ROBUSTNESS OF SEQUENTIAL TESTING OF PARAMETRIC HYPOTHESES

A. Kharin
Belarusian State University
Minsk, BELARUS
e-mail: KharinAY@bsu.by

Abstract

The problems of robustness of parametric hypotheses testing are considered for the cases of simple and composite hypotheses. Conditional error probabilities and expected sample sizes of sequential tests are evaluated. Asymptotic expansions for these characteristics are obtained under the distortions of Tukey–Huber type. Robust sequential tests are constructed with the minimax criterion.

1 Introduction

The sequential approach [1] is intensively used in computer data analysis for hypotheses testing. But the characteristics of sequential tests are problematic to calculate with a given accuracy even for basic hypothetical models [2]. The sequential procedures are applied for the real data sets that are usually not adequate to certain hypothetical model. This hypothetical model is often distorted. Hence the robustness analysis and robust sequential test construction are important problems.

2 The Case of Simple Hypotheses

Let on a measurable space $(\Omega, \mathcal{F})$ discrete random variables $x_1,\ x_2, \ldots$ be defined, $\forall t \in \mathbf{N}$, $x_t \in U = \{u_1, u_2, \ldots, u_M\}$, $M < \infty$, $u_1 < u_2 < \cdots < u_M$. Let these random variables be independent identically distributed, from the discrete probability distribution with a parameter $\theta \in \Theta = \{\theta_0, \theta_1\}$:

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

$a \in \mathbf{N} \setminus \{1\}$; $J(u;\theta)$: $U \times \Theta \longrightarrow \mathbf{N}_0$ is a function satisfying $\sum_{u \in U} a^{-J(u;\theta)} = 1$.

Consider two simple hypotheses w.r.t. the parameter θ:

$$H_0 :\ \theta = \theta_0,\ H_1 :\ \theta = \theta_1. \tag{2}$$

Such a problem appears in applications, where one of the two possible regimes can be realized.

Introduce the notation:

$$\Lambda_n = \Lambda_n(x_1, \ldots, x_n) = \sum\nolimits_{t=1}^{n} \lambda_t;$$

$$\lambda_t = \log_a \left(P(x_t;\theta_1)/P(x_t;\theta_0)\right) = J(x_t;\theta_0) - J(x_t;\theta_1) \in \mathbf{Z}.$$

To test these hypotheses by n $(n = 1, 2\dots)$ observations let us consider the sequential probability ratio test (SPRT) [1]:

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

where $\mathbf{1}_D(\cdot)$ is the indicator function of the set D. The decisions $d_n = 0$ and $d_n = 1$ mean stopping of the observation process and the acceptance of the appropriate hypothesis. The decision $d_n = 2$ means that it is necessary to make the $(n+1)$-th observation. In (3) the thresholds $C_- < C_+$ are the given values (parameters of the test). According to [1], we use

$$C_+ = [\log_a((1-\beta_0)/\alpha_0)], \ C_- = [\log_a(\beta_0/(1-\alpha_0))], \tag{4}$$

where α_0, β_0 are given maximal possible values of the probabilities of type I and type II errors respectively. In fact, the true values α, β for the probabilities of type I and type II errors differ from α_0, β_0.

For $n \in \mathbf{N}$, define the random sequence

$$\xi_n = C_+\mathbf{1}_{[C_+,+\infty)}(\Lambda_n) + C_-\mathbf{1}_{(-\infty,C_-]}(\Lambda_n) + \Lambda_n\mathbf{1}_{(C_-,C_+)}(\Lambda_n) \in \mathbf{Z}. \tag{5}$$

Introduce the notation: $\mathbf{I}_k$ 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.

