Теорія ймовірностей
та математична статистика
Вип. 1(100)/2019, с. 157–166

Teoriya Ĭmovirnosteĭ
ta Matematychna Statystyka
No. 1(100)/2019, pp. 157–166

UDC 519.23

ROBUSTNESS OF SEQUENTIAL HYPOTHESES TESTING FOR HETEROGENEOUS INDEPENDENT OBSERVATIONS

A. YU. KHARIN, T. T. TU

Abstract. The problem of robustness for truncated sequential tests of two simple hypotheses is considered for the model of heterogeneous independent observations under distortions. An approach for performance characteristics calculation is proposed. Asymptotic analysis of robustness is performed. A family of robustified sequential tests is constructed. Numerical examples illustrate the theoretical results.

Key words and phrases. Truncated sequential test, heterogeneous observations, distortions, robustness, error probabilities, expected sample size.

2010 *Mathematics Subject Classification.* 62L10; 62F35.

1. Introduction

In different applications of statistical hypotheses testing the sequential approach [14] can be effectively used to minimize the expected number of observations [2, 12]. The situation where the hypothetical model of data does not exactly describes the real data is quite often in practice [3–5]. The problem of finding analytic dependencies of performance characteristics on the test parameters and the inverse problem (in [9] the latter was studied for the many-alternative sequential test) is still open for most models of data, especially for the purpose of robustness analysis [10]. An approach to solve the problem for the case of independent identically distributed random observations is proposed in [6, 7]. Here we generalize the approach for robustness analysis of sequential tests developed in [8] and consider the model of heterogeneous independent data that often appears in practice.

2. Mathematical model

Let $\{x_1, x_2, \dots\}$ be the observations of a sequence of independent random variables $\{X_n, n \geq 1\}$, defined on the same probability space $(\Omega, \mathbb{F}, \mathbf{P})$ with probability density functions $\{p_n(x, \theta), x \in \mathbb{R}^1, n \geq 1\}$ respectively, where $\theta \in \mathbb{R}^m$ is an unknown vector of parameters. Consider two simple hypotheses:

$$\mathbf{H}_0 : \theta = \theta^0 \text{ vs. } \mathbf{H}_1 : \theta = \theta^1, \tag{1}$$

where $\theta^0, \theta^1 \in \mathbb{R}^m$ are known vectors, $\theta^0 \neq \theta^1$.

Denote the accumulated log-likelihood ratio statistic for n observations:

$$\Lambda_n = \Lambda_n(x_1, x_2, \dots, x_n) = \sum_{i=1}^{n} \lambda_i, \tag{2}$$

where $\lambda_i = \ln\left(\dfrac{p_i(x_i, \theta^1)}{p_i(x_i, \theta^0)}\right)$ is the log-likelihood ratio calculated on the observation x_i, and $p_i(x, \theta)$ is the probability density function of x_i provided the true parameter value is θ.

After n observations one makes the decision:

$$d = \mathbf{1}_{[C_+, +\infty)}(\Lambda_n) + 2 \cdot \mathbf{1}_{(C_-, C_+)}(\Lambda_n), \tag{3}$$

where the thresholds C_- and C_+ are the parameters of the test. According to Wald [14], $C_+ = \ln((1-\beta_0)/\alpha_0)$, $C_- = \ln(\beta_0/(1-\alpha_0))$, where α_0, β_0 are given upper bounds for error probabilities of types I and II respectively. Decisions $d=0$ and $d=1$ mean stopping of the observation process and acceptance of the correspondent hypothesis; decision $d=2$ means that the next observation needs to be made.

Due to the restriction on the number of observations, the sequential probability ratio test (SPRT) [14] may be stopped by truncating at the observation number $n=M$. The Wald truncated SPRT (TSPRT) is formulated as follows: if the SPRT has not lead to terminal decision for $n \le M-1$,

$$\begin{cases} \text{reject } \mathbf{H}_0 & \text{if } \Lambda_M > 0, \\ \text{accept } \mathbf{H}_0 & \text{if } \Lambda_M \le 0. \end{cases} \tag{4}$$

Put $S_1^{(k)}(x) = \mathbf{P}_k\{\Lambda_1 < x\}$, and for $n > 1$:

$$S_n^{(k)}(x) = \mathbf{P}_k\big\{\Lambda_n < x, \text{ and } \Lambda_i \in (C_-, C_+), i = \overline{1, n-1}\big\}, \quad k \in \{0,1\}.$$

The function $S_n^{(k)}(x)$ satisfies the following recurrent relation:

$$S_n^{(k)}(x) = \int_{C_-}^{C_+} F_n^{(k)}(x-y)\, dS_{n-1}^{(k)}(y), \quad n > 1,\ k \in \{0,1\}, \tag{5}$$

where $F_n^{(k)}(x)$ is the cumulative distribution function of λ_n under hypothesis $\mathbf{H}_k$, and $S_1^{(k)}(x) = F_1^{(k)}(x)$.

