ISSN 1561-2430 (Print)/
ISSN 2524-2415 (Online)
УДК 519.2
https://doi.org/10.29235/1561-2430-2018-54-2-179-192

Received 16.03.2018
Поступила в редакцию 16.03.2018

A. Yu. Kharin[1,2], Ton That Tu[1,3]

[1]Belarusian State University, Minsk, Belarus
[2]Research Institute for Applied Problems of Mathematics and Informatics, Minsk, Belarus
[3]Da Nang University of Education, Da Nang, Vietnam

PERFORMANCE ANALYSIS AND ROBUSTNESS EVALUATION OF A SEQUENTIAL PROBABILITY RATIO TEST FOR NON-IDENTICALLY DISTRIBUTED OBSERVATIONS

Abstract. In this article the problem of a sequential test for the model of independent non-identically distributed observations is considered. Based on recursive calculation a new numerical approach to approximate test characteristics for a sequential probability ratio test (SPRT) and a truncated SPRT (TSPRT) is constructed. The problem of robustness evaluation is also studied when the contamination is presented by the distortion of the distributions of all increments of the log-likelihood ratio statistics. The two-side truncated functions are proposed to be used for constructing the robustified SPRT. An algorithm to choose the thresholds of these truncated functions is indicated. The results are applied for a sequential test on parameters of time series with trend. Some kinds of the contaminated models of time series with trend are used to study the robustness of the truncated SPRT. Numerical examples confirming the theoretical results mentioned above are given.

Keywords: sequential test, simple hypotheses, approximation, test characteristics, truncation, non-identically distributed data, robustness evaluation

For citation. Kharin A. Yu., Ton That Tu. Performance analysis and robustness evaluation of a sequential probability ratio test for non-identically distributed observations. *Vestsi Natsyianal'nai akademii navuk Belarusi. Seryia fizika-matematychnykh navuk = Proceedings of the National Academy of Sciences of Belarus. Physics and Mathematics series*, 2018, vol. 54, no. 2, pp. 179–192. https://doi.org/10.29235/1561-2430-2018-54-2-179-192

А. Ю. Харин[1,2], Тон Тхат Ту[1,3]

[1]Белорусский государственный университет, Минск, Беларусь
[2]Научно-исследовательский институт прикладных проблем математики и информатики, Минск, Беларусь
[3]Педагогический университет Дананга, Дананг, Вьетнам

АНАЛИЗ И ИССЛЕДОВАНИЕ РОБАСТНОСТИ ПОСЛЕДОВАТЕЛЬНОГО КРИТЕРИЯ ОТНОШЕНИЯ ВЕРОЯТНОСТЕЙ ДЛЯ МОДЕЛИ НЕЗАВИСИМЫХ НЕОДИНАКОВО РАСПРЕДЕЛЕННЫХ НАБЛЮДЕНИЙ

Аннотация. Рассмотрена проблема последовательного теста для модели независимых неодинаково распределенных наблюдений. На основе рекурсивного расчета построен новый численный подход для аппроксимации тестовых характеристик последовательного критерия отношения вероятностей (ПКОВ) и усеченного ПКОВ (УПКОВ). Исследована проблема анализа робастности, когда «засорение» представлено искажением распределений всех приращений статистики логарифмического отношения правдоподобия. Предложено использование двухсторонних усеченных функций для построения робастного ПКОВ. Указан алгоритм для выбора порогов этих усеченных функций. Результаты применены для последовательной проверки гипотез о параметрах временных рядов с трендом. Для некоторых моделей «засорения» временных рядов с трендом исследована робастность усеченного ПКОВ. Проведенные в работе численные эксперименты подтверждают теоретические выводы.

Ключевые слова: последовательный тест, простые гипотезы, аппроксимация, характеристики теста, усечение, неодинаково распределенные наблюдения, анализ робастности

Для цитирования. Харин, А. Ю. Анализ и исследование робастности последовательного критерия отношения вероятностей для модели независимых неодинаково распределенных наблюдений / А. Ю. Харин, Тон Тхат Ту // Вес. Нац. акад. навук Беларусі. Сер. фіз.-мат. навук. – 2018. – Т. 54, № 2. – С. 179–192. https://doi.org/10.29235/1561-2430-2018-54-2-179-192

Introduction. Sequential analysis was first developed by Abraham Wald [1] and has been widely applied in many fields because of its optimal properties. In practice, the error probabilities α, β of type I and II can be different from the preassigned values α_0, β_0. In addition, the calculation of conditional average number of observations is very important in optimal evaluation of this approach. In the case of independent identically distributed observations, there have been some approaches to approximate the indicated test characteristics. Based on Wald's fundamental identity and likelihood ratio identity, some approximations for the average numbers of observations have been obtained [1–3]. An important improvement in computing these characteristics is that the operating characteristics (OC) and average sample number (ASN) functions were proved to satisfy the Fredholm integral equation of the second kind (FIESK) [3, 4]. Neglecting the conditions on the existence of their solutions, we can resort numerical methods to get the approximations of these characteristics. Another approach to calculate is to use the properties of absorbing Markov chains [5–7]. This approach allows not only to get the approximate values of test characteristics but also to evaluate the robustness of statistical procedures [6, 8, 9]. For the TSPRT, the upper bounds for the error probabilities of type I and II were achieved by using normal approximation for the accumulated log-likelihood ratio statistic when the maximum number of observations is relatively large [1], or in more general case [2]. In the case of non-identical distributed observations, Liu Y. and Li X. R. [10] have shown numerical solutions in some special cases to the OC and ASN functions by constructing the sequence of the FIESK with respect to the sequence of new stopping times. In this paper, another method based on recursive calculations is constructed for approximating the test characteristics of the SPRT and TSPRT as well. Evaluation of robustness for the truncated sequential test is also studied and these results will be applied for sequentially testing the parameters of time series with trend.

1. Mathematical model and auxiliary results

Let $\{X_n, n \geq 1\}$ be a sequence of independent random variables on the same probability space $(\Omega, \mathbf{F}, P)$ with probability density functions $\{p_n(x,\theta), x \in \mathbf{R}^1, n \geq 1\}$ respectively, where θ is an unknown vector of parameters.

Consider two simple hypotheses:

$$H_0 : \theta = \theta^0,\ H_1 : \theta = \theta^1, \tag{1}$$

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

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

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

where $\lambda_i = \ln\left(p_i(x_i, \theta^1) / p_i(x_i, \theta^0)\right)$ 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 [12], C_- and C_+ can be calculated as follows:

$$C_+ = \ln\left((1-\beta_0)/\alpha_0\right),\ C_- = \ln\left(\beta_0/(1-\alpha_0)\right), \tag{4}$$

where α_0, β_0 are the given values for error probabilities of types I and II respectively.

Denote $N = \inf\{n : \Lambda_n \notin (C_-, C_+)\}$, $\alpha = P_0(\Lambda_N \geq C_+)$, $\beta = P_1(\Lambda_N \leq C_-)$, where $P_k(\cdot)$ means the probability measure under $H_k, k \in \{0,1\}$. We will use the following auxiliary results.

L e m m a 1 [11]. *If X is a non-negative, integer valued random variable, then* $E(X) = \sum_{n=1}^{+\infty} P(X \geq n)$.

T h e o r e m 1 [12]. *If f is continuous on* [*a*, *b*] *and g is monotonic on* [*a*, *b*], *then there exists Riemann – Stieltjes integral* $\int_a^b f(x) dg(x)$.

