SEQUENTIAL TESTING OF SIMPLE HYPOTHESES ON PARAMETERS OF MARKOV CHAINS

A. Kharin, D. Kishilov

Department of Probability Theory and Mathematical Statistics, Belarussian State University, Fr. Skoriny av., 4, Minsk, 220050, Belarus. e-mail: *KharinAY@bsu.by*

Abstract. The problem of sequential testing of hypotheses on parameters of Markov chains is considered. The special case of simple hypotheses is analyzed. The expressions for conditional error probabilities and for conditional expected sequence sizes are obtained. The robustness analysis under "contamination" of Tukey—Huber type is performed. The case of Markov chains of the order $K>1$ is studied separately. The results of computer modelling illustrate the theory.

1 Introduction

In different problems of information processing it often appears the necessity to discriminate between two types of random sequences. In many cases these sequences can be described by the models of Markov chains. For on-line discrimination of Markov chains, the sequential testing of statistical hypotheses is an efficient technique. Such a situation is quite typical for medical trials, genetics, and for quality control (see [1]). The sequential probability ratio test (SPRT) proposed by Wald [2] is proved to be optimal [3] w.r.t. expected sequence size minimization. But such optimality is extremely seldom in practice because of the fact, that two simple hypotheses is usually not an adequate model for an applied problem. The hypothetical model is often distorted [4], and the problem of robustness analysis of the SPRT appears to be important. For the case of independent identically distributed observations from a discrete probability distribution of a special type this problem has been solved in [5], also with construction of a "robustified" sequential test. In this paper we analyze how the characteristics of the SPRT change under "contamination" of the Tukey—Huber type for the case where the observed sequences are Markov chains. In some applications the problem of discrimination between Markov chains of a high order arises. In the fourth section of the paper we construct the sequential test for this case. We finalize the paper with the results of computer modelling.

2 Mathematical model and construction of the SPRT

Let a homogeneous Markov chain of the first order x_1, x_2, K, be observed, $x_t \in U = \{0, 1, K, N-1\}$, with a vector of initial probabilities $\pi = (\pi_i)$, $i \in U$, and with a transition probabilities matrix $P = (p_{ij})$, $i, j \in U$: $P\{x_1 = i\} = \pi_i$, $P\{x_n = j \mid x_{n-1} = i\} = p_{ij}$, $i, j \in U$, $n > 1$. There are two hypotheses on parameters of the Markov chain: $H_0 : \pi = \pi^{(0)}$, $P = P^{(0)}$, with an alternative $H_1 : \pi = \pi^{(1)}$, $P = P^{(1)}$, where the hypothetical values $\pi^{(0)}$, $\pi^{(1)}$ and $P^{(0)}$, $P^{(1)}$ are given; $P^{(0)} \neq P^{(1)}$. Introduce the notation: $\lambda_1 = \ln(\pi^{(1)}_{x_1} / \pi^{(0)}_{x_1})$, $\lambda_k = \ln(p^{(1)}_{x_{k-1}, x_k} / p^{(0)}_{x_{k-1}, x_k})$, $k > 1$, $\Lambda_n = \sum_{k=1}^{n} \lambda_k$, $n \geq 1$.

According to the SPRT for testing H_0, H_1, when given $C_-, C_+ \in \Re$, $C_- < 0$, $C_+ > 0$, the hypothesis H_0 is accepted after n observations, if $\Lambda_n \leq C_-$, the hypothesis

H_1 is accepted, if $\Lambda_n \geq C_+$, else an observation number *(n+1)* has to be made. The sequence of random vectors $(\Lambda_n, x_n)'$, $n \in N$, is a Markov chain.

Let $\pi^{(0)}$, $P^{(0)}$, $\pi^{(1)}$, $P^{(1)}$ be such that $\exists a \in \Re$, m_i, $m_{ij} \in Z$, $i, j \in U$, for which $\ln\left(\pi_i^{(1)}/\pi_i^{(0)}\right) = m_i a$, $\ln\left(p_{ij}^{(1)}/p_{ij}^{(0)}\right) = m_{ij} a$; $C_-, C_+ \in Z$. Introduce the notation: $\delta(i,j)$ for the Kroneker delta, $\mathbf{I}_k$ for the identity matrix of the order k, $\mathbf{0}_{m\times n}$ for the $(m \times n)$ matrix with all elements equal to 0, $\mathbf{1}(u)$ for the unit step function, $\mathbf{1}_k$ for the k-vector with all components equal to 1, $t^{(k)}$ for the conditional expected sequence size, under the condition that the hypothesis H_k is true, $k \in \{0,1\}$, α, β for the conditional error probabilities. Denote by $W^{(k)} = (w_{ij}^{(k)}) = \left(\begin{array}{c|c} \mathbf{I}_2 & \mathbf{0}_{2\times N(C_+-C_--1)} \\ --- & ------- \\ R^{(k)} & Q^{(k)} \end{array}\right)$ the matrix with the blocks $R^{(k)}$, $Q^{(k)}$ defined by their elements as

