Evgueni I. Gordienko; Andrey Novikov; Elena Zaitseva

Stability estimating in optimal sequential hypotheses testing

Kybernetika, Vol. 45 (2009), No. 2, 331--344

Persistent URL: http://dml.cz/dmlcz/140072

KYBERNETIKA — VOLUME 45 (2009), NUMBER 2, PAGES 331 – 344

STABILITY ESTIMATING IN OPTIMAL SEQUENTIAL HYPOTHESES TESTING

Evgueni Gordienko, Andrey Novikov and Elena Zaitseva

We study the stability of the classical optimal sequential probability ratio test based on independent identically distributed observations $X_1, X_2, \ldots$ when testing two simple hypotheses about their common density f: $f = f_0$ versus $f = f_1$. As a functional to be minimized, it is used a weighted sum of the average (under f_0) sample number and the two types error probabilities. We prove that the problem is reduced to stopping time optimization for a ratio process generated by $X_1, X_2, \ldots$ with the density f_0. For τ_* being the corresponding optimal stopping time we consider a situation when this rule is applied for testing between f_0 and an alternative $\tilde{f}_1$, where $\tilde{f}_1$ is some approximation to f_1. An inequality is obtained which gives an upper bound for the expected cost excess, when τ_* is used instead of the rule $\tilde{\tau}_*$ optimal for the pair $(f_0, \tilde{f}_1)$. The inequality found also estimates the difference between the minimal expected costs for optimal tests corresponding to the pairs (f_0, f_1) and $(f_0, \tilde{f}_1)$.

Keywords: sequential hypotheses test, simple hypothesis, optimal stopping, sequential probability ratio test, likelihood ratio statistic, stability inequality

AMS Subject Classification: 62L10, 62L15

1. MOTIVATION AND PROBLEM SETTING

In this paper we will find a quantitative estimate of stability of optimal sequential testing a simple hypothesis against a simple alternative. To set the problem, we consider a measurable space $(\Omega, \mathcal{F})$ with given three different probability measures P_0, P_1 and $\tilde{P}_1$. Let $X_1, X_2, \ldots$ be a sequence of random variables on $(\Omega, \mathcal{F})$ independent and identically distributed with respect of each one of the above probabilities. We assume that under P_0, P_1 and $\tilde{P}_1$, the common distributions of X_n, $n \geq 1$, have densities f_0, f_1 and $\tilde{f}_1$ with respect to some σ-finite measure μ on $(\mathbb{R}, \mathscr{B}_{\mathbb{R}})$, and that

$$\mu(\{x : f_0(x) \neq f_1(x)\}) > 0 \quad \text{and} \quad \mu(\{x : f_0(x) \neq \tilde{f}_1(x)\}) > 0.$$

Denote by Supp_0, Supp_1 and $\widetilde{\mathrm{Supp}}_1$ the supports of the distributions of X_1 under P_0, P_1, $\tilde{P}_1$, respectively. In what follows we assume that the densities f_0, f_1 and $\tilde{f}_1$ are positive on the corresponding supports and that

$$\mathrm{Supp}_0 \subset \mathrm{Supp}_1 \cap \widetilde{\mathrm{Supp}}_1. \tag{1}$$

We consider sequential testing problems for two following pairs of simple hypotheses about the density f of distribution of the observed sequence X:

(I) $H_0 : f = f_0$ against $H_1 : f = f_1$, and (II) $H_0 : f = f_0$ against $H_1 : f = \tilde{f}_1$

A sequential test is, by definition, a pair (τ, δ), where τ is a *stopping time* with respect to the filtration $\{\mathscr{F}_n\}_{n\geq 0}$ with $\mathscr{F}_n = \sigma\{X_1, X_2, \ldots, X_n\}$, for $n = 1, 2, \ldots$, and $\mathscr{F}_0 = \{\emptyset, \Omega\}$, and δ is a *terminal decision function*, i. e. $\mathscr{F}_\tau$-measurable function taking values in $\{0, 1\}$. The corresponding error probabilities are defined in Problem I as follows:

- the type I error probability is
$$\alpha(\tau, \delta) = P_0(\delta = 1),$$
- the type II error probability is
$$\beta(\tau, \delta) = P_1(\delta = 0).$$

Similarly (replacing P_1 by $\tilde{P}_1$), the error probabilities $\tilde{\alpha}(\tau, \delta)$, $\tilde{\beta}(\tau, \delta)$ are defined in Problem II.

Let $c_0 > 0$ be a given cost of an observation, and $c_1, c_2 > 0$ be penalties to be paid for corresponding erroneous decisions. We define the expected cost functionals to be minimized over (τ, δ) as

$$W(\tau, \delta) = c_0 \mathrm{E}_0 \tau + c_1 \alpha(\tau, \delta) + c_2 \beta(\tau, \delta), \quad (2)$$
$$\tilde{W}(\tau, \delta) = c_0 \mathrm{E}_0 \tau + c_1 \tilde{\alpha}(\tau, \delta) + c_2 \tilde{\beta}(\tau, \delta), \quad (3)$$

where E_0 stands for the expectation with respect to the probability P_0, $\tau \in \mathscr{T}$, $\delta \in \mathscr{D}_\tau$, being $\mathscr{T}$ the class of all stopping times τ such that $P_0(\tau < \infty) = 1$, and $\mathscr{D}_\tau$ the class of all $\mathscr{F}_\tau$-measurable decision functions.