C o r o l l a r y 1. *If g is monotonic on* [*a*, *b*] *and f is C-Lipschitzian on* [*a*, *b*], *i. e., there exists a positive constant C such that* $|f(x) - f(y)| \leq C|x - y|, \forall x, y \in [a,b]$, *then the following expansion holds:*

$$\int_a^b f(x) dg(x) = \frac{f(a) + f(b)}{2}(g(b) - g(a)) + O(b - a).$$

T h e o r e m 2 [13]. *Let* $f(x), g(x)$ *be two functions defined on* [*a, b*]. *Suppose that f″ and g″ are continuous on* [*a, b*] *and that g is monotonic there. Then, there exist* $\xi,\tau,\eta,\sigma\in(a,b)$ *such that*

$$\int_a^b f dg-\frac{f(a)+f(b)}{2}\left[g(b)-g(a)\right]=\left[g''(\xi)f'(\tau)-f''(\eta)g'(\sigma)\right]\frac{(b-a)^3}{12}.$$

2. Main results

2.1. Numerical approach to calculate the test characteristics. Put $S_1^{(k)}(x)=P_k(\Lambda_1<x)$, and for $n>1$, $S_n^{(k)}(x)=P_k\left(\Lambda_n<x, \text{ and } \Lambda_i\in(C_-,C_+),\ i=\overline{1,n-1}\right)$, $k\in\{0,1\}$.

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

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

where $F_n^{(k)}(x)$ is the cumulative distribution functions of λ_n under hypothesis H_k, and $S_1^{(k)}(x)=F_1^{(k)}(x)$.

Assume that $F_n^{(k)}(x)$, $S_n^{(k)}(x)$, $n\geq 1$, $k\in\{0,1\}$, are continuous functions in $\mathbf{R}$. Then, from the definitions of α, β and Lemma 1 the test characteristics can be expressed as follows:

$$\alpha=G^{(0)}(+\infty)-G^{(0)}(C_+),\ \beta=G^{(1)}(C_-)-G^{(1)}(-\infty),\ E^{(k)}(N)=1+G^{(k)}(C_+)-G^{(k)}(C_-),$$

where $G^{(k)}(x)=\sum\limits_{n=1}^{+\infty}S_n^{(k)}(x)$, $k\in\{0,1\}$, $E^{(k)}(\cdot)$ means expectation under H_k. Since $G^{(1)}(-\infty)=0$, and $G^{(0)}(+\infty)=1+\sum\limits_{n=1}^{+\infty}P_0\left(\Lambda_i\in(C_-,C_+),i=\overline{1,n}\right)=1+G^{(0)}(C_+)-G^{(0)}(C_-)$, we have

$$\alpha=1-G^{(0)}(C_-),\ \beta=G^{(1)}(C_-), \tag{6}$$

$$E^{(k)}(N)=1+G^{(k)}(C_+)-G^{(k)}(C_-),\ k\in\{0,1\}. \tag{7}$$

Assume that $E^{(k)}(N)<+\infty$, $k\in\{0,1\}$. In this case, from Lemma 1 we get

$$p_n^{(k)}=\sum_{j=n}^{\infty}P_k\left(\Lambda_i\in(C_-,C_+),i=\overline{1,j}\right)\to 0 \text{ as } n\to\infty, k\in\{0,1\}.$$

Given a very small positive value ε_0, there exists $n_0\in\mathbf{N}$ such that $p_n^{(k)}\leq\varepsilon_0$, $\forall n\geq n_0$, $k\in\{0,1\}$. Note that $S_{n+1}^{(k)}(x)\leq P_k(\Lambda_i,i=\overline{1,n})$, $\forall x\in\mathbf{R}$, $n\geq 1$, which allows us to approximate $G^{(k)}(x)$ by the new function $\overline{G}^{(k)}(x)$:

$$G^{(k)}(x)\approx\overline{G}^{(k)}(x)=\sum_{i=1}^{n_0}S_i^{(k)}(x),\ \forall x,\ k\in\{0,1\}, \tag{8}$$

where $|G^{(k)}(x)-\overline{G}^{(k)}(x)|\leq\varepsilon_0,\forall x$.

Next, we use a numerical method for approximating the values of functions $S_n^{(k)}(x)$, $n\geq 2$, $k\in\{0,1\}$. Without loss of generality, assume that 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 $t_i=C_-+(i-1)h$, $i=\overline{1,H}$, $h=\dfrac{C_+-C_-}{H-1}$. Using Theorem 2, under some assumptions of the functions $F_n^{(0)}(x)$ and $S_n^{(0)}(x)$, the Riemann-Stieltjes integral $\int_{C_-}^{C_+}F_n^{(0)}(x-y)dS_{n-1}^{(0)}(y)$ can be expanded as $h\to 0$:

$$S_n^{(0)}(x)=\frac{1}{2}\sum_{j=1}^{H-1}\left[F_n^{(0)}\left(x-t_j\right)+F_n^{(0)}\left(x-t_{j+1}\right)\right]\left[S_{n-1}^{(0)}(t_{j+1})-S_{n-1}^{(0)}(t_j)\right]+O(h^2). \tag{9}$$

This can be rewritten:

$$S_n^{(0)}(x) = -\frac{1}{2}\Big[F_n^{(0)}\big(x-t_1\big)+F_n^{(0)}\big(x-t_2\big)\Big]S_{n-1}^{(0)}(t_1)+\frac{1}{2}\Big[F_n^{(0)}\big(x-t_{H-1}\big)+F_n^{(0)}\big(x-t_H\big)\Big]S_{n-1}^{(0)}(t_H)+$$
$$+\frac{1}{2}\sum_{j=2}^{H-1}\Big[F_n^{(0)}\big(x-t_{j-1}\big)+F_n^{(0)}\big(x-t_{j+1}\big)\Big]S_{n-1}^{(0)}(t_j)+O(h^2).$$

Denote $f_i^{(j)} = S_j^{(0)}(t_i),\ j=\overline{1,n_0},\ i=\overline{1,H}.$ For $2 \le n \le n_0,$ we obtain the following systems of linear equations:

$$f_i^{(n)} = -\frac{1}{2}\Big[F_n^{(0)}\big(t_i-t_1\big)+F_n^{(0)}\big(t_i-t_2\big)\Big]f_1^{(n-1)}+\frac{1}{2}\Big[F_n^{(0)}\big(t_i-t_{H-1}\big)+F_n^{(0)}\big(t_i-t_H\big)\Big]f_H^{(n-1)}+$$
$$+\frac{1}{2}\sum_{j=2}^{H-1}\Big[F_n^{(0)}\big(t_i-t_{j-1}\big)+F_n^{(0)}\big(t_i-t_{j+1}\big)\Big]f_j^{(n-1)},\ i=\overline{1,H}. \tag{10}$$

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

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

We get $f^{(n)} = D^{(n)}f^{(n-1)},\ 2\le n\le n_0,$ where $f^{(1)} = \left(f_1^{(1)},...,f_H^{(1)}\right)^T$, and $f_i^{(1)} = F_1^{(0)}(t_i),\ i=\overline{1,H}.$ If the tailed sum $\sum_{i=n_0+1}^{\infty} F_i^{(0)}(x)$ was neglected, the following theorem has been proved.