Define the one-step transition probabilities matrix and the initial probabilities vector for the transient states

$$P^{(k)} = (p_{ij}^k) = \left(\begin{array}{c|c} \mathbf{I}_2 & \mathbf{0}_{2\times(N-2)} \\ --- & ---- \\ R^{(k)} & Q^{(k)} \end{array}\right), \ \pi^{(k)} = \begin{pmatrix} \pi^{(k)}_{C_-+1} \\ \vdots \\ \pi^{(k)}_{C_+-1} \end{pmatrix}, \tag{6}$$

where the blocks $R^{(k)}$, $Q^{(k)}$, and the vector $\pi^{(k)}$ are

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

$$\pi_i^{(k)} = \sum_{u\in U} \delta_{J(u;\theta_0)-J(u;\theta_1),i} P(u;\theta_k), \ i \in (C_-,C_+). \tag{7}$$

For the hypothesis H_k, let $t^{(k)}$ be the expected stopping time of the decision process (expected number of observations), and $B^{(k)}$ be the $((N-2)\times 2)$-matrix of absorption probabilities: its (i,j)-th element equals to the probability of absorption at the state j (acceptance of the hypothesis H_j) starting from the state $\xi_1 = i \in (C_-,C_+)$. Let us denote the i-th column of a matrix W by $W_{(i)}$.

Theorem 1. *If under conditions (1) — (2) the true hypothesis is the hypothesis H_k, and the matrix $S^{(k)} = \mathbf{I}_{N-2} - Q^{(k)}$ is nonsingular, then for the test (3)*

$$t^{(k)} = (\pi^{(k)})'(S^{(k)})^{-1}\mathbf{1}_{N-2} + 1, \ B^{(k)} = (S^{(k)})^{-1}R^{(k)}. \tag{8}$$

Corollary 1. *Under the Theorem 1 conditions the error probabilities of type I and type II are* $\alpha = (\pi^{(0)})' B^{(0)}_{(2)}$, $\beta = (\pi^{(1)})' B^{(1)}_{(1)}$.

Let the hypothetical model (1), (2) be under the distortions of Tukey–Huber type [3]: instead of (1) the obsrvations $x_1, x_2, \ldots$ are taken from the contaminated discrete probability distribution

$$\bar{P}(u;\theta) = \bar{P}_\theta\{x_t = u\} = (1-\varepsilon)P(u;\theta) + \varepsilon \tilde{P}(u;\theta),\ \tilde{P}(u;\theta) = a^{-\tilde{J}(u;\theta)}, \tag{9}$$

$\tilde{J}(u;\theta)$: $U \times \Theta \longrightarrow \mathbf{N}_0$ is a function different from $J(\cdot)$, satisfying $\sum_{u \in U} a^{-\tilde{J}(u;\theta)} = 1$. Define the matrix $\tilde{P}^{(k)}$ analogous to (6), substituting $P(\cdot)$ with $\tilde{P}(\cdot)$.

Theorem 2. *If the hypothetical model (1), (2) is distorted according to (9), and the matrices* $S^{(k)}$, $\tilde{S}^{(k)} = \mathbf{I}_{N-2} - Q^{(k)} - \varepsilon(\tilde{Q}^{(k)} - Q^{(k)})$ *are nonsingular, then the expected number of observations* $\bar{t}^{(k)}$ *and the absorbtion probabilities matrix* $\bar{B}^{(k)}$ *for the distorted model differ from the same characteristics for the hypothetical model by the values of the order* $\mathcal{O}(\varepsilon)$*:*

$$\bar{t}^{(k)} - t^{(k)} = \varepsilon\left((\tilde{\pi}^{(k)} - \pi^{(k)})' + \pi^{(k)}(S^{(k)})^{-1}(\tilde{Q}^{(k)} - Q^{(k)})\right) \times$$
$$\times (S^{(k)})^{-1}(1\ldots1)' + \mathcal{O}(\varepsilon^2);\ \bar{B}^{(k)} - B^{(k)} = \varepsilon(S^{(k)})^{-1} \times$$
$$\times \left((\tilde{Q}^{(k)} - Q^{(k)})(S^{(k)})^{-1}R^{(k)} + \tilde{R}^{(k)} - R^{(k)}\right) + \mathcal{O}(\varepsilon^2),$$

where $\tilde{Q}^{(k)}$, $\tilde{R}^{(k)}$ *are the blocks of the matrix* $\tilde{P}^{(k)}$.