Remark 1. A more traditional, Bayesian, setting uses yet another term, a multiple of $\mathrm{E}_1\tau$, in (2), and has been widely used in sequential hypothesis testing since the seminal paper of Wald and Wolfowitz [15]. The results by Lorden [7] show that minimizing (2) and minimizing the Bayesian risk are equivalent in the sense that the solution of both problems is given by the (respective) sequential probability ratio tests (SPRT). But using criteria (2) and (3) is more convenient in the present stability estimation context, because it makes possible to use only one probability measure, P_0, for the analysis of the related functionals (see details below). On the other hand, the use of only one average sample size functional may sometimes be justified by practical reasons.

To obtain an equivalent setting to optimization problems (2) and (3), which is more convenient for our purposes, we first introduce the standard log-likelihood ratio statistics:

$$S_n = \sum_{k=1}^{n} \ln \left[\frac{f_1(X_k)}{f_0(X_k)} \right], \quad (4)$$

$$\tilde{S}_n = \sum_{k=1}^{n} \ln\left[\frac{\tilde{f}_1(X_k)}{f_0(X_k)}\right], \tag{5}$$

for $n = 1, 2, 3, \dots$, and $S_0 \equiv 0$ and $\tilde{S}_0 \equiv 0$, where, conventionally, $0/0 = 1$ and $\ln(x/0) = \infty$ for $x > 0$.

Second, we define the following functionals on $\mathscr{T}$:

$$V(\tau) = c_0 \mathrm{E}_0 \tau + \mathrm{E}_0 \min\{c_1, c_2 e^{S_\tau}\}, \quad \tau \in \mathscr{T}, \tag{6}$$

$$\tilde{V}(\tau) = c_0 \mathrm{E}_0 \tau + \mathrm{E}_0 \min\{c_1, c_2 e^{\tilde{S}_\tau}\}, \quad \tau \in \mathscr{T}. \tag{7}$$

It follows from Theorem 1 in Section 2 that for any stopping time $\tau \in \mathscr{T}$ there exists a decision rule $\delta_*(\tau) \in \mathscr{D}_\tau$ such that

$$\inf_{\delta \in \mathscr{D}_\tau} W(\tau, \delta) = W(\tau, \delta_*(\tau)) = V(\tau). \tag{8}$$

On the other hand, it is known (see, e. g., [7]), that the infimum over $\tau \in \mathscr{T}$ of the right-hand side of (8) is in fact attained at some $\tau_* \in \mathscr{T}$:

$$V(\tau_*) = \inf_{\tau \in \mathscr{T}} V(\tau). \tag{9}$$

Therefore,

$$\inf_{\tau \in \mathscr{T}, \delta \in \mathscr{D}_\tau} W(\tau, \delta) = W(\tau_*, \delta_*) = V(\tau_*) = \inf_{\tau \in \mathscr{T}} V(\tau). \tag{10}$$

According to [7], the structure of the solution (τ_*, δ_*) of the optimization problem in (10) is as follows: there exist constants A, B, $-\infty < A \le B < \infty$, such that

$$\tau_* = \min\{n \ge 0 : S_n \notin (A, B)\}, \tag{11}$$

$$\delta_* = \begin{cases} 1 & \text{if} \quad S_{\tau_*} \ge B \\ 0 & \text{if} \quad S_{\tau_*} \le A \end{cases} \tag{12}$$

(A. Wald's SPRT [15]).

In a similar manner, for the problem of testing $f = f_0$ versus $f = \tilde{f}_1$ there exist an optimal sequential test $\tilde{\tau}_*$

$$\tilde{\tau}_* = \min\{n \ge 0 : \tilde{S}_n \notin (\tilde{A}, \tilde{B})\}, \tag{13}$$

$$\tilde{\delta}_* = \begin{cases} 1 & \text{if} \quad \tilde{S}_{\tau_*} \ge \tilde{B} \\ 0 & \text{if} \quad \tilde{S}_{\tau_*} \le \tilde{A}, \end{cases} \tag{14}$$

with $-\infty < \tilde{A} \le \tilde{B} < \infty$, such that

$$\tilde{V}(\tilde{\tau}_*) = \inf_{\tau \in \mathscr{T}} \tilde{V}(\tau). \tag{15}$$

Remark 2.

(a) In contrast to (2), (3), the functionals V, $\tilde{V}$ in (6), (7) are evaluated using only the null-hypothesis H_0. This gives certain advantages in setting and solving the stability problem formulated below. In particular, we will make use of the fact that the minimization of V and $\tilde{V}$ in (6), (7) is a standard optimal stopping problem with bounded one-stage and terminal costs.

(b) By virtue of (1) the random variables

$$\ln \frac{f_1(X_k)}{f_0(X_k)} \quad \text{and} \quad \ln \frac{\tilde{f}_1(X_k)}{f_0(X_k)}$$

are finite with P_0-probability one. In fact, we have used this to define V and $\tilde{V}$ in (6) and (7), and will exploit it later.

Remark 3.

(a) Because $\mathscr{F}_0 = \{\Omega, \emptyset\}$, it is obvious that $\mathscr{T} = \{\tau_0\} \cup \mathscr{T}_1$, where $\mathscr{T}_1$ is the class of stopping times τ(with respect to $\{\mathscr{F}_n\}$), such that $\tau \geq 1$, and $\tau_0 \equiv 0$.

