ERROR PROBABILITIES FOR SEQUENTIAL TESTING OF SIMPLE HYPOTHESES UNDER FUNCTIONAL DISTORTIONS

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

Abstract

The problem of error probabilities evaluation for the sequential probability ratio test under functional distortions in the L_2-metrics is analysed. Asymptotic expansions for the type I and II error probabilities are constructed.

1 Introduction

The sequential method of hypotheses testing is widely used for information processing in medicine, statistical quality control, biology and finance. A profit of sequential procedures is that the average number of observations is less than for the equivalent tests procedures based on a fixed number of observations [6]. The sequential probability ratio test (SPRT) proposed by A. Wald [6] is considered in the paper because it is used for practical purposes [5].

The SPRT can be applied effectively if the observed data satisfy the theoretical model. One of the essential sequential approach disadvantages is that the probabilities of types I and II errors are influenced greatly by the distortions in observations. In [1] we develop the approach proposed in [3] for robustness analysis where observed data is contaminated by the outliers of Tukey and Huber type [2]. In this paper we provide the robustness analysis for case of functional distortions of the probability distribution of observations.

2 Mathematical Model

Let random observations x_1, $x_2, \ldots$ be independent and identically distributed according to a probability density function $f(x, \theta)$ with a parameter $\theta \in \Theta = \{\theta_0, \theta_1\}$. The true value of θ is unknown. Let the cumulative distribution function $F(x, \theta)$ corresponds to $f(x, \theta)$.

There are two simple hypotheses on the value of the parameter:

$$H_0 : \theta = \theta_0, \qquad H_1 : \theta = \theta_1. \tag{1}$$

Denote the accumulated likelihood ratio statistic:

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

where

$$\lambda_t = \lambda(x_t) = \ln \frac{f(x_t, \theta_1)}{f(x_t, \theta_0)} \tag{3}$$

is the likelihood ratio statistic calculated for the observation x_t.

To test the hypotheses (1) by $n = 1, 2, \dots$ observations sequentially, the SPRT is used:

$$N = \min\{n \in \mathbf{N} : \Lambda_n \notin (C_-, C_+)\}, \tag{4}$$

$$d = \mathbf{1}_{[C_+, +\infty)}(\Lambda_N), \tag{5}$$

where N is the stopping time, at which the decision d is made according to (5). In (4) the thresholds $C_-, C_+ \in \mathbf{R}$ are the parameters of the test defined according to [6]:

$$C_- = \ln \frac{\beta_0}{1 - \alpha_0}, \quad C_+ = \ln \frac{1 - \beta_0}{\alpha_0}, \tag{6}$$

where α_0, β_0 are given maximal possible values of probabilities of types I and II errors respectively. It is known [6] that α_0, β_0 are just approximate values of the factual error probabilities of types I and II.

Let us make the following assumptions:

A1) the function $f(x, \theta)$ has the finite derivative of the order 1 and 2 on x, and $f(0, \theta) \neq 0$, $\theta \in \Theta$;

A2) the function $\lambda(x)$, defined by (3), is strictly monotone w.r.t. x and has the non-zero derivative.

These assumptions are satisfied, for example, by members of the exponential family of probability distributions that have the following kind of probability density function:

$$f(x, \theta) = a(x)\, b(\theta) \exp\{c(x)\, d(\theta)\},$$

where

1) the functions $a(x)$, $c(x)$ have finite derivatives of the order 1 and 2, also $a(x) \neq 0$ and $b(\theta) \neq 0$;

2) the function $c(x)$ has the derivative of the constant sign, i.e. $c'(x) > 0$, $\forall x$, or $c'(x) < 0$, $\forall x$, and $d(\theta_0) \neq d(\theta_1)$.

Without loss of generality, suppose that the hypothesis H_0 is true, so the value of the probability of type I error is considered.

To approximate the error probability α of the sequential probability ratio test (4), (5), the approach proposed in [4] is applied.

Denote $h = (C_+ - C_-)/m$, where $m \in \mathbf{N}$ is the parameter of a fragmentation (approximation). Introduce the random sequences

$$\Lambda_n^- = \sum_{t=1}^{n} \lambda_t^-, \quad \Lambda_n^+ = \sum_{t=1}^{n} \lambda_t^+;$$

$$\lambda_1^- = C_- + \left[\frac{\lambda_1 - C_-}{h}\right] h, \quad \lambda_t^- = \left[\frac{\lambda_t}{h}\right] h,\ t \geq 2; \quad \lambda_i^+ = \lambda_i^- + h,\ i \in \mathbf{N}.$$

Construct absorbing Markov chains L_n^-, L_n^+ with the states $\{0, 1, \ldots, m, m+1\}$ where 0 and $m+1$ are absorbing ones:

$$L_n^- = \begin{cases} 0, & if \ \Lambda_n^- \in (-\infty, C_- - h], \\ i, & if \ \Lambda_n^- = C_- + (i-1)h, \\ m+1, & if \ \Lambda_n^- \in [C_+, \infty), \end{cases} \qquad L_n^+ = \begin{cases} 0, & if \ \Lambda_n^+ \in (-\infty, C_-], \\ i, & if \ \Lambda_n^+ = C_- + i\,h, \quad i = \overline{1,m}, \\ m+1, & if \ \Lambda_n^+ \in [C_+ + h, \infty). \end{cases} \tag{7}$$