Denote the error probabilities of type I and II and the random number of required observations for TSPRT by α_M, β_M, N_M respectively. Then, we have

$$\begin{aligned} \alpha_M = &\sum_{i=1}^{M-1} \mathbf{P}_0\{\Lambda_j \in (C_-, C_+), j = \overline{1, i-1}, \Lambda_i \ge C_+\} + \\ &+ \mathbf{P}_0\{\Lambda_j \in (C_-, C_+), j = \overline{1, M-1}, \Lambda_M > 0\} = \\ &= 1 - S_1^{(0)}(C_+) + \sum_{i=2}^{M-1} \Big(S_{i-1}^{(0)}(C_+) - S_{i-1}^{(0)}(C_-) - S_i^{(0)}(C_+)\Big) + \\ &+ S_{M-1}^{(0)}(C_+) - S_{M-1}^{(0)}(C_-) - S_M^{(0)}(0) = \\ &= 1 - \sum_{i=1}^{M-1} S_i^{(0)}(C_-) - S_M^{(0)}(0). \end{aligned}$$

Analogously, $\beta_M = \sum_{i=1}^{M-1} S_i^{(1)}(C_-) + S_M^{(1)}(0)$. Using Theorem 12.2 [1] we obtain:

$$\mathbf{E}^{(k)}(N_M) = \sum_{i=1}^{M} \mathbf{P}_k\{N_M \ge i\} = 1 + \sum_{i=1}^{M-1} \Big(S_i^{(k)}(C_+) - S_i^{(k)}(C_-)\Big), \quad k \in \{0,1\}.$$

Next, we will use numerical method for approximating the values of functions $S_n^{(k)}(x)$, $n \ge 2, k \in \{0,1\}$. Without loss of generality, assume that $\mathbf{H}_0$ is true. Let $H > 1$ be a fixed positive integer, and $\{t_i, i = \overline{1, H}\}$ be a partition of $[C_-, C_+]$, where

$$t_i = C_- + (i-1)h,\ i = \overline{1, H}, \quad h = (C_+ - C_-)/(H-1).$$

For the partition $\{t_i, i = \overline{1, H}\}$ defined above, the value t_{i_0} that has the least deviation from 0, is set to be zero. If $F_n^{(k)}(x), n \ge 1, k \in \{0,1\}$, have continuous derivatives of second order in $[C_- - C_+, C_+ - C_-]$, then using Theorem 3 [11] the Riemann–Stieltjes integral $\int_{C_-}^{C_+} F_n^{(0)}(x-y)dS_{n-1}^{(0)}(y)$ can be expanded at $h \to 0$:

$$S_n^{(0)}(x) = \frac{1}{2}\sum_{j=1}^{H-1}\left[F_n^{(0)}(x-t_j) + F_n^{(0)}(x-t_{j+1})\right]\left[S_{n-1}^{(0)}(t_{j+1}) - S_{n-1}^{(0)}(t_j)\right] + O(h^2). \quad (6)$$

Denote $f_i^{(j)} = S_j^{(0)}(t_i), j = \overline{1,M}, i = \overline{1,H}$. For $2 \le n \le M$, we obtain the following systems of linear equations:

$$\begin{aligned} f_i^{(n)} &= -\frac{1}{2}\left[F_n^{(0)}(t_i-t_1) + F_n^{(0)}(t_i-t_2)\right]f_1^{(n-1)} + \\ &+ \frac{1}{2}\left[F_n^{(0)}(t_i-t_{H-1}) + F_n^{(0)}(t_i-t_H)\right]f_H^{(n-1)} + \\ &+ \frac{1}{2}\sum_{j=2}^{H-1}\left[F_n^{(0)}(t_i-t_{j-1}) - F_n^{(0)}(t_i-t_{j+1})\right]f_j^{(n-1)}, \quad i = \overline{1,H}. \end{aligned} \quad (7)$$

Denote $f^{(n)} = (f_1^{(n)}, \dots, f_H^{(n)})^T, n \ge 1$, and $D^{(n)} = \{d_{ij}^n\}_{H\times H}, n \ge 2$, where

$$d_{ij}^n = \begin{cases} \frac{1}{2}\left[F_n^{(0)}(t_i-t_{j-1}) - F_n^{(0)}(t_i-t_{j+1})\right], & i = \overline{1,H},\ j = \overline{2,H-1}, \\ -\frac{1}{2}\left[F_n^{(0)}(t_i-t_1) + F_n^{(0)}(t_i-t_2)\right], & i = \overline{1,H},\ j = 1, \\ \frac{1}{2}\left[F_n^{(0)}(t_i-t_{H-1}) + F_n^{(0)}(t_i-t_H)\right], & i = \overline{1,H},\ j = H. \end{cases}$$

We get $f^{(n)} = D^{(n)}f^{(n-1)}$, $2 \le n \le M$, where $f^{(1)} = \left(f_1^{(1)}, \dots, f_H^{(1)}\right)^T$, and $f_i^{(1)} = F_1^{(0)}(t_i)$, $i = \overline{1,H}$. As a result, we obtain the following theorem.

Theorem 1. *If the functions $F_n^{(k)}(\cdot), 1 \le n \le M, k \in \{0,1\}$, have continuous derivatives of second order in $[C_- - C_+, C_+ - C_-]$, then the following expressions are valid as $h \to 0$:*

$$\alpha_M = 1 - \sum_{i=1}^{M-1} f_1^{(i)} - f_{i_0}^{(M)} + O(h^2), \quad \beta_M = \sum_{i=1}^{M-1} g_1^{(i)} + g_{i_0}^{(M)} + O(h^2),$$

$$\mathbf{E}^{(0)}(N_M) = 1 + \sum_{i=1}^{M-1}\left(f_H^{(i)} - f_1^{(i)}\right) + O(h^2), \quad \mathbf{E}^{(1)}(N_M) = 1 + \sum_{i=1}^{M-1}\left(g_H^{(i)} - g_1^{(i)}\right) + O(h^2),$$

where $g^{(i)} = \left(g_1^{(i)}, \dots, g_H^{(i)}\right)^T$, $i = \overline{1,M}$, are calculated similarly to $f^{(i)}$ replacing $F_i^{(0)}(x)$ with $F_i^{(1)}(x)$ – the distribution function of λ_i under hypothesis $\boldsymbol{H}_1$.

