Error Probabilities Evaluation for Sequential Testing of Simple Hypotheses on Data from Continuous Distributions

Alexey Kharin, Sergey Chernov
Belarusian State University, Independence avenue 4, Minsk, Belarus, e-mail: chernovsy@tut.by

Abstract: *The problem of the error probabilities evaluation for the sequential probability ratio test is considered. Lower and upper estimates for the error probabilities are obtained and the accuracy of the approximation is analysed.*

Keywords**:** Sequential Probability Ratio Test (SPRT), Error Probability, Markov Chain.

1. INTRODUCTION

The sequential method of hypotheses testing is widely used for information processing in medicine, statistical quality control [3], biology [7] and finance [6]. A profit of the sequential procedures is that the average number of observations is less than for the equivalent tests procedures based on a fixed number of observations [1].

The sequential probability ratio test (SPRT) proposed by A.Wald [1] is considered in the paper because it is used quit often for practical purposes [7].

One of the sequential approach disadvantages is that the probabilities of types I and II errors could not be calculated exactly. In [2] an algorithm for approximate calculation of error probabilities is described, but there is no accuracy evaluation. In this paper not only the algorithm of the lower and upper estimates constructing is presented and the convergence of these estimates is proved, but the approximation accuracy is obtained as well.

Theoretical results are illustrated by computer modelling.

2. MATHEMATICAL MODEL

Let random observations $x_1, x_2, \ldots \in \mathbf{R}$ 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:

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

Denote the accumulated likelihood ratio statistic:

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

where

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

is the likelihood ratio statistics calculated on the observation x_t.

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

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

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

where N is the stopping time, at which the decision d is made according to (5). In (4) the thresholds C_-, C_+ are the parameters of the test and defined according to [1]:

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

where α_0, β_0 are given maximal possible values of probabilities of types I and II errors respectively. It is known [1] that α_0, β_0 are only approximate values of the actual error probabilities of types I and II.

Let us make the following assumptions:

A1) the function $f(x,\theta)$ has the finite derivatives 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 1st 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. $c'(x) > 0$ or $c'(x) < 0$, and $d(\theta_0) \neq d(\theta_1)$.

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

3. CONSTRUCTION OF BOUNDARY CHAINS

Denote:

$$h = \frac{C_+ - C_-}{m}, \qquad (6)$$

where $m \in \mathrm{N}$ is the parameter of a fragmentation (approximation); $\Lambda_n^- = \sum_{t=1}^{n} \lambda_t^-$,

where

$$\lambda_1^- = C_- + \left[\frac{\lambda_1 - C_-}{h}\right]h, \qquad \lambda_t^- = \left[\frac{\lambda_t}{h}\right]h, \quad t \geq 2; \qquad (7)$$

here $[\cdot]$ means the integer part of an argument.

As the random variables λ_t^- are independent (because of λ_t independence), the random sequence Λ_n^- is a Markov chain.

Introduce the Markov chain L_n^- with the states $\{0,1,\dots,m+1\}$:

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

and let 0 and $m+1$ be the absorbing states.

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

$$\begin{aligned} &\pi_i^- = F\left(\lambda^{-1}(C_- + ih)\right) - F\left(\lambda^{-1}(C_- + (i-1)h)\right), \\ &\pi_0^- = F\left(\lambda^{-1}(C_-)\right), \qquad \pi_{m+1}^- = 1 - F\left(\lambda^{-1}(C_+)\right), \\ &p_{i,j}^- = F\left(\lambda^{-1}((j-i+1)h)\right) - F\left(\lambda^{-1}((j-i)h)\right), \\ &p_{i,0}^- = F\left(\lambda^{-1}((1-i)h)\right), \\ &p_{i,m+1}^- = 1 - F\left(\lambda^{-1}((m-i+1)h)\right), \quad i, j = \overline{1,m}. \end{aligned}$$

Proof is based on the equivalent transformations of the correspondent random events. •

The Markov chain Λ_n^+ is constructed similarly to Λ_n^-:

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

where

$$\lambda_1^+ = C_- + \left[\frac{\lambda_1 - C_-}{h}\right]h + h, \quad \lambda_t^+ = \left[\frac{\lambda_t}{h}\right]h + h, \ t \geq 2. \qquad (8)$$

Denote the Markov chain L_n^+ as follows:

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

and let 0 and $m+1$ be the absorbing states.

