DOI: https://doi.org/10.46354/i3m.2019.emss.005

QUICKEST CHANGE-POINT DETECTION IN TIME SERIES WITH UNKNOWN DISTRIBUTIONS

S.M. Pergamenshchikov[(a)], Tartakovsky, A.G.[(b)]

[(a)] LMRS, UMR 6085 CNRS, Université de Rouen Normandie, Saint-Etienne du Rouvray Cedex, France International Laboratory SSP & QF, Tomsk State University
[(d)] Space Informatics Laboratory, Moscow Institute of Physics and Technology, Moscow, Russia and AGT StatConsult, Los Angeles, California, USA

[(a)] Serge.Pergamenshchikov@univ-rouen.fr, [(b)] agt@phystech.edu

ABSTRACT
We consider a problem of sequential detection of changes in general time series, in which case the observations are dependent and non-identically distributed, e.g., follow Markov, hidden Markov or even more general stochastic models. It is assumed that the pre-change model is completely known, but the post-change model contains an unknown (possibly vector) parameter. Imposing a distribution on the unknown post-change parameter, we design a mixture Shiryaev-Roberts change detection procedure in such a way that the maximal local probability of a false alarm (MLPFA) in a prespecified time window does not exceed a given level and show that this procedure is nearly optimal as the MLPFA goes to zero in the sense of minimizing the expected delay to detection uniformly over all points of change under very general conditions. These conditions are formulated in terms of the rate of convergence in the strong law of large numbers for the log-likelihood ratios between the "change" and "no-change" hypotheses. An example related to a multivariate Markov model where these conditions hold is given.

Keywords: asymptotic optimality, change-point detection, composite post-change hypothesis, quickest detection, weighted Shiryaev-Roberts procedure

1. INTRODUCTION
The problem of quick detection of abrupt changes in time series arises in different areas related to automatic control, segmentation of signals, biomedical signal processing, quality control engineering, finance, link failure detection in communication networks, intrusion detection in computer systems, and target detection in surveillance systems. See, e.g., Basseville and Nikiforov (1993), Kent (2000), Page (1954), Tartakovsky, Nikiforov, and Basseville (2015) and references therein. One of a challenging application area is intrusion detection in computer networks [Kent (2000), Tartakovsky et al., Rozovskii, Blažek (2006)]. Large scale attacks, such as denial-of-service attacks, occur at unknown points in time and need to be detected in the early stages by observing abrupt changes in the network traffic.

In the change point analysis, a large variety of observation models is used which include i.i.d. sequences of random variables whose distributions change at the disruption time and also different models with dependent observations. Many papers have been devoted to the problem of detecting abrupt changes of the parameters in autoregression and Markov processes, which are widely used in the statistical analysis of time series and statistics of random processes. There is a vast literature on the detection of abrupt parameter changes in Markov time series with known probabilistic characteristics. See, for example, Basseville and Nikiforov (1993), Lai (1998), Pergamenshchikov and Tartakovsky (2018), Yakir (1994) and references therein.

The present paper addresses a general non-i.i.d. model when the post-change distribution contains an unknown parameter. Using the analytical results obtained in Pergamenshchikov and Tartakovsky (2018), we establish very general conditions under which the mixture Shiryaev-Roberts detection procedure is asymptotically optimal, minimizing the expected delay to detection in the class of change detection procedures with the given maximal local probability of a false alarm when this probability is small.

2. PROBLEM
Consider the change-point problem for the general dependent non-i.i.d. model (time series) $(x_k)_{k\ge1}$ specified by the conditional densities of x_k given $(x_1,\dots,x_{k-1})$, denoted as $f_{\theta,k}^{(\nu)}(x_k \mid x_1,\dots,x_{k-1})$, where θ is an unknown parameter. More precisely, we assume that the conditional density changes from
$f_*(x_k \mid x_1,\dots, x_{k-1})$ to $f_\theta(x_k \mid x_1,\dots, x_{k-1})$ at a point ν, i.e.

$$f_{\theta,k}^{(\nu)}\left(x_k \mid x_1,\ldots,x_{k-1}\right) = f_*\left(x_k \mid x_1,\ldots,x_{k-1}\right)\mathbf{1}_{\{k\le\nu\}} + f_\theta\left(x_k \mid x_1,\ldots,x_{k-1}\right)\mathbf{1}_{\{k>\nu\}}. \quad (1)$$

Assume that the conditional densities

$$\left(f_*\left(x_k \mid x_1,\ldots,x_{k-1}\right)\right)_{k\ge 1} \quad (2)$$

are known, the change point ν is a nonrandom unknown integer and one needs to detect the change as soon as possible after it occurs. Introduce the class of detection procedures $\mathbf{M}_{\alpha,m}$ defined by stopping times τ such that

$$\sup_{k\ge 1} \mathbf{P}_*\left(k \le \tau \le k+m\right) \le \alpha, \quad (3)$$

where $\mathbf{P}_*$ is the distribution generated by the family (2), $0<\alpha<1$ is a preassigned upper bound for the false alarm probability and m is a suitably chosen window size. In the sequel, we denote by $\mathbf{P}_{\theta,\nu}$ the distribution of the process $\left(x_k\right)_{k\ge 1}$ defined by the family of the conditional densities (1).