(b) If $S_0 = 0 \notin (A, B)$ in (11) and (12), or if $A = B$, then $\tau_* = \tau_0$, and the optimal value $V(\tau_*) = \min\{c_1, c_2\}$ with

$$\delta_*(\tau_*) = \begin{cases} 1 & \text{if} \quad c_1 \leq c_2 \\ 0 & \text{if} \quad c_1 > c_2. \end{cases} \tag{16}$$

On the other hand, if $A < 0 < B$ then the optimal stopping time τ_* can not take the value 0. The same is true for the optimal stopping time $\tilde{\tau}_*$.

Only the latter case is of practical interest, because in the former case, due to (16), one of the error probabilities is always equal to 1.

Because of this, we will assume in what follows that τ_* in (9) is of type (11) with $A < 0 < B$, and $\tilde{\tau}_*$ in (15) is of type (13) with $\tilde{A} < 0 < \tilde{B}$.

(c) If $c_0 \geq c_2$ then $\tau_* = \tilde{\tau}_* = \tau_0$ as well. Indeed, in this case,

$$V(\tau_0) = \min\{c_1, c_2\} \leq c_0 \leq c_0 \mathrm{E}_0 \tau + \mathrm{E}_0 \min\{c_1, c_2 e^{S_\tau}\} = V(\tau)$$

for any $\tau \in \mathscr{T}_1$. Because, again, this case is not of practical interest, we will be supposing throughout the paper that $c_0 < c_2$.

The stability estimation problem may appear in the following situation. Assume that a statistician has the optimal stopping rule τ_* corresponding to the (f_0, f_1) pair, but he is not quite sure about the density f_1. If, in fact, the true alternative density is $\tilde{f}_1$, then how big is the additional expected cost

$$\Delta = \tilde{V}(\tau_*) - \tilde{V}(\tilde{\tau}_*) \geq 0 \tag{17}$$

he pays because of applying the non-optimal stopping time τ_* instead of the optimal one, $\tilde{\tau}_*$? Here, $\tilde{V}(\tilde{\tau}_*)$ is the optimal expected cost which is attained using the optimal test $(\tilde{\tau}_*, \tilde{\delta}_*)$, and $\tilde{V}(\tau_*)$ is the expected cost corresponding to the test (τ_*, δ_*) (optimal for (f_0, f_1)), when the type II error probability is calculated according to the true alternative density $\tilde{f}_1$ (see (3)).

For instance, $\tilde{f}_1$ of the form

$$\tilde{f}_1 = (1-\varepsilon) f_1 + \varepsilon f', \tag{18}$$

where f' is an unknown "contaminating" density, naturally appears in robust statistics. It is used for judging about the robustness of statistical procedures letting $\varepsilon \to 0$ (see, e. g., [4, 5, 6, 13, 16]).

Our purpose is to estimate the expected cost excess (17) in terms of a suitable measure of discrepancy between f_1 and $\tilde{f}_1$. More precisely, we are interested in an upper bound for Δ defined by (17), which we call "stability index". (Compare with similar definitions of the stability index in [2, 3, 9].) Theorem 2 in Section 2 gives such a bound. Namely, under the conditions of Section 2, we obtain that

$$\Delta \le K d(f_1, \tilde{f}_1), \tag{19}$$

where

$$d(f_1, \tilde{f}_1) = \int \left| \ln \frac{f_1(x)}{\tilde{f}_1(x)} \right| f_0(x) \, \mathrm{d}\mu(x), \tag{20}$$

and K is a constant explicitly calculated using c_0, c_1, c_2 along with means and variances of corresponding likelihood ratios.

Also Theorem 2 gives the following bound for the difference between the respective expected costs for the corresponding optimal tests $(\tau_*, \delta_*(\tau_*))$ and $(\tilde{\tau}_*, \tilde{\delta}_*(\tilde{\tau}_*))$:

$$\left| V(\tau_*) - \tilde{V}(\tilde{\tau}_*) \right| \le \frac{K}{2} d(f_1, \tilde{f}_1). \tag{21}$$

For example, if f_0, f_1 and $\tilde{f}_1$ are densities of exponential distributions with respective parameters λ, μ and $\mu+\varepsilon$, $\varepsilon > 0$, then from (19) and (20) we find that

$$\Delta \le K \frac{\lambda + \mu}{\lambda \mu} \varepsilon$$

(see Example 1 for details).

Remark 4.

(a) In Example 2 of Section 2 we will see that if the distortion of f_1 is given by (18) and $f_1(x)$ and $f'(x)$ have significantly different rates of vanishing, as $x \to \pm\infty$, then $d(f_1, \tilde{f}_1)$ can be infinite. Moreover, by reasons given in Remark 6 it could be conjectured that Δ in (17) does not approach zero as $\varepsilon \to 0$. Such examples explain the necessity of robust modifications of the sequential probability ratio test (see, e. g., [4, 5, 6]).

(b) Essentially, (21) is a quantitative estimation of the sensitivity of the optimal value in the optimal stopping problem (9). A rather general qualitative result on the convergence of optimal value in optimal stopping problems, for continuous-time Markov processes, can be found in [8].

