APPLICATIONES MATHEMATICAE
40,1 (2013), pp. 109–116

E. Gordienko, J. Ruiz de Chávez and A. García (Mexico City)

NOTE ON STABILITY ESTIMATION IN SEQUENTIAL HYPOTHESIS TESTING

Abstract. We introduce a quantitative measure Δ of stability in optimal sequential testing of two simple hypotheses about a density of observations: $f = f_0$ versus $f = f_1$. The index Δ represents an additional cost paid when a stopping rule optimal for the pair (f_0, f_1) is applied to test the hypothesis $f = f_0$ versus a "perturbed alternative" $f = \widetilde{f}_1$. An upper bound for Δ is established in terms of the total variation distance between $f_1(X)/f_0(X)$ and $\widetilde{f}_1(X)/f_0(X)$ with $X \sim f_0$.

1. Problem setting. Consider two probability densities f_0 and f_1 with respect to a given σ-finite measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, and suppose that a sequence $X_1, X_2, \ldots$ of i.i.d. random variables (observations defined on a suitable probability space $(\Omega, \mathcal{F})$) has a common density f_0 or f_1. The sequential hypothesis testing f_0 versus f_1 we deal with consists in finding an a.s. *finite stopping time* τ and a terminal decision function δ which minimize the following cost functional:

$$W(\tau, \delta) := c_0 E_0 \tau + c_1 \alpha(\tau, \delta) + c_2 \beta(\tau, \delta),$$

where E_0 represents the expectation corresponding to the density f_0, α and β are probabilities of errors of type I and type II, respectively, and c_0, c_1, c_2 are given "penalty costs", which are strictly positive.

In [N1] and [GNZ] it was proven (using the results from [L]) that the above problem reduces to the minimization over stopping times of the functional

$$V(\tau) := c_0 E_0 \tau + E_0 \min\{c_1, c_2 e^{S_\tau}\}, \tag{1}$$

2010 *Mathematics Subject Classification*: Primary 62F03; Secondary 62L10.
Key words and phrases: sequential testing of hypotheses, uncertainty about hypothesis, stability index, stability inequality.

DOI: 10.4064/am40-1-7

where $S_0 := 0$ and $S_n := \sum_{k=1}^n \log[f_1(X_k)/f_0(X_k)]$, $n \geq 1$; here X_k, $k \geq 1$, are i.i.d. random variables with density f_0. Furthermore, for certain constants $-\infty < A \leq 0 \leq B < \infty$, the a.s. finite stopping time

$$\tau_* = \inf\{n \geq 0 : S_n \notin (A, B)\} \tag{2}$$

is optimal, that is, $V(\tau_*) = \inf_\tau V(\tau)$.

To set the stability estimating problem we assume that the density f_1 is interpreted as some "theoretical approximation" to a "real" alternative density $\widetilde{f}_1$ about which the statistician is not certain. Such an alternative can be set, for example, as in the problems of robust statistics, namely $\widetilde{f}_1 := (1-\varepsilon)f_1 + \varepsilon f'$ (see e.g. [H] and [J], [K1], [K2] on robust procedures for sequential hypothesis testing).

For the pair of densities $(f_0, \widetilde{f}_1)$ the cost functional $\widetilde{V}(\tau)$ is defined as in (1) with S_τ replaced by $\widetilde{S}_\tau$:

$$\widetilde{V}(\tau) := c_0 E_0 \tau + E_0 \min\{c_1, c_2 e^{\widetilde{S}_\tau}\}, \tag{3}$$

where

$$\widetilde{S}_0 := 0, \qquad \widetilde{S}_n := \sum_{k=1}^n \log\left[\frac{\widetilde{f}_1(X_k)}{f_0(X_k)}\right], \qquad X_k \sim f_0, \quad n \geq 1.$$

Again, there exist constants $-\infty < \widetilde{A} \leq 0 \leq \widetilde{B} < \infty$ such that the stopping time

$$\widetilde{\tau}_* := \inf\{n \geq 0 : \widetilde{S}_n \notin (\widetilde{A}, \widetilde{B})\} \tag{4}$$

is optimal in testing f_0 against $\widetilde{f}_1$, i.e. $\widetilde{V}(\widetilde{\tau}_*) = \inf_\tau \widetilde{V}(\tau)$.