Our goal is to find an optimal change-point detection procedure which minimizes the conditional average delay time, i.e.,

$$\inf_{\tau\in\mathbf{M}_{\alpha,m}} \mathbf{E}_{\theta,\nu}\left(\tau-\nu \mid \tau\ge\nu\right) \quad \text{for all } \nu\ge 0$$

where $\mathbf{E}_{\theta,\nu}$ is the expectation with respect to the distribution $\mathbf{P}_{\theta,\nu}$. However, finding a strictly optimal detection procedure in the problem (4) is very difficult, if at all possible. For this reason, we consider an asymptotic problem of finding a first-order asymptotically optimal rule that satisfies

$$\lim_{\alpha\to 0} \frac{\inf_{\tau\in\mathbf{M}_{\alpha,m}} \mathbf{E}_{\theta,\nu}\left(\tau-\nu \mid \tau>.\nu\right)}{\mathbf{E}_{\theta,\nu}\left(\tau-\nu \mid \tau>\nu\right)} = 1. \quad (4)$$

3. MAIN RESULTS

3.1. The Information Lower Bound

To study the optimality properties for the detection procedures we use the lower bound obtained in Pergamenshchikov and Tartakovsky (2018):

$$\liminf_{\alpha\to 0} \frac{\inf_{\tau\in\mathbf{M}_{\alpha,m}} \mathbf{E}_{\theta,\nu}\left(\tau-\nu \mid \tau\ge\nu\right)}{\left|\ln\alpha\right|} \ge \frac{1}{I(\theta)}, \quad (5)$$

where $I(\theta)$ is the generalized Kullback-Leibler information number. This asymptotic lower bound holds whenever the log-likelihood ratio (LLR) obeys the strong law of large numbers:

$$\frac{1}{n}\log \sum_{t=k+1}^{k+n} \frac{f_\theta\left(x_t \mid x_1,\ldots,x_{t-1}\right)}{f_*\left(x_t \mid x_1,\ldots,x_{t-1}\right)} \to I(\theta) \quad (6)$$

as $n\to\infty$ $\mathbf{P}_{\theta,k}$ – a.s.

3.2. The Mixture Shiryaev-Roberts Procedure

Moreover, in this paper using the modified Shiryaev - Roberts procedures proposed in Pergamenshchikov, S. M. and Tartakovsky, A.G. (2018) we construct a special weighted procedure T^* which belongs to the class $\mathbf{M}_{\alpha,m}$.

Let $W(\theta)$ be a distribution on the parameter space Θ. Define the likelihood ratio (LR) mixture as

$$\Lambda_n^k(W) = \int_\Theta \prod_{i=k+1}^{n} \frac{f_{\theta,i}\left(X_i \mid X_1,\ldots,X_{i-1}\right)}{f_{0,i}\left(X_i \mid X_1,\ldots,X_{i-1}\right)} dW(\theta), \; n>k.$$

In what follows, we assume that $W(\theta)$ is quite arbitrary satisfying the condition

$\left(\mathbf{C}_W\right)$ For any $\delta>0$, the measure W is positive on $\left\{u\in\Theta : \left|u-\theta\right|<\delta\right\}$ for any $\theta\in\Theta$, i.e.,

$$W\left\{u\in\Theta : \left|u-\theta\right|<\delta\right\} > 0.$$

This condition means that we do not consider parameter values of θ from Θ of the measure null.

Introduce the Shiryaev-Roberts (SR) statistic

$$R_n(\theta) = \sum_{k=1}^{n}\prod_{i=k}^{n} \frac{f_{\theta,i}\left(X_i \mid X_1,\ldots,X_{i-1}\right)}{f_{0,i}\left(X_i \mid X_1,\ldots,X_{i-1}\right)}.$$

Note that it is tuned to $\theta\in\Theta$. In this paper we use the mixture SR statistic

$$R_n^W = \sum_{k=1}^{n} \Lambda_n^k(W) \equiv \int_\Theta R_n(\theta)dW(\theta), \quad n\ge 1,$$

$R_0^W = 0$.