T h e o r e m 3. *Assume that $E^{(k)}(N) < +\infty,\ k\in\{0,1\}$. If $F_n^{(k)}(x),\ n\ge 1,\ k\in\{0,1\},$ have continuous derivatives of second order in $[C_- - C_+, C_+ - C_-],$ then the following asymptotic expansions hold at $h\to 0,\ \varepsilon_0\to 0$:*

$$\alpha = 1-\sum_{i=1}^{n_0^{(0)}} f_1^{(i)} + O(h^2)+O(\varepsilon_0),\ \beta = \sum_{i=1}^{n_0^{(1)}} g_1^{(i)} + O(h^2)+O(\varepsilon_0),$$

$$E^{(0)}(N) = 1+\sum_{i=1}^{n_0^{(0)}}\left(f_H^{(i)}-f_1^{(i)}\right)+O(h^2)+O(\varepsilon_0),\ E^{(1)}(N) = 1+\sum_{i=1}^{n_0^{(1)}}\left(g_H^{(i)}-g_1^{(i)}\right)+O(h^2)+O(\varepsilon_0),$$

where $n_0^{(k)} = \min\left\{n : p_n^{(k)} \le \varepsilon_0\right\},\ k\in\{0,1\},\ g^{(i)} = \left(g_1^{(i)},...,g_H^{(i)}\right)^T,\ i=\overline{1,n_0^{(1)}},$ are calculated similarly to $f^{(i)}$ replacing $F_i^{(0)}(x)$ with $F_i^{(1)}(x)$ – the distribution function of λ_i under hypothesis H_1.

P r o o f. Note that by the way of selecting $n_0^{(k)}$, we have $|G^{(k)}(x)-\bar{G}^{(k)}(x)|\le\varepsilon_0, \forall x\in\mathbf{R}, k\in\{0,1\}.$ The result is directly derived from (6), (7), (10) and Theorem 2.

R e m a r k 1. In practice, it is not easy to determine $n_0^{(k)}$ theoretically with respect to a given value ε_0. However, if we know $q_n^{(k)} = P_k\left(\Lambda_n \in (C_-, C_+)\right)\to 0$ as $n\to+\infty,$ then $n_0^{(k)}$ can be chosen from the weaker condition: $n_0^{(k)} = \min\{n : q_n^{(k)}\le\varepsilon_0\}.$ This condition seems to be reasonable: in this case, all probabilities of the form $P_k\left(\Lambda\in(C_-,C_+), i=\overline{1,n}\right),\ n\ge n_0^{(k)},$ are much less than ε_0 and the test will terminate finitely with probability 1 as well.

R e m a r k 2. In general, there is still a problem of calculating the probability $q_n = P_k\left(\Lambda_n \in (C_-, C_+)\right)$ because of the difficulty in getting theoretically the probability distribution for the sum of independent random variables $\lambda_n,\ n \geq 1$. Note that $P_0\left(\Lambda_i \in (C_-, C_+), i = \overline{1,n}\right) = f_H^{(n)} - f_1^{(n)} + O(h^2)$. Therefore, if the way of finding index $n_0^{(k)}$ in Remark 1 is not feasible, these indices can be possibly chosen from the following conditions:

$$n_0^{(0)} = \inf\left\{n \geq 1 : f_H^{(n)} - f_1^{(n)} \leq \varepsilon_0\right\},\ n_0^{(1)} = \inf\left\{n \geq 1 : g_H^{(n)} - g_1^{(n)} \leq \varepsilon_0\right\}.$$

R e m a r k 3. In the case of independent identically distributed observations, due to Stein's lemma [3] a sufficient condition for $E^{(k)}(N) < +\infty,\ k \in \{0,1\}$, is $P_k(\lambda_1 = 0) < 1,\ k \in \{0,1\}$.

Next, we modify the method above to approximate error probabilities of type I and II for the TSPRT. Let M be the maximal number of observations that we may measure. The Wald's TSPRT is formulated as follows. If the sampling process has progressed to the n-th stage ($n < M$):

$$\begin{cases} \text{reject } H_0 \text{ if } \Lambda_n \geq C_+, \\ \text{accept } H_0 \text{ if } \Lambda_n \leq C_-, \end{cases} \tag{11}$$

and takes one more observation if $\Lambda\ \in (C_-, C_+)$. If the SPRT does not lead to the terminal decision for $n < M$, then

$$\begin{cases} \text{reject } H_0 \text{ if } \Lambda_M > 0, \\ \text{accept } H_0 \text{ if } \Lambda_M < 0. \end{cases} \tag{12}$$

For the partition $\{t_i, i = \overline{1,H}\}$ defined above, we set the value t_{i0} with respect to the smallest absolute value to be zero. Denote type I, II error probabilities and the number of observations used in TSPRT at the stage M by α_M, β_M and N_M respectively.

T h e o r e m 4. *If the functions* $F_n^{(k)}(x),\ n \geq 1,\ k \in \{0,1\}$, *have continuous derivatives of second order in* $[C_- - C_+, C_+ - C_-]$, *then the following expressions are valid:*

$$\alpha_M = 1 - \sum_{i=1}^{M-1} f_1^{(i)} - f_{i0}^{(M)} + O(h^2),\ \ \beta_M = \sum_{i=1}^{M-1} g_1^{(i)} + g_{i0}^{(M)} + O(h^2),$$

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

P r o o f. We have:

$$\sum_{n=1}^{M-1} P_0\left(\Lambda_i \in (C_-, C_+),\ i = \overline{1, n-1},\ \Lambda_n \geq C_+\right) =$$

$$= 1 - P_0(\Lambda_1 < C_+) + \sum_{n=1}^{M-2} P_0\left(\Lambda_i \in (C_-, C_+),\ i = \overline{1,n}\right) - \sum_{n=2}^{M-1} P_0\left(\Lambda_i \in (C_-, C_+),\ i = \overline{1, n-1},\ \Lambda_n < C_+\right) =$$

$$= 1 - f_H^1 + \sum_{n=1}^{M-2}\left(f_H^n - f_1^n\right) - \sum_{n=2}^{M-1} f_H^n + O(h^2) = 1 - f_H^{(M-1)} - \sum_{n=1}^{M-2} f_1^{(n)} + O(h^2), \tag{13}$$

$$P_0\left(\Lambda_i \in (C_-, C_+),\ i = \overline{1, M-1},\ \Lambda_M > 0\right) =$$

$$= P_0\left(\Lambda_i \in (C_-, C_+),\ i = \overline{1, M-1}\right) - P_0\left(\Lambda_i \in (C_-, C_+),\ i = \overline{1, M-1},\ \Lambda_M \leq 0\right) =$$

$$= f_H^{(M-1)} - f_1^{(M-1)} - f_{i0}^{(M)} + O(h^2). \tag{14}$$

From (11)–(12) and (13)–(14), we obtain $\alpha_M = 1 - \sum_{i=1}^{M-1} f_1^{(i)} - f_{i0}^{(M)} + O(h^2)$. Furthermore, we also have

$$P_0(N_M = 1) = P_0(\Lambda_1 \leq C_-) + P_0(\Lambda_1 \geq C_+) = 1 + f_1^{(1)} - f_H^{(1)} + O(h^2),$$

$$P_0(N_M = M) = P_0\left(\Lambda_i \in (C_-, C_+), i = \overline{1, M-1}\right) = f_H^{(M-1)} - f_1^{(M-1)} + O(h^2),$$

and for $2 \le i \le M-1$,

$$\begin{aligned} P_0(N_M = i) &= P_0\left(\Lambda_j \in (C_-, C_+), j = \overline{1, i-1},\ \Lambda_i \le C_-\right) + P_0\left(\Lambda_j \in (C_-, C_+), j = \overline{1, i-1},\ \Lambda_i \ge C_+\right) = \\ &= f_1^{(i)} + P_0\left(\Lambda_j \in (C_-, C_+), j = \overline{1, i-1}\right) - P_0\left(\Lambda_j \in (C_-, C_+), j = \overline{1, i-1},\ \Lambda_i \le C_+\right) + O(h^2) = \\ &= f_1^{(i)} - f_H^{(i)} + f_H^{(i-1)} - f_1^{(i-1)} + O(h^2). \end{aligned}$$

From that we get:

$$\begin{aligned} E^{(0)}(N_M) &= \sum_{i=1}^{M} i P_0(N_M) = \\ &= 1 + f_1^{(1)} - f_H^{(1)} + M\left(f_H^{(M-1)} - f_1^{(M-1)}\right) + \sum_{i=2}^{M-1} i\left(f_1^{(i)} - f_H^{(i)} + f_H^{(i-1)} - f_1^{(i-1)}\right) + O(h^2) = \\ &= 1 + \sum_{i=1}^{M-1}\left(f_H^{(i)} - f_1^{(i)}\right) + O(h^2). \end{aligned}$$

The rest part is proved similarly.