Remark 1. Due to Theorem 6.8 and the definition of Riemann–Stieltjes integral [13], the main terms in the right hand sides of formulae in Theorem 1 can also be used to approximate the expressions in the left hand sides respectively in case of continuous functions $F_n^{(0)}(\cdot), n \ge 1, k \in \{0,1\}$. In this situation, it is complicated to make any conclusion on the accuracy orders of these approximation formulae. However, a value of the parameter H can be increased to improve the accuracy.

3. Robustness evaluation

In practice, there is often the case that the observed data do not follow the hypothetical model exactly, e. g. the hypothetical model is distorted. This will lead to the distortion in the distributions of increments λ_n of log-likelihood statistic Λ_n. In this section, we will study the case where these influences can be described in the form of contaminated model of Huber type ([3]) for each increment λ_n as follows: $\bar{F}_n(x) = (1-\delta)F_n(x) + \delta\tilde{F}_n(x), n \ge 1$, where $\tilde{F}_n(x)$ is a contaminating CDF, and $\delta \in [0, 1/2)$ is the level of contamination.

Introduce the notation: $\bar{f}^{(n)}$, $\bar{D}^{(n)}$, $\bar{\alpha}_M$ are the elements calculated similarly to $f^{(n)}$, $D^{(n)}$, α_M replacing $F_n^{(0)}(x)$ with $\bar{F}_n^{(0)}(x)$, $n \ge 1$, $\bar{N}_M$ is the new stopping times for

TSPRT; $\hat{D}^{(n)}$ are the elements also calculated analogously to $D^{(n)}$ replacing $F_n^{(0)}(x)$ with $\tilde{F}_n^{(0)}(x) - F_n^{(0)}(x)$, $n \geq 1$. Put $Q^{(1)} = \hat{f}^{(1)}$ that is computed similarly to $f^{(1)}$ replacing $F_1^{(0)}(x)$ with $\tilde{F}_1^{(0)}(x) - F_1^{(0)}(x)$, and for $n \geq 2$:

$$Q^{(n)} = \hat{D}^{(n)} D^{(n-1)} \ldots D^{(2)} f^{(1)} + \ldots + D^{(n)} D^{(n-1)} \ldots \hat{D}^{(2)} f^{(1)} + D^{(n)} D^{(n-1)} \ldots D^{(2)} \hat{f}^{(1)}.$$

Theorem 2. *If functions $F_n^{(k)}(x)$ and $\tilde{F}_n^{(k)}(x), n = \overline{1, M}, k \in \{0, 1\}$, have continuous derivatives of the second order in $[C_- - C_+, C_+ - C_-]$, then the following expressions hold at $h \to 0, \delta \to 0$:*

$$\bar{\alpha}_M - \alpha_M = -\delta \left(\sum_{i=1}^{M-1} Q_1^{(i)} + Q_{i_0}^{(M)} \right) + O(h^2) + O(\delta^2),$$

$$\bar{\beta}_M - \beta_M = \delta \left(\sum_{i=1}^{M-1} R_1^{(i)} + R_{i_0}^{(M)} \right) + O(h^2) + O(\delta^2),$$

$$\mathbf{E}^{(0)}(\bar{N}_M) - \mathbf{E}^{(0)}(N_M) = \delta \sum_{i=1}^{M-1} \left(Q_H^{(i)} - Q_1^{(i)} \right) + O(h^2) + O(\delta^2),$$

$$\mathbf{E}^{(1)}(\bar{N}_M) - \mathbf{E}^{(1)}(N_M) = \delta \sum_{i=1}^{M-1} \left(R_H^{(i)} - R_1^{(i)} \right) + O(h^2) + O(\delta^2),$$

where $R^{(n)}$, $n \geq 1$, are calculated similarly to $Q^{(n)}$ replacing $F_n^{(0)}(x)$, $\tilde{F}_n^{(0)}(x)$ with $F_n^{(1)}(x)$, $\tilde{F}_n^{(1)}(x)$.

Proof. Note that

$$\bar{f}^{(1)} = f^{(1)} + \delta \hat{f}^{(1)}, \quad \bar{D}^{(n)} = D^{(n)} + \delta \hat{D}^{(n)}, \quad n \geq 2,$$

$$f^n = D^{(n)} f^{(n-1)}, \quad \bar{f}^n = \bar{D}^{(n)} \bar{f}^{(n-1)}, \quad n \geq 2.$$

From that, we have

$$\bar{f}^{(2)} = \bar{D}^{(2)} \bar{f}^{(1)} = \big(D^{(2)} + \delta \hat{D}^{(2)}\big)\big(f^{(1)} + \delta \hat{f}^{(1)}\big) = f^{(2)} + \delta Q^{(2)} + O_H(\delta^2),$$

$$\bar{f}^{(3)} = \bar{D}^{(3)} \bar{f}^{(2)} = \big(D^{(3)} + \delta \hat{D}^{(3)}\big)\big(f^{(2)} + \delta Q^{(2)}\big) + O_H(\delta^2) = f^{(3)} + \delta Q^{(3)} + O_H(\delta^2),$$

where $O_H(\delta^2)$ is an H-dimensional column vector with all elements that are $O(\delta^2)$.

By induction, we get $\bar{f}^{(n)} = f^{(n)} + \delta Q^{(n)} + O_H(\delta^2)$, $n \geq 1$. The rest parts of proof are derived from Thereom 1. □

4. Construction of the robust TSPRT

To reduce the influence of outliers on λ_n, we can "cut" the values of λ_n with the following function (Figure 1a):

$$f_{g_-}^{g_+}(x) = g_- \cdot \mathbf{1}_{(-\infty, g_-]}(x) + x \cdot \mathbf{1}_{(g_-, g_+)}(x) + g_+ \cdot \mathbf{1}_{[g_+, +\infty)}(x), \tag{8}$$

where g_-, g_+ are two given values, $g_- < 0 < g_+$.

