6. *Let $0 < A < B < \infty$ be any real numbers, and let for any $n = 1, 2, \ldots$*

$$I_{\{z_n \notin [A,B]\}} \le \psi_n \le I_{\{z_n \notin (A,B)\}} \tag{4.49}$$

μ^n-almost everywhere on $C_n^\psi \cap (\{f_0^n > 0\} \cup \{f_1^n > 0\})$. And let ϕ be a decision rule defined by

$$\phi_n = I_{\{z_n \ge B\}}, \tag{4.50}$$

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

Then for any (ψ', ϕ') such that

$$\alpha(\psi', \phi') \le \alpha(\psi, \phi) \quad \text{and} \quad \beta(\psi', \phi') \le \beta(\psi, \phi) \tag{4.51}$$

it holds

$$E_0\tau_\psi \le E_0\tau_{\psi'} \quad \text{and} \quad E_1\tau_\psi \le E_1\tau_{\psi'}. \tag{4.52}$$

Both inequalities in (4.52) are strict if at least one of the inequalities in (4.51) is strict.

If there are equalities in all of the inequalities in (4.51) and in one of the inequalities in (4.52), then there are equalities in both inequalities in (4.52), and ψ' satisfies (4.49) (with ψ'_n instead of ψ_n) μ^n-almost everywhere on $C_n^{\psi'} \cap (\{f_0^n > 0\} \cup \{f_1^n > 0\})$ for any $n = 1, 2, \ldots$, and ϕ' satisfies (4.50) (with ϕ'_n instead of ϕ_n) μ^n-almost everywhere on $S_n^{\psi'} \cap (\{f_0^n > 0\} \cup \{f_1^n > 0\})$ for any $n = 1, 2, \ldots$.

Proof. Follows from the proof of Theorem 6 in Novikov (2008a). □

Remark 4.7. *It is not difficult to see that, in the same way, the results of Section 4.1 in Novikov (2008a) allow to find optimal sequential tests of two simple hypotheses about the location parameter in any case of autoregressive process of finite order $p \geq 1$, AR(p).*

Remark 4.8. *In essence, Theorem 6 in Novikov (2008a) states that any SPRT conserves its Wald-Wolfowitz optimality property when the distribution of the first observation does not necessarily match the common distribution of all other observations, provided that all of them (starting from the first one) are independent. Under some additional conditions, this fact was noted by Schmitz (1973), a reference which should have been added into the reference list in Novikov (2008a).*

5 APPENDIX. PROOFS OF THE MAIN RESULTS

5.1 Proof of Theorem 2.2

Let us prove first the following Lemma which will also be helpful in other proofs below.

Lemma 5.1. *Let ϕ, F_1, F_2 be some measurable functions on a measurable space with a measure μ, such that*

$$0 \leq \phi(x) \leq 1, \quad F_1(x) \geq 0, \quad F_2(x) \geq 0,$$

and

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

Then

$$\int (\phi(x)F_1(x) + (1 - \phi(x))F_2(x)) d\mu(x) \geq \int \min\{F_1(x), F_2(x)\} d\mu(x) \tag{5.1}$$

with an equality if and only if

$$I_{\{F_1(x) < F_2(x)\}} \leq \phi(x) \leq I_{\{F_1(x) \leq F_2(x)\}} \tag{5.2}$$

μ-almost anywhere.

Proof. To prove (5.1) it suffices to show that

$$\int [(\phi(x)F_1(x) + (1 - \phi(x))F_2(x)) - \min\{F_1(x), F_2(x)\}] d\mu(x) \geq 0 \tag{5.3}$$

which is trivial because the function under the integral sign is non-negative.

Being so, there is an equality in (5.3) if and only if

$$\phi(x)F_1(x) + (1 - \phi(x))F_2(x) = \min\{F_1(x), F_2(x)\},$$

or

$$\phi(x)(F_1(x) - F_2(x)) = \min\{F_1(x), F_2(x)\} - F_2(x),$$

μ-almost anywhere, which is only possible if (5.2) holds true. □

To prove (2.10), let us give to the left-hand side of it the form

$$\lambda_0\alpha(\psi,\phi)+\lambda_1\beta(\psi,\phi))=\sum_{n=1}^{\infty}\int(1-\psi_1)\dots(1-\psi_{n-1})\psi_n[\phi_n\lambda_0 f_0^n+(1-\phi_n)\lambda_1 f_1^n]d\mu^n \tag{5.4}$$

(see (1.3) and (1.4)).

