

REPORT DOCUMENTATION PAGE			Form Approved OMB NO. 0704-0188		
The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA, 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.					
1. REPORT DATE (DD-MM-YYYY) 19-01-2016		2. REPORT TYPE Final Report		3. DATES COVERED (From - To) 1-Jun-2014 - 30-Nov-2015	
4. TITLE AND SUBTITLE Final Report: General Multidecision Theory: Hypothesis Testing and Changepoint Detection-Classification with Applications to Homeland Security			5a. CONTRACT NUMBER W911NF-14-1-0246		
			5b. GRANT NUMBER		
			5c. PROGRAM ELEMENT NUMBER 611102		
6. AUTHORS Alexander Tartakovsky			5d. PROJECT NUMBER		
			5e. TASK NUMBER		
			5f. WORK UNIT NUMBER		
7. PERFORMING ORGANIZATION NAMES AND ADDRESSES University of Connecticut - Storrs 438 Whitney Road Ext., Unit 1133 Storrs, CT 06269 -1133			8. PERFORMING ORGANIZATION REPORT NUMBER		
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES) U.S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211			10. SPONSOR/MONITOR'S ACRONYM(S) ARO		
			11. SPONSOR/MONITOR'S REPORT NUMBER(S) 65367-MA.5		
12. DISTRIBUTION AVAILABILITY STATEMENT Approved for Public Release; Distribution Unlimited					
13. SUPPLEMENTARY NOTES The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.					
14. ABSTRACT We have addressed all objectives planned in the proposal. First, we proved asymptotic optimality of the Generalized SLRT and the Adaptive SLRT for testing multiple composite hypotheses and very general non-iid stochastic models as the probabilities of errors become small. The results are indeed very general and include Markov, hidden Markov, state-space, and autoregression models as particular cases. Second, we developed computationally efficient and nearly optimal tests for detecting unstructured and structured patterns in multi-stream (sensor, channel) systems assuming that data between channels are mutually independent but may be of a very					
15. SUBJECT TERMS asymptotic optimality of Generalized SLRT and Adaptive SLRT					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT UU	b. ABSTRACT UU	c. THIS PAGE UU			Alexander Tartakovsky
					19b. TELEPHONE NUMBER 860-486-5804

Report Title

Final Report: General Multidecision Theory: Hypothesis Testing and Changepoint Detection-Classification with Applications to Homeland Security

ABSTRACT

We have addressed all objectives planned in the proposal. First, we proved asymptotic optimality of the Generalized SLRT and the Adaptive SLRT for testing multiple composite hypotheses and very general non-iid stochastic models as the probabilities of errors become small. The results are indeed very general and include Markov, hidden Markov, state-space, and autoregression models as particular cases. Second, we developed computationally efficient and nearly optimal tests for detecting unstructured and structured patterns in multi-stream (sensor, channel) systems assuming that data between channels are mutually independent but may be of a very general non-iid structure in channels, and that the number of affected channels is unknown and may vary from small to large. Third, we developed a general Bayesian theory of quickest changepoint detection for general non-iid stochastic models assuming a certain stability of the log-likelihood ratio (LLR) process expressed via the r -complete convergence of the LLR to a finite and positive number which can be regarded as the Kullback–Leibler information number. Fourth, we developed a similar minimax change detection theory modifying and relaxing previous results of Lai (1998) to complete convergence of the LLR and considering novel classes of detection procedures that confine local maximal conditional probability of a false alarm.

Enter List of papers submitted or published that acknowledge ARO support from the start of the project to the date of this printing. List the papers, including journal references, in the following categories:

(a) Papers published in peer-reviewed journals (N/A for none)

Received

Paper

TOTAL:

Number of Papers published in peer-reviewed journals:

(b) Papers published in non-peer-reviewed journals (N/A for none)

Received

Paper

TOTAL:

Number of Papers published in non peer-reviewed journals:

(c) Presentations

Number of Presentations: 0.00

Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received Paper

TOTAL:

Number of Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received Paper

TOTAL:

Number of Peer-Reviewed Conference Proceeding publications (other than abstracts):

