

Robustness Evaluation in Sequential Testing of Composite Hypotheses

Alexey Kharin

Belarusian State University, Minsk

Abstract: The problem of sequential testing of composite hypotheses is considered. Asymptotic expansions are constructed for the conditional error probabilities and expected sample sizes under “contamination” of the probability distribution of observations. To obtain these results a new approach based on approximation of the generalized likelihood ratio statistic by a specially constructed Markov chain is proposed. The approach is illustrated numerically.

Keywords: Sequential Test, Approximation, Asymptotic Expansion.

1 Introduction

Sequential testing of hypotheses (Wald, 1947) is used as an adequate statistical methodology in applications where not only accuracy (percentage of correct decisions), but also the number of observations used is important: medicine (Jennison and Turnbull, 2000), quality control, finance (Ghosh and Sen, 1991), etc.

In the applications sequential methods are often applied to “contaminated” data (see Huber, 1981), and this fact results in significant increasing of percentage of incorrect decisions (A. Kharin, 2002); the optimal properties of sequential procedures (see Aivazian, 1959) become to be broken. Some results on robustness analysis of sequential tests for simple hypotheses case are presented in Quang (1985); A. Kharin (2002); A. Y. Kharin and Kishylau (2005). The robust sequential tests are constructed in case of simple hypotheses for several models of data in A. Kharin (2002); A. Kharin and Kishylau (2005).

In this paper we evaluate robustness of the conditional error probabilities and expected sample sizes under “contamination” of the probability distribution of data for composite hypotheses.

2 Mathematical Model

Let on a measurable space (Ω, \mathcal{F}) a random sequence $x_1, x_2, \dots \in \mathbf{R}$ with an n -variate probability density function (pdf) $p_n(x_1, \dots, x_n \mid \theta)$ be observed, where $n \in \mathbf{N} = \{1, 2, \dots\}$, $\theta \in \Theta \subseteq \mathbf{R}^k$ is an unknown value of the random vector of parameters. The pdf $p(\theta)$ of the parameter θ is supposed to be known. Consider two composite hypotheses:

$$\mathcal{H}_0 : \theta \in \Theta_0, \mathcal{H}_1 : \theta \in \Theta_1, \quad (1)$$

where $\Theta_0 \cup \Theta_1 = \Theta$, $\Theta_0 \cap \Theta_1 = \emptyset$.

To construct a sequential test for the hypotheses (1) in the considered Bayesian setting, let us use the method of weight functions proposed in Wald (1947). Introduce the notation:

$$\mathbf{1}_S(s) = \begin{cases} 1, & s \in S, \\ 0, & s \notin S, \end{cases}$$

is the indicator function of a set S ;

$$W_i = \int_{\Theta_i} p(\theta) d\theta, \quad w_i(\theta) = \frac{1}{W_i} p(\theta) \mathbf{1}_{\Theta_i}(\theta), \quad \theta \in \Theta, \quad i = 0, 1;$$

$$\Lambda_n = \Lambda_n(x_1, \dots, x_n) = \log \frac{\int_{\Theta} w_1(\theta) p_n(x_1, \dots, x_n | \theta) d\theta}{\int_{\Theta} w_0(\theta) p_n(x_1, \dots, x_n | \theta) d\theta}. \quad (2)$$

The following parametric family of sequential tests is used in the notation (2) for the hypotheses (1):

$$N = \min \{n \in \mathbf{N} : \Lambda_n \notin (C_-, C_+)\}, \quad (3)$$

$$d = \mathbf{1}_{[C_+, +\infty)}(\Lambda_N), \quad (4)$$

where N is the random number of observation (stopping time), at which the decision d is made according to (4). The decision $d = i$ means that the hypothesis \mathcal{H}_i , $i = 0, 1$, is accepted; $C_- < 0$, $C_+ > 0$ are two parameters of the test (3), (4). In Wald (1947) the following expressions are used for these parameters:

$$C_- = \log \frac{\beta_0}{1 - \alpha_0}, \quad C_+ = \log \frac{1 - \beta_0}{\alpha_0}, \quad (5)$$

where α_0, β_0 are maximal admissible values for the error type I (accept \mathcal{H}_1 provided \mathcal{H}_0 is really true) and type II (accept \mathcal{H}_0 provided \mathcal{H}_1 is really true) probabilities, $\beta_0 < 1 - \alpha_0$.

The true values α, β of the error type I and II probabilities differ from α_0, β_0 (see e.g. A. Kharin, 2002), and calculation of these characteristics is an important problem (Lai, 2001) in the aspect of quantitative robustness analysis. It is also important to get expressions for calculation of conditional mathematical expectation of the sample size N for the sequential test (3), (4) with the fixed parameters (5).