$$w_{Ni+s\,Nj+t}^{(k)} = \delta(j-i, m_{st}) p_{st}^{(k)},\ i,j \in (C_-, C_+), s,t \in U,$$

$$w_{Ni+s\,Nj}^{(k)} = \begin{cases} \sum\limits_{t\in U} \mathbf{1}(i + m_{st} - C_+), & j = C_+, i \in (C_-, C_+), s \in U, \\ \sum\limits_{t\in U} \mathbf{1}(C_- - i - m_{st}), & j = C_-, i \in (C_-, C_+), s \in U. \end{cases} \quad (1)$$

Introduce the vector $\omega^{(k)} = (\omega_i^{(k)})$, $i \in \{NC_- + 1, \ldots, NC_+ - 1\}$:

$$\omega_{Ni+s}^{(k)} = \delta(i, m_s)\pi_s^{(k)},\ i \in (C_-, C_+), s \in U; \quad (2)$$

$$\omega_{NC_-}^{(k)} = \sum_{s\in U} \mathbf{1}(C_- - m_s)\pi_s^{(k)},\ \omega_{NC_+}^{(k)} = \sum_{s\in U} \mathbf{1}(m_s - C_+)\pi_s^{(k)}. \quad (3)$$

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

Theorem 1 *Under the model described above, if* $\left|S^{(k)}\right| \neq 0$, $k \in \{0,1\}$, *then for the SPRT characteristics the following equations hold:*

$$\alpha = (\omega^{(0)})' B_{(2)}^{(0)} + \omega_{NC_+}^{(0)},\ \beta = (\omega^{(1)})' B_{(1)}^{(1)} + \omega_{NC_-}^{(1)},\ t^{(k)} = (\omega^{(k)})' (S^{(k)})^{-1} \mathbf{1}_{N(C_+-C_--1)} + 1.$$

Proof is based on the theory of absorbing Markov chains [6]; it should be applied to the sequence

$$\xi_n = NC_- \mathbf{1}_{(-\infty, C_-]}(\Lambda_n/a) + NC_+ \mathbf{1}_{[C_+, +\infty)}(\Lambda_n/a) + ((\Lambda_n/a)N + x_n)\mathbf{1}_{(C_-, C_+)}(\Lambda_n/a), n \in N, \quad (4)$$

which is a homogeneous Markov chain with $N(C_+ - C_- - 1) + 2$ states, and two of them, NC_- and NC_+, are absorbing, the transition probabilities matrix is given by (1), and the initial probabilities vector is given by (2), (3). ■

3 Robustness analysis of the SPRT under "contaminations"

Let the hypothetical model described above be distorted: the initial probabilities vector and the transition probabilities matrix are given by the mixtures [4]

$$\bar{\pi}^{(k)} = (1-\varepsilon)\pi^{(k)} + \varepsilon\tilde{\pi}^{(k)},\ \bar{P}^{(k)} = (1-\varepsilon)P^{(k)} + \varepsilon\tilde{P}^{(k)},\ k \in \{0,1\}, \quad (5)$$

where $\widetilde{\pi}^{(k)} \neq \pi^{(k)}$ and $\widetilde{P}^{(k)} \neq P^{(k)}$ are the initial probabilities vector and the transition probabilities matrix for the "contaminating" Markov chain, $\varepsilon \in [0,0.5)$ is a probability of "contamination".

Introduce $\widetilde{W}^{(k)}$, $\widetilde{Q}^{(k)}$, $\widetilde{R}^{(k)}$, $\widetilde{\omega}^{(k)}$, $\widetilde{\omega}^{(k)}_{NC_\pm}$, $k \in \{0,1\}$, similarly to the hypothetical case, replacing $p_{ij}^{(k)}$ and $\pi_i^{(k)}$ with $\widetilde{p}_{ij}^{(k)}$ and $\widetilde{\pi}_i^{(k)}$, $i, j \in \{NC_- + 1, \ldots, NC_+ - 1\}$. Define the matrix $\widetilde{S}^{(k)} = \mathbf{I}_{N(C_+ - C_- - 1)} - Q^{(k)} - \varepsilon(\widetilde{Q}^{(k)} - Q^{(k)})$, $k \in \{0,1\}$.