Theorem 2. *If assumptions A1 and A2 are fulfilled for the considered model, then the initial probabilities and the one-step transition probabilities of the Markov chain* L_n^+ *are:*

$$\begin{aligned} &\pi_i^+ = F\left(\lambda^{-1}(C_- + ih)\right) - F\left(\lambda^{-1}(C_- + (i-1)h)\right), \\ &\pi_0^+ = F\left(\lambda^{-1}(C_-)\right), \qquad \pi_{m+1}^+ = 1 - F\left(\lambda^{-1}(C_+)\right), \\ &p_{i,j}^+ = F\left(\lambda^{-1}((j-i)h)\right) - F\left(\lambda^{-1}((j-i-1)h)\right), \\ &p_{i,0}^+ = F\left(\lambda^{-1}(-ih)\right), \\ &p_{i,m+1}^+ = 1 - F\left(\lambda^{-1}((m-i)h)\right), \quad i, j = \overline{1,m}. \end{aligned}$$

Proof is similar to the proof of Theorem 1 •.

Let us notice that $\pi_i^- = \pi_i^+$, $i = \overline{0, m+1}$. So denote $\pi_i = \pi_i^- = \pi_i^+$. Probabilities $p_{i,j}^{\pm}$ are the elements of the matrices $P^{\pm} \in R^{m \times m}$ and $R^{\pm} \in R^{m \times 2}$:

$$\begin{aligned} &(P^{\pm})_{i,j} = p_{i,j}^{\pm}, \\ &(R^{\pm})_{i,1} = p_{i,0}^{\pm}, \ (R^{\pm})_{i,2} = p_{i,m+1}^{\pm}, \ i, j = \overline{1,m}. \end{aligned} \qquad (9)$$

Let us notice that the transition probabilities matrices $P^{\pm}$ are Toeplitz [8], therefore effective algorithms can be applied to perform calculations with them.

Theorem 3. The *Markov chain* Λ_n *is bounded by* Λ_n^- *and* Λ_n^+*:*

$$\forall n \in N: \quad \Lambda_n^- \leq \Lambda_n \leq \Lambda_n^+.$$

Proof is based on the properties of the integer part of the real number and on the definitions of Markov chains Λ_n^-, Λ_n^+ (7), (8). •

The constructed Markov chains Λ_n^- и Λ_n^+ are called here by *boundary chains*.

4. ASYMPTOTIC PROPERTIES

Let $O_{p \times q}(h^k)$, $k \in \mathbf{N}$, be the matrix of the size $p \times q$, with each element being $O(h^k)$, $h \to 0$.

Lemma 1. *If assumptions A1 and A2 are fulfilled for the considered model, then the matrices* $P^{\pm}$ *and* $R^{\pm}$ *satisfy the following equalities at* $h \to 0$*:*

$$P^+ - P^- = O_{m \times m}(h^2), \quad R^+ - R^- = O_{m \times 2}(h).$$

Proof. Denote $f(y) = f(y \mid \theta_0)$ and consider the

(ij) -th element of the matrix P^+. From Theorem 2 we get:

$$p_{ij}^+ = \int_{\lambda^{-1}((j-i-1)h)}^{\lambda^{-1}((j-i)h)} f(y)dy, \; i,j=\overline{1,m}.$$

The function $f(y)$ has finite derivatives of the order 1 and 2 because assumption A1 is fulfilled. Using the middle rectangle formula [9] and then expanding $\lambda^{-1}(\cdot)$ and $f(\cdot)$ in the Taylor series, we get:

$$p_{ij}^+ = f\left(\lambda^{-1}((j-i)h) - \frac{A}{2}\right)A + O(A^3),$$

where $A = \lambda^{-1}((j-i)h) - \lambda^{-1}((j-i-1)h)$, $i,j=\overline{1,m}$.