3 Approximation of the Random Sequence Λ_n by a Markov Chain

Let us split the state space \mathbf{R} of the sequence Λ_n on $m + 2$ “cells”:

$$A_0 = (-\infty, C_-); \quad A_i = [C_{i-1}, C_i), \quad i = 1, \dots, m; \quad A_{m+1} = [C_+, +\infty),$$

where $m \in \mathbf{N}$,

$$h = \frac{C_+ - C_-}{m}; \quad C_- = C_0 < C_1 < \dots < C_{m-1} < C_m = C_+, \quad C_i = C_- + i \cdot h.$$

Introduce the notation: $S_1^m = \{A_1, \dots, A_m\}$ is the set of the intervals; $p_{\Lambda_n}(u)$ is the pdf of the generalized likelihood ratio statistic (2); $p_{\Lambda_{n+1}|\Lambda_n}(u | y)$ is the conditional pdf of Λ_{n+1} provided $\Lambda_n = y$, $n \in \mathbf{N}$.

Theorem 1 Let Λ_n be a Markov sequence, $\forall \theta \in \Theta$ the pdf's $p_{\Lambda_1}(u)$, $p_{\Lambda_{n+1}|\Lambda_n}(u|y)$, $n \in \mathbf{N}$, be differentiable functions of the variable $u \in [C_-, C_+]$, and $\exists c_1, c_2$, $0 < c_1, c_2 < \infty$:

$$|p'_{\Lambda_1}(u)| \leq c_1; \quad |p'_{\Lambda_{n+1}|\Lambda_n}(u|y)| \leq c_2, \quad u, y \in [C_-, C_+], \quad n \in \mathbf{N}. \quad (6)$$

Then for $n \in \mathbf{N}$, $k \in \{1, \dots, n-1\}$, satisfying

$$\mathbb{P} \left\{ \Lambda_{n-k} \in A^{n-k}, \dots, \Lambda_n \in A^n \right\} > 0, \quad A^{n-k}, \dots, A^n \in S_1^m,$$

the asymptotic expansion at $h \rightarrow 0$ ($m \rightarrow \infty$) holds:

$$\mathbb{P} \left\{ \Lambda_{n+1} \in A^{n+1} \mid \Lambda_n \in A^n, \dots, \Lambda_{n-k} \in A^{n-k} \right\} = \mathbb{P} \left\{ \Lambda_{n+1} \in A^{n+1} \mid \Lambda_n \in A^n \right\} + R_\Lambda(h), \quad (7)$$

where

$$A^{n-k}, \dots, A^n \in S_1^m, \quad R_\Lambda(h) = \begin{cases} \mathcal{O}(h^2), & \text{if } A^{n+1} \in S_1^m, \\ \mathcal{O}(h), & \text{if } A^{n+1} \in \{A_0, A_{m+1}\}. \end{cases}$$

Proof. Denote the left and right bounds of the interval A^j , $j \in \mathbf{N}$, by a_L^j , a_R^j , respectively. First, prove (7) for $k = 1$, $A^{n+1} \in S_1^m$:

$$\mathbb{P} \left\{ \Lambda_{n+1} \in A^{n+1} \mid \Lambda_n \in A^n, \Lambda_{n-1} \in A^{n-1} \right\} = \mathbb{P} \left\{ \Lambda_{n+1} \in A^{n+1} \mid \Lambda_n \in A^n \right\} + \mathcal{O}(h^2). \quad (8)$$

From the definition,

$$\mathbb{P} \left\{ \Lambda_{n+1} \in A^{n+1} \mid \Lambda_n \in A^n, \Lambda_{n-1} \in A^{n-1} \right\} = \frac{\mathbb{P} \left\{ \Lambda_{n+1} \in A^{n+1}, \Lambda_n \in A^n, \Lambda_{n-1} \in A^{n-1} \right\}}{\mathbb{P} \left\{ \Lambda_n \in A^n, \Lambda_{n-1} \in A^{n-1} \right\}}. \quad (9)$$