2. ASSUMPTIONS, RESULTS AND EXAMPLES

Recall that the functionals W, $\tilde{W}$, V and $\tilde{V}$ were defined, respectively, in (2), (3), (6) and (7). The following theorem is implicit in [7]. It is proved in [10] in a much more general situation.

Theorem 1. For each stopping time $\tau \in \mathscr{T}_1$ there exist decision rules $\delta_*(\tau)$, $\tilde{\delta}_*(\tau) \in \mathscr{D}_\tau$ such that

$$\inf_{\delta \in \mathscr{D}_\tau} W(\tau, \delta) = W(\tau, \delta_*(\tau)) = V(\tau), \tag{22}$$

$$\inf_{\delta \in \mathscr{D}_\tau} \tilde{W}(\tau, \delta) = \tilde{W}(\tau, \tilde{\delta}_*(\tau)) = \tilde{V}(\tau). \tag{23}$$

As was explained in the preceding Section, Theorem 1 allows to apply the general theory of optimal stopping ([1, 14], see also [7]) to get the following

Corollary 1.

(a) There exist stopping times τ_* of type (11) and $\tilde{\tau}_*$ of type (13) such that

$$V(\tau_*) = \inf_{\tau \in \mathscr{T}} V(\tau), \tag{24}$$

$$\tilde{V}(\tilde{\tau}_*) = \inf_{\tau \in \mathscr{T}} \tilde{V}(\tau). \tag{25}$$

(b) The optimal decision functions $\delta_* = \delta_*(\tau_*)$, $\tilde{\delta}_* = \tilde{\delta}_*(\tilde{\tau}_*)$ are defined by (12) and (14), respectively.

Because of Corollary 1, we will deal in what follows with the optimization problem given in (24)–(25). To estimate the stability in this problem, we will use the following additional condition. Let X be a generic random variable with the distribution identical to that of $X_1, X_2, \ldots$, under P_0.

Assumption 1. There exists a constant $\gamma > 0$ such that

$$\mathrm{E}_0 \exp\left\{\gamma \left|\ln \frac{f_1(X)}{f_0(X)}\right|\right\} < \infty, \tag{26}$$

$$\mathrm{E}_0 \exp\left\{\gamma \left|\ln \frac{\tilde{f}_1(X)}{f_0(X)}\right|\right\} < \infty. \tag{27}$$

Inequalities (26) and (27) imply that the following means and variances are finite:

$$a = \mathrm{E}_0 \left[\ln \frac{f_1(X)}{f_0(X)}\right] < 0, \quad \tilde{a} = \mathrm{E}_0 \left[\ln \frac{\tilde{f}_1(X)}{f_0(X)}\right] < 0$$

($-a$ and $-\tilde{a}$ being the corresponding Kullback–Leibler information numbers), and

$$\sigma^2 = \mathrm{E}_0 \left[\ln \frac{f_1(X)}{f_0(X)} \right]^2 - a^2 > 0, \quad \tilde{\sigma}^2 = \mathrm{E}_0 \left[\ln \frac{\tilde{f}_1(X)}{f_0(X)} \right]^2 - \tilde{a}^2 > 0.$$

Let us introduce the random variables

$$Y = \ln \frac{f_1(X)}{f_0(X)} - a, \quad \tilde{Y} = \ln \frac{\tilde{f}_1(X)}{f_0(X)} - \tilde{a}.$$

From Assumption 1 it follows that there exists a number $T > 0$ such that

$$\mathrm{E}_0 e^{tY} \le e^{\max\{\sigma^2, \tilde{\sigma}^2\} t^2}, \quad \mathrm{E}_0 e^{t\tilde{Y}} \le e^{\max\{\sigma^2, \tilde{\sigma}^2\} t^2} \tag{28}$$

for all $0 \le t \le T$ (see the proof of Lemma 5, Chapt. III in [12]). Let us define:

$$g = \max \left\{ -\frac{a}{T}, -\frac{\tilde{a}}{T}, 2\sigma^2, 2\tilde{\sigma}^2 \right\} \tag{29}$$

(the terms of $-\frac{a}{T}$ and $-\frac{\tilde{a}}{T}$ will be needed for Theorem 2 below). Then for all t, $0 \le t \le T$, we get that

$$\max \left\{ \mathrm{E}_0 e^{tY}, \mathrm{E}_0 e^{t\tilde{Y}} \right\} \le \exp \left\{ \frac{g}{2} t^2 \right\}. \tag{30}$$

We have done all needed preparations to formulate our stability estimation result.

Theorem 2. Suppose that Assumption 1 holds. Then for any $\tilde{f}_1$

$$\Delta \le 2c_1, \tag{31}$$

and if

$$d(f_1, \tilde{f}_1) \le \left(\frac{c_0}{c_2} \right)^{-\frac{\max\{a, \tilde{a}\}}{2g}} \tag{32}$$

then

$$\Delta \le K d(f_1, \tilde{f}_1) \max\{1, \ln^2 d(f_1, \tilde{f}_1)\}, \tag{33}$$

where

$$K = 2c_1 \left[\left(\frac{c_0}{c_2} \right)^{\frac{\min\{a, \tilde{a}\}}{g}} + \frac{3g}{(\max\{a, \tilde{a}\})^2} + \frac{2g^2}{(\max\{a, \tilde{a}\})^4} + 1 \right]. \tag{34}$$

Corollary 2. Under Assumption 1 and (32) we get that

$$|V(\tau_*) - \tilde{V}(\tilde{\tau}_*)| \le \frac{K}{2} d(f_1, \tilde{f}_1) \max\{1, \ln^2 d(f_1, \tilde{f}_1)\} \tag{35}$$

Remark 5. Under Assumption 1, "the deviation measure" $d(f_1, \tilde{f}_1)$ is finite.

