

On Robustifying of the Sequential Probability Ratio Test for a Discrete Model under “Contaminations”

Alexey Kharin
Belarussian State University, Minsk

Abstract: The problem of robustifying of the sequential probability ratio test is considered for a discrete hypothetical model. Exact values for error probabilities and for conditional expected sample sizes are obtained. Asymptotic robustness analysis for these characteristics is performed under “contaminations”. A two-parametric family of modified sequential probability ratio tests is proposed and analyzed to get the robust test by the minimax risk criterion. Numerical experiments illustrate the theoretical results.

Keywords: Sequential Probability Ratio Test, Robustness, Distortions, Risk.

1 Introduction

The sequential approach (Wald, 1947; Siegmund, 1985) is often used in many applications, especially in medical trials (see Whitehead, 1997; Bauer and Röhmel, 1995), and quality control (see Cowden, 1957), because of two reasons. 1) For some observation schemes it is natural to follow the sequential approach to construct a probability model. 2) Sequential statistical procedures have some optimal properties, e.g. the sequential probability ratio test (SPRT) minimizes the expected sample size (ESS) (see Ghosh, 1970).

The optimal properties of the SPRT are valid only in frames of certain hypothetical models. In practice, model assumptions are usually not satisfied, a hypothetical model is distorted. That is why it is important to analyze robustness (Huber, 1981; Stockinger and Dutter, 1987; Rieder, 1994) of the SPRT, and to construct robust sequential tests, having some information on admissible distortions. To solve these two problems one needs to have exact equations for the main characteristics of the SPRT. To obtain such equations is complicated even for very simple hypothetical models (see Siegmund, 1975).

Some preliminary results on asymptotic performance analysis under distortions are given in Kharin (2001a,b), for the case of two simple hypotheses testing, in frames of a discrete hypothetical model. In the paper we propose a family of modified sequential tests for robustifying of the SPRT, obtain the exact equation for the guaranteed upper risk functional, analyze its asymptotic behaviour and present an algorithm for construction of the robust sequential test by the minimax risk criterion.

2 Hypothetical Model

Let discrete random variables x_1, x_2, \dots be defined on a measurable space (Ω, \mathcal{F}) , $\forall t \in \mathbf{N}$, $x_t \in U = \{u_1, u_2, \dots, u_M\}$, $M < +\infty$, $u_1 < u_2 < \dots < u_M$. Let these random variables be independent identically distributed, according to a discrete probability

distribution with a parameter $\theta \in \Theta = \{\theta_0, \theta_1\}$:

$$P(u; \theta) = P_\theta\{x_t = u\} = a^{-J(u; \theta)}, \quad t \in \mathbf{N}, u \in U, \quad (1)$$

$a \in \mathbf{Q}$, $a > 1$, where $J(u; \theta): U \times \Theta \longrightarrow \mathbf{N}_0$ is a function satisfying the condition: $\sum_{u \in U} a^{-J(u; \theta)} = 1$; \mathbf{N} , \mathbf{N}_0 , \mathbf{Z} , \mathbf{Q} , \mathbf{R} are the sets of natural, nonnegative integer, integer, rational and real numbers respectively.

Denote the accumulated likelihood ratio statistic:

$$\Lambda_n = \Lambda_n(x_1, \dots, x_n) = \sum_{t=1}^n \lambda_t, \quad (2)$$

where

$$\lambda_t = \log_a \frac{P(x_t; \theta_1)}{P(x_t; \theta_0)} = J(x_t; \theta_0) - J(x_t; \theta_1) \in \mathbf{Z} \quad (3)$$

is the likelihood ratio statistic calculated on the observation x_t . It follows from (1) — (3), that

$$\Lambda_n = \Lambda_{n-1} + \lambda_n,$$

λ_n does not depend on Λ_{n-1} , so $\{\Lambda_n\}$ is a Markov chain with discrete time and with countable state space (see Kemeni and Snell, 1959).

Consider the case of two simple hypotheses with respect to the parameter θ :

$$H_0 : \theta = \theta_0, \quad H_1 : \theta = \theta_1. \quad (4)$$

Such a problem appears often when it is necessary to test the hypotheses on the order of elements in some sequence, for example, in genetics (Waterman, 1989), (Durbin, 1998), in quality control (Cowden, 1957), and in sequential signal processing.

To test these hypotheses after n ($n = 1, 2, \dots$) observations one makes the decision

$$d = \mathbf{1}_{[C_+, +\infty)}(\Lambda_n) + 2 \cdot \mathbf{1}_{(C_-, C_+)}(\Lambda_n) \quad (5)$$

according to the SPRT (see Wald, 1947; Siegmund, 1985). The decisions $d = 0$ and $d = 1$ mean stopping of the observation process and the acceptance of the appropriate hypothesis. The decision $d = 2$ means that it is necessary to make the $(n + 1)$ -th observation. In (5) the thresholds C_- , C_+ are the parameters of the test. Following Wald (1947), we may use $C_+ = [\log_a ((1 - \beta_0)/\alpha_0)]$, $C_- = [\log_a (\beta_0/(1 - \alpha_0))]$, where α_0 , β_0 are the given (“desired”) probabilities of the errors of types I and II respectively, and $[\cdot]$ is used for the integer part of an argument.