The associated detection procedure, which we will call the *Mixture Shiryaev-Roberts* (MiSR) detection procedure, is given by the stopping time

$$T_a = \inf\left\{n\ge 1 : \log R_n^W \ge a\right\}, \quad \inf\{\varnothing\} = +\infty \quad (7)$$

where $a>-\infty$ is a threshold controlling for the false alarm risk. Write $T^* = T_{a^*}$, where a^* is some function of α which goes to ∞ as $\alpha\to 0$. Using a left-tail complete convergence condition in the strong law (6), which usually holds under "concentration" conditions for the LLR process, it can be shown that along with the lower bound (5) the following upper bound holds:

$$\limsup_{\alpha\to 0} \frac{\mathbf{E}_{\theta,\nu}\left(T^* - \nu \mid \tau \ge \nu\right)}{|\ln \alpha|} \le \frac{1}{I(\theta)}. \quad (8)$$

Note that it follows from the bounds (5) and (8) that the proposed MiSR procedure is asymptotically optimal since the asymptotic equality (4) holds for $T = T^*$.

As an example for which we check the strong law of large numbers (6) and "concentration" conditions for the LLR process, we consider the change point detection problem for the autoregressive model. Specifically, assume that $(x_k)_{k\ge 1}$ is the autoregressive process of order p:

$$x_k = a_1 x_{k-1} + \ldots + a_p x_{k-p} + \varepsilon_k, \quad (9)$$

before the change time ν, i.e. for $k \le \nu$, and

$$x_k = \theta_1 x_{k-1} + \ldots + \theta_p x_{k-p} + \varepsilon_k, \quad (10)$$

after ν. Here $(\varepsilon_k)_{k\ge 1}$ is i.i.d. sequence of random Gaussian variables with the parameter (0,1). The parameters $a = (a_1, \ldots, a_p)$ are known and, the parameters $\theta = (\theta_1, \ldots, \theta_p) \ne a$ are unknown. We assume that both processes (9) and (10) are stable, that is all roots of the corresponding characteristic polynomials lie inside the unit circle of complex plane. For this example, the procedure T^* is asymptotically optimal with

$$I(\theta) = \frac{(\theta - a)' F^{-1} (\theta - a)}{2},$$

where F is the covariance matrix of order p for the stationary process (10) which is given in Example 5 in Pergamenshchikov and Tartakovsky (2018).

4. MONTE CARLO SIMULATIONS

In this section, through the Python software we provide Monte Carlo (MC) simulations for the AR(1) model, which is a particular case of (9) – (10) for $p = 1$. Specifically, let the pre-change value $a = 0$ and the post-change value $\theta \in \Theta = \{\theta_1, \ldots, \theta_N\}$, $-1 < \theta_1 < \ldots < \theta_N < 1$, $\theta_i \ne 0$. We set

$$L_n^\theta(X_n, X_{n-1}) = \exp\left\{\theta X_n X_{n-1} - \frac{\theta^2 X_{n-1}^2}{2}\right\}, \; n \ge 1.$$

The MiSR stopping time is written as

$$T_a = \inf\left\{n \ge 1 : \sum_{j=1}^{N} W(\theta_j) R_n(\theta_j) \ge e^a\right\},$$

where the SR statistic $R_n(\theta)$ satisfies the recursion

$$R_{n+1}(\theta) = [1 + R_n(\theta)] L_{n+1}^\theta(X_n, X_{n+1}), \; n \ge 1, \; R_0(\theta) = 0.$$

Thus, the MiSR procedure can be easily implemented. The generalized information number $I(\theta) = \theta^2/[2\,(1-\theta^2)]$, so the first-order approximation yields the following approximate formula for the average delay to detection $ADD_{\nu,\theta}(T_a) = \mathbf{E}_{\nu,\theta}(T_a - \nu \mid T_a > \nu)$:

$$ADD_{\nu,\theta}(T_a) \approx ADD_{\nu,\theta}^{app}(T_a) = \frac{2(1-\theta^2)a}{\theta^2}. \quad (11)$$

In the MC simulations, we set

$$\Theta = \{-0.9, -0.8, -0.7, -0.6, -0.5, -0{,}4, -0.3, -0.2, -0.1, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$$

and uniform prior $W(\theta_j) = 1/18$. The results are presented in Table 1 for the upper bound on the maximal local probability of false alarm (LPFA) $\beta = 0.01$ and the number of MC runs 10^5. In the table, we compare operating characteristics of the MiSR rule T_a with that of the SR rule

$$\tau_B = \inf\{n \ge 1 : R_n(\theta) \ge B\}.$$

The thresholds a and B (shown in the table) were selected in such a way that the maximal probabilities of false alarm of both rules ($LPFA(T_a)$ and $LPFA(T_B^*)$) were practically the same.