Uncertainty about $\widetilde{f}_1$ does not allow one to find $\widetilde{A}$ and $\widetilde{B}$. However, we suppose that for the pair of densities (f_0, f_1) the constants (A, B) in (2) can be obtained (at least "theoretically"), and that the stopping rule determined by (A, B) is applied to $\widetilde{S}_n$. Thus, in order to test f_0 against $\widetilde{f}_1$, instead of the inaccessible stopping rule (4), the following stopping rule is exercised: stop on first exit of the sequence $\widetilde{S}_n$, $n \geq 1$, from the interval (A, B). This procedure generates the stopping time $\widetilde{\tau}$:

$$\widetilde{\tau} := \inf\{n \geq 0 : \widetilde{S}_n \notin (A, B)\}. \tag{5}$$

The following *stability index* (see, for instance, [GY]) measures an extra cost paid for using $\widetilde{\tau}$ instead of the optimal stopping time $\widetilde{\tau}_*$:

$$\Delta = \widetilde{V}(\widetilde{\tau}) - \widetilde{V}(\widetilde{\tau}_*) \geq 0, \tag{6}$$

where the functional $\widetilde{V}$ is as defined in (3).

The goal of this note is to prove the following "stability inequality":

$$\Delta \le \mathbf{Var}\Big(\frac{f_1(X)}{f_0(X)}, \frac{\widetilde{f}_1(X)}{f_0(X)}\Big)\Big\{(c_0+c_1)\Big[\log_\beta\Big(\frac{1}{4}\mathbf{Var}\Big(\frac{f_1(X)}{f_0(X)}, \frac{\widetilde{f}_1(X)}{f_0(X)}\Big)\Big)\Big]^2 + b\log_\beta\Big(\frac{1}{4}\mathbf{Var}\Big(\frac{f_1(X)}{f_0(X)}, \frac{\widetilde{f}_1(X)}{f_0(X)}\Big)\Big)\Big\}, \tag{7}$$

where $\beta \in (0,1)$ and $b < \infty$ are certain constants (which can be evaluated for some particular classes of densities), X has density f_0, and **Var** denotes the total variation distance.

REMARK 1. In the paper [GNZ], which dealt with the same problem, a different "stability inequality" was obtained. After the publication of the paper, it was realized that the proof given in [GNZ] is correct only for another definition of the stability index Δ, which is different from (and less practical than) that given in (6). Moreover, for such a proof rather restrictive assumptions on densities have to be used.

2. Stability index estimation

ASSUMPTION 1. Let X denote a random variable with density f_0. We assume that there exists a Borel set $D \subset \mathbb{R}$ such that $\int_D f_0(x)\,dx = 1$, and $f_0(x), f_1(x), \widetilde{f}_1(x) > 0$ for all $x \in D$.

In what follows we denote $Y = f_1(X)/f_0(X)$ and $\widetilde{Y} = \widetilde{f}_1(X)/f_0(X)$.

THEOREM 1. *There are constants $\beta \in (0,1)$ and $b < \infty$ such that*

$$\Delta \le \mathbf{Var}(Y,\widetilde{Y})\Big\{(c_0+c_1)\Big[\log_\beta\Big(\frac{1}{4}\mathbf{Var}(Y,\widetilde{Y})\Big)\Big]^2 + b\log_\beta\Big(\frac{1}{4}\mathbf{Var}(Y,\widetilde{Y})\Big)\Big\}. \tag{8}$$

REMARK 2. The proof in Section 4 shows that if in (7) and (8), $\mathbf{Var}(Y,\widetilde{Y})$ is small enough, then β and b are completely determined by the densities (f_0, f_1), and so can be chosen independently of $\widetilde{f}_1$.