Corollary 2. *Under the Theorem 2 conditions the error probabilities* $\bar{\alpha}$, $\bar{\beta}$ *of types I and II for the distorted model differ from the same characteristics for the hypothetical model by the values of the order* $\mathcal{O}(\varepsilon)$*:*

$$\bar{\alpha} - \alpha = \varepsilon\Bigg((\pi^{(0)})'\big((S^{(0)})^{-1}\big((\tilde{Q}^{(0)} - Q^{(0)})(S^{(0)})^{-1}R^{(0)} +$$
$$+ \tilde{R}^{(0)} - R^{(0)}\big)\big)_{(2)} + (\tilde{\pi}^{(0)} - \pi^{(0)})B^{(0)}_{(2)}\Bigg) + \mathcal{O}(\varepsilon^2),$$
$$\bar{\beta} - \beta = \varepsilon\Bigg((\pi^{(1)})'\big((S^{(1)})^{-1}\big((\tilde{Q}^{(1)} - Q^{(1)})(S^{(1)})^{-1}R^{(1)} +$$
$$+ \tilde{R}^{(1)} - R^{(1)}\big)\big)_{(1)} + (\tilde{\pi}^{(1)} - \pi^{(1)})B^{(1)}_{(1)}\Bigg) + \mathcal{O}(\varepsilon^2).$$

Using the theory presented above the minimax robust sequential test is constructed in [4]. The generalizations of the results for the case of arbitrary dicrete distributions and for dependent observations are discussed in [5], [6].

3 Testing of Composite Hypotheses

Suppose a sequence $x_1, x_2, \ldots$ of i.i.d. random variables is observed from a continuous ditribution with the p.d.f. $p(x \mid \theta)$, where $\theta \in \Theta \subseteq \mathbf{R}$ is an unknown value of random parameter. Consider two composite hypotheses

$$H_0:\ \theta \in \Theta_0,\ H_1:\ \theta \in \Theta_1; \tag{10}$$

$\Theta_0, \Theta_1 \in \Theta$, $\Theta_0 \cap \Theta_1 = \emptyset$. Assume that the prior p.d.f. $p(\theta)$ is known.

One of the possible techniques to test the hypotheses (10) is using of weight functions proposed by Wald [1]. Introduce the notation:

$$W_i = \int_{\Theta_i} p(\theta)d\theta,\ w_i(\theta) = \frac{1}{W_i} \cdot p(\theta) \cdot \mathbf{1}_{\Theta_i}(\theta),\ \theta \in \Theta,\ i = 0, 1; \tag{11}$$

$$\Lambda_n = \Lambda_n(x_1, \ldots, x_n) = \ln \frac{\int_\Theta w_1(\theta) \prod_{i=1}^n p(x_i \mid \theta)d\theta}{\int_\Theta w_0(\theta) \prod_{i=1}^n p(x_i \mid \theta)d\theta}. \tag{12}$$

For testing the hypotheses (10), under the notation (11), (12) the following parametric family of tests is used:

$$N = \min\{n \in \mathbf{N}:\ \Lambda_n \notin (C_-, C_+)\}, \tag{13}$$

$$d = \mathbf{1}_{[C_+, +\infty)}(\Lambda_N), \tag{14}$$

where (13) gives the stopping rule, N is the random number of the observation, at which the decision d is made according to (14); $d = i$ means that the hypothesis H_i, $i = 0, 1$, is accepted; $C_- < 0$, $C_+ > 0$ are parameters of the test, which are usually choosen in practice according to (4).

Introduce the discretization parameters

$$m \in \mathbf{N},\ h = \frac{C_+ - C_-}{m},\ C_- = C_0 < C_1 < C_2 < \cdots < C_{m-1} < C_m = C_+,\ C_i = C_- + i \cdot h;$$

$$A_i = [C_{i-1}, C_i),\ i = 1, \ldots, m,\ A_- = (-\infty, C_-),\ A_+ = [C_+, +\infty). \tag{15}$$