In fact, the true values for the probabilities of the errors of types I and II differ from α_0 , β_0 . Let us obtain the exact equations for the true values of characteristics mentioned above and for the conditional ESS $t^{(0)}$, $t^{(1)}$.

3 Performance Analysis for the Hypothetical Model

Let $\mathbf{1}_D(\cdot)$ be the indicator function of the set D . Introduce the notation: $\delta_{i,j} = \{1, \text{ if } i = j, 0, \text{ if } i \neq j\}$ is the Kroneker symbol; \mathbf{I}_k is the identity matrix of the k -th order; $\mathbf{0}_{m \times n}$

is the $(m \times n)$ -matrix all elements of which are equal to 0, $\mathbf{1}(u)$ is the unit step function; $\mathbf{1}_k$ is the k -vector-column all elements of which are equal to 1.

For the true hypothesis H_k , let $t^{(k)}$ be the expected time before stopping of the decision process (ESS). Let α, β be the true values of probabilities of the errors of types I and II respectively for the test (5). Let $C_-, C_+ \in \mathbf{Z}, N = C_+ - C_- + 1$.

Let us denote for $k \in \{0, 1\}$ the $(N \times N)$ -matrix

$$P^{(k)} = (p_{ij}^{(k)}) = \left(\begin{array}{c|c} \mathbf{I}_2 & \mathbf{0}_{2 \times (N-2)} \\ \hline R^{(k)} & Q^{(k)} \end{array} \right), \quad (6)$$

where the blocks $R^{(k)}, Q^{(k)}$ are defined as

$$p_{ij}^{(k)} = \begin{cases} \sum_{u \in U} \delta_{J(u; \theta_0) - J(u; \theta_1), j-i} P(u; \theta_k), & i, j \in (C_-, C_+), \\ \sum_{u \in U} \mathbf{1}(C_- - i + J(u; \theta_1) - J(u; \theta_0)) P(u; \theta_k), & i \in (C_-, C_+), j = C_-, \\ \sum_{u \in U} \mathbf{1}(J(u; \theta_0) - J(u; \theta_1) + i - C_+) P(u; \theta_k), & i \in (C_-, C_+), j = C_+, \end{cases} \quad (7)$$

and the vector

$$\pi^{(k)} = (\pi_i^{(k)}), \quad \pi_i^{(k)} = \sum_{u \in U} \delta_{J(u; \theta_0) - J(u; \theta_1), i} P(u; \theta_k), \quad i \in (C_-, C_+); \quad (8)$$

$$\begin{aligned} \pi_{C_+}^{(k)} &= \sum_{i \geq C_+} \sum_{u \in U} \delta_{J(u; \theta_0) - J(u; \theta_1), i} P(u; \theta_k), \\ \pi_{C_-}^{(k)} &= \sum_{i \leq C_-} \sum_{u \in U} \delta_{J(u; \theta_0) - J(u; \theta_1), i} P(u; \theta_k). \end{aligned}$$

Define also the matrices $S^{(k)} = \mathbf{I}_{N-2} - Q^{(k)}, B^{(k)} = (S^{(k)})^{-1} R^{(k)}$. Let $W_{(i)}$ denotes the i -th column of a matrix W .

Theorem 1 (Kharin, 2001b) *If under conditions (1) — (4) the true hypothesis is the hypothesis H_k , and $|S^{(k)}| \neq 0, k \in \{0, 1\}$, then for the test (5)*

$$t^{(k)} = (\pi^{(k)})' (S^{(k)})^{-1} \mathbf{1}_{N-2} + 1, \quad \alpha = (\pi^{(0)})' B_{(2)}^{(0)} + \pi_{C_+}^{(0)}, \quad \beta = (\pi^{(1)})' B_{(1)}^{(1)} + \pi_{C_-}^{(1)}. \quad (9)$$

4 Asymptotic Performance Analysis under “Contaminations”

Let the hypothetical model (1), (4) be under “contaminations” of Tukey–Huber type (see Huber, 1981); instead of (1) the observations x_1, x_2, \dots belong to a mixed discrete probability distribution

$$\bar{P}(u; \theta) = \bar{P}_\theta \{x_t = u\} = (1 - \varepsilon) P(u; \theta) + \varepsilon \tilde{P}(u; \theta), \quad t \in \mathbf{N}, \quad u \in U, \quad (10)$$

where $\varepsilon \in [0, \frac{1}{2})$ is a probability of “contamination” presence; the “contaminating” probability distribution is

$$\tilde{P}(u; \theta) = a^{-\tilde{J}(u; \theta)}, \quad (11)$$

and $\tilde{J}(u; \theta): U \times \Theta \rightarrow \mathbf{N}_0$ is a function different from $J(\cdot)$ and satisfying the norm condition $\sum_{u \in U} a^{-\tilde{J}(u; \theta)} = 1$.