To prove (8) we use the well-known exponential inequalities (from [S2]) for the probabilities of "tails" of stopping times defined in (2), (4) and (5). In general, these inequalities are very rough. This implies that the constant $\beta < 1$ in (8) can be too close to 1. (And in some examples this is indeed the case.) For this reason, inequality (8) does not have much quantitative value, and the simplest exponential bound on "tail probabilities" is used only to indicate the way to obtain inequalities of this type. The advantage of the inequalities in [S2] is that they hold without any additional restrictions. Using more accurate bounds on "tail probabilities", which hold under certain restrictions on the densities f_0, f_1 and $\widetilde{f}_1$, one can prove more precise versions of (8) by applying the same arguments as in Section 4. For example, assuming the existence of exponential moments of Y and $\widetilde{Y}$, the inequalities for "tail probabilities" of stopping times can be derived from exponential inequalities for sums of i.i.d. random variables (see for instance [S1], [PU]).

In any case, inequality (8) ensures that if the "approximating density" f_1 approaches to $\widetilde{f}_1$ in the sense that $\mathbf{Var}(Y, \widetilde{Y}) \to 0$, then the stability index $\varDelta$ (defined in (6)) tends to zero. To grasp what the rate of convergence $\varDelta \to 0$ could be, we present in Section 3 some evaluations of $\varDelta$ in the simplest example of exponential densities. These data are obtained by simulations and numerical optimization.

3. Calculations of stability index for exponential densities. In functionals (1) and (3), we have chosen $c_0 = 1$ and $c_1 = c_2 = 100$. Let $f_0 \sim \mathrm{Exp}(\lambda = 1)$, $f_1 \sim \mathrm{Exp}(\lambda = 2)$, $\widetilde{f}_{1,\varepsilon} \sim \mathrm{Exp}(\lambda = 2+\varepsilon)$, $\varepsilon > 0$. We denote $Y = f_1(X)/f_0(X)$, $\widetilde{Y}_\varepsilon = \widetilde{f}_{1,\varepsilon}(X)/f_0(X)$, where $X \sim f_0$. For this example it is easy to calculate $\mathbf{Var}(Y, \widetilde{Y}_\varepsilon)$ in (8) (see Table 1). The constants β and b in (8) can also be evaluated, but they are of little help for estimating the stability index $\varDelta$ quantitatively (since β is too close to 1). Table 1 displays the results of numerical calculations of $\varDelta = \varDelta(\varepsilon)$. To carry out these calculations we used Monte-Carlo optimizations of the corresponding stopping rules in (2), (4) and (5).

Table 1. Numerical evaluation of $\varDelta$

(f_0, f_1)		$V(\tau_*) = \inf_\tau V(\tau) = 15.1995$			
$(f_0, \widetilde{f}_{1,\varepsilon})$	ε	$\mathbf{Var}(Y, \widetilde{Y}_\varepsilon)$	$\widetilde{V}(\widetilde{\tau}_*) = \inf_\tau \widetilde{V}(\tau)$	$\widetilde{V}(\widetilde{\tau})$	$\varDelta = \varDelta(\varepsilon)$
	0.1	0.1464	13.6410	13.8312	0.1902
	0.01	0.01448	15.0676	15.2301	0.1626
	10^{-3}	0.001446	15.2509	15.3564	0.1055
	10^{-5}	0.00001446	15.2966	15.2966	0.1024
	10^{-6}	1.1931×10^{-6}	15.2597	15.3289	0.0692

4. Proof of inequality (8). For the random variable X with density f_0 we denote

$$\mu(f_1, \widetilde{f}_1) := \mathbf{Var}\left(\frac{f_1(X)}{f_0(X)}, \frac{\widetilde{f}_1(X)}{f_0(X)}\right) \equiv \mathbf{Var}(Y, \widetilde{Y}). \tag{9}$$

In view of Assumption 1, when calculating $\mathbf{Var}(Y, \widetilde{Y})$ in (9) we can restrict ourselves to the set D where $Y, \widetilde{Y} \in (0, \infty)$. Then using the definition of the total variation distance and the monotonicity of log, it can be easily shown that $\mu(f_1, \widetilde{f}_1) = \mathbf{Var}(\log Y, \log \widetilde{Y})$.