Remark 6. By virtue of definition (20) and condition (1), the measure of discrepancy between f_1 and $\tilde{f}_1$, $d(f_1, \tilde{f}_1)$, is small if $\tilde{f}_1$ does not deviate considerably from f_1 on the support of f_0, Supp_0. In this case the stability index Δ is relatively small due to (33). At the same time it does not matter the behaviour of f_1 and $\tilde{f}_1$ outside of Supp_0. For example, let

$$f_0(x) = \lambda e^{-\lambda x}, x \geq 0, \quad f_1(x) = \frac{1}{2}\mu e^{-\mu|x|}, \ \ x \in \mathbb{R}, \quad \text{and}$$

$$\tilde{f}_1(x) = \begin{cases} \frac{1}{2}\mu e^{-\mu x}, & x > 0, \\ \frac{1}{2}, & -1 \leq x \leq 0, \\ 0, & x < -1. \end{cases}$$

In this case $d(f_1, \tilde{f}_1) = 0$ and $\Delta = 0$, which means that there is no difference between the stopping rules τ_* and $\tilde{\tau}_*$ from the point of view of minimization of (2), (3) (see also (3.17) in [11]).

Example 1. (Exponential densities) Let f_0, f_1 and $\tilde{f}_1$ be the exponential densities with the respective parameters λ, μ, $\mu + \varepsilon$ (inverse to the means), $\lambda < \mu$, $0 < \varepsilon < 1$. Then

$$\ln \frac{f_1(x)}{f_0(x)} = \ln \frac{\mu}{\lambda} + (\lambda - \mu)x,$$

$$\ln \frac{\tilde{f}_1(x)}{f_0(x)} = \ln \frac{\mu + \varepsilon}{\lambda} + (\lambda - \mu - \varepsilon)x \quad (x \geq 0),$$

and therefore Assumption 1 is satisfied for any $\gamma < \lambda/(\mu + 1 - \lambda)$. Some constants involved in (34) can be calculated explicitly. For instance,

$$a = \ln \frac{\mu}{\lambda} - \frac{\mu - \lambda}{\lambda}, \quad \tilde{a} = \ln \frac{\mu + \varepsilon}{\lambda} - \frac{\mu + \varepsilon - \lambda}{\lambda}.$$

In this example, inequality (33) holds for all sufficiently small ε with

$$d(f_1, \tilde{f}_1) \leq \frac{\lambda + \mu}{\lambda \mu}\varepsilon.$$

Example 2. ("Contamination" with a "long tail" distribution) Let f_1 be the exponential density with parameter 1,

$$f_0(x) = \begin{cases} \frac{2}{\pi(1+x^2)}, & x \geq 0, \\ 0, & x < 0, \end{cases}$$

and

$$\tilde{f}_1 = (1 - \varepsilon)f_1 + \varepsilon f_0. \tag{36}$$

Since for $x \geq 0$

$$\left|\ln \frac{f_1(x)}{\tilde{f}_1(x)}\right| = \left|\ln \left((1-\varepsilon) + \frac{2\varepsilon}{\pi} \frac{e^x}{1+x^2}\right)\right| \sim x, \quad x \to \infty,$$

we obtain that $d(f_1, \tilde{f}_1) = \infty$.

Remark 7.

(a) We make a conjecture that the problem of optimal testing in this example is unstable. This means that it could be expected that the stability index Δ does not approach zero as $\varepsilon \to 0$. A reason supporting this could be the fact that in this case, as easily seen, $a = -\infty$, $\mathrm{E}_0 \left[\ln\left(f_1(X)/f_0(X)\right)\right]^2 = \infty$, while $\tilde{a}$ and $\tilde{\sigma}^2$ are finite.

(b) By a plain verification one can see that Assumption 1 is satisfied when the "tails" of all of the densities f_0, f_1 and $\tilde{f}_1$ have power orders (possibly different) of vanishing.

(c) The simplest example of an unstable testing problem can be given admitting that the error costs c_1, c_2 may depend on a "small parameter" $\varepsilon > 0$. Let μ be the counting measure on $\{0, 1\}$, f_0 and f_1 be the unit masses concentrated, respectively, at 0 and 1, and let $\tilde{f}_1(0) = \varepsilon$, $\tilde{f}_1(1) = 1 - \varepsilon$. Assume that $c_0 = 1$ and let $c_1 = c_1(\varepsilon) = c_2 = c_2(\varepsilon)$ be defined as specified below. Let $\tilde{A}$ and $\tilde{B}$, $\tilde{A} < 0 < \tilde{B}$, be any two constants. In [7] it is proved (see the proof of Theorem 2 [7]) that there exist c_1 and c_2 such that the sequential probability ratio test (13)–(14) with the constants $\tilde{A}$ and $\tilde{B}$ is a solution of the optimization problem (15). But in this case it is easy to see that, under the null-hypothesis, $\tilde{\tau}_* = \left]\tilde{A}/\ln\varepsilon\right[$, where $]a[$ stands for the minimum integer greater or equal than a. If we choose now $\tilde{A} = -\ln^2\varepsilon$, and its corresponding c_1 and c_2 as above, then $\mathrm{E}_0\tilde{\tau}_* \to \infty$, as $\varepsilon \to 0$, and thus $\tilde{V}(\tilde{\tau}_*) \geq \mathrm{E}_0\tilde{\tau}_*$ tends to infinity, as $\varepsilon \to 0$.