R e m a r k 4. From Theorem 1 and Corollary 1, we have:

(*i*) In the case that functions $F_n^{(k)}(x)$, $k \in \{0,1\}$, are C-Lipschitzian on $[C_- - C_+, C_+ - C_-]$, the formulas in Theorem 3 and Theorem 4 are still valid with the order of accuracy $O(1)$.

(*ii*) By the definition of the Riemann – Stieltjes integral and Theorem 1, these formulas in Theorem 3 and Theorem 4 are still applicable in the case of continuous functions $F_n^{(k)}(x)$, $k \in \{0,1\}$, without any conclusion about the order of accuracy. For the TSPRT, due to the limited number of terms in the sum we can increase the number H to get better approximation.

2.2. 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 [14]. This leads to the distortion in the distributions of increments λ_n of log-likelihood statistic Λ_n. In this section, we study the case where these influences can be described in the form of contaminated model of Huber type [15] for each increment λ_n as follows:

$$\overline{F}_n(x) = (1-\delta)F_n(x) + \delta\tilde{F}_n(x),\ n \ge 1,$$

where $\tilde{F}_n(x)$ is a contaminating CDF, and $\delta \in [0, 1/2)$ is the level of contamination.

Introduce the notation: $\overline{p}_n^{(k)}, \overline{f}^{(n)}, \overline{D}^{(n)}, \overline{\alpha}, \overline{\alpha}_M$ are the elements calculated similarly to $p_n^{(k)}$, $f^{(n)}$, $D^{(n)}, \alpha, \alpha_M$ replacing $F_n^{(0)}(x)$ with $\overline{F}_n^{(0)}(x)$, $n \ge 1$, $\overline{N}$ and $\overline{N}_M$ are the new stopping times for the SPRT and TSPRT respectively; $\hat{D}^{(n)}$ are the elements also calculated analogously to $D^{(n)}$ replacing $F_n^{(0)}(x)$ with $\tilde{F}_n^{(0)}(x) - F_n^{(0)}(x)$, $n \ge 1$. 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 $n \ge 2$:

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

T h e o r e m 5. *Assume that* $E^{(k)}(N) < +\infty$ *and* $E^{(k)}(\overline{N}) < +\infty$, $k \in \{0,1\}$. *If the functions* $F_n^{(k)}(x)$ *and* $\tilde{F}_n^{(k)}(x)$, $n \ge 1$, $k \in \{0,1\}$, *have continuous derivatives of second order in* $[C_- - C_+, C_+ - C_-]$, *then the following expressions hold:*

$$\overline{\alpha} - \alpha = -\delta \sum_{i=1}^{n_0^{(0)}} Q_1^{(i)} + O(h^2) + O(\delta^2) + O(\varepsilon_0),\ \overline{\beta} - \beta = \delta \sum_{i=1}^{n_0^{(1)}} R_1^{(i)} + O(h^2) + O(\delta^2) + O(\varepsilon_0),$$

$$E^{(0)}(\bar{N}) - E^{(0)}(N) = \delta \sum_{i=1}^{n_0^{(0)}} (Q_H^{(i)} - Q_1^{(i)}) + O(h^2) + O(\delta^2) + O(\varepsilon_0),$$

$$E^{(1)}(\bar{N}) - E^{(1)}(N) = \delta \sum_{i=1}^{n_0^{(1)}} (R_H^{(i)} - R_1^{(i)}) + O(h^2) + O(\delta^2) + O(\varepsilon_0),$$

where $n_0^{(k)} = \inf\{n \ge 1 : p_n^{(k)} \le \varepsilon_0$ and $\bar{p}_n^{(k)} \le \varepsilon_0\}$, $k \in \{0,1\}$, $R^{(n)}$, $n \ge 1$, *are calculated similarly to* $Q^{(n)}$ *replacing* $F_n^{(0)}(x)$, $\tilde{F}_n^{(0)}(x)$ *with* $F_n^{(1)}(x)$, $\tilde{F}_n^{(1)}(x)$.

P r o o f. Note that $\bar{f}^{(1)} = f^{(1)} + \delta \hat{f}^{(1)}$, $\bar{D}^{(n)} = D^{(n)} + \delta \hat{D}^{(n)}$, $f^n = D^{(n)} f^{(n-1)}$, $\bar{f}^n = \bar{D}^{(n)} \bar{f}^{(n-1)}$, $n \ge 2$. From that, we have:

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

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

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

By induction, we get: $\bar{f}^{(n)} = f^{(n)} + \delta Q^{(n)} + O_H(\delta^2)$, $n \ge 1$. The rest parts of proof are derived from the proof of Theorem 3.

Similarly, we also have the following result for the TSPRT.

T h e o r e m 6. *If the functions* $F_n^{(k)}(x)$ *and* $\tilde{F}_n^{(k)}(x)$, $n = \overline{1, M}$, $k \in \{0,1\}$, *have continuous derivatives of second order in* $[C_- - C_+, C_+ - C_-]$, *then the following expressions hold:*

$$\bar{\alpha}_M - \alpha_M = -\delta \left(\sum_{i=1}^{M-1} Q_1^{(i)} + Q_{i0}^{(M)} \right) + O(h^2) + O(\delta^2), \ \bar{\beta}_M - \beta_M = \delta \left(\sum_{i=1}^{M-1} R_1^{(i)} + R_{i0}^{(M)} \right) + O(h^2) + O(\delta^2),$$

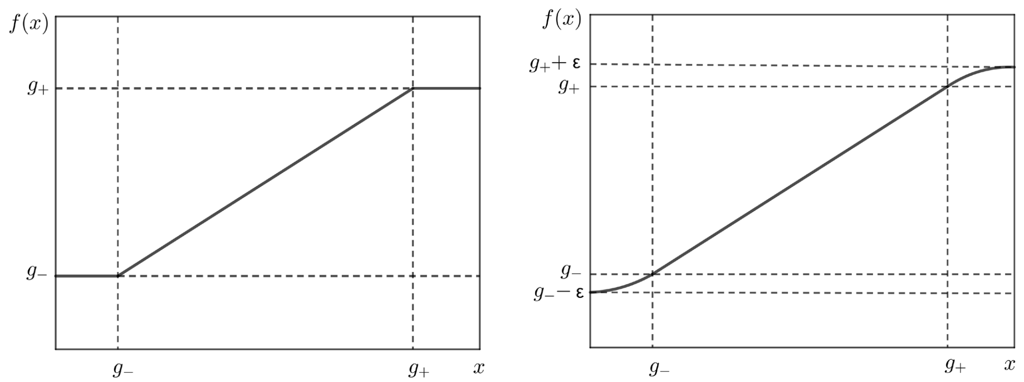
$$E^{(0)}(\bar{N}_M) - E^{(0)}(N_M) = \delta \sum_{i=1}^{M-1} (Q_H^{(i)} - Q_1^{(i)}) + O(h^2) + O(\delta^2),$$