Theorem 1. *If assumptions A1 and A2 are hold for the considered model, then the initial probabilities and the one-step transition probabilities of the Markov chains L_n^- and L_n^+ are*

$$\begin{aligned}
&\pi_i^\pm = F(\lambda^{-1}(C_- + i\,h)) - F(\lambda^{-1}(C_- + (i-1)\,h)), \\
&\pi_0^\pm = F(\lambda^{-1}(C_-)), \\
&\pi_{m+1}^\pm = 1 - F(\lambda^{-1}(C_+)); \\
&p_{i,j}^- = p_{j-i}^- = F(\lambda^{-1}((j-i+1)h)) - F(\lambda^{-1}((j-i)h)), \\
&p_{i,0}^- = p_{-i}^- = F(\lambda^{-1}((1-i)h)), \\
&p_{i,m+1}^- = p_{m+1-i}^- = 1 - F(\lambda^{-1}((m-i+1)h)), \ \ i,j = \overline{1,m}; \\
&p_{k,l}^+ = p_{l-k}^+ = p_{l-k-1}^-, \ \ k = \overline{1,m}, \ \ l = \overline{0,m+1},
\end{aligned}$$

where $\lambda^{-1}(\cdot)$ is the inverse function to the function $\lambda(\cdot)$ specified by (3).

Denote $\pi = \pi^\pm$. Probabilities $p_{i,j}^\pm$ are the elements of the matrices $P^\pm \in \mathbf{R}^{m \times m}$ and $R^\pm \in \mathbf{R}^{m \times 2}$, let $S^\pm = I - P^\pm$.

Let α be the probability of type I error for the SPRT (4), (5). Let α_m^- and α_m^+ be the probabilities of absorption in the state $m+1$ for the Markov chains L_n^- and L_n^+ respectively.

It is proved in [4] that the probabilities α_m^-, α and α_m^+ satisfy the inequality $\alpha_m^- \leq \alpha \leq \alpha_m^+$ and the asymptotic equality $\alpha_m^+ - \alpha_m^- = O(h)$, $h \to 0$, is also fulfilled. So as the approximation of an unknown value of α, one can take $\hat{\alpha}_m = (\alpha_m^+ + \alpha_m^-)/2$ The deviation from this value is $|\alpha - \hat{\alpha}_m| \leq (\alpha_m^+ - \alpha_m^-)/2$.

3 Error Probabilities under Distortions

Let the observations x_t in the SPRT (4), (5) have a probability density function $f_1(x,\theta)$ that is unknown and different from the theoretical one $f(x,\theta)$, but the distance between $f_1(x,\theta)$ and $f(x,\theta)$ in the L_2 metrics is supposed to be not more than ε:

$$\left(\int_{\mathbf{R}} |f(x,\theta) - f_1(x,\theta)|^2 dx \right)^{\frac{1}{2}} \leq \varepsilon.$$

Let π_1, $P_1^\pm$ and $R_1^\pm$ be the vector of initial probabilities and the matrices of one-step transition and absorption probabilities of the Markov chains $L_n^\pm$ (7) in case where observations x_t are distributed by the probability density function $f_1(x,\theta)$. Let $\alpha_m^\pm(\varepsilon)$ be the probabilities of absorption of the Markov chains $L_n^\pm$.

Lemma 1. *If assumptions A1 and A2 are hold for the hypothetical and distorted models, then the following inequalities are satisfied element-by-element:*

$$|\pi_1 - \pi| \leq \tilde{\pi}\, \varepsilon, \quad |P_1^{\pm} - P^{\pm}| \leq \tilde{P}^{\pm}\, \varepsilon, \quad |R_1^{\pm} - R^{\pm}| \leq \tilde{R}^{\pm}\, \varepsilon,$$

where $\tilde{\pi}$, $\tilde{P}^{\pm}$ *and* $\tilde{R}^{\pm}$ *are the constant matrices that are independent of* $f_1(x, \theta)$.

Theorem 2. *The probabilities* $\alpha_m^-(\varepsilon)$ *and* $\alpha_m^+(\varepsilon)$ *at* $\varepsilon \to 0$ *satisfy the asymptotic expansions*

$$\alpha_m^{\pm}(\varepsilon) = \alpha_m^{\pm} + a_1^{\pm}\varepsilon + O(\varepsilon^2),$$

where

$$|a_1^{\pm}| \leq \tilde{\pi}(S^{\pm})^{-1}R^{\pm} + \pi(S^{\pm})^{-1}\tilde{P}^{\pm}(S^{\pm})^{-1}R^{\pm} + \pi(S^{\pm})^{-1}\tilde{R}^{\pm}.$$

Theorem 2 states that the difference between the type I error probability of the distorted model and the type I error probability of the hypothetical model has the first order on ε, and gives the upper estimate of the appropriate coefficient.

References

[1] Charnou S. (2009). Sequential test robustifications for simple hypotheses under outliers. *Abstracts of the International Conference on Robust Statistics*. University of Parma, p. 22.

[2] Huber P. (2004). *Robust Statistics: Theory and Methods*. John Wiley and Sons, New York.

[3] Kharin A. (2002). An approach to performance analysis of the SPRT for simple hypotheses testing. *Proc. of the Belarusian State University*. No. **1**, pp. 92-96. (Russian language)

[4] Kharin A., Chernov S. (2009). Error Probabilities Evaluation for Sequential Testing of Simple Hypotheses on Data from Continuous Distribution. *Pattern Recognition and Information Processing. Proceedings of the International Conference*. Belarusian State University, Minsk, pp. 63-66.

[5] Mukhopadhyay N., Datta S., Chattopadhyay S. (2004). *Applied Sequential Methodologies*. Marcel Dekker, New York.

[6] Wald A. (1947). *Sequential Analysis*. John Wiley & Sons, New York.