Let $\tilde{P}^{(k)}$, $\tilde{\pi}^{(k)}$ be defined similarly to $P^{(k)}$, $\pi^{(k)}$ according to (6) — (8) by substitution $\tilde{P}(u; \theta)$ instead of $P(u; \theta)$, and be not depend on ε . Let $\tilde{Q}^{(k)}$, $\tilde{R}^{(k)}$ be the blocks of the matrix $\tilde{P}^{(k)}$, $k \in \{0, 1\}$. Let $\tilde{\pi}_{C_{\pm}}^{(k)}$ be calculated similarly to $\pi_{C_{\pm}}^{(k)}$, $k \in \{0, 1\}$, with substitution $\tilde{P}(u; \theta_k)$ instead of $P(u; \theta_k)$. Define the matrix $\hat{S}^{(k)} = \mathbf{I}_{N-2} - Q^{(k)} - \varepsilon(\tilde{Q}^{(k)} - Q^{(k)})$.

Theorem 2 *If the hypothetical model (1), (4) is distorted according to (10), (11), the true hypothesis is H_k , and $|S^{(k)}| \neq 0$, $|\hat{S}^{(k)}| \neq 0$, $k \in \{0, 1\}$, then the conditional ESS $\bar{t}^{(k)}$ and the true probabilities $\bar{\alpha}$, $\bar{\beta}$ of the errors of the types I and II for the distorted model, differ from the corresponding characteristics for the hypothetical model by the values of the order $\mathcal{O}(\varepsilon)$:*

$$\bar{t}^{(k)} - t^{(k)} = \varepsilon \left((\tilde{\pi}^{(k)} - \pi^{(k)})' + (\pi^{(k)})'(S^{(k)})^{-1}(\tilde{Q}^{(k)} - Q^{(k)}) \right) \times (S^{(k)})^{-1} \mathbf{1}_{N-2} + \mathcal{O}(\varepsilon^2); \quad (12)$$

$$\bar{\alpha} - \alpha = \varepsilon \left((\pi^{(0)})' \left((S^{(0)})^{-1} \left((\tilde{Q}^{(0)} - Q^{(0)})(S^{(0)})^{-1} R^{(0)} + \tilde{R}^{(0)} - R^{(0)} \right) \right)_{(2)} + (\tilde{\pi}^{(0)} - \pi^{(0)})' B_{(2)}^{(0)} + \tilde{\pi}_{C_+}^{(0)} - \pi_{C_+}^{(0)} \right) + \mathcal{O}(\varepsilon^2), \quad (13)$$

$$\bar{\beta} - \beta = \varepsilon \left((\pi^{(1)})' \left((S^{(1)})^{-1} \left((\tilde{Q}^{(1)} - Q^{(1)})(S^{(1)})^{-1} R^{(1)} + \tilde{R}^{(1)} - R^{(1)} \right) \right)_{(1)} + (\tilde{\pi}^{(1)} - \pi^{(1)})' B_{(1)}^{(1)} + \tilde{\pi}_{C_-}^{(1)} - \pi_{C_-}^{(1)} \right) + \mathcal{O}(\varepsilon^2). \quad (14)$$

Proof. It consists of six stages. 1) Define the random sequence

$$\xi_n = C_+ \mathbf{1}_{[C_+, +\infty)}(\Lambda_n) + C_- \mathbf{1}_{(-\infty, C_-]}(\Lambda_n) + \Lambda_n \mathbf{1}_{(C_-, C_+)}(\Lambda_n) \in \mathbf{Z}, \quad n \in \mathbf{N}. \quad (15)$$

2) Prove that the sequence (15) is a Markov chain with discrete time and a countable state space. 3) Obtain the matrix of the transition probabilities and the vector of the initial probabilities of the nonabsorbing states for the Markov chain (15) in the form

$$\bar{P}^{(k)} = P^{(k)} + \varepsilon(\tilde{P}^{(k)} - P^{(k)}), \quad \bar{\pi}^{(k)} = \pi^{(k)} + \varepsilon(\tilde{\pi}^{(k)} - \pi^{(k)}), \quad k \in \{0, 1\}, \quad (16)$$

respectively. 4) Get (12) from (9) and (16). 5) Construct the asymptotic expansion w.r.t. ε for the difference between the absorption probabilities matrices:

$$\bar{B}^{(k)} - B^{(k)} = \varepsilon(S^{(k)})^{-1} \left((\tilde{Q}^{(k)} - Q^{(k)})(S^{(k)})^{-1} R^{(k)} + \tilde{R}^{(k)} - R^{(k)} \right) + \mathcal{O}(\varepsilon^2). \quad (17)$$

6) Come to (13), (14) by using (17), (16), and (9). ■

Using this result one can approximate the characteristics of the SPRT under distortions with the linear accuracy by ε . Some numerical results relating such an approximation are given by Kharin (2001a).