Transform the numerator of (9) using the Markov property of Λ_n :

$$\begin{aligned} & \mathbb{P} \left\{ \Lambda_{n+1} \in A^{n+1}, \Lambda_n \in A^n, \Lambda_{n-1} \in A^{n-1} \right\} = \\ & \mathbb{P} \left\{ a_L^{n-1} \leq \Lambda_{n-1} < a_R^{n-1}, a_L^n \leq \Lambda_n < a_R^n, a_L^{n+1} \leq \Lambda_{n+1} < a_R^{n+1} \right\} = \\ & \int_{a_L^{n-1}}^{a_R^{n-1}} p_{\Lambda_{n-1}}(u) \int_{a_L^n}^{a_R^n} p_{\Lambda_n|\Lambda_{n-1}}(y|u) \int_{a_L^{n+1}}^{a_R^{n+1}} p_{\Lambda_{n+1}|\Lambda_n}(z|y) dz dy du. \end{aligned} \quad (10)$$

From the property of a Markov sequence and the condition (6) it follows that $\exists c > 0$:

$$\forall n \in \mathbf{N}, \quad |p'_{\Lambda_n}(u)| \leq c < \infty, \quad u \in [C_-, C_+].$$

Using the rectangular approximation formula (see, e.g. Bahvalov, 1973), we get

$$\begin{aligned} & \mathbb{P} \left\{ \Lambda_{n+1} \in A^{n+1}, \Lambda_n \in A^n, \Lambda_{n-1} \in A^{n-1} \right\} = \\ & h^3 \cdot p_{\Lambda_{n-1}} \left(\frac{a_R^{n-1} + a_L^{n-1}}{2} \right) \cdot p_{\Lambda_n|\Lambda_{n-1}} \left(\frac{a_R^n + a_L^n}{2} \mid \frac{a_R^{n-1} + a_L^{n-1}}{2} \right) \times \\ & p_{\Lambda_{n+1}|\Lambda_n} \left(\frac{a_R^{n+1} + a_L^{n+1}}{2} \mid \frac{a_R^n + a_L^n}{2} \right) + \mathcal{O}(h^4). \end{aligned} \quad (11)$$

For the denominator of (9) we get analogously:

$$\begin{aligned} & \mathbb{P} \left\{ \Lambda_n \in A^n, \Lambda_{n-1} \in A^{n-1} \right\} = \\ & h^2 \cdot p_{\Lambda_{n-1}} \left(\frac{a_R^{n-1} + a_L^{n-1}}{2} \right) \cdot p_{\Lambda_n | \Lambda_{n-1}} \left(\frac{a_R^n + a_L^n}{2} \middle| \frac{a_R^{n-1} + a_L^{n-1}}{2} \right) + \mathcal{O}(h^2). \end{aligned} \quad (12)$$

Using (11), (12) in (9) and then performing asymptotic expansion, we find

$$\begin{aligned} & \mathbb{P} \left\{ \Lambda_{n+1} \in A^{n+1} \mid \Lambda_n \in A^n, \Lambda_{n-1} \in A^{n-1} \right\} = \\ & h \cdot p_{\Lambda_{n+1} | \Lambda_n} \left(\frac{a_R^{n+1} + a_L^{n+1}}{2} \middle| \frac{a_R^n + a_L^n}{2} \right) + \mathcal{O}(h^2). \end{aligned} \quad (13)$$

By similar transformations for the conditional probability from the right side of (8) we find

$$\mathbb{P} \left\{ \Lambda_{n+1} \in A^{n+1} \mid \Lambda_n \in A^n \right\} = h \cdot p_{\Lambda_{n+1} | \Lambda_n} \left(\frac{a_R^{n+1} + a_L^{n+1}}{2} \middle| \frac{a_R^n + a_L^n}{2} \right) + \mathcal{O}(h^2). \quad (14)$$

Comparing (13) and (14), we get (8).

For $A^{n+1} \in \{A_0, A_{m+1}\}$ the scheme of the proof is the same, except integration on z in (10), where the rectangular approximation formula is not applied. For this case we get:

$$\mathbb{P} \left\{ \Lambda_{n+1} \in A^{n+1} \mid \Lambda_n \in A^n, \Lambda_{n-1} \in A^{n-1} \right\} = \mathbb{P} \left\{ \Lambda_{n+1} \in A^{n+1} \mid \Lambda_n \in A^n \right\} + \mathcal{O}(h). \quad (15)$$

Summarizing (8) and (15), we get (7) for $k = 1$.

The scheme of the proof of the result (7) for $k = 2, 3, \dots$ is the same as for $k = 1$. ■

Note, that the main term in the asymptotic expansion (7) has the order $\mathcal{O}(h)$ for $A^{n+1} \in S_1^m$, and $\mathcal{O}(1)$ for $A^{n+1} \in \{A_0, A_{m+1}\}$.