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

2.3. Robustifying the TSPRT. To reduce the influence of outliers in λ_n, we can truncate the values of λ_n by the following function (Figure *a*):

$$f_{g_-}^{g_+}(x) = g_- \cdot \mathbf{1}_{(-\infty, g_-]}(x) + x \cdot \mathbf{1}_{(g_-, g_+)}(x) + g_+ \cdot \mathbf{1}_{[g_+, +\infty)}(x), \tag{15}$$

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



a *b*

Plots of truncated functions

Put $\overline{\lambda}_n = f_{g_-}^{g_+}(\lambda_n),\ \overline{\Lambda}_n = \overline{\lambda}_1 + ... + \overline{\lambda}_n,\ \overline{N} = \inf\{n : \overline{\Lambda}_n \notin (C_-, C_+)\},\ \overline{\alpha} = P_0(\overline{\Lambda}_{\overline{N}} \ge C_+),\ \overline{\beta} = P_1(\overline{\Lambda}_{\overline{N}} \le C_-)$.

L e m m a 2. *If* $g_- \le C_- - C_+$ *and* $g_+ \ge C_+ - C_-$, *then:*

(*i*) $P_k(N = i) = P_k(\overline{N} = i),\ i \ge 1,\ k \in \{0,1\}$,

(*ii*) $\alpha = \overline{\alpha},\ \beta = \overline{\beta}$.

P r o o f.

(*i*) Clearly, $\overline{\lambda}_t = \lambda_t$ if and only if $\lambda_t \in [g_-, g_+]$, and $P_k(\overline{\lambda}_t > x) = P_k(\lambda_t > x),\ \forall x \in [g_-, g_+],\ k \in \{0,1\}$. Additionally, if $x, x + y \in (C_-, C_+)$ then $|y| < C_+ - C_-$. Therefore,

$$P_k(\overline{N} = 1) = P_k(\overline{\lambda}_1 \ge C_+) + P_k(\overline{\lambda}_1 \le C_-) = P_k(\lambda_1 \ge C_+) + P_k(\lambda_1 \le C_-) = P_k(N = 1).$$

For $i > 1$, we get

$$P_k(\overline{N} = i) = P_k\left(\overline{\Lambda}_j \in (C_-, C_+), j = \overline{1, i-1},\ \overline{\Lambda}_i \notin (C_-, C_+)\right) =$$

$$= P_k\left(\Lambda_j \in (C_-, C_+), j = \overline{1, i-1},\ \Lambda_{i-1} + \overline{\lambda}_i \notin (C_-, C_+)\right) =$$

$$= P_k\left(\Lambda_j \in (C_-, C_+), j = \overline{1, i-1}\right) - P_k\left(\Lambda_j \in (C_-, C_+), j = \overline{1, i-1},\ \Lambda_i \in (C_-, C_+)\right) =$$

$$= P_k\left(\Lambda_j \in (C_-, C_+), j = \overline{1, i-1},\ \Lambda_i \notin (C_-, C_+)\right).$$

So, N and $\overline{N}$ have the same probability distributions.

(*ii*) Similarly, $P_0(\overline{\Lambda}_1 \ge C_+) = P_0(\overline{\lambda}_1 \ge C_+) = P_0(\lambda_1 \ge C_+) = P_0(\Lambda_1 \ge C_+)$ and

$$P_0\left(\overline{\Lambda}_j \in (C_-, C_+), j = \overline{1, i-1},\ \overline{\Lambda}_i \ge C_+\right) = P_0\left(\Lambda_j \in (C_-, C_+), j = \overline{1, i-1},\ \overline{\lambda}_i \ge C_+ - \Lambda_{i-1}\right) =$$

$$= \int_{C_-}^{C_+} dx_1 \int_{C_-}^{C_+} dx_2 \int_{C_-}^{C_+} dx_{i-2} \int_{C_-}^{C_+} f_{\Lambda_1, \ldots, \Lambda_{i-1}}(x_1, \ldots, x_{i-1}) P_0(\overline{\lambda}_i \ge C_+ - x_{i-1}) dx_{i-1} =$$

$$= \int_{C_-}^{C_+} dx_1 \int_{C_-}^{C_+} dx_2 . \int_{C_-}^{C_+} dx_{i-2} \int_{C_-}^{C_+} f_{\Lambda_1, \ldots, \Lambda_{i-1}}(x_1, \ldots, x_{i-1}) P_0(\lambda_i \ge C_+ - x_{i-1}) dx_{i-1} =$$

$$= P_0\left(\Lambda_j \in (C_-, C_+), j = \overline{1, i-1},\ \lambda_i \ge C_+ - \Lambda_{i-1}\right) = P_0\left(\Lambda_j \in (C_-, C_+), j = \overline{1, i-1},\ \Lambda_i \ge C_+\right).$$

Therefore, $P_0(\overline{\Lambda}_{\overline{N}} \ge C_+) = P_0(\Lambda_N \ge C_+)$. Similarly, we obtain $P_1(\overline{\Lambda}_{\overline{N}} \le C_-) = P_1(\Lambda_N \le C_-)$.

C o r o l l a r y 2. *The results of Lemma* 2 *are still valid for the TSPRT.*

R e m a r k 5. There are some remarks for choosing the thresholds g_- and g_+:

(*i*) If $g_- \ge 0$, then $\beta_M = 0$; if $g_+ \le 0$, 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. If g_+ decreases, there is 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 (15), the distribution function of $\overline{\lambda}_n$ is:

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

which is generally a discontinuous function. Therefore, the numerical results in Theorem 3 and Theorem 4 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 (15) in the following form (Figure *b*):

$$f_{g_-}^{g_+}(x) = \begin{cases} \dfrac{\varepsilon g_-}{x} + g_- - \varepsilon, & x \le g_-, \\ x, & g_- < x < g_+, \\ -\dfrac{\varepsilon g_+}{x} + g_+ + \varepsilon, & x \ge g_+. \end{cases} \tag{16}$$

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

$$F_{\overline{\lambda}_n}(x) = \begin{cases} 0, & x \le g_- - \varepsilon, \\ F_{\lambda_n}\left(\dfrac{\varepsilon g_-}{x - g_- + \varepsilon}\right), & g_- - \varepsilon < x \le g_-, \\ F_{\lambda_n}(x), & g_- < x < g_+, \\ F_{\lambda_n}\left(\dfrac{\varepsilon g_+}{g_+ + \varepsilon - x}\right), & g_+ \le x < g_+ \varepsilon, \\ 1, & x \ge g_+ + \varepsilon. \end{cases}$$

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 (15). In this case, both error probabilities can increase again.

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

– choose a positive value $K \in \mathbf{N}$ and a small value $\varepsilon > 0$;

– 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 $g_-(i) = -ih,\ g_+(i) = ih,\ h = \dfrac{C_+ - C_-}{K+1}$;

– for each pair $(g_-(i), g_+(j))$ calculate $\alpha_M(i,j)$ and $\beta_M(i,j)$ using Theorem 4 and truncated function (16);

– choose $(g_-(i), g_+(j))$ such that $\alpha_M(i,j) + \beta_M(i,j)$ is minimal.