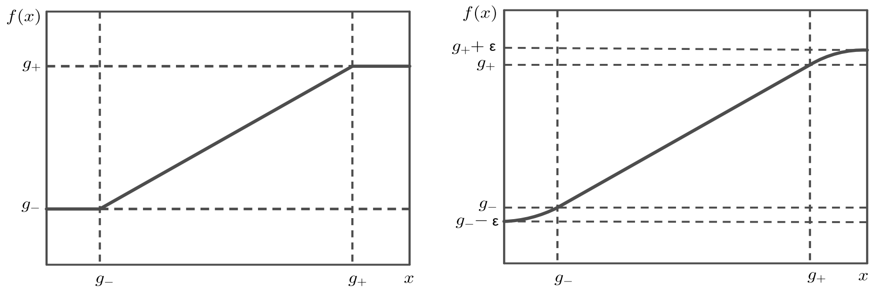
Put $\bar{\lambda}_n = f_{g_-}^{g_+}(\lambda_n), \bar{\Lambda}_n = \bar{\lambda}_1 + \ldots + \bar{\lambda}_n$. Let $\bar{\alpha}_M, \bar{\beta}_M, \bar{N}_M$ be the new test characteristics with respect to new increments $\bar{\lambda}_n$.

Lemma 1. *If $g_- \leq C_- - C_+$ and $g_+ \geq C_+ - C_-$, then*

(i) $\mathbf{P}_k\{N_M = i\} = \mathbf{P}_k\{\bar{N}_M = i\}, i = \overline{1, M}, k \in \{0, 1\}$,

(ii) $\alpha_M = \bar{\alpha}_M$ *and* $\beta_M = \bar{\beta}_M$.

Proof. Clearly, $\bar{\lambda}_t = \lambda_t$ if and only if $\lambda_t \in [g_-, g_+]$, and $\mathbf{P}_k\{\bar{\lambda}_t > x\} = \mathbf{P}_k\{\lambda_t > x\}$, $\forall x \in [g_-, g_+]$, $k \in \{0, 1\}$. Additionally, if $x, x + y \in (C_-, C_+)$ then $|y| < C_+ - C_-$. The rest part of proof is based on identity transformation and independence of random variables $\lambda_n, n = \overline{1, M}$. □



(a) (b)

FIGURE 1. Plots of cutting functions

Remark 2. Some proposals for choosing the thresholds g_- and g_+ are given below:

(i) If $g_- \geq 0$, then $\beta_M = 0$. If $g_+ \leq 0$, then $\alpha_M = 0$. Therefore, the possible choice is that we should select $g_- \in (C_- - C_+, 0)$ and $g_+ \in (0, C_+ - C_-)$.

(ii) If g_- increases, β_M will decrease, but α_M will increase provided g_+ is fixed. If g_+ decreases, there will be an opposite picture. So, the possible and reasonable criterion for choosing g_- and g_+ is to minimize the sum $\alpha_M + \beta_M$ for the TSPRT.

Using the truncated function (8), the distribution function of $\bar{\lambda}_n$ is as follows:

$$F_{\bar{\lambda}_n}(x) = \mathbf{P}(\bar{\lambda}_n < x) = \begin{cases} 0, & x \leq g_-, \\ F_{\lambda_n}(x), & g_- < x \leq g_+, \\ 1, & x > g_+, \end{cases}$$

which is generally a discontinuous function. Therefore, Theorem 1 cannot be applied for calculating the test characteristics. To make use of the proposed numerical approach, we can use a modified version of the function (8) in the following form (Figure 1b):

$$f_{g_-}^{g_+}(x) = \begin{cases} \frac{\varepsilon g_-}{x} + g_- - \varepsilon, & x \leq g_-, \\ x, & g_- < x < g_+, \\ -\frac{\varepsilon g_+}{x} + g_+ + \varepsilon, & x \geq g_+. \end{cases} \tag{9}$$

In this case, when $F_{\lambda_n}(x)$ is continuous, the distribution function of $\bar{\lambda}_n$ is also continuous and has the following form:

$$F_{\bar{\lambda}_n}(x) = \begin{cases} 0, & x \leq g_- - \varepsilon, \\ F_{\lambda_n}\left(\frac{\varepsilon g_-}{x - g_- + \varepsilon}\right), & g_- - \varepsilon < x \leq g_-, \\ F_{\lambda_n}(x), & g_- < x < g_+, \\ F_{\lambda_n}\left(\frac{\varepsilon g_+}{g_+ + \varepsilon - x}\right), & g_+ \leq x < g_+ \varepsilon, \\ 1, & x \geq g_+ + \varepsilon. \end{cases}$$

When $|g_+ - g_-|$ is small, we have to take more observations for the sequential test (e. g. the number of observations tends to the maximum number M). This means that we have more information for the test and this leads to the downward trend of both error probabilities. However, when $|g_+ - g_-|$ is sufficiently small, the number of observations are mostly M and we have to make the final decision according to (4). In this case, both error probabilities can increase again.

The following algorithm can be used to choose thresholds g_- and g_+:

Step 0: Choose a positive value $K \in \mathbb{N}$ and a small value $\varepsilon > 0$.

Step 1: Split $[C_- - C_+, 0]$ and $[0, C_+ - C_-]$ into cells by points $\{g_-(i), i = \overline{1, K}\}$ and $\{g_+(i), i = \overline{1, K}\}$ respectively, where $g_-(i) = -ih, g_+(i) = ih, h = (C_+ - C_-)/(K + 1)$.

Step 2: For each pair $(g_-(i), g_+(j))$ calculate $\alpha_M(i,j)$ and $\beta_M(i,j)$ using Theorem 1 and truncated function (9).