Applying Lemma 5.1 to each summand in (5.4) we immediately have:

$$\lambda_0\alpha(\psi,\phi)+\lambda_1\beta(\psi,\phi)\geq\sum_{n=1}^{\infty}\int(1-\psi_1)\dots(1-\psi_{n-1})\psi_n\min\{\lambda_0 f_0^n,\lambda_1 f_1^n\}d\mu^n \tag{5.5}$$

with an equality if and only if ϕ_n satisfies (2.11) μ^n-almost everywhere on S_n^ψ, for any $n=1,2,\dots$

5.2 Proof of Lemma 3.1

Obviously, (3.2) is equivalent to

$$\int s_k^\psi(kf^k+l_k)d\mu^k+\int c_{k+1}^\psi\left((k+1)f^{k+1}+v_{k+1}\right)d\mu^{k+1}\geq\int c_k^\psi\left(kf^k+v_k\right)d\mu^k. \tag{5.6}$$

By the Fubini theorem, the left-hand side of (5.6) is equal to

$$\int s_k^\psi(kf^k+l_k)d\mu^k+\int c_{k+1}^\psi\left(\int\left((k+1)f^{k+1}+v_{k+1}\right)d\mu(x_{k+1})\right)d\mu^k$$

$$=\int c_k^\psi[\psi_k(kf^k+l_k)+(1-\psi_k)\int\left((k+1)f^{k+1}+v_{k+1}\right)d\mu(x_{k+1})]d\mu^k. \tag{5.7}$$

Because $f^{k+1}(x_1,\dots,x_{k+1})$ is a joint density function of $(X_1,\dots,X_{k+1})$, we have

$$\int f^{k+1}(x_1,\dots,x_{k+1})d\mu(x_{k+1})=f^k(x_1,\dots,x_k),$$

so that the right-hand side of (5.7) transforms to

$$\int c_k^\psi\left[kf^k+\psi_k l_k+(1-\psi_k)\left(f^k+\int v_{k+1}d\mu(x_{k+1})\right)\right]d\mu^k. \tag{5.8}$$

Applying Lemma 5.1 we see that (5.8) is greater than or equal to

$$\int c_k^\psi\left[kf^k+\min\{l_k,f^k+\int v_{k+1}d\mu(x_{k+1})\}\right]d\mu^k=\int c_k^\psi\left(kf^k+v_k\right)d\mu^k, \tag{5.9}$$

by the definition of v_k in (3.3).

Moreover, by the same Lemma 5.1, (5.8) is equal to (5.9) if and only if (3.4) holds μ^k-almost everywhere on C_k^ψ.

5.3 Proof of Lemma 3.2

Let us prove first that if (3.14) is satisfied for any ψ such that $E\tau_\psi < \infty$ and for any $\lambda_0 > 0$ and $\lambda_1 > 0$, then the problem (P, P_0, P_1) is truncatable.

Let ψ be any stopping rule such that $P(\tau_\psi < \infty) = 1$. Let us show that (3.14) is fulfilled for any $\lambda_0 > 0$ and $\lambda_1 > 0$.

Let $L(\psi) = L(\psi; \lambda_0, \lambda_1) < \infty$, leaving the possibility $L(\psi) = \infty$ till the end of the proof. 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 (2.12) and (3.12)

$$L(\psi) - L_N(\psi) = \sum_{n=N}^{\infty} \int s_n^\psi (nf^n + l_n) d\mu^n - \int c_N^\psi \left(Nf^N + l_N \right) d\mu^N. \tag{5.10}$$

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

$$\int c_N^\psi l_N d\mu^N \to 0$$

as $N \to \infty$, because of (3.14).