In practice, due to the limitation of time and capacity of computation we can consider only the symmetric case $g_- = -g_+$ and select $(g_-(i), g_+(i))$ such that $\alpha_M(i,i) + \beta_M(i,i)$ is minimal.

2.4. Application for sequential testing on parameters of time series with trend. Let $x_1, x_2, ...$ be the observed time series with a trend in the following form [8]:

$$x_t = \theta^T \psi(t) + \xi_t,\ t \ge 1, \tag{17}$$

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

Consider two simple hypotheses concerning the trend coefficients:

$$H_0 : \theta = \theta^0, H_1 : \theta = \theta^1,$$

where $\theta^0, \theta^1 \in \mathbf{R}^m$ are two given vectors, $\theta^0 \ne \theta^1$.

For all $t \ge 1$ we have: $x_t \sim N\left(\theta^T \psi(t); \sigma^2\right),\ t \ge 1,\ p_t(x, \theta) = \dfrac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\dfrac{1}{2\sigma^2}\left(x - \theta^T \psi(t)\right)^2\right\}$ and

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

Put $V_n = \sum_{t=1}^{n} \psi(t)\psi^T(t)$. Due to the properties of the normal distribution, λ_t and Λ_n also have the normal distributions with the following parameters:

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

$$E(\Lambda_n) = -\frac{1}{2\sigma^2}\left\{2(\theta^0 - \theta^1)^T V_n\theta + (\theta^1)^T V_n\theta^1 - (\theta^0)^T V_n\theta^0\right\},$$

$$D(\lambda_t) = \frac{(\theta^0 - \theta^1)^T \psi(t)\psi^T(t)(\theta^0 - \theta^1)}{\sigma^2},\ D(\Lambda_n) = \frac{(\theta^0 - \theta^1)^T V_n(\theta^0 - \theta^1)}{\sigma^2}.$$

Introduce the notation: $\Gamma = (\theta^0 - \theta^1)(\theta^0 - \theta^1)^T$, $\sigma_n^2 = D^{(0)}(\lambda_n) = D^{(1)}(\lambda_n) = \dfrac{(\theta^0 - \theta^1)^T \psi(n)\psi^T(n)(\theta^0 - \theta^1)}{\sigma^2}$

and for $k \in \{0,1\}$

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

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

Without loss of generality assume that hypothesis H_0 is true and we are interested in studying type I error probability α and the average number of observations $E^{(0)}(N)$.

2.4.1. Calculation of the test characteristics. A sufficient condition for the termination of the test can be found in [16].

T h e o r e m 7 [16]. *If* $\mathrm{tr}(\Gamma V_n) \to +\infty$ *as* $n \to +\infty$, *then the test* (3)–(4) *terminates finitely with probability* 1.

Furthermore, in this case we know the exact probability distribution of Λ_n, $\Lambda_n \sim N\left(m_n^{(k)}, s_n^2\right)$. When hypothesis H_k is true, the index $n_0^{(k)}$, $k \in \{0,1\}$, can be chosen from the condition:

$$n_0^{(k)} = \inf\left\{n \ge 1 : P_k\left(\Lambda_n \in (C_-, C_+)\right) \le \varepsilon_0\right\},\ k \in \{0,1\},$$

where ε_0 is a given small positive value.

Next, we can use Theorem 3 and Theorem 4 for calculating the test characteristics for the SPRT and TSPRT as well, where $F_n^{(k)}(x)$, $n \ge 1$, are the normal distribution functions $N(\mu_n^{(k)}, \sigma_n^2)$, and the index $n_0^{(k)}$ can be calculated following Remarks 1, 2. In practice, the condition $E^{(k)}(N) < +\infty$ of Theorem 3 can be neglected because under the condition $\mathrm{tr}(\Gamma V_n) \to +\infty$ as $n \to +\infty$, we have [16]:

$$\lim_{n \to +\infty} P_k\left(\Lambda_n \in (C_-, C_+)\right) = 0, k \in \{0,1\}.$$

2.4.2. Robustness evaluation for the TSPRT. In this section, we will use the results of Theorem 4 and Theorem 6 for evaluating the robustness of the TSPRT with the maximum number of observations M for model (17) under the distortion on its different components.

Case 1. Distortion in the error component ξ_t. Instead of hypothetical model (17) we consider the following contaminated model:

$$\overline{x}_t = \theta^T \psi(t) + \overline{\xi}_t, t \ge 1, \tag{18}$$

where $\overline{\xi}_t = (1 - \tau_t)\xi_t + \tau_t\tilde{\xi}_t$, $t \ge 1$, $\{\tilde{\xi}_t, t \ge 1\}$ is a sequence of independent random variables, $\{\tau_t, t \ge 1\}$ is a sequence of independent identically distributed random variables, $P(\tau_t = 0) = 1 - \delta, P(\tau_t = 1) = \delta$, $\tau_t, \xi_t, \tilde{\xi}_t$ are independent and $\delta \in [0, 1/2)$ is the level of contamination.

Let $\overline{\alpha}_M$ be the error probability of type I when replacing λ_t by $\overline{\lambda}_t$, where $\overline{\lambda}_t = \lambda_t(\overline{x}_t)$, $t \ge 1$, and $\overline{N}_M$ is the new stopping time for the TSPRT at stage M.

T h e o r e m 8. *For the model* (18) *and the TSPRT* (3), (11)–(12), *the following expressions are valid:*

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

P r o o f. Under hypothesis H_0, we have: $\bar{\lambda}_t = -\frac{\sigma_t^2}{2} - \frac{(\theta^0-\theta^1)^T \psi(t)}{\sigma^2}\bar{\xi}_t,\ t \ge 1$. From that we get:

$$\begin{aligned}\bar{F}_n^{(0)}(x) &= P_0(\bar{\lambda}_n < x) = P_0(\bar{\lambda}_n < x, \tau_n = 0) + P_0(\bar{\lambda}_n < x, \tau_n = 1) = \\ &= (1-\delta)P_0\left(-\frac{\sigma_n^2}{2} - \frac{(\theta^0-\theta^1)^T\psi(n)}{\sigma^2}\xi_n < x\right) + \delta P_0\left(-\frac{\sigma_n^2}{2} - \frac{(\theta^0-\theta^1)^T\psi(n)}{\sigma^2}\tilde{\xi}_n < x\right) = \\ &= (1-\delta)F_n^{(0)}(x) + \delta\tilde{F}_n^{(0)}(x),\end{aligned} \tag{19}$$

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 (19) and Theorem 6.

C a s e 2. Distortion in the basic function of trend $\psi(t)$. We consider the following model:

$$\bar{x}_t = \theta^T\tilde{\psi}(t) + \xi_t,\ t \ge 1, \tag{20}$$

where $\tilde{\psi}(t) = \left(\tilde{\psi}_1(t),...,\tilde{\psi}_m(t)\right)^T$ is a basic function of trend such that with a given positive δ, $\|\tilde{\psi}(t) - \psi(t)\| = \max_{1\le i\le m}\sup_{t\ge 1}|\tilde{\psi}_i(t) - \psi_i(t)| \le \delta$.