Step 3: Choose $(g_-(i), g_+(j))$ such that $\alpha_M(i,j) + \beta_M(i,j)$ is minimal.
In practice, we often use the symmetric case $g_-(i) = -g_+(i)$. Then, it is recommended to select $(g_-(i), g_+(i))$ such that $\alpha_M(i,i) + \beta_M(i,i)$ is minimal.

5. Sequential testing on parameters of time series with trend

Let $x_1, x_2, \dots$ be the observed time series with a trend in the following form:

$$x_t = \theta^T \psi(t) + \xi_t,\ t = 1, 2, \dots, M, \tag{10}$$

where $\psi(t) = (\psi_1(t), \psi_2(t), \dots, \psi_m(t))^T \in \mathbb{R}^m$, is the vector of linearly independent basic functions of trend, $\theta = (\theta_1, \theta_2, \dots, \theta_m)^T \in \mathbb{R}^m$ is an unknown vector of coefficients, and $\{\xi_t, t \geq 1\}$ is the sequence of independent identically distributed random variables, $\xi_t \sim \mathcal{N}(0, \sigma^2)$, σ is a given positive constant.

Consider two simple hypotheses concerning the trend coefficients: $\mathbf{H}_0 : \theta = \theta^0$, $\mathbf{H}_1 : \theta = \theta^1$, where $\theta^0, \theta^1 \in \mathbb{R}^m$ are two given vectors, $\theta^0 \neq \theta^1$. For all $t \geq 1$ we have:

$$x_t \sim N(\theta^T \psi(t); \sigma^2), \quad p_t(x, \theta) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}(x - \theta^T\psi(t))^2\right\}, \quad t \geq 1,$$

$$\lambda_t = \lambda_t(x_t) = -\frac{1}{2\sigma^2}\left\{2x_t(\theta^0 - \theta^1)^T\psi(t) + (\theta^1)^T\psi(t)\psi^T(t)\theta^1 - (\theta^0)^T\psi(t)\psi^T(t)\theta^0\right\}.$$

Introduce the notation: For $n \geq 1$, $k \in \{0, 1\}$

$$\sigma_n^2 = \mathbf{D}(\lambda_n) = \frac{(\theta^0 - \theta^1)^T\psi(n)\psi^T(n)(\theta^0 - \theta^1)}{\sigma^2},$$

$$\mu_n^{(k)} = \mathbf{E}^{(k)}(\lambda_n) = \frac{(-1)^{k+1}}{2\sigma^2}(\theta^0 - \theta^1)^T\psi(n)\psi^T(n)(\theta^0 - \theta^1) = \frac{(-1)^{k+1}\sigma_n^2}{2},$$

$$s_n^2 = \sum_{t=1}^n \sigma_t^2,\ m_n^{(k)} = \sum_{t=1}^n \mu_t^{(k)} = \frac{(-1)^{k+1}s_n^2}{2}.$$

Due to the properties of the normal distribution, λ_n and Λ_n have also the normal distributions $\mathcal{N}\big(\mu_n^{(k)}, \sigma_n^2\big)$, $\mathcal{N}\big(m_n^{(k)}, s_n^2\big)$ respectively under hypothesis $\mathbf{H}_k$, $k \in \{0, 1\}$. Theorem 1 can be applied for calculating the test characteristics of the TSPRT. Next, we will study the robustness of the TSPRT in the cases of three contaminated models. Without loss of generality, assume that hypothesis $\mathbf{H}_0$ is true.

Case 1: Distortion of the error component ξ_t

Instead of hypothetical model (10) consider the following contaminated model:

$$\bar{x}_t = \theta^T\psi(t) + \bar{\xi}_t, \quad t \geq 1, \tag{11}$$

where $\bar{\xi}_t = (1 - \tau_t)\xi_t + \tau_t\tilde{\xi}_t$, $t \geq 1$, $\{\tilde{\xi}_t, t \geq 1\}$ is a sequence of independent random variables with the distribution functions having continuous derivative of second order in $[C_- - C_+, C_+ - C_-]$, $\{\tau_t, t \geq 1\}$ is a sequence of independent identically distributed random variables, $\mathbf{P}(\tau_t = 0) = 1 - \delta$, $\mathbf{P}(\tau_t = 1) = \delta$, the variables $\tau_t, \xi_t, \tilde{\xi}_t$ are independent, and $\delta \in [0, 1/2)$ is the level of contamination. Let $\bar{\alpha}_M$ be the error probability of type I when replacing λ_t by $\bar{\lambda}_t = \lambda_t(\bar{x}_t)$, and $\bar{N}_M$ is the new stopping time for the TSPRT.

Theorem 3. *For the model* (11) *and the TSPRT* (2)–(4)*, the following expressions are valid as $h \to 0$, $\delta \to 0$:*

$$\bar{\alpha}_M = \alpha_M + O(h^2) + O(\delta), \quad \mathbf{E}^{(0)}(\bar{N}_M) = \mathbf{E}^{(0)}(N_M) + O(h^2) + O(\delta).$$

Proof. Under hypothesis $\mathbf{H}_0$, we have

$$\bar{\lambda}_t = -\frac{\sigma_t^2}{2} - \frac{(\theta^0 - \theta^1)^T \psi(t)}{\sigma^2} \bar{\xi}_t, \quad t \geq 1.$$