(d) Manuscripts

Received Paper

07/11/2014	2.00	Aleksey S. Polunchenko, Grigory Sokolov, Alexander G. Tartakovsky. Optimal Design and Analysis of the ExponentiallyWeighted Moving Average Chart for Exponential Data, Sri Lankan Journal of Applied Statistics (06 2014)
07/11/2014	3.00	Alexander G. Tartakovsky. NEARLY OPTIMAL SEQUENTIAL TESTS OF COMPOSITEHYPOTHESES – REVISITED, Proceedings of the Steklov Institute of Mathematics (Invited Paper for Special Issue in honor of the 80th birthday of Professor Albert Shiryaev) (05 2014)

TOTAL: 2

Number of Manuscripts:

Books

Received Book

07/11/2014 1.00 Alexander Tartakovsky, Igor Nikiforov, Michele Basseville. Sequential Analysis: Hypothesis Testing and
Changepoint Detection, Boca Raton, FL: Chapman & Hall/CRC, (08 2014)

TOTAL: 1

Received Book Chapter

TOTAL:

Patents Submitted

Patents Awarded

Awards

Graduate Students

<u>NAME</u>	<u>PERCENT SUPPORTED</u>
FTE Equivalent:	
Total Number:	

Names of Post Doctorates

NAME

PERCENT SUPPORTED

FTE Equivalent:

Total Number:

Names of Faculty Supported

NAME

PERCENT SUPPORTED

FTE Equivalent:

Total Number:

Names of Under Graduate students supported

NAME

PERCENT SUPPORTED

FTE Equivalent:

Total Number:

Student Metrics

This section only applies to graduating undergraduates supported by this agreement in this reporting period

The number of undergraduates funded by this agreement who graduated during this period: 0.00

The number of undergraduates funded by this agreement who graduated during this period with a degree in science, mathematics, engineering, or technology fields:..... 0.00

The number of undergraduates funded by your agreement who graduated during this period and will continue to pursue a graduate or Ph.D. degree in science, mathematics, engineering, or technology fields:..... 0.00

Number of graduating undergraduates who achieved a 3.5 GPA to 4.0 (4.0 max scale):..... 0.00

Number of graduating undergraduates funded by a DoD funded Center of Excellence grant for Education, Research and Engineering:..... 0.00

The number of undergraduates funded by your agreement who graduated during this period and intend to work for the Department of Defense 0.00

The number of undergraduates funded by your agreement who graduated during this period and will receive scholarships or fellowships for further studies in science, mathematics, engineering or technology fields:..... 0.00

Names of Personnel receiving masters degrees

NAME

Total Number:

Names of personnel receiving PHDs

NAME

Total Number:

Names of other research staff

NAME

PERCENT SUPPORTED

FTE Equivalent:

Total Number:

Sub Contractors (DD882)

Inventions (DD882)

Scientific Progress

See attached.

Technology Transfer

N/A

**GENERAL MULTIDECISION THEORY:
HYPOTHESIS TESTING AND CHANGEPOINT DETECTION
WITH APPLICATIONS TO HOMELAND SECURITY**

**FINAL TECHNICAL REPORT
GRANT # W911NF-14-1-0246
REPORT # FTR-1-16
DATES COVERED: 05/19/2014–01/18/2016**

**AUTHOR:
Alexander Tartakovsky**

PERFORMING ORGANIZATION:

**Department of Statistics
University of Connecticut
Storrs, CT 06269-4120**

Principal Investigator: Alexander Tartakovsky

SPONSORING/MONITORING AGENCY:

**The U.S. Army Research Office
Mathematical Sciences Division
P.O. Box 12211
RTP, NC 27709**

POC: Dr. Joseph Myers

Approved for Public Release; Distribution Unlimited

Contents

1	SUMMARY OF ADDRESSED TASKS AND ACCOMPLISHMENTS	3
2	MAIN RESULTS	3
2.1	Asymptotic Optimality Properties of the Multihypothesis Sequential Tests	3
2.1.1	The Multihypothesis Generalized Sequential Likelihood Ratio Test	4
2.1.2	Near Optimality of the GSLRT	5
2.2	Detection of Structured and Unstructured Patterns in Multiple Data Streams	7
2.2.1	Asymptotic Optimality of the G-SLRT	10
2.2.2	Asymptotic Optimality of the M-SLRT	14
2.2.3	Feasibility	15
2.3	Asymptotic Bayesian Theory of Quickest Changepoint Detection	16
2.3.1	A General Changepoint Model	18
2.3.2	Optimality Criteria	19
2.3.3	Change Detection Procedures	19
2.3.4	r -Quick Convergence Versus r -Complete Convergence	21
2.3.5	Asymptotic Operating Characteristics and Optimality of the Shiryaev Procedure	22
2.4	Asymptotic Pointwise and Minimax Theory of Quickest Changepoint Detection	24
2.4.1	Novel Optimality Criteria	26
2.4.2	Asymptotic Optimality of the SR Procedure	27
3	POTENTIAL IMPACTS	29
4	SCIENTIFIC PERSONNEL SUPPORTED BY THIS PROJECT	30
5	REQUIRED NUMERICAL DATA RELATED TO THIS GRANT	30
	REFERENCES	34