T h e o r e m 9. *For the model* (20) *and the TSPRT* (3), (11)–(12)*, the following expressions are valid:*

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

P r o o f. Put $\eta(t) = (\theta^0-\theta^1)^T\tilde{\psi}(t) - (\theta^0-\theta^1)^T\psi(t),\ t\ge 1$, then $|\eta(t)| \le \delta\sum_{i=1}^m|\theta_i^0 - \theta_i^1|$. Under hypothesis H_0, we have $\lambda_n \sim N\left(\mu_n^{(0)}, \sigma_n^2\right)$, $\bar{\lambda}_n \sim N\left(\tilde{\mu}_n^{(0)}, \tilde{\sigma}_n^2\right)$. Let $\varphi(x)$ and $\Phi(x)$ be the standard normal PDF and CDF. Therefore, for all $x \in [C_- - C_+, C_+ - C_-]$, $\bar{F}_n^{(0)}(x) - F_n^{(0)}(x) = \Phi\left(\frac{x-\tilde{\mu}_n^{(0)}}{\tilde{\sigma}_n}\right) - \Phi\left(\frac{x-\mu_n^{(0)}}{\sigma_n}\right)$. Using mean value theorem, there exits $\zeta \in \mathbf{R}$ such that

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

On the other hand, $|\tilde{\sigma}_n - \sigma_n| = \frac{\||(\theta^0-\theta^1)^T\tilde{\psi}(n)| - |(\theta^0-\theta^1)^T\psi(n)|\|}{\sigma} \le \frac{|\eta(n)|}{\sigma}$. From that, we get:

$$\bar{F}_n^{(0)}(x) - F_n^{(0)}(x) = O(\delta),\ \forall x \in [C_- - C_+, C_+ - C_-],\ n \ge 1,$$

which implies $\bar{f}^{(1)} = f^{(1)} + O_H(\delta)$, $\bar{D}^{(n)} = D^{(n)} + O_{H\times H}(\delta)$, $n \ge 2$. Therefore, $\bar{f}^{(n)} = f^{(n)} + O_H(\delta)$, $n \ge 1$. The rest part of proof is derived from Theorem 3.

C a s e 3. Joint distortion in both components $\psi(t)$ and ξ_t. Consider the following mixed model:

$$\bar{x}_t = \theta^T\tilde{\psi}(t) + (1-\tau_t)\xi_t + \tau_t\tilde{\xi}_t,\ t \ge 1, \tag{21}$$

where $\{\tau_t, t \ge 1\}$ is a sequence of independent identically distributed random variables, $P(\tau_t = 0) = 1 - \delta_1$, $P(\tau_t = 1) = \delta_1$, and $\tau_t, \xi_t, \tilde{\xi}_t$ are independent, $\|\tilde{\psi}(t) - \psi(t)\| \le \delta_2$, δ_1 and δ_2 are given positive constants, $\delta_1 \in (0,1)$.

T h e o r e m 10. *For the model* (21) *and the TSPRT* (3), (11)–(12), *the following expressions are valid:*

$$\overline{\alpha}_M = \alpha_M + O(h^2) + O(\delta_1) + O(\delta_2),\ E^{(0)}(\overline{N}_M) = E^{(0)}(N_M) + O(h^2) + O(\delta_1) + O(\delta_2).$$

P r o o f. Denote:

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

For $n \ge 1$, we have

$$\overline{F}_n^{(0)}(x) = P_0(\overline{\lambda}_n < x, \tau_n = 0) + P_0(\overline{\lambda}_n < x, \tau_n = 1) = (1-\delta_1)P_0\left(h(\tilde{\psi},\xi,n) < x\right) + \delta_1 P_0\left(h(\tilde{\psi},\tilde{\xi},n) < x\right) =$$
$$= P_0\left(h(\tilde{\psi},\xi,n) < x\right) + \delta_1\left[P_0\left(h(\tilde{\psi},\tilde{\xi},n) < x\right) - P_0\left(h(\tilde{\psi},\xi,n) < x\right)\right].$$

From the proof of Theorem 9, we knew $P_0\left(h(\tilde{\psi},\xi,n) < x\right) = F_n^{(0)}(x) + O(\delta_2), \forall x \in [C_- - C_+, C_+ - C_-]$. Therefore, $\overline{F}_n^{(0)}(x) = F_n^{(0)}(x) + O(\delta_1) + O(\delta_2)$. The rest part of the proof is similar to the proof of Theorem 9.

3. Numerical examples

The probability model (17) was considered and the hypotheses (2) were tested with the following values of parameters: $\sigma = 10$, $\theta^0 = (1,2,2,2)^T$, $\theta^1 = (1,1,2,1)^T$, $\psi(t) = (1, t/10, t^2/10, 1/t)$. The thresholds C_- and C_+ were calculated according to Wald [1]. Denote the sample estimate of a characteristic γ with Monte-Carlo method by $\hat{\gamma}$. The number of repetitions used in Monte-Carlo simulation was 100 000. The index $n_0^{(0)}$ was chosen according to Remark 1 with $\varepsilon_0 = 10^{-5}$. The approximate values $\overline{\alpha}, \overline{t}_0$ constructed as main terms in Theorem 3 and Monte-Carlo estimates $\hat{\alpha}, \hat{t}_0$ are presented in Tab. 1 for different values of partition number *H*, where $t_0 = E^{(0)}(N)$.

Table 1. **Approximate values of the test characteristics for SPRT**

α_0	β_0	$n_0^{(0)}$	$\hat{\alpha}$	$\hat{t}_0$	H	$\overline{\alpha}$	$\overline{t}_0$
0.1	0.1	134	0.07896	46.37639	50	0.08345	46.13523
					100	0.07940	46.35240
0.1	0.05	136	0.07482	51.49942	50	0.08376	51.15358
					100	0.078362	51.43005

With very small value $\varepsilon_0 = 10^{-5}$, the change in value of index n_0 is negligible corresponding to different values of α_0 and β_0. When the value *H* increases, the approximate values $\overline{\alpha}$ and $\overline{t}_0$ are much closer to their Monte-Carlo estimates $\hat{\alpha}$ and $\hat{t}_0$ respectively. To get better approximate values, we can increase $n_0^{(0)}$ or *H*, or both of them, but we should consider the possible amount of time used for calculating as well as computation capacity of the machine.

Next, we choose *H* to be 200. The approximate values of test characteristics calculated according Theorem 4 and Monte-Carlo estimates for the TSPRT are shown in Tab. 2 with different possible numbers of observations *M*.

Table 2. **Approximate values of the test characteristics for TSPRT**

α_0	β_0	M	$\hat{\alpha}_M$	$\overline{\alpha}_M$	$\hat{t}_0(M)$	$\overline{t}_0(M)$
0.1	0.1	40	0.22447	0.22427	38.08305	38.05922
		50	0.15096	0.15226	43.14025	43.11857
0.1	0.05	40	0.22345	0.22472	39.22610	39.22269
		50	0.15435	0.15490	45.97958	45.95119

For the TSPRT, there is no requirement of determining the index n_0, and the maximum number of observations M is usually not too large. Due to these advantages we can possibly increase the number of partitions H to get better accuracy of approximation. In Tab. 2, with the same levels of α_0, β_0 when the value M increases, the error probability α_M decreases but the average number of observations $E^{(0)}(N_M)$ increases. This can easily be understood because the more observations we have, the higher accuracy of the test is. In addition, the average number of observation has an upward trend with respect to M to reach its real expected values in Tab. 1. With $H = 200$, the approximate values $\overline{\alpha}_M$ and $\overline{t}_0(M)$ are relatively close to their Monte-Carlo estimates. Furthermore, compared with Tab. 1, the limitation of maximum number of observations leads to so remarkable change in error probabilities of the test.