Taking into account that

$$\lambda^{-1}((j-i-1)h) = \lambda^{-1}((j-i)h) - (\lambda^{-1})'((j-i)h)h + O(h^2),$$

we have

$$\begin{aligned} &A = (\lambda^{-1})'((j-i)h)h + O(h^2) = O(h), \\ &p_{ij}^+ = f\left(\lambda^{-1}((j-i)h) + O(h)\right) \times (\lambda^{-1})'((j-i)h) + \\ &+ O(h^2) = \\ &= f\left(\lambda^{-1}((j-i)h)\right) \times (\lambda^{-1})'((j-i)h)h + O(h^2). \end{aligned} \quad (10)$$

Considering the (ij) -th element of the matrix P^- is made similarly:

$$\begin{aligned} &p_{ij}^- = f\left(\lambda^{-1}((j-i+1)h) + O(h)\right)(\lambda^{-1})'((j-i)h)h + \\ &+ O(h^2) = \\ &= f\left(\lambda^{-1}((j-i)h) + O(h)\right) \times (\lambda^{-1})'((j-i)h)h + O(h^2) \\ &= f\left(\lambda^{-1}((j-i)h)\right) \times (\lambda^{-1})'((j-i)h)h + O(h^2). \end{aligned} \quad (11)$$

Subtracting p_{ij}^- from p_{ij}^+ yields $p_{ij}^+ - p_{ij}^- = O(h^2)$. So $P^+ - P^- = O_{m\times m}(h^2)$. The equality $R^+ - R^- = O_{m\times 2}(h)$ is obtained similarly. •

Lemma 2. *Under the conditions of Lemma 1, the matrices* $P^\pm$, $R^\pm$ *and* π *satisfy the following asymptotics at* $h \to 0$*:*

$$P^\pm = O_{m\times m}(h), \quad R^\pm = O_{m\times 2}(1), \quad \pi = O_{1\times m}(h).$$

Proof follows from (10), (11) and similar results for the elements of $R^\pm$ and π. •

Lemma 3. *Under the conditions of Lemma 1, the following asymptotic results hold at* $h \to 0$*:*

$$\begin{aligned} &(I-P^\pm)^{-1} = I + O_{m\times m}(h), \\ &(I-P^+)^{-1} - (I-P^-)^{-1} = O_{m\times m}(h^2). \end{aligned}$$

Proof follows from the expansion of the function $g(\lambda) = \dfrac{1}{1-\lambda}$, $|\lambda|<1$, in the Taylor series and the result of Lemma 1. •

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. From the construction of Markov chains L_n^- and L_n^+, it follows that

$$\alpha_m^- \le \alpha \le \alpha_m^+. \quad (12)$$

Theorem 4. *For the considered model the following asymptotic result holds:*

$$\alpha_m^+ - \alpha_m^- = O(h).$$

Proof. Let $R_j^\pm$ be the j -th column of the matrix $R^\pm$, $j=1,2$. Then according to the definition

$$\begin{aligned} &\alpha_m^+ = P(L_n^+ \ge C_+ \mid H_0) = \\ &= \pi_{m+1} + \pi\sum_{i=0}^{\infty}(P^+)^i R_2^+ = \pi_{m+1} + \pi\left(I-P^+\right)^{-1}R_2^+. \end{aligned}$$

For α_m^- we obtain in the same way

$$\alpha_m^- = \pi_{m+1} + \pi\left(I-P^-\right)^{-1}R_2^-.$$

Subtracting α_m^- from α_m^+ yields

$$\begin{aligned} &\alpha_m^+ - \alpha_m^- = \pi\left(I-P^+\right)^{-1}R_2^+ - \pi\left(I-P^-\right)^{-1}R_2^- = \\ &= \pi\left((I-P^+)^{-1} - (I-P^-)^{-1}\right)R_2^+ + \\ &+ \pi\left(I-P^-\right)^{-1}\left(R_2^+ - R_2^-\right). \end{aligned}$$

Using the results of Lemmas 1-3 and taking in mind (6), we have:

$$\begin{aligned} &\pi\left((I-P^+)^{-1} - (I-P^-)^{-1}\right)R_2^+ + \\ &\quad + \pi\left(I-P^-\right)^{-1}\left(R_2^+ - R_2^-\right) = \\ &= O_{1\times m}(h)O_{m\times m}(h^2)O_{m\times 1}(1) + \\ &\quad + O_{1\times m}(h)(I + O_{m\times m}(h))O_{m\times 1}(h) = \\ &= O_{1\times m}(h)O_{m\times 1}(h) + O_{1\times m}(h)O_{m\times 1}(h) = O(h). \end{aligned}$$

So $\alpha_m^+ - \alpha_m^- = O(h)$. •

From Theorem 4 and (12) it follows that α_m^+ tends to the value $\alpha+0$, and α_m^- tends to $\alpha-0$ at $h\to 0$ ($m\to\infty$).