5 Robustness Characteristics and the Robust Sequential Test

To robustify the SPRT w.r.t. large deviations from the hypothetical distributions (see Huber, 1981) of λ_t , $t \in \mathbf{N}$ consider a family of tests, based on the modification of the SPRT:

$$d_n = \mathbf{1}_{[C_+, +\infty)}(\Lambda_n^g) + 2 \cdot \mathbf{1}_{(C_-, C_+)}(\Lambda_n^g), \quad (18)$$

$$\Lambda_n^g = \sum_{i=1}^n g(\lambda_i), \quad \lambda_i = J(u; \theta_0) - J(u; \theta_1), \quad (19)$$

where $g(\cdot)$ is a function, such that

$$g(z) = g_- \cdot \mathbf{1}_{(-\infty, g_-]}(z) + g_+ \cdot \mathbf{1}_{[g_+, +\infty)}(z) + z \cdot \mathbf{1}_{(g_-, g_+)}(z), \quad (20)$$

and $g_-, g_+ \in \mathbf{R}$ are some parameters, $g_- < g_+$. Note, that if $g_- < C_- - C_+$, and $g_+ > C_+ - C_-$, then the test (18) — (20) turns into the SPRT.

We assume prior probabilities π_0, π_1 of the hypotheses H_0 and H_1 respectively to be known. Let $\bar{t}^{(k)}(g(\cdot))$ be the conditional ESS under the true hypothesis H_k . Let us denote by $\bar{\alpha}(g(\cdot)), \bar{\beta}(g(\cdot))$ probabilities of errors of type I and II respectively for the test (18) — (20) under distortions (10), (11). Let $w_0, w_1 \geq 0$ be the losses, caused by the errors of type I and type II respectively.

Introduce the robustness characteristics: 1) the risk functional for the distorted model as the loss function

$$r(g(\cdot); \bar{P}(\cdot, \cdot)) = w_0 \pi_0 \bar{\alpha}(g(\cdot)) + w_1 \pi_1 \bar{\beta}(g(\cdot)), \quad (21)$$

2) the guaranteed upper risk functional as

$$r_*(g(\cdot)) = \sup_{\bar{P}(\cdot, \cdot)} r(g(\cdot); \bar{P}(\cdot, \cdot)), \quad (22)$$

where $g(\cdot)$ is defined by (20).

Consider a random sequence

$$\xi_n(g(\cdot)) = C_- \mathbf{1}_{(-\infty, C_-]}(\Lambda_n^g) + C_+ \mathbf{1}_{[C_+, +\infty)}(\Lambda_n^g) + \Lambda_n^g \mathbf{1}_{(C_-, C_+)}(\Lambda_n^g), \quad (23)$$

where Λ_n^g is defined in (19).

Lemma 1 *For the considered model the random sequence (23) is a Markov chain with N states, the states C_- and C_+ are absorbing; if the true hypothesis is H_k , then the one-step transition probabilities matrix has the form*

$$\bar{P}^{(k)}(g(\cdot)) = (\bar{p}_{ij}^{(k)}(g(\cdot))) = \begin{pmatrix} \mathbf{I}_2 & | & \mathbf{0}_{2 \times (N-2)} \\ \hline \bar{R}^{(k)}(g(\cdot)) & | & \bar{Q}^{(k)}(g(\cdot)) \end{pmatrix}, \quad (24)$$

where the blocks $\bar{R}^{(k)}(g(\cdot))$, $\bar{Q}^{(k)}(g(\cdot))$ are given by

$$\bar{P}_{ij}^{(k)}(g(\cdot)) = \begin{cases} \sum_{u \in U} \delta_{g(J(u, \theta_0) - J(u, \theta_1), j-i)} \bar{P}(u; \theta_k), & j \in (C_-, C_+), \\ \sum_{u \in U} \mathbf{1}_{(-\infty, C_-]}(g(J(u, \theta_0) - J(u, \theta_1)) + i) \bar{P}(u; \theta_k), & j = C_-, \\ \sum_{u \in U} \mathbf{1}_{[C_+, +\infty)}(g(J(u, \theta_0) - J(u, \theta_1)) + i) \bar{P}(u; \theta_k), & j = C_+, \end{cases} \quad (25)$$

$$i \in (C_-, C_+),$$

and the vector of initial probabilities of nonabsorbing states is given by

$$\bar{\pi}^{(k)}(g(\cdot)) = (\bar{\pi}_i^{(k)}(g(\cdot))) = \begin{pmatrix} \bar{\pi}_{C_-+1}^{(k)}(g(\cdot)) \\ \vdots \\ \bar{\pi}_{C_+-1}^{(k)}(g(\cdot)) \end{pmatrix},$$

$$\bar{\pi}_i^{(k)}(g(\cdot)) = \sum_{u \in U} \delta_{g(J(u; \theta_0) - J(u; \theta_1), i)} \bar{P}(u; \theta_k), \quad i \in (C_-, C_+). \quad (26)$$

Proof. To prove Lemma 1 it is enough to apply the aggregation criterion for the Markov chain Λ_n^g with the countable set of states (see Kemeni and Snell, 1959). ■