For $k = 1, 2, \ldots$ we set

$$Z_k = \log\left[\frac{f_1(X_k)}{f_0(X_k)}\right], \qquad \widetilde{Z}_k = \log\left[\frac{\widetilde{f}_1(X_k)}{f_0(X_k)}\right], \tag{10}$$

with Z and $\widetilde{Z}$ being the corresponding generic random variables. In (10), $X_1, X_2, \ldots$ are i.i.d. random variables with density f_0. Also as before we denote

$$S_n = \sum_{k=1}^{n} Z_k, \qquad \widetilde{S}_n = \sum_{k=1}^{n} \widetilde{Z}_k, \qquad n \geq 1, \qquad S_0 = \widetilde{S}_0 := 0.$$

Let (A, B) and $(\widetilde{A}, \widetilde{B})$ be the pairs of constants in (2) and (4), respectively. The stopping times obtained using rules as in (2) applied to $\{Z_k\}$ and $\{\widetilde{Z}_k\}$ will be denoted by $\Psi(Z)$ and $\Psi(\widetilde{Z})$ respectively. In particular, $\Psi(\widetilde{Z})$ is the stopping time defined in (5). On the other hand, the stopping times obtained from rules as in (4) applied to $\{Z_k\}$ and $\{\widetilde{Z}_k\}$ will be denoted by $\widetilde{\Psi}(Z)$ and $\widetilde{\Psi}(\widetilde{Z})$ respectively. Considering the cases of practical interest, we can assume that $c_0 < c_1$. Using the results obtained in [N1], [N2] it can be proven that $A, \widetilde{A} \geq A_* := \log(c_0/c_1)$, and that the constant B_* can be chosen independently of $\widetilde{f}_1$ in such a way that $B_* \geq B, \widetilde{B}$. Since $f_0 \neq f_1$ and $f_0 \neq \widetilde{f}_1$, it can be easily shown that there are constants $\gamma, \widetilde{\gamma}, \delta, \widetilde{\delta} > 0$ such that

$$P(Z \leq -\gamma) \geq \delta, \qquad P(\widetilde{Z} \leq -\widetilde{\gamma}) \geq \widetilde{\delta}. \tag{11}$$

We suppose that for the densities under consideration, constants γ_* δ_* are given for which the following inequalities hold: $0 < \gamma_* \leq \min(\gamma,\, \widetilde{\gamma})$ and $0 < \delta_* \leq \min(\delta, \widetilde{\delta})$.

In view of (4)–(6) and the above definition of Ψ, we have

$$\begin{aligned} \Delta &= \widetilde{V}[\Psi(\widetilde{Z})] - \widetilde{V}[\widetilde{\Psi}(\widetilde{Z})] \\ &= \widetilde{V}[\Psi(\widetilde{Z})] - V[\Psi(Z)] + \max_{\Psi' \in \{\Psi, \widetilde{\Psi}\}} V[\Psi'(Z)] - \max_{\Psi' \in \{\Psi, \widetilde{\Psi}\}} \widetilde{V}[\Psi'(\widetilde{Z})] \\ &\leq 2 \max_{\Psi' \in \{\Psi, \widetilde{\Psi}\}} |V[\Psi'(Z)] - \widetilde{V}[\Psi'(\widetilde{Z})]|. \end{aligned} \tag{12}$$

Estimation of each term under the “ max ” sign in (12) is accomplished in the same manner. Choose, for example, $\Psi' = \Psi$. Denoting $g(x) = \min\{c_1, c_2 e^x\}$, for $x \in [0, \infty)$, $\tau = \Psi(Z)$ and $\widetilde{\tau} = \Psi(\widetilde{Z})$, we obtain (see (1) and (3)) for any $n \geq 1$ that