Let $[x]$ mean the integer part of x (maximal integer number, which is not greater than x). For the random sequence Λ_n introduce the discrete random sequence Z_n^m , $n \in \mathbb{N}$, with the finite state space $V = \{0, 1, \dots, m+1\}$:

$$Z_n^m = \begin{cases} 0, & \text{if } Z_{n-1}^m = 0, \\ m+1, & \text{if } Z_{n-1}^m = m+1, \\ \left(\left[\frac{\Lambda_n - C_-}{h} \right] + 1 \right) \cdot \mathbf{1}_{(C_-, C_+)}(\Lambda_n) + (m+1) \cdot \mathbf{1}_{[C_+, +\infty)}(\Lambda_n), & \text{otherwise;} \end{cases} \quad (16)$$

$$Z_0^m = 0.$$

Would the remainder term $R_\Lambda(h)$ in (7) be equal to 0, the random sequence (16) should be a Markov chain. Theorem 1 states that the Markov property is approximately valid for the sequence (16), and gives the accuracy for this approximation.

Introduce the matrix of the size $(m+2) \times (m+2)$ of the conditional probabilities for the random sequence (16):

$$P^{(n)}(\theta) = (p_{ij}^n(\theta)) = \left(\mathbb{P} \left\{ Z_{n+1}^m = j \mid Z_n^m = i \right\} \right), \quad i, j \in V, \quad n \in \mathbb{N}.$$

Note, that the theory of calculation of the characteristics for Markov chains is much better developed than for Markov sequences, so let us consider an approximation of Z_n^m by the Markov chain $z_n^m \in V$, $n \in \mathbf{N}$, that has the initial probability distribution and the transition probabilities matrix $P^{(n)}(\theta)$ the same as for the sequence (16).

Because $\{0\}$, $\{m+1\}$ are absorbing states, it is convenient to renumerate the states: $V = \{\{0\}, \{m+1\}, \{1\}, \dots, \{m\}\}$. The matrix $P^{(n)}(\theta)$ after renumeration can be represented in the form:

$$P^{(n)}(\theta) = \left(\begin{array}{c|c} \mathbf{I}_2 & \mathbf{0}_{2 \times m} \\ \hline R^{(n)}(\theta) & Q^{(n)}(\theta) \end{array} \right), \quad \forall \theta \in \Theta, \quad (17)$$

where $R^{(n)}(\theta)$ and $Q^{(n)}(\theta)$ are blocks of the size $m \times 2$ and $m \times m$ respectively, \mathbf{I}_k is the identity matrix of the size k , $\mathbf{0}_{2 \times m}$ is the matrix of the size $2 \times m$ with all elements equal to 0. Introduce the notation: $\pi(\theta) = (\pi_i(\theta))$ is the vector of initial probabilities of the states $1, \dots, m$ for the sequence (16), $\pi_0(\theta)$ and $\pi_{m+1}(\theta)$ are the initial probabilities of the absorbing states 0 and $m+1$ respectively for (16); $\mathbf{1}_m$ is the vector of the size m with all components equal to 1; $\gamma_{\mathcal{H}_i}(\theta) = P\{d = i \mid \theta\}$, $i = 0, 1$, is the conditional probability to accept the hypothesis \mathcal{H}_i provided the parameter takes the value $\theta \in \Theta$; matrices $S(\theta)$ of the size $m \times m$ and $B(\theta)$ of the size $m \times 2$ are

$$S(\theta) = \mathbf{I}_m + \sum_{i=1}^{\infty} \prod_{j=1}^i Q^{(j)}(\theta), \quad B(\theta) = R^{(1)}(\theta) + \sum_{i=1}^{\infty} \prod_{j=1}^i Q^{(j)}(\theta) R^{(i+1)}(\theta);$$

$B_{(j)}(\theta)$ is the j -th column of the matrix $B(\theta)$, $j = 1, 2$; $t(\theta) = E\{N \mid \theta\}$, $t_i = E\{N \mid \theta \in \Theta_i\}$ ($i = 0, 1$) are the conditional sample sizes; $t = E\{N\}$ is the unconditional sample size.

Theorem 2 *Under the Theorem 1 conditions, if a sequential test from the family (3), (4) is used for the hypotheses (1), then $\forall \theta \in \Theta$ at $h \rightarrow 0$ the probabilistic characteristics of the test satisfy the following asymptotic expansions:*

$$t(\theta) = 1 + (\pi(\theta))' \cdot S(\theta) \cdot \mathbf{1}_m + \mathcal{O}(h),$$

$$\gamma_{\mathcal{H}_0}(\theta) = (\pi(\theta))' B_{(1)}(\theta) + \pi_0(\theta) + \mathcal{O}(h), \quad \gamma_{\mathcal{H}_1}(\theta) = (\pi(\theta))' B_{(2)}(\theta) + \pi_{m+1}(\theta) + \mathcal{O}(h).$$

Proof. The proof consists of two stages.