Lemma. *Let for some $n \in \mathbf{N}$ the parameters of discretization (15) and the p.d.f. $p(x \mid \theta)$ satisfy the condition:*

$$P\{\Lambda_1 \in A^1, \ldots, \Lambda_n \in A^n\} > 0,\ A^j \in \{A_1, \ldots, A_m\},\ j = 1, \ldots, n,$$

and the statistic (12) can be presented in the form: $\Lambda_n = \Psi_n(\bar{x}^{(n)})$, $\Psi_n(\cdot)$: $\mathbf{R} \to \mathbf{R}$ is a (strictly) increasing function. Then $\forall k \in \{1, \ldots, n-1\}$:

$$P\{\Lambda_{n+1} \in A^{n+1} \mid \Lambda_n \in A^n, \ldots, \Lambda_{n-k} \in A^{n-k}\} = P\{\Lambda_{n+1} \in A^{n+1} \mid \Lambda_n \in A^n\} + R_\Lambda(h), \tag{16}$$

where $A^j \in \{A_1, \ldots, A_m\}$, $j = 1, \ldots, n$, $R_\Lambda(h) = \begin{cases} \mathcal{O}(h^2), & A^{n+1} \in \{A_1, \ldots, A_m\}, \\ \mathcal{O}(h), & A^{n+1} \in \{A_-, A_+\}. \end{cases}$

Proof is based on integral presentation of the left and right parts of (16) and on the integral theorem about a mean value. ■

The Lemma states that the Markov property holds for the random sequence Λ_n approximately, and gives the accuracy of the approximation.

Introduce the random sequences: $\bar{Z}_n^m = [\frac{\Psi_n(\bar{x}^{(n)}) - C_-}{h}]$, $n \in \mathbf{N}$;

$$Z_n^m = \bar{Z}_n^m \cdot \mathbf{1}_{(0, m+1)}(\bar{Z}_n^m) + (m+1) \cdot \mathbf{1}_{[m+1, +\infty)}(\bar{Z}_n^m),\ n \in \mathbf{N}. \tag{17}$$

Represent the matrix of conditional state-to-state probabilities of the sequence Z_n^m at the step n in the form:

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

where $R^{(n)}(\theta)$ and $Q^{(n)}(\theta)$ are matrices of the sizes $m \times 2$ and $m \times m$ respectively. Denote by $\pi(\theta) = (\pi_i(\theta))$, $i = 1, \ldots, m$, the vector of initial probabilities of the states $1, \ldots, m$ of the sequence Z_n^m. Define the matrices:

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

Denote by $\gamma_{H_i}(\theta)$ the conditional probability of acceptance of the hypothesis H_i , provided the parameter takes the value $\theta \in \Theta$.

Theorem 3. *Under the Lemma conditions, $\forall \theta \in \Theta$ the following asymptotic expansions hold at $h \to 0$:*

$$E\{N \mid \theta\} = 1 + (\pi(\theta))' \cdot S(\theta) \cdot \mathbf{1}_m + \mathcal{O}(h),$$

$$\gamma_{H_i}(\theta) = (\pi(\theta))' B_{(i+1)}(\theta) + \mathcal{O}(h), \; i = 0, 1.$$

Proof. The proof uses the theory of absorbing Markov chains [7], the results of Lemma and the relation (15) between m and h. ∎

Corollary 3. *Under the Theorem 3 conditions, the error probabilities of type I and II respectively satisfy the asymptotic expansions:*

$$\alpha = \frac{1}{W_0} \cdot \int_{\Theta_0} (\pi(\theta))' B_{(2)}(\theta) p(\theta) d\theta + \mathcal{O}(h);$$

$$\beta = \frac{1}{W_1} \cdot \int_{\Theta_1} (\pi(\theta))' B_{(1)}(\theta) p(\theta) d\theta + \mathcal{O}(h).$$

Corollary 4. *Under the Theorem 3 conditions, the following asymptotic expansions hold for the mathematical expectations of the sample size:*

$$E\{N \mid \theta \in \Theta_i\} = 1 + \frac{1}{W_i} \cdot \int_{\Theta_i} (\pi(\theta))' \cdot S(\theta) \cdot \mathbf{1}_m \cdot p(\theta) d\theta + \mathcal{O}(h);$$

$$E\{N\} = 1 + \int_{\Theta} (\pi(\theta))' \cdot S(\theta) \cdot \mathbf{1}_m \cdot p(\theta) d\theta + \mathcal{O}(h).$$