$$(13)\quad |V[\Psi(Z)] - \widetilde{V}[\Psi(\widetilde{Z})]| \le c_0|E_0\tau - E_0\widetilde{\tau}| + |E_0 g(S_\tau) - E_0 g(\widetilde{S}_{\widetilde{\tau}})|$$
$$\le c_0|E_0\tau I_{\{\tau\le n\}} - E_0\widetilde{\tau} I_{\{\widetilde{\tau}\le n\}}| + |E_0 g(S_\tau) I_{\{\tau\le n\}} - E_0 g(\widetilde{S}_{\widetilde{\tau}}) I_{\{\widetilde{\tau}\le n\}}|$$
$$+ c_0 E_0\tau I_{\{\tau>n\}} + c_0 E_0\widetilde{\tau} I_{\{\widetilde{\tau}>n\}} + E_0 g(S_\tau) I_{\{\tau>n\}} + E_0 g(\widetilde{S}_{\widetilde{\tau}}) I_{\{\widetilde{\tau}>n\}}$$
$$=: I_1 + I_2 + I_3 + I_4.$$

On the right-hand side of (13), $\tau I_{\{\tau\le n\}}$ and $g(S_\tau) I_{\{\tau\le n\}}$ are nonnegative bounded functions of the random vector $\bar{Z}_n = (Z_1,\ldots,Z_n)$. Thus denoting $\widetilde{\bar{Z}}_n = (\widetilde{Z}_1,\ldots,\widetilde{Z}_n)$ we find that in (13),

$$(14)\qquad I_1 \le c_0(n/2)\,\mathbf{Var}(\bar{Z}_n, \widetilde{\bar{Z}}_n), \quad I_2 \le (1/2)\,\mathbf{Var}(\bar{Z}_n, \widetilde{\bar{Z}}_n).$$

Let us show by induction that

$$(15)\qquad \mathbf{Var}(\bar{Z}_n, \widetilde{\bar{Z}}_n) \le n\,\mathbf{Var}(Z, \widetilde{Z}), \qquad n = 1, 2, \ldots.$$

Indeed, by the triangle inequality,

$$(16)\quad \mathbf{Var}[(Z_1,\ldots,Z_n,Z_{n+1}),(\widetilde{Z}_1,\ldots,\widetilde{Z}_n,\widetilde{Z}_{n+1})]$$
$$\le \mathbf{Var}[(Z_1,\ldots,Z_n,Z_{n+1}),(\widetilde{Z}_1,\ldots,\widetilde{Z}_n,Z_{n+1})]$$
$$+ \mathbf{Var}[(\widetilde{Z}_1,\ldots,\widetilde{Z}_n,Z_{n+1}),(\widetilde{Z}_1,\ldots,\widetilde{Z}_n,\widetilde{Z}_{n+1})].$$

For every function $\varphi : \mathbb{R}^{n+1} \to \mathbb{R}$ bounded by 1 we obtain (by independence)

$$|E_0\varphi(Z_1,\ldots,Z_n,Z_{n+1}) - E_0\varphi(\widetilde{Z}_1,\ldots,\widetilde{Z}_n,Z_{n+1})|$$
$$\le \int_{\mathbb{R}} |E_0\varphi(Z_1,\ldots,Z_n,z) - E_0\varphi(\widetilde{Z}_1,\ldots,\widetilde{Z}_n,z)|\,P_Z(dz)$$
$$\le \mathbf{Var}(\bar{Z}_n, \widetilde{\bar{Z}}_n) \le n\,\mathbf{Var}(Z, \widetilde{Z}).$$

Consequently, the first summand on the right-hand side of (16) is less than $n\,\mathbf{Var}(Z, \widetilde{Z})$. Applying again independence and conditioning on $\widetilde{Z}_1,\ldots,\widetilde{Z}_n$, we conclude that the second term on the right-hand side of (16) is less than $\mathbf{Var}(Z, \widetilde{Z})$.