Introduce the initial probabilities for accepting of hypotheses

$$\bar{\pi}_{C_+}^{(k)} = \sum_{i \geq C_+} \sum_{u \in U} \delta_{g(J(u; \theta_0) - J(u; \theta_1), i)} \bar{P}(u; \theta_k), \quad \bar{\pi}_{C_-}^{(k)} = \sum_{i \leq C_-} \sum_{u \in U} \delta_{g(J(u; \theta_0) - J(u; \theta_1), i)} \bar{P}(u; \theta_k), \quad (27)$$

and the matrices $\bar{S}^{(k)}(g(\cdot)) = \mathbf{I}_{N-2} - \bar{Q}^{(k)}(g(\cdot))$, $\bar{B}^{(k)}(g(\cdot)) = (\bar{S}^{(k)}(g(\cdot)))^{-1} \bar{R}^{(k)}(g(\cdot))$, $k \in \{0, 1\}$.

Theorem 3 For the distorted model (2) — (4), (10), (11), if the test (18) — (20) is used, and $|\bar{S}^{(k)}(g(\cdot))| \neq 0$, $k \in \{0, 1\}$, then

$$\bar{t}^{(k)}(g(\cdot)) = (\bar{\pi}^{(k)}(g(\cdot)))' (\bar{S}^{(k)}(g(\cdot)))^{-1} \mathbf{1}_{N-2} + 1,$$

$$\bar{\alpha}(g(\cdot)) = (\bar{\pi}^{(0)}(g(\cdot)))' \bar{B}_{(2)}^{(0)}(g(\cdot)) + \bar{\pi}_{C_+}^{(0)}(g(\cdot)),$$

$$\bar{\beta}(g(\cdot)) = (\bar{\pi}^{(1)}(g(\cdot)))' \bar{B}_{(1)}^{(1)}(g(\cdot)) + \bar{\pi}_{C_-}^{(1)}(g(\cdot)).$$

Proof. It is done by applying the absorbing Markov chain theory (Kemeni and Snell, 1959) and Lemma 1 to the sequence (23). ■

Let $\tilde{P}^{(k)}(g(\cdot))$, $\tilde{\pi}^{(k)}(g(\cdot))$, and $\tilde{\pi}_{C_{\pm}}^{(k)}(g(\cdot))$ be calculated similarly to (24) — (27), respectively, by substitution $\tilde{P}(\cdot; \cdot)$ instead of $\bar{P}(\cdot; \cdot)$, and $\tilde{Q}^{(k)}(g(\cdot))$, $\tilde{R}^{(k)}(g(\cdot))$ be the blocks of the matrix $\tilde{P}^{(k)}(g(\cdot))$, $k \in \{0, 1\}$. Introduce the matrix $\hat{S}^{(k)}(g(\cdot)) = \mathbf{I}_{N-2} - \tilde{Q}^{(k)} - \varepsilon(\tilde{Q}^{(k)}(g(\cdot)) - Q^{(k)})$.

Corollary 1 Under the Theorem 3 conditions, if $|\hat{S}^{(k)}(g(\cdot))| \neq 0$, $k \in \{0, 1\}$, then

$$\bar{t}^{(k)}(g(\cdot)) - t^{(k)} = \varepsilon \left((\tilde{\pi}^{(k)}(g(\cdot)) - \pi^{(k)})' + (\pi^{(k)})' (S^{(k)})^{-1} (\tilde{Q}^{(k)}(g(\cdot)) - Q^{(k)}) \right) \times (S^{(k)})^{-1} \mathbf{1}_{N-2} + \mathcal{O}(\varepsilon^2);$$

$$\bar{\alpha}(g(\cdot)) - \alpha = \varepsilon \left((\pi^{(0)})' \left((S^{(0)})^{-1} \left((\tilde{Q}^{(0)}(g(\cdot)) - Q^{(0)}) (S^{(0)})^{-1} R^{(0)} + \tilde{R}^{(0)}(g(\cdot)) - R^{(0)} \right) \right)_{(2)} + (\tilde{\pi}^{(0)}(g(\cdot)) - \pi^{(0)})' B_{(2)}^{(0)} + \tilde{\pi}_{C_+}^{(0)}(g(\cdot)) - \pi_{C_+}^{(0)} \right) + \mathcal{O}(\varepsilon^2),$$

$$\bar{\beta}(g(\cdot)) - \beta = \varepsilon \left((\pi^{(1)})' \left((S^{(1)})^{-1} \left((\tilde{Q}^{(1)}(g(\cdot)) - Q^{(1)}) (S^{(1)})^{-1} R^{(1)} + \tilde{R}^{(1)}(g(\cdot)) - R^{(1)} \right) \right)_{(1)} + (\tilde{\pi}^{(1)}(g(\cdot)) - \pi^{(1)})' B_{(1)}^{(1)} + \tilde{\pi}_{C_-}^{(1)}(g(\cdot)) - \pi_{C_-}^{(1)} \right) + \mathcal{O}(\varepsilon^2).$$