1. Calculation of the main terms in the asymptotic expansions for $E\{N \mid \theta\}$ and $\gamma_{\mathcal{H}_i}(\theta)$, $i = 0, 1$. These terms are calculated as the appropriate characteristics for the non-homogeneous Markov chain z_n^m , $n \in \mathbf{N}$ (see, e.g. Kemeni and Snell, 1959) with the state space V (two of the states are absorbing: $\{0\}$ and $\{m+1\}$), initial probabilities vector of transient states $\pi(\theta)$, initial probabilities of absorbing states $\pi_0(\theta)$, $\pi_{m+1}(\theta)$, and the transition probabilities matrix $P^{(n)}(\theta)$ presented in the form (17).

2. Asymptotic analysis of the remainders at $h \rightarrow 0$. The result of Theorem 1 is applied to analyze the differences between the characteristics of the random sequence (16) and the Markov chain z_n^m ; the relation $h = (C_+ - C_-)/m$ is used, where m is the size of the square matrix $Q^{(n)}(\theta)$ and the number of rows in the matrix $R^{(n)}(\theta)$. ■

Corollary 1 Under the conditions of Theorem 2, the error type I and II probabilities satisfy the asymptotic expansions:

$$\begin{aligned}\alpha &= \int_{\Theta} \left((\pi(\theta))' B_{(2)}(\theta) + \pi_{m+1}(\theta) \right) w_0(\theta) d\theta + \mathcal{O}(h); \\ \beta &= \int_{\Theta} \left((\pi(\theta))' B_{(1)}(\theta) + \pi_0(\theta) \right) w_1(\theta) d\theta + \mathcal{O}(h).\end{aligned}\quad (18)$$

Corollary 2 Under the conditions of Theorem 2, the following asymptotic expansions for conditional and unconditional mathematical expectations of the sample size hold:

$$\begin{aligned}t_i &= 1 + \int_{\Theta} (\pi(\theta))' \cdot S(\theta) \cdot \mathbf{1}_m \cdot w_i(\theta) d\theta + \mathcal{O}(h), \quad i = 0, 1; \\ t &= 1 + \int_{\Theta} (\pi(\theta))' \cdot S(\theta) \cdot \mathbf{1}_m \cdot p(\theta) d\theta + \mathcal{O}(h).\end{aligned}\quad (19)$$

4 Case of Independent Observations with the Gaussian Probability Distribution

Consider the model of data frequently used in practice, where the observations x_1, x_2, \dots are independent identically distributed random variables with the Gaussian probability distribution having the pdf

$$p(x | \theta) = n_1(x; \theta, \sigma_x^2) = (2\pi\sigma_x^2)^{-\frac{1}{2}} e^{-\frac{1}{2\sigma_x^2}(x-\theta)^2}, \quad x \in \mathbf{R}. \quad (20)$$

Suppose the random parameter θ also has the Gaussian probability distribution with pdf

$$p(\theta) = n_1(\theta; \mu, \sigma_\theta^2), \quad \theta \in \Theta = \mathbf{R}. \quad (21)$$

The values $\sigma_x > 0$, $\sigma_\theta > 0$, $\mu \in \mathbf{R}$ are known. The hypotheses $\mathcal{H}_0, \mathcal{H}_1$ correspond to the sets

$$\Theta_0 = (-\infty, \bar{\theta}), \quad \Theta_1 = [\bar{\theta}, +\infty). \quad (22)$$

Introduce the notation: $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{y^2}{2}} dy$ is the standard Gaussian probability distribution function;

$$B = \log \frac{W_0}{W_1}, \quad D_- = C_- - B, \quad \gamma^2 = \frac{\sigma_x^2}{\sigma_\theta^2};$$

$$l_n(y) = \sqrt{n} \frac{y - \bar{\theta} + \gamma^2(\mu - \bar{\theta})/n}{\sigma_x \sqrt{1 + \frac{\gamma^2}{n}}}, \quad \Psi_n(y) = \log \frac{\Phi(l_n(y))}{\Phi(-l_n(y))}, \quad y \in \mathbf{R};$$

$$\Xi(\theta) = \left(\Phi \left(\frac{\sqrt{n} \Psi_n^{-1}(ih + D_-) - \theta}{\sigma_x} \right) - \Phi \left(\frac{\sqrt{n} \Psi_n^{-1}((i-1)h + D_-) - \theta}{\sigma_x} \right) \right)^{-1};$$

$$\phi_{i,j}^{(n)}(\theta) = \Xi(\theta) \int_{\Psi_n^{-1}((i-1)h + D_-)}^{\Psi_n^{-1}(ih + D_-)} n_1 \left(y; \theta, \frac{\sigma_x^2}{n} \right) \Phi \left(\frac{(n+1) \Psi_{n+1}^{-1}(jh + D_-) - ny - \theta}{\sigma_x} \right) dy,$$

$$\phi_{i,m+1}^{(n)}(\theta) = 1, \quad \phi_{i,-1}^{(n)}(\theta) = 0, \quad \theta \in \Theta, \quad i, j = 1, \dots, m.$$