It remains to show that

$$\int c_N^\psi N f^N d\mu^N = NP(\tau_\psi \geq N) \to 0 \text{ as } N \to \infty. \tag{5.11}$$

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

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

This means that

$$\sum_{n=1}^{\infty} \int s_n^\psi (nf^n + l_n) d\mu^n = \infty$$

which immediately implies by (3.12) that

$$L_N(\psi) \geq \sum_{n=1}^{N-1} \int s_n^\psi (nf_\theta^n + l_n) d\mu^n \to \infty.$$

The "if"-part of Lemma 3.2 is proved.

Supposing now that the problem (P, P_0, P_1) is truncatable and that ψ is any stopping rule with $E\tau_\psi < \infty$, from (3.13) it follows that the right-hand side of (5.10) tends to 0, as $N \to \infty$, and because, again, the first summand in (5.10) tends to zero, for the same reason as above, and so does $\int c_N^\psi N f^N d\mu^N$, as $N \to \infty$, (3.14) follows.

5.4 Proof of Corollary 3.2

Let us suppose that *i)* is satisfied. Let ψ be any stopping rule such that $E\tau_\psi < \infty$. Then for any $i = 0, 1$ such that

$$P_i(\tau_\psi < \infty) = 1,$$

we have

$$\int c_n^\psi \min\{\lambda_0 f_0^n, \lambda_1 f_1^n\} d\mu^n \leq \lambda_i \int c_n^\psi f_i^n d\mu^n = \lambda_i P_i(\tau_\psi \geq n) \to 0,$$

as $n \to \infty$. That is, condition (3.14) is fulfilled. By Lemma 3.2 the problem (P, P_0, P_1) is truncatable.

Let us show that under condition *iii)* of Corollary 3.2 the problem (P, P_0, P_1) is truncatable as well.

Let

$$Z_n = \frac{f_1^n(X_1, \dots, X_n)}{f_0^n(X_1, \dots, X_n)}, \quad n = 1, 2, \dots$$

.

If *iii)* is satisfied, then for any $\lambda_0 > 0$ and $\lambda_1 > 0$

$$\int \min\{\lambda_0 f_0^n, \lambda_1 f_1^n\} d\mu^n = E_0 \min\{\lambda_0, \lambda_1 Z_n\} \to 0, \tag{5.12}$$

as $n \to \infty$, because $x \mapsto \min\{\lambda_0, \lambda_1 x\}$ is a non-negative continuous and bounded function of x. Because $c_n^\psi \le 1$ for any ψ and for any $n = 1, 2, \dots$, (3.14) now follows from (5.12). Again, by Lemma 3.2 the problem (P, P_0, P_1) is truncatable.

If *ii)* holds, then

$$E_0 \min\{\lambda_0, \lambda_1 Z_n\} = \int \min\{\lambda_0 f_0^n, \lambda_1 f_1^n\} d\mu^n \to 0,$$

as $n \to \infty$. Thus, $\min\{\lambda_0, \lambda_1 Z_n\} \to 0$ in P_0-probability. For any $0 < \epsilon < \lambda_0$ we have:

$$P_0(Z_n > \epsilon/\lambda_1) = P_0(\lambda_1 Z_n > \epsilon) = P_0(\min\{\lambda_0, \lambda_1 Z_n\} > \epsilon) \to 0,$$

as $n \to \infty$, which implies that $Z_n \to 0$ as $n \to \infty$ in P_0-probability. Therefore, condition *iii)* of the Lemma is satisfied. Above we proved that it implies (3.14). By Lemma 3.2 it follows that the problem (P, P_0, P_1) is truncatable.

5.5 Proof of Lemma 3.5

Let us denote

$$U = \inf_\psi L(\psi), \quad U_N = \inf_{\psi \in \mathscr{D}^N} L(\psi).$$

By Corollary 3.1, for any $N = 1, 2, \dots$

$$U_N = 1 + R_0^N.$$

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

$$\lim_{N \to \infty} U_N \ge U. \tag{5.13}$$

Let us show first that there is an equality in (5.13).