Theorem 2 *If the hypothetical model is distorted according to (5),* $|S^{(k)}| \neq 0$, $|\widetilde{S}^{(k)}| \neq 0$, $k \in \{0,1\}$, *then the characteristics* $\overline{\alpha}$, $\overline{\beta}$, $\overline{t}^{(k)}$ *of the SPRT differ from the correspondent hypothetical characteristics by the values of the order* $O(\varepsilon)$:

$$\overline{t}^{(k)} - t^{(k)} = \varepsilon\left(\left(\widetilde{\omega}^{(k)} - \omega^{(k)}\right)' + \left(\omega^{(k)}\right)'\left(S^{(k)}\right)^{-1}\left(\widetilde{Q}^{(k)} - Q^{(k)}\right)\right) \times$$

$$\times \left(S^{(k)}\right)^{-1} \mathbf{1}_{N(C_+ - C_- - 1)} + O(\varepsilon^2),$$

$$\overline{\alpha} - \alpha = \varepsilon\left(\left(\omega^{(0)}\right)'\left(\left(S^{(0)}\right)^{-1}\left(\left(\widetilde{Q}^{(0)} - Q^{(0)}\right)\left(S^{(0)}\right)^{-1} R^{(0)} + \widetilde{R}^{(0)} - R^{(0)}\right)\right)_{(2)} +$$

$$+ (\widetilde{\omega}^{(0)} - \omega^{(0)})' B^{(0)}_{(2)} + \widetilde{\omega}^{(0)}_{NC_+} - \omega^{(0)}_{NC_+}) + O(\varepsilon^2),$$

$$\overline{\beta} - \beta = \varepsilon\left(\left(\omega^{(1)}\right)'\left(\left(S^{(1)}\right)^{-1}\left(\left(\widetilde{Q}^{(1)} - Q^{(1)}\right)\left(S^{(1)}\right)^{-1} R^{(1)} + \widetilde{R}^{(1)} - R^{(1)}\right)\right)_{(1)} +$$

$$+ (\widetilde{\omega}^{(1)} - \omega^{(1)})' B^{(1)}_{(1)} + \omega^{(1)}_{NC_-} - \omega^{(1)}_{NC_-}) + O(\varepsilon^2).$$

Proof is based on the fact that under the distortions (5) the matrix of transition probabilities and the vector of initial probabilities for the sequence (4) have the form $\overline{W}^{(k)} = W^{(k)} + \varepsilon(\widetilde{W}^{(k)} - W^{(k)})$, $\overline{\omega}^{(k)} = \omega^{(k)} + \varepsilon(\widetilde{\omega}^{(k)} - \omega^{(k)})$. ■

Some numerical results on linear approximation of the characteristics of the SPRT are presented in [5] for the case of independent observations.

4 The case of high order Markov chains

Let $x_1, x_2, \ldots$ be an observable Markov chain of the order $K>1$ with the set of values $U = \{0,1,\ldots,N-1\}$. Let the probability distribution for the K initial observations be $P\{x_1 = X_1, \ldots, x_K = X_K\} = \Pi(X_1, \ldots, X_K)$, $X_i \in U$. Introduce the notation for the conditional distribution $P\{x_n = X_{K+1} \mid x_{n-1} = X_K, \ldots, x_{n-K} = X_1\} = P(X_1, \ldots, X_{K+1})$, $X_i \in U$, $n \geq K+1$. We consider the hypothesis $H_0 : \Pi = \Pi^{(0)}$, $P = P^{(0)}$ with the alternative $H_1 : \Pi = \Pi^{(1)}$, $P = P^{(1)}$, where the values $\Pi^{(0)}$, $\Pi^{(1)}$, $P^{(0)}$, $P^{(1)}$ are given.

Introduce the function $I : U^K \to V = \{0,1,\ldots,N^K - 1\}$, $I(X_1, X_2, \ldots, X_K) = \sum_{j=0}^{K-1} X_{K-j} N^j$, $X_i \in U$, $i \in \{1, \ldots, K\}$; I is a bijection, and there exists the inverse function $I^{-1}(\cdot)$. Introduce the sequence of vectors $I_m = I(x_{(m-1)K+1}, \ldots, x_{mK})$, $m = 1,2,\ldots$, which is a

Markov chain. Define the vector $\tilde{\pi}^{(k)} = (\tilde{\pi}_i^{(k)})$ and the matrix $\tilde{P}^{(k)} = (\tilde{p}_{ij}^{(k)})$, $i, j \in V$, $k \in \{0,1\}$, as

$$\tilde{\pi}_i^{(k)} = \Pi^{(k)}(I^{-1}(i)), i \in V,$$

$$\tilde{p}_{ij}^{(k)} = P^{(k)}(X_1,\ldots,X_K,Y_1) \cdot P(X_2,X_3,\ldots,Y_1,Y_2) \cdot \ldots \cdot P(X_K,Y_1,\ldots,Y_K),\ i,j \in V,$$

where $(X_1,\ldots,X_K) = I^{-1}(i)$, $(Y_1,\ldots,Y_K) = I^{-1}(j)$. Under the hypothesis $H^{(k)}$ being true, the vector of initial probabilities $\tilde{\pi}$ and the transition probabilities matrix $\tilde{P}$ of the Markov chain I_m are equal to $\tilde{\pi}^{(k)}$ и $\tilde{P}^{(k)}$ respectively.