Theorem 3 *If the model (20) – (22) is valid, then $\forall \theta \in \mathbf{R}$ the conditional probabilities for the random sequence (16) in the blocks $Q^{(n)}(\theta)$, $R^{(n)}(\theta)$ of the matrix (17) are*

$$p_{ij}^{(n)}(\theta) = \phi_{i,j}^{(n)}(\theta) - \phi_{i,j-1}^{(n)}(\theta), \quad i = 1, \dots, m, \quad j = 0, 1, \dots, m+1; \quad (23)$$

the initial probabilities for the sequence (16) are given by the expressions:

$$\begin{aligned} \pi_i(\theta) &= \Phi\left(\frac{\Psi_1^{-1}(D_- + ih) - \theta}{\sigma_x}\right) - \Phi\left(\frac{\Psi_1^{-1}(D_- + (i-1)h) - \theta}{\sigma_x}\right), \quad i = 1, \dots, m, \\ \pi_0(\theta) &= \Phi\left(\frac{\Psi_1^{-1}(D_-) - \theta}{\sigma_x}\right), \quad \pi_{m+1}(\theta) = 1 - \Phi\left(\frac{\Psi_1^{-1}(D_- + mh) - \theta}{\sigma_x}\right). \end{aligned} \quad (24)$$

Proof. For the model (20) – (22) we get the following representation of the generalized likelihood ratio statistic (2) by equivalent transformations:

$$\Lambda_n = B + \Psi_n(\bar{x}^{(n)}), \quad \bar{x}^{(n)} = \frac{1}{n} \sum_{i=1}^n x_i. \quad (25)$$

Note, that $\Psi_n\{\cdot\}$ is a strictly increasing function. Representing the random events related to the random sequence Λ_n as the random events related to $\bar{x}^{(n)}$, $n \in \mathbf{N}$, and using the properties of the function $\Phi(\cdot)$, we come to (23), (24). ■

Let us note that the conditions of Theorem 2 are satisfied for this model, because the statistic $\bar{x}^{(n)}$ is a Markov sequence, and from (25) it follows that Λ_n is also a Markov sequence.

5 Results of Numerical Experiments

A series of numerical experiments was performed to illustrate the results presented by Corollaries 1, 2 and Theorem 3. The probability model (20) – (22) of observations was considered, and the hypotheses (22) were tested for the following values of parameters:

$$\bar{\theta} = 0, \quad \mu \in \{-0.5, -0.25\}, \quad \sigma_x = 1, \quad \sigma_\theta = 0.5.$$

The Monte-Carlo estimates were computed for the type I error probability α and for the conditional mathematical expectation of the sample size t_0 , when the test (3), (4) is used (these estimates are denoted by $\hat{\alpha}$ and \hat{t}_0 respectively); the number of replications for both values of μ was equal to 10000. The parameters C_- , C_+ of the test (3) were calculated according to (5), where $\alpha_0 = \beta_0 = 0.1$. The results are presented in Table 1.

For comparison, the approximations of α and t_0 were computed as the main terms in the asymptotic expansions (18), (19) using (23), (24) for $m \in \{2, 4, 6, 8, 10\}$; these approximations are denoted by α_m and $(t_0)_m$ respectively. The results are also given in Table 1.

It can be seen from Table 1 that even with small values of $m \geq 8$ the results of approximation are very close to the Monte-Carlo estimates, and could be used in practice for calculation of the characteristics of the sequential test (3), (4). Similar results were also obtained for the error type II probability β and the conditional mathematical expectation of the sample size t_1 .