References

1. Wald A. *Sequential Analysis*. New York, John Wiley and Sons, 1947. 212 p.

2. Govindarajulu Z. *Sequential Statistics*. Singapore, World Scientific Publishing Co. Pte. Ltd, 2004. 680 p. https://doi.org/10.1142/5575

3. Tartakovsky A., Nikiforov I., Basseville M. *Sequential Analysis: Hypothesis Testing and Changepoint Detection*. CRC Press, 2015. 603 p. https://doi.org/10.1201/b17279

4. Cox D. R., Miller H. D. *The theory of stochastic processes*. New York, John Wiley and Sons, 1965. 398 p.

5. Kharin A. An approach to performance analysis of the sequential probability ratio test for the simple hypotheses testing. *Vestnik BGU. Seriya 1, Fizika. Matematika. Informatika = Proceedings of the Belarusian State University*, 2002, no. 1, pp. 92–96 (in Russian).

6. Kharin A. Performance and robustness evaluation in sequential hypotheses testing. *Communications in Statistics – Theory and Methods,* 2016, vol. 45, no. 6, pp. 1693–1709. https://doi.org/10.1080/03610926.2014.944659

7. Kharin A. Robustness evaluation in sequential testing of composite hypotheses. *Austrian Journal of Statistics*, 2008, vol. 37, no. 1, pp. 51–60. https://doi.org/10.17713/ajs.v37i1.286

8. Kharin A. Y. *Robustness of Bayesian and Sequential Statistical Decision Rules*. Minsk, BSU, 2013, 207 p. (in Russian).

9. Kharin A. Robustness analysis for Bayesian sequential testing of composite hypotheses under simultaneous distortions of priors and likelihoods. *Austrian Journal of Statistics*, 2011, vol. 40, no. 1–2, pp. 65–73. https://doi.org/10.17713/ajs.v40i1&2.198

10. Liu Y., Li X. R. Performance analysis of sequential probability ratio test. *Sequential Analysis*, 2013, vol. 32, no. 4, pp. 469–497. https://doi.org/10.1080/07474946.2013.843329

11. Gut A. *Probability: A Graduate Course*. New York, Springer Science-Business Media, 2005. 603 p. https://doi.org/10.1007/b138932

12. Rudin W. *Principles of Mathematical Analysis*. McGraw-Hill, 1976. 342 p.

13. Mercer P. R. Hadamard's inequality and Trapezoid Rules for the Riemann-Stieltjes integral. *Journal of Mathematical Analysis and Applications,* 2008, vol. 334, no. 2, pp. 921–926. https://doi.org/10.1016/j.jmaa.2008.03.026

14. Kharin A., Galinskij V. On minimax robustness of Bayesian statistical prediction. *Probability Theory and Mathematical Statistics*. Vilnius, TEV, 1999, pp. 259–266.

15. Huber P. *Robust Statistics*. New York, John Wiley and Sons, 1981. 308 p. https://doi.org/10.1002/0471725250

16. Kharin A., Ton That Tu. Performance and robustness analysis of sequential hypotheses testing for time series with trend. *Austrian Journal of Statistics*, 2017, vol. 46, no. 3–4, pp. 23–36. https://doi.org/10.17713/ajs.v46i3-4.668

Список использованных источников

1. Wald, A. Sequential Analysis / A. Wald. – New York: John Wiley and Sons, 1947. – 212 p.

2. Govindarajulu, Z. Sequential Statistics / Z. Govindarajulu. – Singapore: World Scientific Publishing Co. Pte. Ltd, 2004. – 680 p. https://doi.org/10.1142/5575

3. Tartakovsky, A. Sequential Analysis: Hypothesis Testing and Changepoint Detection / A. Tartakovsky, I. Nikiforov, M. Basseville. – New York: Chapman and Hall/CRC, 2014. – 603 p. https://doi.org/10.1201/b17279

4. Cox, D. R. The theory of stochastic processes / D. R. Cox, H. D. Miller. – New York: John Wiley and Sons, 1965. – 398 p.

5. Харин, А. Ю. Об одном подходе к анализу последовательного критерия отношения правдоподобия для различения простых гипотез / А. Харин // Вестн. БГУ. Сер. 1, Физика. Математика. Информатика. – 2002. – № 1. – С. 92–96.

6. Kharin, A. Performance and robustness evaluation in sequential hypotheses testing / A. Kharin // Communications in Statistics – Theory and Methods. – 2016. – Vol. 45, № 6. – P. 1693–1709. https://doi.org/10.1080/03610926.2014.944659

7. Kharin, A. Robustness evaluation in sequential testing of composite hypotheses / A. Kharin // Aust. J. Stat. – 2008. – Vol. 37, № 1. – P. 51–60. https://doi.org/10.17713/ajs.v37i1.286

8. Харин, А. Ю. Робастность байесовских и последовательных статистических решающих правил / А. Ю. Харин. – Минск: БГУ, 2013. – 207 с.

9. Kharin, A. Robustness analysis for Bayesian sequential testing of composite hypotheses under simultaneous distortions of priors and likelihoods / A. Kharin // Aust. J. Stat. – 2011. – Vol. 40, № 1/2. – P. 65–73. https://doi.org/10.17713/ajs.v40i1&2.198

10. Liu, Y. Performance analysis of sequential probability ratio test / Y. Liu, X. R. Li // Sequential Analysis. – 2013. – Vol. 32, № 4. – P. 469–497. https://doi.org/10.1080/07474946.2013.843329

11. Gut, A. Probability: A Graduate Course / A. Gut. – New York: Springer Science-Business Media, 2005. – 603 p. https://doi.org/10.1007/b138932

12. Rudin, W. Principles of mathematical analysis / W. Rudin. – McGraw-Hill, 1976. – 342 p.

13. Mercer, P. R. Hadamard's inequality and Trapezoid Rules for the Riemann-Stieltjes integral / P. R. Mercer // J. Math. Analysis and Applications. – 2008. – Vol. 334, № 2. – P. 921–926. https://doi.org/10.1016/j.jmaa.2008.03.026

14. Kharin, A. On minimax robustness of Bayesian statistical prediction / A. Kharin, V. Galinskij // Probability Theory and Mathematical Statistics. – Vilnius: TEV, 1999. – P. 259–266.

15. Huber, P. Robust Statistics / P. Huber. – New York: John Wiley and Sons, 1981. – 308 p. https://doi.org/10.1002/0471725250

16. Kharin, A. Performance and robustness analysis of sequential hypotheses testing for time series with trend / A. Kharin, Ton That Tu // Aust. J. Stat. – 2017. – Vol. 46, № 3&4. – P. 23–36. https://doi.org/10.17713/ajs.v46i3-4.668

Information about the authors

Alexey Yu. Kharin – Ph. D. (Physics and Mathematics), Associate Professor, Head of the Department of Probability Theory and Mathematical Statistics, Belarusian State University (4, Nezavisimosti Ave., 220030, Minsk, Republic of Belarus). E-mail: kharinAY@bsu.by. https://orcid.org/0000-0002-5790-1956

Ton That Tu – Postgraduate Student, Belarusian State University (4, Nezavisimosti Ave., 220030, Minsk, Republic of Belarus). E-mail: tthattu@gmail.com

Информация об авторах

Харин Алексей Юрьевич – кандидат физико-математических наук, доцент, заведующий кафедрой теории вероятностей и математической статистики, Белорусский государственный университет (пр. Независимости, 4, 220030, г. Минск, Республика Беларусь). E-mail: kharinAY@bsu.by. https://orcid.org/0000-0002-5790-1956

Тон Тхат Ту – соискатель, Белорусский государственный университет (пр. Независимости, 4, 220030, г. Минск, Республика Беларусь). E-mail: tthattu@gmail.com