Table 1: Operating Characteristics of the MiSR and SR Detection Procedures

$\beta = 0.01$, $\nu = 0$			
θ	e^a	$ADD_{\nu,\theta}(T_a)$	$LPFA(T_a)$
0.9	395	11.74	0.0080
0.8	420	14.72	0.0073
0.7	440	18.97	0.0070
0.6	470	25.32	0.0065
0.5	595	36.35	0.0049
0.4	1040	59.57	0.0024

$\nu = 0$, $B = 791$			
θ	$ADD_{\nu,\theta}(\tau_B)$	$LPFA(\tau_B)$	$ADD_{\nu,\theta}^{app}(T_a)$
0.9	11.08	0.0079	2.81
0.8	13.72	0.0073	6.80

0.7	17.52	0.0071	12.67
0.6	23.15	0.0065	21.88
0.5	31.84	0.0049	38.33
0.4	45.88	0.0025	72.94

$\beta = 0.01$, $\nu = 10$			
θ	e^a	$ADD_{\nu,\theta}(T_a)$	$LPFA(T_a)$
0.9	395	10.05	0.0080
0.8	420	12.72	0.0073
0.7	440	16.59	0.0070
0.6	470	22.55	0.0065
0.5	595	32.96	0.0049
0.4	1040	55.34	0.0024

$\nu = 10$, $B = 791$			
θ	$ADD_{\nu,\theta}(\tau_B)$	$LPFA(\tau_B)$	$ADD^{app}_{\nu,\theta}(T_a)$
0.9	9.62	0.0079	2.81
0.8	11.98	0.0073	6.80
0.7	15.30	0.0071	12.67
0.6	20.34	0.0065	21.88
0.5	28.01	0.0049	38.33
0.4	40.83	0.0025	72.94

It is seen that for relatively large values of the post-change parameter, $\theta \geq 0.6$, the SR rule only slightly outperforms the MiSR rule, but for small parameter values (i.e., for close hypotheses) the difference becomes quite substantial. The worst change point is $\nu = 0$, as expected. Also, the first-order approximation (11) is not too accurate, especially for small and large parameter values.

5. REMARKS

1. Despite the fact that the MiSR procedure is first-order asymptotically optimal for practically arbitrary weight function $W(\theta)$ that satisfies condition ($\mathbf{C}_W$), for practical purposes its choice may be important. In fact, selection of the weight W affects higher-order asymptotic performance, and therefore, real performance of the detection procedure. For example, if the set Θ is continuous, one has to avoid $W(\theta)$ that concentrates in the vicinity of a specific parameter value θ_1 since in this case the MiSR procedure will be nearly optimal at and in the vicinity of θ_1 but will not have a good performance for other parameter values. The choice of $W(\theta)$ is also related to the computational issue. It is reasonable to select the weight as to be in the class of conjugate priors, if possible, or to select a uniform prior if Θ is compact. A substantial simplification occurs when $\Theta = \{\theta_1, .., \theta_N\}$ is a finite discrete set. If the observations are i.i.d., then in the discrete case, it is possible to find an optimal (in a certain sense) weight using the approach proposed by Fellouris and Tartakovsky (2013) for the hypothesis testing problem.

2. The traditional constraint on the false alarm risk in minimax changepoint detection problems is the lower bound on the average run length to false alarm (ARL2FA) $\mathbf{E}_\infty[\tau] \geq \gamma \geq 1$. This measure of false alarms makes sense when the distribution of the stopping time τ (in our case of T_a) is approximately geometric. This is typically the case (at least asymptotically as $a \to \infty$) for i.i.d. data models [Pollak and Tartakovsky (2009), Yakir (1995)]. However, apart for the i.i.d. case there is no any result on the asymptotic distribution of the stopping time T_a (as $a \to \infty$), so for general non-i.i.d. models of interest in the present paper this is not necessarily true. Therefore, the usefulness of the ARL2FA is under the question, as discussed in detail in Tartakovsky et al. (2015). In fact, in general, large values of the ARL2FA do not guarantee small values of the maximal local PFA $\sup_{k \geq 1} \mathbf{P}_\infty(\tau < k + m \mid \tau \geq k)$. But the opposite is always true since the maximal local PFA is a more stringent false alarm measure in the sense that if it is small, then the ARL2FA is necessarily large. This argument motivated us considering the maximal local PFA instead of the conventional ARL2FA.