1. SUMMARY OF ADDRESSED TASKS AND ACCOMPLISHMENTS

We have addressed all objectives planned in the proposal. First, we proved asymptotic optimality of the Generalized SLRT and the Adaptive SLRT for testing multiple composite hypotheses and very general non-iid stochastic models as the probabilities of errors become small. The results are indeed very general and include Markov, hidden Markov, state-space, and autoregression models as particular cases. Second, we developed computationally efficient and nearly optimal tests for detecting unstructured and structured patterns in multi-stream (sensor, channel) systems assuming that data between channels are mutually independent but may be of a very general non-iid structure in channels, and that the number of affected channels is unknown and may vary from small to large. Third, we developed a general Bayesian theory of quickest changepoint detection for general non-iid stochastic models assuming a certain stability of the log-likelihood ratio (LLR) process expressed via the r -complete convergence of the LLR to a finite and positive number which can be regarded as the Kullback–Leibler information number. Fourth, we developed a similar minimax change detection theory modifying and relaxing previous results of Lai (1998) to complete convergence of the LLR and considering novel classes of detection procedures that confine local maximal conditional probability of a false alarm.

2. MAIN RESULTS

2.1. Asymptotic Optimality Properties of the Multihypothesis Sequential Tests

Consider the following problem of testing multiple composite hypotheses associated with general non-iid stochastic models. Let $(\Omega, \mathcal{F}, \mathcal{F}_n, P_\theta)$, $n = 1, 2, \dots$, be a filtered probability space with standard assumptions about monotonicity of the σ -algebras $\mathcal{F}_n$. The vector parameter $\theta = (\theta_1, \dots, \theta_\ell)$ belongs to a subset $\tilde{\Theta}$ of ℓ -dimensional Euclidean space. The sub- σ -algebra $\mathcal{F}_n = \mathcal{F}_n^X = \sigma(\mathbf{X}_1^n)$ of $\mathcal{F}$ is generated by the stochastic process $\mathbf{X}_1^n = (X_1, \dots, X_n)$ observed up to time n . The hypotheses to be tested are “ $H_i : \theta \in \Theta_i$ ”, $i = 0, 1, \dots, N$ ($N \geq 1$), where Θ_i are disjoint subsets of $\tilde{\Theta}$. We will also suppose that there is an *indifference* zone $I_{\text{in}} \in \tilde{\Theta}$ in which there are no constraints on the probabilities of errors imposed. The indifference zone, where any decision is acceptable, is usually introduced keeping in mind that the correct action is not critical and often not even possible when the hypotheses are too close, which is perhaps the case in most, if not all, practical applications. However, in principle I_{in} may be an empty set. The probability measures P_θ and $P_{\tilde{\theta}}$ are assumed to be locally mutually absolutely continuous. By $p_\theta(X_n | \mathbf{X}_1^{n-1})$, $n \geq 1$ we denote corresponding conditional densities which may depend on n .

A multihypothesis sequential test $\delta = (T, d)$ consists of the pair (T, d) , where T is a stopping time with respect to the filtration $\{\mathcal{F}_n\}_{n \geq 0}$, and $d = d_T(\mathbf{X}_1^T) \in \{0, 1, \dots, N\}$ is an $\mathcal{F}_T$ -measurable (terminal) decision rule specifying which hypothesis is to be accepted once observations have stopped. Specifically, the hypothesis H_i is accepted if $d = i$ and rejected if $d \neq i$, i.e., $\{d = i\} = \{T < \infty, \delta \text{ accepts } H_i\}$. The quality of a sequential test is judged on the basis of its error probabilities and expected sample sizes or more generally on the moments of the sample size. Let $\alpha_{ij}(\delta, \theta) = P_\theta(d = j), \theta \in \Theta_i$ ($i \neq j, i, j = 0, 1, \dots, N$) be the probability of accepting the hypothesis H_j by the test δ when the true value of the parameter θ is fixed and belongs to the subset Θ_i and let $\beta_i(\delta, \theta) = P_\theta(d \neq i), \theta \in \Theta_i$ be the probability of rejecting the hypotheses H_i