To bound the terms I_3 and I_4 in (13) we use the simple exponential inequalities for $P_0(\tau > n)$, $P_0(\widetilde{\tau} > n)$ obtained in [S2]. Denoting

$$m := [(B_* - A_*)/\gamma_*] + 1, \quad \beta := (1 - \delta_*^m)^{1/m},$$

($[\cdot]$ stands for integer part) we can see that, from the result in [S2],

$$(17)\quad \max\{P_0(\tau > n), P_0(\widetilde{\tau} > n)\} \le \gamma\beta^n,\ n = 1, 2, \ldots, \text{ where } \gamma = \beta^{-m}.$$

Using these inequalities, after simple calculations we find that

$$(18)\qquad I_3 \le \beta^{n+1} n\left(2c_0\frac{\gamma(2-\beta)}{\beta(1-\beta)}\right), \qquad n = 1, 2, \ldots.$$

Applying (17) again we see that

$$I_4 \le \beta^{n+1}(2c_1\gamma/\beta), \quad n = 1, 2, \ldots. \tag{19}$$

Combining inequalities (12)–(15), (18) and (19) we establish that for every $n \ge 1$,

$$\begin{aligned} \Delta &\le 2\bigg[\frac{c_0}{2} n^2 \mu(f_1, \widetilde{f}_1) + \frac{c_1}{2} n\mu(f_1, \widetilde{f}_1) \\ &\quad + \beta^{n+1} n\bigg(2c_0 \frac{\gamma(2-\beta)}{\beta(1-\beta)}\bigg) + \beta^{n+1}(2c_1\gamma/\beta)\bigg] \\ &\le (c_0 + c_1) n^2 \mu(f_1, \widetilde{f}_1) + \beta^{n+1} n\bigg[\frac{4\gamma}{\beta}\bigg(c_0 \frac{2-\beta}{1-\beta} + c_1\bigg)\bigg]. \end{aligned} \tag{20}$$

Now we choose in (20) $n = [\log_\beta \frac{1}{4}\mu(f_1, \widetilde{f}_1)]$ (the integer part). Then inequality (8) follows from (20). ■

REMARK 3. In view of (11), if $\mathbf{Var}(Z, \widetilde{Z})$ is small enough, the constants γ_* and δ_* can be chosen in such a way that they do not depend on the density $\widetilde{f}_1$.

Acknowledgements. The authors wish to thank the anonymous reviewers for suggestions and comments on an earlier draft. The authors are grateful to Professor A. Novikov for fruitful discussions on the subject of this paper.

References

[GNZ] E. Gordienko, A. Novikov and E. Zaitseva, *Stability estimates in optimal sequential hypothesis testing*, Kybernetika 45 (2009), 331–344.

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

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

[J] J. Jurečková and P. K. Sen, *Robust Statistical Procedures. Asymptotics and Interrelations*, Wiley, New York, 1996.

[K1] A. Y. Kharin, *On robustifying of the sequential probability ratio test for discrete model under "contaminations"*, Austral. J. Statist. 3 (2002), 267–277.

[K2] A. Y. Kharin and D. V. Kishilov, *Analysis of the accuracy and stability of the sequential testing of hypotheses on the parameters of a Markov chain*, Vesti Nats. Akad. Navuk Belarusi Ser. Fiz.-Mat. Navuk 2005, no. 4, 30–35.

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

[N1] A. Novikov, *Optimal sequential tests for two simple hypotheses based on independent observations*, Int. J. Pure Appl. Math. 45 (2008), 291–314.

[N2] A. Novikov, *Optimal sequential tests for two simple hypotheses*, Sequential Anal. 28 (2009), 188–217.

[PU] I. S. Pinelis and S. A. Utev, *Sharp exponential estimates for sums of independent random variables*, Theory Probab. Appl. 34 (1989), 340–346.
[S1] W. Stadje, *Inequalities for first-exit probabilities and expected first-exit times of a random walk*, Comm. Statist. Stoch. Models 12 (1996), 103–120.
[S2] C. Stein, *A note on cumulative sums*, Ann. Math. Statist. 17 (1946), 498–499.

E. Gordienko, J. Ruiz de Chávez, A. García
Departamento de Matemáticas
Universidad Autónoma Metropolitana-Iztapalapa
San Rafael Atlixco 186
Col. Vicentina
C.P. 09340, Ciudad de México, D.F., México
E-mail: gord@xanum.uam.mx
jrch@xanum.uam.mx
agar@xanum.uam.mx

Received on 13.7.2012;
revised version on 16.11.2012 (2143)