Proof. The scheme of this proof is similar to the proof of Theorem 2. ■

Let us denote:

$$\begin{aligned} U_{g_+} &= \{u \in U : u = \arg \max_{u \in U} g(J(u; \theta_0) - J(u; \theta_1))\}, \\ U_{g_-} &= \{u \in U : u = \arg \min_{u \in U} g(J(u; \theta_0) - J(u; \theta_1))\}. \end{aligned} \tag{28}$$

Theorem 4 *If the hypothetical model (1) — (4) is under distortions (10), (11) and the test (18) — (20) is used, then the guaranteed upper risk functional (22) is*

$$r_*(g(\cdot)) = r(g(\cdot), \tilde{P}_*(\cdot; \cdot)), \tag{29}$$

where “the worst probability distribution” $\tilde{P}_*(\cdot; \cdot)$ satisfies the conditions

$$\sum_{u \in U_{g_+}} \tilde{P}_*(u; \theta_0) = 1, \quad \sum_{u \in U_{g_-}} \tilde{P}_*(u; \theta_1) = 1. \tag{30}$$

Proof. To maximize the risk functional (21) it is sufficient to maximize error probabilities of the type I and II. By choosing $\tilde{P}(u; \theta_0)$ we maximize $\bar{\alpha}(g(\cdot))$, and by $\tilde{P}(u; \theta_1)$ — $\bar{\beta}(g(\cdot))$. Having solved these maximization problems, we come to (28), (30), and therefore, we get (29). ■

Denote by $g_e(\cdot)$ the identity function: $g_e(z) \equiv z, z \in \mathbf{Z}$. Let us indicate by the subindex $(\cdot)_*$ the fact that the “worst probability distribution” $\tilde{P}_*(\cdot; \cdot)$ has been used for calculation of the characteristic (\cdot) .

Corollary 2 *Under the Theorem 4 and Corollary 1 conditions the guaranteed upper risk functional (22) satisfies the asymptotic expansion*

$$\begin{aligned} r_*(g(\cdot)) &= r(g_e(\cdot); P(\cdot; \cdot)) + \\ \varepsilon &\left(w_0 \pi_0 \left((\pi^{(0)})' \left((S^{(0)})^{-1} \left((\tilde{Q}_*^{(0)}(g(\cdot)) - Q^{(0)}) (S^{(0)})^{-1} R^{(0)} + \tilde{R}_*^{(0)}(g(\cdot)) - R^{(0)} \right) \right)_{(2)} + \right. \\ &\quad \left. (\tilde{\pi}_*^{(0)}(g(\cdot)) - \pi^{(0)})' B_{(2)}^{(0)} + \tilde{\pi}_{*C_+}^{(0)}(g(\cdot)) - \pi_{C_+}^{(0)} \right) + \\ w_1 \pi_1 &\left((\pi^{(1)})' \left((S^{(1)})^{-1} \left((\tilde{Q}_*^{(1)}(g(\cdot)) - Q^{(1)}) (S^{(1)})^{-1} R^{(1)} + \tilde{R}_*^{(1)}(g(\cdot)) - R^{(1)} \right) \right)_{(1)} + \right. \\ &\quad \left. (\tilde{\pi}_*^{(1)}(g(\cdot)) - \pi^{(1)})' B_{(1)}^{(1)} + \tilde{\pi}_{*C_-}^{(1)}(g(\cdot)) - \pi_{C_-}^{(1)} \right) \Big) + \mathcal{O}(\varepsilon^2). \end{aligned} \tag{31}$$

Proof. To prove Corollary 2 it is sufficient to represent (29) in the form (21). Using it and the result of Corollary 1, we come to (31). ■

Corollary 2 states, that the increment of the guaranteed upper risk functional under distortions of a level ε , w.r.t. the risk functional of the SPRT without distortions, has the first order w.r.t. ε :

$$r_*(g(\cdot)) - r(g_e(\cdot), P(\cdot, \cdot)) = \mathcal{O}(\varepsilon).$$

It is reasonable to minimize the coefficient at ε , indicated in Corollary 2, by choosing of g_-, g_+ .

Using the result of Theorem 4, and Corollary 2, let us propose the criterion for the robust test construction. For a fixed parameter $1 \leq C_t < +\infty$, define the robust test by the minimax risk criterion, as a solution of the extremal problem

$$\begin{cases} r_*(g(\cdot)) \longrightarrow \min_{g(\cdot)}, \\ \pi_0 \bar{t}^0(g(\cdot)) + \pi_1 \bar{t}^1(g(\cdot)) \leq C_t(\pi_0 t^0 + \pi_1 t^1), \end{cases} \quad (32)$$

where a set of admissible functions $g(\cdot) : \mathbf{Z} \longrightarrow \mathbf{Z}$ is given by (20). The restriction on the ESS means that the ESS of the robust test should be not more than C_t times greater than the ESS for the SPRT for the hypothetical case. This problem can be solved numerically, varying parameters g_-, g_+ through the set of admissible values. One of the possible algorithms is simply to try all the integer values between C_- and C_+ , which has the computational complexity $\mathcal{O}(N^2)$. If there is no solution, then the value of C_t should be increased.