The problem of discrimination between H_0, H_1 is reducible to the problem of testing of $\tilde{H}_0$: $\tilde{\pi} = \tilde{\pi}^{(0)}, \tilde{P} = \tilde{P}^{(0)}$ with the alternative $\tilde{H}_1$: $\tilde{\pi} = \tilde{\pi}^{(1)}, \tilde{P} = \tilde{P}^{(1)}$. Let Λ_m be the statistic constructed for the sequence I_m similarly to Section 2. Let the parameters $C_-, C_+ \in \Re$, $C_- < 0$, $C_+ > 0$ be fixed. According to the SPRT, the hypothesis H_0 is accepted after mK observations, if $\Lambda_m \le C_-$; H_1 is accepted, if $\Lambda_m \ge C_+$; else the observation process is to be continued, it means that the next K observations $x_{mK+1},\ldots,x_{(m+1)K}$ have to be made. For the analysis of statistical properties of this test the results presented in Sections 2 and 3 could be applied.

5 Results of computer modelling

In Table 1 the estimates for the characteristics of the SPRT are presented. These estimates are obtained by statistical simulation. The case where *K=4, N=2* was considered. The hypothesis H_0 formed by the equations:

$$\Pi(X_1,\ldots,X_4) \equiv \frac{1}{2^4},$$

$$P(X_1,X_2,X_3,X_4,X_5) = \begin{cases} 0.6, & \text{if } (X_1,X_2,X_3,X_4,X_5) = (1,1,1,0,0) \\ 0.4, & \text{if } (X_1,X_2,X_3,X_4,X_5) = (1,1,1,0,1); \\ 0.5, & \text{otherwise} \end{cases}$$

for the hypothesis H_1:

$$\Pi(X_1,\ldots,X_4) \equiv \frac{1}{2^4},$$

$$P(X_1,X_2,X_3,X_4,X_5) = \begin{cases} 0.4, & \text{if } (X_1,X_2,X_3,X_4,X_5) = (0,0,0,1,1) \\ 0.6, & \text{if } (X_1,X_2,X_3,X_4,X_5) = (0,0,0,1,0). \\ 0.5, & \text{otherwise} \end{cases}$$

Let α_0, β_0 be the given values of conditional error probabilities, $\tilde{\alpha}$, $\tilde{\beta}$, $\tilde{t}^{(0)}$, $\tilde{t}^{(1)}$ be the estimates of the conditional error probabilities and of the conditional expected sequence sizes, which are obtained in computer modelling; according to [1], $C_+ = \log\frac{1-\beta_0}{\alpha_0}$, $C_- = \log\frac{\beta_0}{1-\alpha_0}$. The number of experiments was equal to 10000 with every set of parameters.

α_0	β_0	$\tilde{\alpha}$	$\tilde{\beta}$	$\tilde{t}^{(0)}$	$\tilde{t}^{(1)}$
0.01	0.01	0.0075	0.0085	178.886	180.292
0.01	0.05	0.0071	0.0348	169.438	124.348
0.01	0.10	0.0070	0.0807	157.394	94.925
0.05	0.05	0.0327	0.0294	119.911	118.335
0.05	0.10	0.0332	0.0831	110.291	86.094
0.10	0.10	0.0833	0.0812	79.425	79.260

Table 1 Results of computer modelling

This paper is a result of research and academic partnership between the Belarussian State University, Faculty of Applied Mathematics and Informatics, and the Vienna University of Technology, Institute for Statistics and Probability Theory.

References

[1] Wu, T.-J., Hsieh, Ya.-Ch., and Li, L.-A. (2001). Statistical measures of DNA sequence dissimilarity under Markov chain models of base composition. *Biometrics,* **57**, 441-448.

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

[3] Siegmund, D. (1975). Error probabilities and average sample number of the sequential probability ratio test. *Journal of the Royal Stat. Soc., Ser. B.* Vol. 37, pp. 394–401.

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

[5] Kharin, A. (2002). On robustifying of the sequential probability ratio test for a discrete model under "contaminations". *Austrian Journal of Statistics,* **31** (4), 267-277.

[6] Kemeni, J.G., and Snell, J.L. (1959). *Finite Markov Chains.* Wiley, New York.