From that we get:

$$\begin{aligned}
\bar{F}_n^{(0)}(x) &= \mathbf{P}_0(\bar{\lambda}_n < x) = \mathbf{P}_0(\bar{\lambda}_n < x, \tau_n = 0) + \mathbf{P}_0(\bar{\lambda}_n < x, \tau_n = 1) = \\
&= (1-\delta)\mathbf{P}_0\left(-\frac{\sigma_n^2}{2} - \frac{(\theta^0 - \theta^1)^T \psi(n)}{\sigma^2} \xi_n < x\right) + \\
&\quad + \delta \mathbf{P}_0\left(-\frac{\sigma_n^2}{2} - \frac{(\theta^0 - \theta^1)^T \psi(n)}{\sigma^2} \tilde{\xi}_n < x\right) = \\
&= (1-\delta) F_n^{(0)}(x) + \delta \tilde{F}_n^{(0)}(x), \qquad (12)
\end{aligned}$$

where $\tilde{F}_n^{(0)}(x)$ is the distribution function of random variable $\zeta_n = -\frac{\sigma_n^2}{2} - \frac{(\theta^0-\theta^1)^T\psi(n)}{\sigma^2}\tilde{\xi}_n$. The rest part of proof is directly derived from (12) and Theorem 2. □

Case 2: Distortion of the basic function of trend $\psi(t)$

Consider the following model:

$$\bar{x}_t = \theta^T \tilde{\psi}(t) + \xi_t, \quad t \geq 1, \qquad (13)$$

where $\tilde{\psi}(t) = \big(\tilde{\psi}_1(t), \ldots, \tilde{\psi}_m(t)\big)^T$ is a basic function of trend such that $\big\|\tilde{\psi}(t) - \psi(t)\big\| = \max_{1 \leq i \leq m} \sup_{1 \leq t \leq M} \big|\tilde{\psi}_i(t) - \psi_i(t)\big| \leq \delta$.

Theorem 4. *For the model* (13) *and the TSPRT* (2)–(4)*, the following expressions are valid as* $h \to 0$*,* $\delta \to 0$*:*

$$\bar{\alpha}_M = \alpha_M + O(h^2) + O(\delta), \quad \mathbf{E}^{(0)}(\bar{N}_M) = \mathbf{E}^{(0)}(N_M) + O(h^2) + O(\delta).$$

Proof. Put $\eta(t) = (\theta^0 - \theta^1)^T \tilde{\psi}(t) - (\theta^0 - \theta^1)^T \psi(t)$, $t \geq 1$, then

$$|\eta(t)| = \left|\left(\theta^0 - \theta^1\right)^T \left(\tilde{\psi}(t) - \psi(t)\right)\right| \leq \sum_{i=1}^m \left|\theta_i^0 - \theta_i^1\right| \left\|\tilde{\psi}(t) - \psi(t)\right\| \leq \delta \sum_{i=1}^m \left|\theta_i^0 - \theta_i^1\right|.$$

Under the hypothesis $\mathbf{H}_0$, we have $\lambda_n \sim N\big(\mu_n^{(0)}, \sigma_n^2\big)$, $\bar{\lambda}_n \sim N\big(\tilde{\mu}_n^{(0)}, \tilde{\sigma}_n^2\big)$. Therefore, for all $x \in [C_- - C_+, C_+ - C_-]$

$$\bar{F}_n^{(0)}(x) - F_n^{(0)}(x) = \Phi\left(\frac{x - \tilde{\mu}_n^{(0)}}{\tilde{\sigma}_n}\right) - \Phi\left(\frac{x - \mu_n^{(0)}}{\sigma_n}\right),$$

where $\Phi(x)$ is the standard normal CDF. Using the mean value theorem, we get that there exists $\zeta \in \mathbb{R}$ such that

$$\bar{F}_n^{(0)}(x) - F_n^{(0)}(x) = \left(\frac{x - \tilde{\mu}_n^{(0)}}{\tilde{\sigma}_n} - \frac{x - \mu_n^{(0)}}{\sigma_n}\right)\varphi(\zeta) = (\tilde{\sigma}_n - \sigma_n)\left(\frac{1}{2} - \frac{x}{\sigma_n \tilde{\sigma}_n}\right)\varphi(\zeta),$$

where $\varphi(x)$ is the standard normal PDF.

On the other hand,

$$|\tilde{\sigma}_n - \sigma_n| = \frac{\Big|\big|(\theta^0 - \theta^1)^T \tilde{\psi}(n)\big| - \big|(\theta^0 - \theta^1)^T \psi(n)\big|\Big|}{\sigma} \leq \frac{|\eta(n)|}{\sigma}.$$

From that, we get $\bar{F}_n^{(0)}(x) - F_n^{(0)}(x) = O(\delta)$, $\forall x \in [C_- - C_+, C_+ - C_-]$, $n \geq 1$, which implies $\bar{f}^{(1)} = f^{(1)} + O_H(\delta)$, $\bar{D}^{(n)} = D^{(n)} + O_{H \times H}(\delta)$, $n \geq 2$. Therefore, $\bar{f}^{(n)} = f^{(n)} + O_H(\delta)$, $n \geq 1$. The rest part of proof is derived from Theorem 1. □

Case 3: Distortion of both components $\psi(t)$ and ξ_t

Consider the following mixed model:

$$\bar{x}_t = \theta^T \tilde{\psi}(t) + (1-\tau_t)\xi_t + \tau_t \tilde{\xi}_t, \quad t \geq 1, \tag{14}$$

where $\{\tau_t, t \geq 1\}$ is a sequence of i. i. d. random variables, $\mathbf{P}(\tau_t = 0) = 1 - \delta_1$, $\mathbf{P}(\tau_t = 1) = \delta_1$, and τ_t, ξ_t, $\tilde{\xi}_t$ are independent, and $\left\|\tilde{\psi}(t) - \psi(t)\right\| \leq \delta_2$.