Let us also mention a simple approach to robustify the SPRT for the situation, where $J(u; \theta) > 0$, $u \in U$, $\theta \in \Theta$. To avoid large influences of a ‘‘contaminating’’ distribution, we can use the following function, without solving (32),

$$g(z) = z(1 - \mathbf{1}_{\{\lambda_{max}, \lambda_{min}\}}(z)) + \delta_{\lambda_{max}, z} - \delta_{\lambda_{min}, z}, \quad (33)$$

where $\lambda_{max} = \max_{u \in U} (J(u; \theta_0) - J(u; \theta_1))$, and $\lambda_{min} = \min_{u \in U} (J(u; \theta_0) - J(u; \theta_1))$. For a nonsingular case $\lambda_{max} \geq 1$, $\lambda_{min} \leq -1$. The test (33) is not the robust test by the minimax criterion (32), but it is easy constructed. The formula (33) means that the influence of a summand into (19) under distortions should be less than the maximal possible influence without distortions.

6 Results of Experiments

Let us illustrate the theoretical results by the results of numerical experiments in the example: $a = 2$, $M = 10$, $U = \{1, 2, \dots, 10\}$, the functions $J(u; \theta_0)$, $J(u; \theta_1)$ are defined by the following table.

u	1	2	3	4	5	6	7	8	9	10
$J(u; \theta_0)$	1	2	3	4	5	6	7	8	9	9
$J(u; \theta_1)$	3	2	4	5	6	7	8	9	9	1

Let the contaminating distribution be $\tilde{P}(10; \theta_0) = 1$, $\tilde{P}(u; \theta_0) = 0$, $u \in U \setminus \{10\}$; $\tilde{P}(1; \theta_1) = 1$, $\tilde{P}(u; \theta_1) = 0$, $u \in U \setminus \{1\}$. For this case $\lambda_{max} = 8$, $\lambda_{min} = -2$. Choose $\alpha_0 = \beta_0 = 0.01$. For this case $C_- = -6$, $C_+ = 7$.

In Figure 1 we present the results of statistical simulation for the sum of conditional error probabilities $\bar{\alpha}(g(\cdot)) + \bar{\beta}(g(\cdot))$. The results are presented for three tests: for the classical Wald SPRT, and for two its modifications. The first modification corresponds to the the robust sequential test by the minimax risk criterion, obtained as the solution of (32) for small values of ε . This test is correspondent to the values $g_- = -1, g_+ = 2$. The second modification is obtained with the values $g_- = -1, g_+ = 5$, and is robust by the minimax risk criterion for $\varepsilon > 0.33$.

In Figure 2 the results of statistical simulation for the sum of conditional expected sample sizes $\bar{t}^{(0)}(g(\cdot)) + \bar{t}^{(1)}(g(\cdot))$ are given for the same three tests.

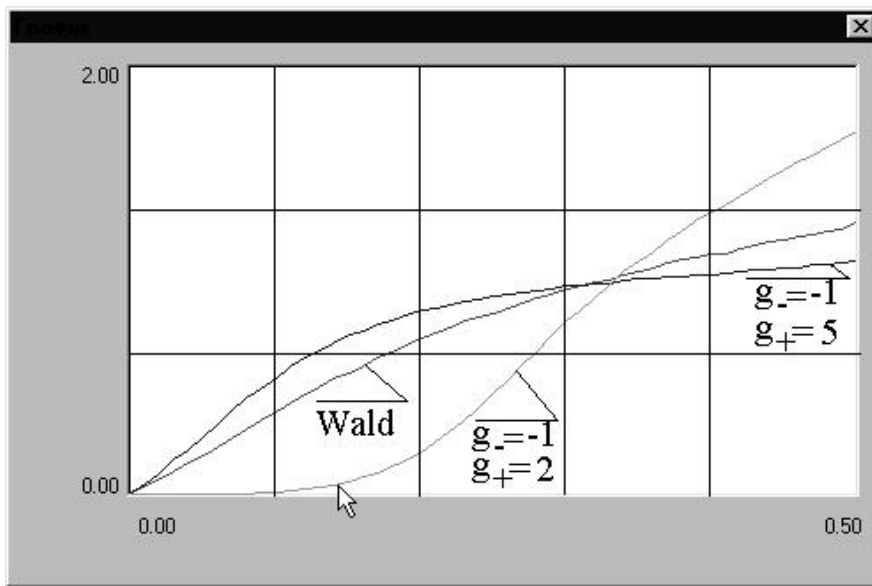


Figure 1: Dependencies of estimates of $\bar{\alpha}(g(\cdot)) + \bar{\beta}(g(\cdot))$ on ε