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

$$U_N \ge U + 3\epsilon \tag{5.14}$$

for all sufficiently large N.

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

Because, for a truncated problem, $L_N(\psi) \to L(\psi)$, as $N \to \infty$, we have that

$$L_N(\psi) \le U + 2\epsilon \tag{5.15}$$

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 (5.14).

Now, (3.23) follows from the monotone convergence theorem, because

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

5.6 Proof of Theorem 3.2

Let ψ be any stopping rule. By Lemma 3.4 for any fixed $k \geq 1$ the following inequalities hold:

$$\begin{aligned} L(\psi) &\geq \sum_{n=1}^{k} \int s_n^\psi (nf^n + l_n)d\mu^n + \int c_{k+1}^\psi \left((k+1)f^{k+1} + V_{k+1}\right) d\mu^{k+1} && (5.16)\\ &\geq \sum_{n=1}^{k-1} \int s_n^\psi (nf^n + l_n)d\mu^n + \int c_k^\psi \left(kf^k + V_k\right) d\mu^k && (5.17)\\ &\geq \dots \\ &\geq \int s_1^\psi (f^1 + l_1)d\mu^1 + \int c_2^\psi \left(2f^2 + V_2\right) d\mu^2 && (5.18)\\ &\geq 1 + \int V_1(x)d\mu(x) = 1 + R_0. && (5.19) \end{aligned}$$

Supposing (3.24), by Lemma 3.5 we have that there are equalities in all the inequalities (5.16) – (5.19).

Applying Lemma 3.1, we see that (3.25) is fulfilled μ^k-almost everywhere on C_k^ψ, for any $k = 1, 2, \dots$, because, by (3.20), it is a necessary condition of equality in the respective inequality in (5.16) – (5.19). The "only if"-part of Theorem 3.2 is proved.

Let now ψ satisfy (3.25) μ^k-almost everywhere on C_k^ψ, for any $k = 1, 2, \dots$.

It follows from Lemma 3.1 and (3.20) that all the inequalities in (5.17) – (5.19) are in fact equalities, i.e.

$$\sum_{n=1}^{k} \int s_n(nf^n + l_n)d\mu^n + \int c_{k+1}^\psi \left((k+1)f^{k+1} + V_{k+1}\right) d\mu^{k+1} = 1 + R_0 \qquad (5.20)$$

for any $k = 0, 1, 2, \dots$.

In particular, this means that

$$\int c_{k+1}^\psi \left((k+1)f^{k+1}\right) d\mu^{k+1} = (k+1)P(\tau_\psi \geq k+1) \leq 1 + R_0.$$

Because of this,

$$P(\tau_\psi \geq k+1) \leq (1 + R_0)/(k+1) \to 0,$$

as $k \to \infty$, and hence

$$P(\tau_\psi < \infty) = 1. \qquad (5.21)$$

It follows by Corollary 2.1 now that

$$L(\psi) = \lim_{k\to\infty} \sum_{n=1}^{k} \int s_n^\psi (nf^n + l_n)d\mu^n,$$

which implies, by virtue of (5.20), that $L(\psi) \leq 1 + R_0$.

On the other hand, by Lemma 3.4, $L(\psi) \geq 1 + R_0$.

Thus, $L(\psi) = 1 + R_0$.

ACKNOWLEDGMENTS

The author thanks an anonymous referee for his valuable comments and helpful suggestions on the improvement of the early version of the article.

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.

REFERENCES

Berk, R. H. (1975). Locally Most Powerful Sequential Tests, *Annals of Statistics* 3: 373-381.

Castillo E. and García, J. (1983). Necessary Conditions for Optimal Truncated Sequential Tests. Simple Hypotheses (in Spanish), *Stochastica* 7: 63-81.

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

Cochlar, J. and Vrana, I. (1978). On the Optimum Sequential Test of Two Hypotheses for Statistically Dependent Observations, *Kybernetika* 14: 57-69.

DeGroot, M. H. (1970). *Optimal Statistical Decisions*, New York: McGraw-Hill.