Consider now for example the case of Gaussian probability distributions, where

$$p(x \mid \theta) = n_1(x; \theta, \sigma_x^2) = (2\pi\sigma_x^2)^{-\frac{1}{2}} e^{-\frac{1}{2\sigma_x^2}(x-\theta)^2}, \; x \in \mathbf{R}, \; p(\theta) = n_1(\theta; \mu, \sigma_\theta^2), \; \theta \in \mathbf{R}; \quad (18)$$

$\sigma_x > 0$, $\sigma_\theta > 0$, $\mu \in \mathbf{R}$ are known values;

$$\Theta_0 = (-\infty, \bar{\theta}),\ \Theta_1 = [\bar{\theta}, +\infty). \tag{19}$$

Introduce the notation: $D_\pm = C_\pm - \ln \frac{W_0}{W_1}$; $\bar{x}^{(n)} = \frac{1}{n}\sum_{i=1}^{n} x_i$; $\gamma^2 = \frac{\sigma_x^2}{\sigma_\theta^2}$;

$$l_n(\bar{x}^{(n)}) = \sqrt{n} \cdot \frac{\bar{x}^{(n)} - \bar{\theta} + \frac{\gamma^2}{n}(\mu - \bar{\theta})}{\sigma_x \cdot \sqrt{1 + \frac{\gamma^2}{n}}},\ \Psi_n(\bar{x}^{(n)}) = \ln \frac{\Phi(l_n(\bar{x}^{(n)}))}{\Phi(-l_n(\bar{x}^{(n)}))}, \tag{20}$$

where $\Phi(\cdot)$ is a distribution function of the standard normal law.

Theorem 4. *For the model (18), (19), in the notation (20), the state-to-state probabilities for transient states of the sequence Z_n^m at the step n equal to*

$$p_{ij}^{(n)}(\theta) = \frac{\int_{\Psi_n^{-1}((i-1)h+D_-)}^{\Psi_n^{-1}(ih+D_-)} n_1(y;\theta,\frac{\sigma_x^2}{n}) \int_{(n+1)\Psi_{n+1}^{-1}((j-1)h+D_-)-n\bar{x}^{(n)}}^{(n+1)\Psi_{n+1}^{-1}(jh+D_-)-n\bar{x}^{(n)}} n_1(z;\theta,\sigma_x^2) dz dy}{\int_{\Psi_n^{-1}((i-1)h+D_-)}^{\Psi_n^{-1}(ih+D_-)} n_1(y;\theta,\frac{\sigma_x^2}{n}) dy},$$

$$i, j = 1, \ldots, m.$$

Proof consists of direct calculations of the indicated probabilities using properties of the normal distribution. ■

Theorem 4 gives the expression for the matrix $Q^{(n)}(\theta)$ in the explicit form. The matrix $R^{(n)}(\theta)$ and the vector $\pi(\theta)$ are calculated in the analogous way.

Using Theorem 3 and Corollaries 3, 4, the robustness analysis under distortions of Tukey–Huber type can be performed following the scheme as in the simple hypotheses case.

The research is supported by the Belarusian Science Foundation, project F06M-072.

References

[1] Wald A. (1947). *Sequential Analysis*. Wiley, New York.

[2] Siegmund D. (1975). Error Probabilities and Average Sample Number of the Sequentional Probability Ratio Test. *J. Roy. Statist. Soc. Ser. B.* Vol. **37**, pp. 394-401.

[3] Huber, P. (1981). *Robust Statistics*. Wiley, New York.

[4] Kharin A. (2002). On Robustifying of the SPRT for a Discrete Model under "Contaminations". *Austrian Journal of Statistics*. Vol. **31**, No. 4, pp. 267-277.

[5] Kharin A., Kishylau D. (2005). Robust Sequential Testing of Hypotheses on Discrete Distributions. *Austrian Journal of Statistics*. Vol. **34**, No. 2, pp. 153-162.

[6] Kharin A.Yu., Kishylau D.V. (2005). Performance and Robustness Analysis for Sequential Testing of Hypotheses on Parameters of Markov Chains. *Proceedings of the National Academy of Sciences of Belarus*. No. 4, pp. 30-35. (In Russian)

[7] Kemeny J.G., Snell J.L. (1959). *Finite Markov Chains*. Wiley, New York.