Table 1: Results of numerical experiments

| Approximations | m | | | | | Monte-Carlo Estimates $\hat{\alpha}, \hat{t}_0$ |
|-----------------------|------|------|------|------|------|---|
| | 2 | 4 | 6 | 8 | 10 | |
| Case of $\mu = -0.5$ | | | | | | |
| α_m | 0.13 | 0.10 | 0.09 | 0.07 | 0.07 | 0.07 |
| $(t_o)_m$ | 11.6 | 12.3 | 13.9 | 15.5 | 15.3 | 14.8 |
| Case of $\mu = -0.25$ | | | | | | |
| α_m | 0.06 | 0.03 | 0.05 | 0.09 | 0.09 | 0.09 |
| $(t_o)_m$ | 12.8 | 15.3 | 19.9 | 17.0 | 19.2 | 18.8 |

6 Robustness Evaluation under “Contamination” of Hypothetical Probability Distributions of Observations

In practice the observed data usually do not follow the hypothetical model exactly, the hypothetical model is distorted (see Hampel, Ronchetti, Rousseeuw, and Stahel, 1986). The model of “contamination” of the hypothetical probability distributions (see Huber, 1981) is often used to analyze robustness of the statistical procedures.

Suppose the hypothetical model considered in Section 2 is distorted: the data observed are really obtained from a “contaminated” probability distribution with a pdf

$$\bar{p}_n(x_1, \dots, x_n | \theta) = (1 - \varepsilon)p_n(x_1, \dots, x_n | \theta) + \varepsilon\tilde{p}_n(x_1, \dots, x_n | \theta), \quad x_i \in \mathbf{R}, \theta \in \Theta, \quad (26)$$

where $\tilde{p}(x_1, \dots, x_n | \theta)$ is a “contaminating” pdf, $\varepsilon \in [0, 1/2)$ is a probability of “contamination” presence (“contamination” level).

Introduce the notation: $\tilde{\pi}(\theta), \tilde{\pi}_0(\theta), \tilde{\pi}_{m+1}(\theta), \tilde{Q}^{(n)}(\theta), \tilde{R}^{(n)}(\theta)$ are the elements calculated analogously to $\pi(\theta), \pi_0(\theta), \pi_{m+1}(\theta), Q^{(n)}(\theta), R^{(n)}(\theta)$ replacing the hypothetical pdf $p_n(x_1, \dots, x_n)$ with the “contaminating” pdf $\tilde{p}_n(x_1, \dots, x_n)$ in the probability distribution of the random sequence (2); $\Delta\pi_0(\theta) = \tilde{\pi}_0(\theta) - \pi_0(\theta)$, $\Delta\pi_1(\theta) = \tilde{\pi}_{m+1}(\theta) - \pi_{m+1}(\theta)$; $\bar{t}(\theta)$ and $\bar{\gamma}_{\mathcal{H}_i}(\theta)$ ($i = 0, 1$) are the conditional mathematical expectation of the sample size and the conditional probability of acceptance of the hypothesis \mathcal{H}_i respectively, provided the parameter value is θ , for the distorted model (26).

Theorem 4 For the distorted model (26), if the conditions of Theorem 2 are valid, then $\forall \theta \in \Theta$, the following asymptotic expansions hold at $\varepsilon \rightarrow 0, h \rightarrow 0$:

$$\begin{aligned} \bar{t}(\theta) - t(\theta) &= \varepsilon \left((\tilde{\pi}(\theta) - \pi(\theta))' S(\theta) + (\pi(\theta))' \sum_{l=1}^{\infty} \sum_{j=1}^l \prod_{k=1}^{j-1} Q^{(k)}(\theta) \times \right. \\ &\quad \left. (\tilde{Q}^{(j)}(\theta) - Q^{(j)}(\theta)) \prod_{k=j+1}^l Q^{(k)}(\theta) \right) \mathbf{1}_m + \mathcal{O}(\varepsilon^2) + \mathcal{O}(h); \\ \bar{\gamma}_{\mathcal{H}_i}(\theta) - \gamma_{\mathcal{H}_i}(\theta) &= \varepsilon \left(\Delta\pi_i(\theta) + (\tilde{\pi}(\theta) - \pi(\theta))' B_{(i+1)}(\theta) + \tilde{R}^{(1)}(\theta) - R^{(1)}(\theta) + \right. \\ &\quad \left. \sum_{l=1}^{\infty} \left(\sum_{j=1}^l \prod_{k=1}^{j-1} Q^{(k)}(\theta) (\tilde{Q}^{(j)}(\theta) - Q^{(j)}(\theta)) \prod_{k=j+1}^l Q^{(k)}(\theta) R^{(l+1)}(\theta) + \right. \right. \end{aligned}$$

$$\prod_{j=1}^l Q^{(j)}(\theta) (\tilde{R}^{(l+1)}(\theta) - R^{(l+1)}(\theta)) \Big) + \mathcal{O}(\varepsilon^2) + \mathcal{O}(h), \quad i = 0, 1.$$

Proof. The proof consists of two stages.