At the same time, it is obvious that $\tau_* \equiv 1$ and hence $V(\tau_*) = 1$.

Example 3. (Normal densities) Let f_0, f_1 and $\tilde{f}_1$ be the normal densities with zero means and standard deviations σ_0, σ_1 and $\sigma_1 + \varepsilon$, respectively. It is easy to see that Assumption 1 is satisfied. Consequently, the inequalities of Theorem 2 apply. In addition, one can calculate that

$$d(f_1, \tilde{f}_1) \leq \frac{1}{\sigma_1}\left(1 + \sigma_0^2/\sigma_1^2\right)\varepsilon.$$

Indeed, by (20),

$$d(f_1, \tilde{f}_1) = \mathrm{E}\left|\frac{X^2}{2}\left(\frac{1}{(\sigma_1+\varepsilon)^2} - \frac{1}{\sigma_1^2}\right) + \ln\left(1 + \frac{\varepsilon}{\sigma_1}\right)\right|,$$

where X is an $\mathcal{N}(0, \sigma_0^2)$-random variable. Thus,

$$d(f_1, \tilde{f}_1) \leq \frac{\sigma_0^2}{2}\left(\frac{1}{\sigma_1^2} - \frac{1}{(\sigma_1+\varepsilon)^2}\right) + \ln\left(1 + \frac{\varepsilon}{\sigma_1}\right) \leq \sigma_0^2\frac{\varepsilon}{\sigma_1^3} + \frac{\varepsilon}{\sigma_1}.$$

3. THE PROOFS

To prove Theorem 1 we need the following almost obvious lemma (see also [10]).

Lemma 1. Let $F_1 : \mathscr{X} \to \mathbb{R}$, $F_2 : \mathscr{X} \to \mathbb{R}$ be some measurable non-negative functions on a measurable space $\mathscr{X}$ with a measure μ, and let A be a measurable subset of $\mathscr{X}$.

Then for any measurable function $\phi : \mathscr{X} \to [0,1]$

$$\int_A (\phi(x)F_1(x) + (1-\phi(x))F_2(x))\,\mathrm{d}\mu(x) \geq \int_A \min\{F_1(x), F_2(x)\}\,\mathrm{d}\mu(x), \tag{37}$$

with an equality if

$$I_{\{F_1(x)<F_2(x)\}} \leq \phi(x) \leq I_{\{F_1(x)\leq F_2(x)\}} \tag{38}$$

for $x \in A$.

The proof of Theorem 1. Let us only prove (22): the proof of (23) is the same.

Let $\tau \in \mathscr{T}_1$ be any stopping time. For any $n = 1, 2, \ldots$ let us denote:

$$x^n = (x_1, \ldots, x_n) \in \mathbb{R}^n, \quad f_j^n(x^n) = \prod_{i=1}^n f_j(x_i), \quad j = 0, 1,$$

$$L(\tau, \delta) = c_1 P_0(\delta = 1) + c_2 P_1(\delta = 0).$$

Then

$$L(\tau, \delta) = \sum_{n=1}^\infty \int_{\{\tau=n\}} \left[c_1 f_0^n(x^n) I_{\{\delta=1\}} + c_2 f_1^n(x^n) I_{\{\delta=0\}}\right] \mathrm{d}\mu^n(x^n). \tag{39}$$

Applying Lemma 1 to each summand in (39) (with $\phi = I_{\{\delta=1\}}$), we have that

$$L(\tau, \delta) \geq \sum_{n=1}^\infty \int_{\{\tau=n\}} \min\{c_1 f_0^n(x^n), c_2 f_1^n(x^n)\} \mathrm{d}\mu^n(x^n), \tag{40}$$

with an equality if

$$\delta = \delta_*(\tau) = \sum_{n=1}^\infty I_{\{c_1 f_0^n(x^n) \leq c_2 f_1^n(x^n)\}} I_{\{\tau=n\}}. \tag{41}$$

It is easy to see that

$$\int_{\{\tau=n\}} \min\{c_1 f_0^n(x^n), c_2 f_1^n(x^n)\}\,\mathrm{d}\mu^n(x^n) = \mathrm{E}_0 \min\{c_1, c_2 Z_n\} I_{\{\tau=n\}},$$

where

$$Z_n = \prod_{k=1}^{n} \frac{f_1(X_k)}{f_0(X_k)}, \quad n = 1, 2, \dots,$$

so it follows from (40) that

$$L(\tau, \delta) \geq \mathrm{E}_0 \min\{c_1, c_2 Z_\tau\},$$

or, because of (2) and (6),

$$W(\tau, \delta) \geq V(\tau). \tag{42}$$

There is an equality in (42) if (41) is fulfilled. Thus, (22) follows. □

Remark 8. Because the optimal stopping rule τ_* in (10) has form (11) with $A < \ln(c_1/c_2) < B$ (see Section 3 in [11]), it follows that the optimal decision function δ defined in (41) is equivalent to (12) if $\tau = \tau_*$.