ACKNOWLEDGMENTS

This work was supported by the RSF grant 17-11-01049 (National Research Tomsk State University), by the research project no. 2.3208.2017/4.6 (the Ministry of Education and Science of the Russian Federation). The work of Alexander Tartakovsky was partially supported by the Russian Science Foundation under grant 18-19-00452.

REFERENCES

Basseville, M. and Nikiforov, I.V, 1993. Detection of Abrupt Changes: Theory and Applications. Englewood Cliffs: Prentice Hall.

Girardin, V., Konev, V. and Pergamenshchikov, S. M., 2018. Kullback-Leibler Approach to CUSUM Quickest Detection Rule for Markovian Time Series. Sequential Analysis, 37 (3), 322-341.

Fellouris, G., Tartakovsky, A. G., 2013. Almost optimal sequential tests of discrete composite hypotheses. Statistica Sinica, 23, 1717-1741.

Kent, S., 2000. On the trial of intrusions into information systems. IEEE Spectrum, 37, Issue 12, 52-56.

Lai, T.L., 1998. Information bounds and quick detection of parameters changes in stochastic systems. IEEE Transactions on Information Theory, 44(7), 2917-2929.

Page, E.S., 1954. Continuous inspection schemes. Biometrika, 41, 100-115.

Pollak, M., Tartakovsky, A., 2009. Optimality properties of the Shiryaev-Roberts procedure. Statistica Sinica, 19, 1729-1739.

Pergamenshchikov, S. M., Tartakovsky, A.G., 2018. Asymptotically optimal pointwise and minimax quickest change-point detection for dependent

data. Statistical inference for stochastic processes, 21 (1), 217-259.
Tartakovsky, A.G., Rozovskii, B.L., Blažek, R.B., and Kim, H., 2006. Detection of intrusions in information systems by sequential change-point methods. Statistical Methodology, 3, 252-293.
Tartakovsky, A., Nikiforov, I. and Basseville, M. Sequential Analysis: Hypothesis Testing and Changepoint Detection. Chapman & Hall book, Taylor & Francis Group, 2015.
Yakir, B., 1994. Optimal detection of a change in distribution when observations form a Markov chain with a finite state space. In: Change-Point problems (E. Carlstein, H. Muller and D. Siegmund, Eds), Hayward, CA: Inst. Math. Statist., 346-358.
Yakir, B., 1995. A note on the run length to false alarm of a change-point detection policy. Annals of Statistics, 23, 272-281.

S.M. Pergamenshchikov is professor at Laboratory of Mathematics Rafael Salem of University of Rouen and scientific director of International Laboratory of Statistics of Stochastic processes and Quantitative Finance of Russian National Research Tomsk State University. He works in three fields: statistics, probability and financial mathematics. In statistics his research interests include sequential analysis, nonparametric estimation, model selection, signal and image processing. In probability this is asymptotic analysis for stochastic differential equations, renewal theory for general Markov processes, uniform geometric ergodicity and concentration inequalities. In financial mathematics it is asymptotic hedging problems with transaction costs, optimal investment and consumption and ruin probability problems. During the research work S.M. Pergamenshchikov published more than 100 papers and one monograph in Springer, 72 of them being indexed in the MathSciNet, Web of Science, Scopus databases. He has publications in main high rated journals such as “Annals of Probability”, “Annals of Applied Probability”, “Annals of Statistics”, “Finance and Stochastics”, etc. He obtained several awards, including “Russian Federal Professor in Mathematics” (2016), “Scientific Excellence Award” (2011) in France. He obtained Ph.D (1986) in the Institute of Mathematics, USSR Academy of Sciences in Novosibirsk and Dr. Sci. (1994) in the Institute of Mathematics and Mechanics of Russian Academy of Sciences in Ekaterinburg.

A.G. Tartakovsky is a head of Space Informatics Laboratory at the Moscow Institute of Physics and Technology. His research interests include theoretical and applied statistics; applied probability; sequential analysis; changepoint detection phenomena; and a variety of applications including statistical image and signal processing; video surveillance and object detection and tracking; information integration/fusion; intrusion detection and network security; detection and tracking of malicious activity; and mathematical/engineering finance applications. He is the author of two books (the third is in preparation), several book chapters, and over 150 conference and journal papers. Dr. Tartakovsky is a Fellow of the Institute of Mathematical Statistics and a senior member of IEEE. He obtained several awards, including a 2007 Abraham Wald Award in Sequential Analysis. Dr. Tartakovsky obtained Ph.D. (1981) and Dr. Sci. (1990) degrees from the Moscow Institute of Physics and Technology, Russia (FizTech).