1. Calculation of the characteristics for the Markov chain which has the state space V , the initial probabilities vector $(\bar{\pi}_0(\theta), \bar{\pi}(\theta), \bar{\pi}_{m+1}(\theta))'$ and the transition probabilities matrix

$$\bar{P}^{(n)}(\theta) = \begin{pmatrix} \mathbf{I}_2 & | & \mathbf{0}_{2 \times m} \\ \hline \bar{R}^{(n)}(\theta) & | & \bar{Q}^{(n)}(\theta) \end{pmatrix},$$

where

$$\bar{\pi}_i(\theta) = (1 - \varepsilon)\pi_i(\theta) + \varepsilon\tilde{\pi}_i(\theta), \quad i \in V,$$

$$\bar{R}^{(n)}(\theta) = (1 - \varepsilon)R^{(n)}(\theta) + \varepsilon\tilde{R}^{(n)}(\theta), \quad \bar{Q}^{(n)}(\theta) = (1 - \varepsilon)Q^{(n)}(\theta) + \varepsilon\tilde{Q}^{(n)}(\theta).$$

2. Asymptotic analysis at $\varepsilon \rightarrow 0$, $h \rightarrow 0$ of the correspondent differences between the calculated characteristics under “contamination” and their hypothetical values, using the results of Theorem 2. ■

Note, that it follows from Theorem 4, that the indicated differences of the characteristics $\bar{t}(\theta)$, $\bar{\gamma}_{\mathcal{H}_i}(\theta)$, $i = 0, 1$, from the hypothetical values $t(\theta)$, $\gamma_{\mathcal{H}_i}(\theta)$, for the sequential test (3), (4) under the “contamination” (26), have the first order by ε .

Using the result of Theorem 4 one can approximate the characteristics of the sequential test (3), (4) under the distortion (26), and use these approximation in construction of the robust sequential test by the minimax criterion (see A. Kharin, 2002).

7 Conclusion

In this paper the approach to calculate with a given accuracy the characteristics of sequential tests is proposed for composite hypotheses. This approach is used for quantitative robustness analysis of sequential tests under “contamination” of the probability distribution of observations. The results can be used for minimax robust sequential test construction.

Acknowledgements

This research was supported by the Belarusian National Science Foundation, grant No. F06M-072.

The author appreciates the help of the student of the Belarusian State University Sergey Chernov in performing numerical experiments.

References

- Aivazian, S. A. (1959). Comparison of optimal properties of Neyman–Pearson and Wald tests. *Probability Theory and its Applications*, 4, 86-93.
- Bahvalov, N. S. (1973). *Numerical Methods (in Russian)*. Moscow: Nauka.

- Ghosh, B. K., and Sen, P. K. (1991). *Handbook of Sequential Analysis*. New York, Basel, Hong Kong: Marcel Dekker.
- Hampel, F., Ronchetti, E., Rousseeuw, P., and Stahel, W. (1986). *Robust Statistics. The Approach Based on Influence Functions*. New York: John Wiley and Sons.
- Huber, P. (1981). *Robust Statistics*. New York: John Wiley and Sons.
- Jennison, C., and Turnbull, B. W. (2000). *Group Sequential Methods with Applications to Clinical Trials*. Boca Raton: Chapman & Hall / CRC.
- Kemeni, J. G., and Snell, J. L. (1959). *Finite Markov Chains*. New York: John Wiley and Sons.
- Kharin, A. (2002). On robustifying of the sequential probability ratio test for a discrete model under “contaminations”. *Austrian Journal of Statistics*, 31, 267-277.
- Kharin, A., and Kishylau, D. (2005). Robust sequential testing of hypotheses on discrete probability distributions. *Austrian Journal of Statistics*, 34, 153-162.
- Kharin, A. Y., and Kishylau, D. V. (2005). Performance and robustness analysis for sequential testing of hypotheses on parameters of Markov chains (in Russian). *Proceedings of the National Academy of Sciences of Belarus*, 4, 30-35.
- Lai, T. L. (2001). Sequential analysis: Some classical problems and new challenges. *Statistica Sinica*, 11, 303-408.
- Quang, P. X. (1985). Robust sequential testing. *Annals of Statistics*, 13, 638-649.
- Wald, A. (1947). *Sequential Analysis*. New York: John Wiley and Sons.

Author's Address:

Alexey Kharin

Department of Probability Theory and Mathematical Statistics

Faculty of Applied Mathematics and Informatics

Belarusian State University

Independence avenue, 4

Minsk 220030

Belarus

e-mail: KharinAY@bsu.by