The proof of Theorem 2. According to (17) and (6) we have:

$$\begin{aligned}
\Delta = \tilde{V}(\tau_*) - \tilde{V}(\tilde{\tau}_*) &= \Big(\tilde{V}(\tau_*) - V(\tau_*)\Big) + \Big(V(\tau_*) - \tilde{V}(\tilde{\tau}_*)\Big) \\
&\leq \Big|\tilde{V}(\tau_*) - V(\tau_*)\Big| + \Big|\min_{\tau \in \{\tau_*, \tilde{\tau}_*\}} V(\tau) - \min_{\tau \in \{\tau_*, \tilde{\tau}_*\}} \tilde{V}(\tau)\Big| \\
&\leq 2 \max_{\tau \in \{\tau_*, \tilde{\tau}_*\}} \Big|V(\tau) - \tilde{V}(\tau)\Big| \\
&= 2 \max_{\tau \in \{\tau_*, \tilde{\tau}_*\}} \Big|\mathrm{E}_0 \min\{c_1, c_2 e^{S_\tau}\} - \mathrm{E}_0 \min\{c_1, c_2 e^{\tilde{S}_\tau}\}\Big|.
\end{aligned} \tag{43}$$

Let $\tau = \tau_*$ or $\tau = \tilde{\tau}_*$ and $r(x) = \min\{c_1, c_2 e^x\}$, $x \in \mathbb{R}$, $I = \mathrm{E}_0|r(S_\tau) - r(\tilde{S}_\tau)|$.

For any $n \geq 1$ we have: $I = I_1 + I_2$, where

$$I_1 = \sum_{k=1}^{n} \mathrm{E}_0\{|r(S_\tau) - r(\tilde{S}_\tau)|; \tau = k\}, \quad I_2 = \mathrm{E}_0\{|r(S_\tau) - r(\tilde{S}_\tau)|; \tau > n\}. \tag{44}$$

It is easy to see that the function r satisfies the Lipschitz condition with the constant c_1. Thus, for $n \geq 1$

$$I_1 \leq c_1 \sum_{k=1}^{n} \mathrm{E}_0 \Big| \sum_{i=1}^{k} \xi_i - \sum_{i=1}^{k} \tilde{\xi}_i \Big|,$$

where

$$\xi_i = \ln \frac{f_1(X_i)}{f_0(X_i)}, \quad \tilde{\xi}_i = \ln \frac{\tilde{f}_1(X_i)}{f_0(X_i)},$$

so

$$I_1 \leq c_1 \sum_{k=1}^{n} k \mathrm{E}_0 |\xi_1 - \tilde{\xi}_1| = c_1 \frac{n(n+1)}{2} \mathrm{E}_0 |\xi_1 - \tilde{\xi}_1|. \tag{45}$$

Thus, by (44), (45) and (20),

$$I \leq c_1 \left[\frac{n(n+1)}{2} d(f_1, \tilde{f}_1) + P(\tau > n) \right]. \tag{46}$$

Therefore, (43), (46) give that

$$\Delta \leq c_1 \left[n(n+1) d(f_1, \tilde{f}_1) + 2 \max_{\tau \in \{\tau_*, \tilde{\tau}_*\}} P(\tau > n) \right]. \tag{47}$$

To bound $P(\tau > n)$ in (47) we need first the following simple

Lemma 2. If τ_* of type (11) is a solution of the optimization problem (10), then

$$A \geq \ln(c_0/c_2). \tag{48}$$

Proof. From Theorem 5 [11] it follows that $z = e^A$ is a solution of the equation

$$c_0 + \mathrm{E}_0 \rho \left(z \frac{f_1(X)}{f_0(X)} \right) = c_2 z,$$

where $\rho(z)$ is some non-negative function. Thus,

$$c_2 z \geq c_0,$$

or

$$e^A \geq c_0/c_2,$$

so (48) follows.

Remark 9. It is obvious that if $\tilde{\tau}_*$ of type (13) is a solution of the optimization problem (9) then $\tilde{A} \geq \ln(c_0/c_2)$ as well.

Let us introduce further

$$Y_k = \ln \frac{f_1(X_k)}{f_0(X_k)} - a, \quad \tilde{Y}_k = \ln \frac{\tilde{f}_1(X_k)}{f_0(X_k)} - \tilde{a}, \quad k = 1, 2, \ldots$$

Let $A_0 = \ln(c_0/c_2)$ (which is negative, see Remark 3 (c)). Let us define:

$$\tau_{A_0} = \min\{n \geq 1 : S_n \leq A_0\}, \quad \tilde{\tau}_{A_0} = \min\{n \geq 1 : \tilde{S}_n \leq A_0\},$$

and

$$n_0 = \left] \frac{A_0}{a} \right[.$$

For $\tau = \tau_*$ we obtain:

$$P_0(\tau > n) \leq P_0(\tau_{A_0} > n) = P_0 \left(\min_{1 \leq k \leq n} S_k > A_0 \right)$$

$$\leq P_0(S_n > A_0) = P_0\left(\sum_{k=1}^{n} Y_k > A_0 - na\right). \tag{49}$$