Theorem 5. *For the model* (14) *and the TSPRT* (2)–(4)*, the following expressions are valid as $h \to 0$, $\delta_1 \to 0$, $\delta_2 \to 0$:*

$$\bar{\alpha}_M = \alpha_M + O(h^2) + O(\delta_1) + O(\delta_2), \tag{15}$$

$$\mathbf{E}^{(0)}(\bar{N}_M) = \mathbf{E}^{(0)}(N_M) + O(h^2) + O(\delta_1) + O(\delta_2). \tag{16}$$

Proof. Denote

$$h(\psi, \xi, t) = -\frac{((\theta^0 - \theta^2)^T \psi(t))^2}{2\sigma^2} - \frac{(\theta^0 - \theta^1)^T \psi(t)}{\sigma^2}\xi_t, \quad t \geq 1.$$

With $n \geq 1$, we have

$$\begin{aligned}\bar{F}_n^{(0)}(x) &= \mathbf{P}_0(\bar{\lambda}_n < x) = \mathbf{P}_0(\bar{\lambda}_n < x, \tau_n = 0) + \mathbf{P}_0(\bar{\lambda}_n < x, \tau_n = 1) = \\ &= (1-\delta_1)\mathbf{P}_0\Big(h\big(\tilde{\psi}, \xi, n\big) < x\Big) + \delta_1 \mathbf{P}_0\Big(h\big(\tilde{\psi}, \tilde{\xi}, n\big) < x\Big) = \\ &= \mathbf{P}_0\Big(h\big(\tilde{\psi}, \xi, n\big) < x\Big) + \delta_1\Big[\mathbf{P}_0\Big(h\big(\tilde{\psi}, \tilde{\xi}, n\big) < x\Big) - \mathbf{P}_0\Big(h\big(\tilde{\psi}, \xi, n\big) < x\Big)\Big].\end{aligned}$$

From the proof of Theorem 4, we knew

$$\mathbf{P}_0\Big(h\big(\tilde{\psi}, \xi, n\big) < x\Big) = F_n^{(0)}(x) + O(\delta_2), \quad \forall x \in [C_- - C_+, C_+ - C_-].$$

Therefore, $\bar{F}_n^{(0)}(x) = F_n^{(0)}(x) + O(\delta_1) + O(\delta_2)$. The rest part of the proof is similar to the proof of Theorem 4. □

6. Results of computer experiments

The probability model (10) was considered and the hypotheses (2) were tested with the following values of parameters:

$$\sigma = 10, \quad \theta^0 = (1,2,2,2)^T, \quad \theta^1 = (1,1,2,1)^T, \quad \psi(t) = (1, t/10, t^2/10, 1/t)^{\cdot}$$

The thresholds C_- and C_+ were calculated according to Wald ([14]). Denote the estimate of a characteristic γ with Monte-Carlo method by $\hat{\gamma}$. The number of repetitions used in Monte-Carlo simulation was 100 000. Figure 2 illustrates the dependence of test characteristics on the level of contamination δ for the TSPRT according to Theorem 3 with $\alpha_0 = \beta_0 = 0.1$, $H = 100$, $M = 40$, $\tilde{\xi}_t \sim N(0, \tilde{\sigma}^2)$, $\tilde{\sigma} = 50$. When the level δ increases, the number of observations contaminated increases as well. This will lead to the upward trend of error probabilities. However, in Figure 2 the average number of observations has a downward trend with respect to δ. This means that contaminated observations in the data make the test terminate earlier.

The model (11) is continued to be used as an outlier model for robustification example. In the Figure 3, there are shown the results of the algorithm for choosing the optimal thresholds g_- and g_+ in Remark 2 in the symmetric case $g_- = -g_+$ with $\alpha_0 = 0.05$, $\beta_0 = 0.01$, $\delta = 0.1$, $\varepsilon = 0.001$, $M = 60$, $K = 50$. Numerical approach in Theorem 1 is used for calculating the test characteristics with $H = 200$. The modified truncated function (9) is used for adjusting the sequential test.

When $|g_+(i) - g_-(i)|$ decreases, e. g. the index i decreases, then the total sum $t_0(M) + t_1(M)$ increases to the maximum $2M = 120$, but the sum of errors $\alpha_M + \beta_M$ drops to its minimum at $i = 3$ before going up again. We can use Monte Carlo method

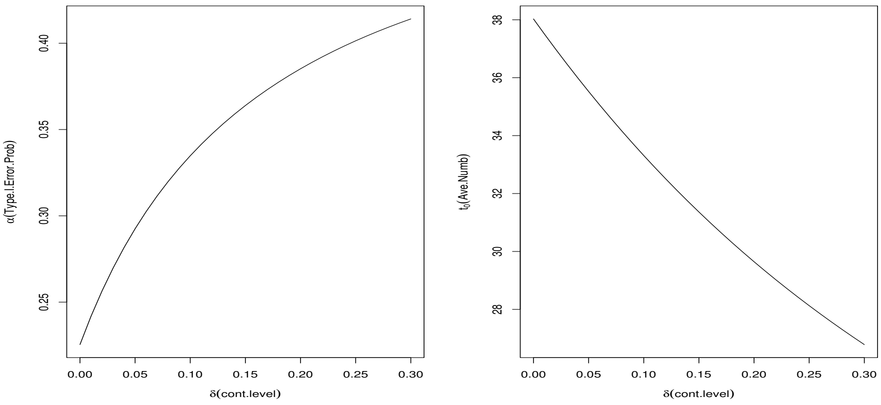


FIGURE 2. Dependence of test characteristics on the contamination level

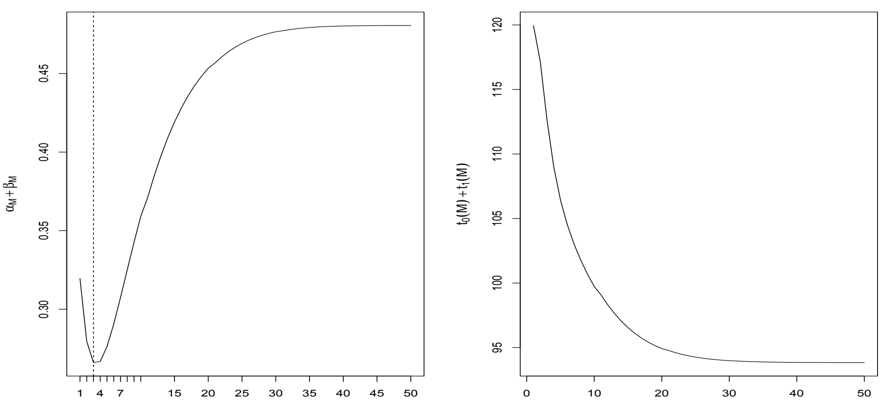


FIGURE 3. Sums of error probabilities and conditional expected numbers of observations