Denote by

$$\tilde{\alpha}_m = \frac{1}{2}\left(\alpha_m^+ + \alpha_m^-\right). \tag{13}$$

Corollary 1. *The approximation* ***Ошибка! Источник ссылки не найден.*** *converges to the value of probability of type I error with the rate $O(h)$, and the absolute deviation from this value is not more then a half of the segment $[\alpha_m^-, \alpha_m^+]$:*

$$|\,\alpha - \tilde{\alpha}_m\,| \le \frac{1}{2}\left(\alpha_m^+ - \alpha_m^-\right).$$

5. NUMERICAL EXPERIMENTS

To illustrate the theoretical results, we performed computer modelling. The case $\alpha_0 = \beta_0 = 0.1$ was considered.

Let $\tilde{\alpha}_{MC}$ be the Monte-Carlo estimate for the probability of type I error for the SPRT (4), (5). For each set of parameters 5 000 000 replications were performed.

The observations $x_1, x_2, \ldots$ were normally distributed with the probability density function $f(x,\theta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\theta)^2}$. The analysed hypotheses were $H_0: \theta = 0,\ H_1: \theta = \theta_1$.

The results of computer modelling are presented in Table 1.

Table 1. Results of numerical experiments

θ_1	m	α_m^-	$\tilde{\alpha}_m$	$\tilde{\alpha}_{MC}$	α_m^+
1.0	50	0.04906	0.05981	0.05843	0.07055
	100	0.05362	0.05896		0.06431
	500	0.05762	0.05869		0.05976
	1000	0.05814	0.05868		0.05922
	2000	0.05841	0.05868		0.05895
0.5	100	0.05239	0.08206	0.07667	0.11172
	500	0.07108	0.07691		0.08274
	1000	0.07384	0.07675		0.07966
	2000	0.07525	0.07671		0.07817
	3000	0.07573	0.07670		0.07767
0.3	100	0.02960	0.12765	0.08535	0.22570
	500	0.06928	0.08703		0.10478
	1000	0.07692	0.08576		0.09460
	2000	0.08104	0.08545		0.08986
	3000	0.08245	0.08539		0.08833
0.2	500	0.05649	0.09857	0.08993	0.14065
	1000	0.07143	0.09214		0.11285
	2000	0.08022	0.09053		0.10084
	3000	0.08336	0.09023		0.09710
	4000	0.08498	0.09013		0.09527

These results show that the lower and upper bounds of the probability of type I error are precise enough for reasonable values of m. So the proposed methodology can be used in practise to approximate the type I and II error probabilities for sequential testing of data from a continuous distribution.

6. REFERENCES

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

[2] A. Kharin. An approach to performance analysis of the SPRT for simple hypotheses testing. *Proc. of the Belarusian State University*, 1 (2002), p. 92-96. (In Russian)

[3] *Handbook of Sequential Analysis.* Eds.: B.K. Ghosh et al. Marcel Dekker. New York, 1991.

[4] A. Kharin, D. Kishylau. Robust sequential testing of hypotheses on discrete probability distributions, *Austrian Journal of Statistics*, 34 (2) (2005), p. 153-162.

[5] J.G. Kemeni, J.L. Snell. *Finite Markov Chains.* Wiley, New York, 1959.

[6] T.L. Lai. Sequential analysis: Some classical problems and new challenges. *Statistica Sinica,* 11, (2001), p. 303-408.

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

[8] V.V. Voevodin, E.E. Tyrtyshnikov. *Computational processes with Toeplitz matrices.* Nauka, Moscow, 1987. (In Russian)

[9] V.M. Verbzhyckii. *Numerical methods. Mathematical analysis and ordinary differential equations.* Vysshaya Shkola, Moscow, 2001. (In Russian)