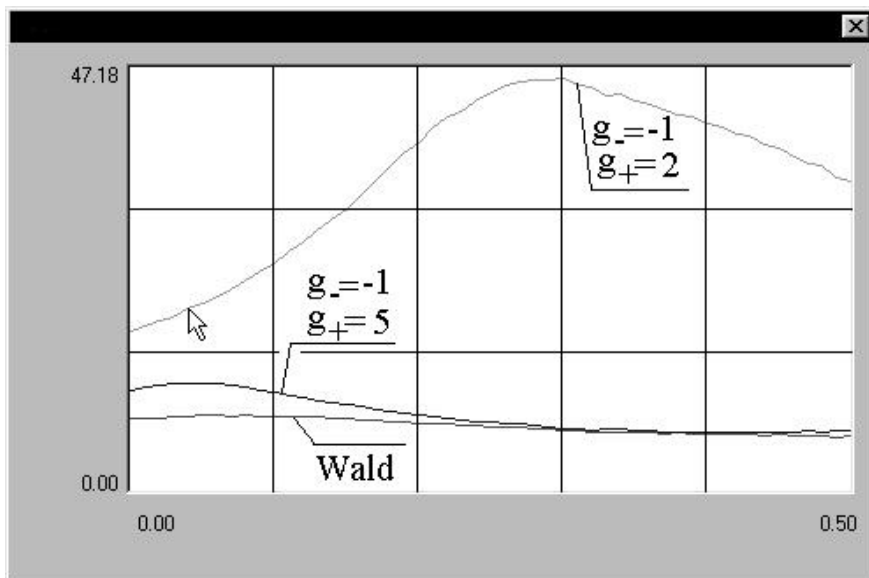


Figure 2: Dependencies of estimates of $\bar{t}^{(0)}(g(\cdot)) + \bar{t}^{(1)}(g(\cdot))$ on ε

7 Conclusion

The results of the paper give a possibility to calculate the robustness characteristics of a sequential test from the considered family of tests, and to construct the robust test under “contaminations” w.r.t. the minimax criterion. In practice, the exact value of the “contamination” probability is unknown, but usually an information of the type $\varepsilon \leq \varepsilon_+$ is available. For such a case, using the monotonicity of the risk functional w.r.t. the value of ε , one can get the upper estimates for the error probabilities, for conditional ESSs, and for the guaranteed upper risk functional. These results are obtained for the discrete probability distributions of observations. The same methodology is shown to be useful for the absolutely continuous case (see Kharin, 2002).

Acknowledgments

The research was supported by the INTAS Fellowship Program for Young Scientists, grant No. YSF-01/1-75. The author is thankful to Professor Rudolf Dutter, personally, and to his colleagues from the Institute of Statistics and Probability Theory, for wonderful conditions to prepare the paper in the Vienna University of Technology.

The author is also thankful to the Referee for useful comments which helped to improve the paper.

References

- P. Bauer and J. Röhmel. An adaptive method for establishing a dose response relationship. *Statistics in Medicine*, 14:1595–1607, 1995.
- J. Cowden. *Statistical Methods in Quality Control*. Prentice-hall, Englewood Cliffs, 1957.
- R. Durbin. *Biological Sequence Analysis*. Cambridge University Press, Cambridge, 1998.
- B.K. Ghosh. *Sequential Tests of Statistical Hypotheses*. Addison-Wesley, Reading, 1970.
- P. Huber. *Robust Statistics*. Wiley, New York, 1981.
- J.G. Kemeni and J.L. Snell. *Finite Markov Chains*. Wiley, New York, 1959.
- A. Kharin. On asymptotic robustness and performance analysis for sequential hypotheses testing. *Proceedings of the International Conference "Applied Stochastic Models and Data Analysis"*, pages 56–61, 2001a.
- A. Kharin. Robustness of the sequential probability ratio test for discrete contaminated data. *Proceedings of the VI-th International Conference "Computer Data Analysis and Modeling"*, pages 185–191, 2001b.
- A. Kharin. An approach to performance analysis of the sequential probability ratio test for simple hypotheses. *Proc. of the Belarussian State University*, 1:92–96, 2002.

- H. Rieder. *Robust Asymptotic Statistics*. Springer–Verlag, New York, 1994.
- D. Siegmund. Error probabilities and average sample number of the sequential probability ratio test. *J. Roy. Statist. Soc. Ser. B*, 37:394–401, 1975.
- D. Siegmund. *Sequential Analysis. Tests and Confidence Intervals*. Springer–Verlag, New York, 1985.
- N. Stockinger and R. Dutter. Robust time series analysis: a survey. *Kybernetika*, 23, 1987.
- A. Wald. *Sequential Analysis*. John Wiley and Sons, New York, 1947.
- M.S. Waterman. *Mathematical Methods for DNA Sequences*. CRC Press, Boca Raton, 1989.
- J. Whitehead. *The Design and Analysis of Sequential Clinical Trials*. John Wiley and Sons, New York, 1997.

Author's address:

Assist. Prof. Dr. Alexey Kharin
Department of Probability Theory and Mathematical Statistics
Belarussian State University
Fr. Skoriny av. 4
Minsk 220050
Belarus

Tel. +375 17 2095394
Fax +375 17 2095054
E-mail: KharinAY@bsu.by
<http://www.fpmi.bsu.by/>