to check the efficiency of the robust test, and the result is presented in Table 1 for $g_- = g_-(3) = -0.4435$, $g_+ = g_+(3) = 0.4435$.

Table 1 illustrates that the robust sequential test leads to reduction of error probabilities. Although this modified test will increase the conditional expected numbers of observations, these values are always in the given ranges.

TABLE 1. The efficiency of the robust test

Type of test	$\hat{\alpha}_M$	$\hat{\beta}_M$	$\hat{\alpha}_M + \hat{\beta}_M$	$\hat{t}_0(M)$	$\hat{t}_1(M)$	$\hat{t}_0(M) + \hat{t}_1(M)$
Non-robustified	0.28533	0.19768	0.48301	48.43908	45.56539	94.00447
Robustified	0.13474	0.13165	0.26639	58.29301	54.22145	112.51446

ACKNOWLEDGMENT

The authors gratefully acknowledge the comments of the Referee that have helped to improve the paper exposition.

References

1. A. Gut, *Probability: a graduate course*, Springer, New York, 2005.
2. *Handbook of sequential analysis*, (B. Ghosh, P. K. Sen, ed.), Marcel Dekker, New York, 1991.
3. P. Huber, *Robust statistics*, John Wiley and Sons, New York, 1981.
4. A. Kharin, *Robust Bayesian prediction under distortions of prior and conditional distributions*, Journal of Mathematical Sciences, **126** (2005), no. 1, 992–997.
5. A. Kharin, P. Shlyk, *Robust multivariate Bayesian forecasting under functional distortions in the chi-square metric*, Journal of Statistical Planning and Inference, **139** (2009), 3842–3846.
6. A. Kharin, *Performance and robustness evaluation in sequential hypotheses testing*, Communications in Statistics – Theory and Methods, **45** (2016), no. 6, 1693–1709.
7. A. Yu. Kharin, *Robustness of sequential testing of hypotheses on parameters of M-valued random sequences*, Journal of Mathematical Sciences, **189** (2013), no. 6, 924–931.
8. A. Kharin, Ton That Tu, *Performance and robustness analysis of sequential hypotheses testing for time series with trend*, Austrian Journal of Statistics, **46** (2017), no. 3-4, 23–36.
9. G. G. Kosenko, V. P. Harchenko, A. G. Kukush, *Threshold choice in many-alternative subsequent rule for given mean risk*, Radioelectronics and Communications Systems, **39** (1996), no. 8, 38–42.
10. T. Lai, *Sequential analysis: some classical problems and new challenges*, Statistica Sinica, **11** (2001), no. 2, 303–408.
11. P. R. Mercer, *Hadamard's inequality and trapezoid rules for the Riemann–Stieltjes integral*, Journal of Mathematical Analysis and Applications, **344** (2008), 921–926.
12. N. Mukhopadhyay, S. Datta, S. Chattopadhyay, *Applied sequential methodologies*, Marcel Dekker, New York, 2004.
13. W. Rudin, *Principles of mathematical analysis*, McGraw-Hill, USA, 1976.
14. A. Wald, *Sequential analysis*, John Wiley and Sons, New York, 1947.

Belarusian State University
Current address: Independence av. 4, Minsk 220030, Belarus
E-mail address: KharinAY@bsu.by

The University of Danang – University of Science and Education
Current address: Danang, Vietnam
E-mail address: tttu@ued.udn.vn

Received 16.03.2019

РОБАСТНІСТЬ ПОСЛІДОВНОЇ ПЕРЕВІРКИ ГІПОТЕЗ ДЛЯ НЕОДНОРІДНИХ НЕЗАЛЕЖНИХ СПОСТЕРЕЖЕНЬ

А. Ю. ХАРІН, Т. Т. ТУ

Анотація. Розглядається проблема робастності для усічених послідовних тестів перевірки двох простих параметричних гіпотез у моделі неоднорідних незалежних спостережень за наявності спотворень. Запропоновано підхід для обчислення характеристик ефективності. Проведено асимптотичний аналіз робастності. Побудовано сім'ю послідовних тестів, яка дозволила підвищити робастність рішень, що приймаються. Чисельні приклади ілюструють одержані теоретичні результати.

РОБАСТНОСТЬ ПОСЛЕДОВАТЕЛЬНОЙ ПРОВЕРКИ ГИПОТЕЗ ДЛЯ НЕОДНОРОДНЫХ НЕЗАВИСИМЫХ НАБЛЮДЕНИЙ

А. Ю. ХАРИН, Т. Т. ТУ

Аннотация. Рассматривается проблема робастности для усеченных последовательных тестов проверки двух простых параметрических гипотез в модели неоднородных независимых наблюдений при наличии искажений. Предложен подход для вычисления характеристик эффективности. Проведен асимптотический анализ робастности. Построено семейство последовательных тестов, позволившее повысить робастность принимаемых решений. Численные примеры иллюстрируют полученные теоретические результаты.