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

Ghosh, B. K. (1970). *Sequential Tests of Statistical Hypotheses*, Reading, Mass.-London-Don Mills, Ont.: Addison-Wesley.

Ghosh, M., Mukhopadhyay, N., and Sen, P.K. (1997). *Sequential Estimation*, New York: Wiley.

Hawix, A. and Schmitz, N. (1998). Remark on the Modified Kiefer-Weiss Problem for Exponential Families, *Sequential Analysis*, 17: 297-303.

Kiefer, J. and Weiss, L. (1957). Some Properties of Generalized Sequential Probability Ratio Tests, *Annals of Mathematical Statistics* 28: 57-75.

Lai, T. L. (1997). On Optimal Stopping Problems in Sequential Hypothesis Testing, *Statistica Sinica* 7: 33-51.

Lai, T. L. (2001). Sequential Analysis: Some Classical Problems and New Challenges, *Statistica Sinica* 11: 303-408.

Lehmann, E. L. (1959). *Testing statistical hypotheses*, New York: Wiley; London: Chapman & Hall.

Liu, Y. and Blostein, D. (1992). Optimality of the Sequential Probability Ratio Test for Nonstationary Observations, *IEEE Transactions on Information Theory* 28: 177-182.

Lorden, G. (1980). Structure of Sequential Tests Minimizing an Expected Sample Size, *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 51: 291-302.

Müller-Funk, U., Pukelsheim, F., and Witting, H. (1985). Locally Most Powerful Tests for Two-Sided Hypotheses, in *Probability and statistical decision theory, Vol. A (Bad Tatzmannsdorf, 1983)*, pp. 31–56, Dordrecht: Reidel.

Novikov, A. (2009). Optimal Sequential Multiple Hypothesis Tests, *Kybernetika* 45, no. 2, 309-330, preprint arXiv:0811.1297v1 [math.ST] (http://arxiv.org/abs/0811.1297).

Novikov, A. (2008a). Optimal Sequential Tests for Two Simple Hypotheses Based on Independent Observations, *International Journal of Pure and Applied Mathematics* 45: 291-314.

Novikov, A. (2008b). Optimal Sequential Procedures With Bayes Decision Rules. To appear in *Kybernetika*, preprint arXiv:0812.0159v1 [math.ST] (http://arxiv.org/abs/0812.0159)

Schmitz, N. (1968). Likelihoodquotienten-Sequenztests bei homogenen Markoffschen Ketten (in German), *Biometrische Zeitschrift* 10: 231-247.

Schmitz, N. (1973). Zur Optimalität des Likelihoodquotienten-Sequenztests, (in German), in *Transactions of the Sixth Prague Conference on Information Theory, Statistical Decision Functions, Random Processes (Tech. Univ. Prague, Prague, 1971; dedicated to the memory of Antonín Špaček)*, pp. 781-791, Prague: Academia.

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

Schmitz, N. (1995). Comment on: “Optimality of the Sequential Probability Ratio Test for Nonstationary Observations” by Y. Liu and S.D. Blostein, *IEEE Trans. Inform. Theory* 41: 861.

Schmitz, N. and Süselbeck, B. (1983). Sequential Probability Ratio Tests for Homogeneous Markov Chains, *Lecture Notes in Statistics* 20: 181-202.

Schwarz, G. (1962). Asymptotic Shapes of Bayes Sequential Testing Regions, *Annals of Mathematical Statistics* 33: 224-236.

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

Wald, A. (1945). Sequential Tests of Statistical Hypotheses. *Annals of Mathematical Statistics* 16: 117-186.

Wald, A. (1950). *Statistical Decision Functions*, New York: Wiley.

Wald, A. and Wolfowitz, J. (1948). Optimum Character of the Sequential Probability Ratio Test, *Annals of Mathematical Statistics* 19: 326-339.

Weiss, L. (1962). On Sequential Tests which Minimize the Maximum Expected Sample Size, *Journal of American Statistical Assocciation* 57: 551-566.

Zacks, S. (1971). *The Theory of Statistical Inference. Wiley Series in Probability and Mathematical Statistics*, New York: Wiley.