To find an upper bound for the right-hand side of (49), we use the following inequality:

$$P_0\left(\sum_{k=1}^{n} Y_k \geq x\right) \leq \exp\left\{-\frac{x^2}{2gn}\right\}, \tag{50}$$

which is valid under conditions (30) for $0 \leq x \leq ngT$, (see Theorem 15, Chapt. III in [12]). Using the choice of the constant g as in (29) we see that for any $n \geq n_0$

$$0 \leq A_0 - na \leq -na = n\frac{-a}{T}T \leq ngT.$$

Thus, applying (50) with $x = A_0 - na$ to (49) we obtain for $n \geq n_0$:

$$P_0\left(\sum_{k=1}^{n} Y_k > A_0 - na\right) \leq e^{aA_0/g}e^{-a^2 n/(2g)}. \tag{51}$$

Analogously,

$$P_0\left(\sum_{k=1}^{n} \tilde{Y}_k > A_0 - n\tilde{a}\right) \leq e^{\tilde{a}A_0/g}e^{-\tilde{a}^2 n/(2g)} \tag{52}$$

for all $n \geq \left]\frac{A_0}{\tilde{a}}\right[$.

Let $\lambda = \min\{a, \tilde{a}\}$, $\mu = \max\{a, \tilde{a}\}$, $n_* = \left]\frac{A_0}{\mu}\right[$, $K_0 = 2\exp\{\frac{\lambda A_0}{g}\}$, $\kappa = \frac{\mu^2}{2g}$ (recall that A_0, λ, μ are negative numbers).

Then, combining (47)–(50) together we conclude that for all $n \geq n_*$

$$\Delta \leq c_1\left[n(n+1)d + K_0 e^{-\kappa n}\right], \tag{53}$$

where $d = d(f_1, \tilde{f}_1)$.

The first affirmation of Theorem 2 immediately follows from (43).

Let now

$$d \leq \exp\left\{-\frac{A_0\kappa}{\mu}\right\}. \tag{54}$$

Then choosing

$$n = \left]\frac{1}{\kappa}\ln\frac{1}{d}\right[, \tag{55}$$

it is easy to see that $n \geq n_*$. Making use of (55) in (53), we get:

$$\begin{aligned}\Delta &\leq c_1\left[(\frac{1}{\kappa}\ln\frac{1}{d}+1)(\frac{1}{\kappa}\ln\frac{1}{d}+2)d + K_0 d\right]\\ &\leq c_1 d\max\left\{1, \ln^2 d\right\}\left[\frac{1}{\kappa^2}+\frac{3}{\kappa}+2+K_0\right].\end{aligned} \tag{56}$$

The last inequality yields the desired stability inequality (33) with the constant defined in (34). Inequality (35) in Corollary 2 follows from (43) and (56). □

ACKNOWLEDGEMENTS

The work of E. Gordienko and A. Novikov was partially supported by Mexico's National Council for Science and Technology (CONACyT), under Grant CB-2005-C01-49854-F.

The authors wish to thank the anonymous referees for valuable suggestions on improvement of the previous version of the paper.

(Received April 8, 2008.)

REFERENCES

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

[2] E. I. Gordienko and F. S. Salem: Estimates of stability of Markov control processes with unbounded costs. Kybernetika *36* (2000), 195–210.

[3] E. I. Gordienko and A. A. Yushkevich: Stability estimates in the problem of average optimal switching of a Markov chain. Math. Methods Oper. Res. *57* (2003), 345–365.

[4] P. J. Huber: A robust version of the probability ratio test. Ann. Math. Statist. *36* (1965), 1753–1758.

[5] A. Kharin: On robustifying of the sequential probability ratio test for a discrete model under "contaminations". Austrian J. Statist. *3* (2002), 4, 267–277.

[6] A. Kharin: Robust sequential testing of hypotheses on discrete probability distributions. Austrian J. Statist. *34* (2005), 2, 153–162.

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

[8] V. Mackevičius: Passage to the limit in problems of optimal stopping of Markov processes (in Russian). Litovsk. Mat. Sb. (Russian) *13* (1973), 1, 115–128, 236.

[9] R. Montes-de-Oca, A. Sakhanenko, and F. Salem-Silva: Estimates for perturbations of general discounted Markov control chains. Appl. Math. *30* (2003), 287–304.

[10] A. Novikov: Optimal sequential tests for two simple hypotheses. Sequential Analysis *28* (2009), No. 2.

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

[12] V. V. Petrov: Sums of Independent Random Variables. Springer, New York 1975.

[13] P. X. Quang: Robust sequential testing. Ann. Statist. *13* (1985), 638–649.

[14] A. N. Shiryayev: Statistical Sequential Analysis. Nauka, Moscow 1969. (In Russian.)

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

[16] J. Whitehead: The Design and Analysis of Sequential Clinical Trials. Wiley, New York 1997.

Evgueni Gordienko, Andrey Novikov, and Elena Zaitseva, Departamento de Matemáticas, Universidad Autónoma Metropolitana – Unidad Iztapalapa, San Rafael Atlixco 186, col. Vicentina, México D.F. Mexico.

e-mails: gord@xanum.uam.mx, an@xanum.uam.mx, lenagordi@hotmail.com