when it is true. Introduce the following two classes of tests

$$\begin{aligned}\mathbb{C}([\alpha_{ij}]) &= \left\{ \delta : \sup_{\boldsymbol{\theta} \in \Theta_i} \alpha_{ij}(\delta, \boldsymbol{\theta}) \leq \alpha_{ij} \text{ for all } i, j = 0, 1, \dots, N, i \neq j \right\}, \\ \mathbb{C}(\boldsymbol{\beta}) &= \left\{ \delta : \sup_{\boldsymbol{\theta} \in \Theta_i} \beta_i(\delta, \boldsymbol{\theta}) \leq \beta_i \text{ for all } i = 0, 1, \dots, N \right\},\end{aligned}\tag{1}$$

for which maximal error probabilities do not exceed the given numbers α_{ij} and β_i .

The goal is to find tests that are nearly (asymptotically) optimal as $\alpha_{ij} \rightarrow 0$ and $\beta_i \rightarrow 0$ in the sense of minimizing the expected sample size $E_{\boldsymbol{\theta}}T$ or more generally higher moments of the stopping time $E_{\boldsymbol{\theta}}T^m$, $m \geq 1$ for all parameter values $\boldsymbol{\theta} \in \Theta$.

In the IPR for the grant at USC Tartakovsky (2013a), we designed an adaptive matrix sequential likelihood ratio test (AMSLRT) based on one-stage delayed estimators of the unknown parameters and proved its asymptotic optimality assuming the strong law of large numbers (SLLN) for the log-likelihood ratio (LLR) processes. The advantage of this adaptive test over the generalized sequential likelihood ratio test (GSLRT), which we consider below, is that the error probabilities are easily controlled (upper-bounded). However, obviously the AMSLRT is inferior to the GSLRT since there is loss of information at each stage, and this is expected to influence its performance degradation especially in the vector case where the dimensionality of the parameter ℓ is relatively large.

Below we show that the GSLRT is also asymptotically optimal.

2.1.1. The Multihypothesis Generalized Sequential Likelihood Ratio Test

Define the generalized LR statistics

$$\hat{\Lambda}_n^i = \frac{\sup_{\boldsymbol{\theta} \in \Theta} \prod_{k=1}^n p_{\boldsymbol{\theta}}(X_k | \mathbf{X}_1^{k-1})}{\sup_{\boldsymbol{\theta} \in \Theta_i} \prod_{k=1}^n p_{\boldsymbol{\theta}}(X_k | \mathbf{X}_1^{k-1})} = \frac{\prod_{k=1}^n p_{\boldsymbol{\theta}_n^*}(X_k | \mathbf{X}_1^{k-1})}{\sup_{\boldsymbol{\theta} \in \Theta_i} \prod_{k=1}^n p_{\boldsymbol{\theta}}(X_k | \mathbf{X}_1^{k-1})}, \quad i = 0, 1, \dots, N, \tag{2}$$

where $\boldsymbol{\theta}_n^* = \arg \sup_{\boldsymbol{\theta} \in \Theta} p_{\boldsymbol{\theta}}(\mathbf{X}_1^n)$ is the MLE estimator. The *Multihypothesis Generalized Sequential Likelihood Ratio Test* (MGSLRT) is of the form

$$\text{stop at the first } n \geq 1 \text{ such that for some } i \quad \hat{\Lambda}_n^j \geq A_{ji} \quad \text{for all } j \neq i \tag{3}$$

and accept the (unique) H_i that satisfies these inequalities, where A_{ij} are positive and finite numbers (thresholds).

Note that the MGSLRT $\hat{\delta} = (\hat{T}, \hat{d})$ given by (3) can be also represented as follows:

$$\hat{T} = \min_{0 \leq i \leq N} \hat{T}_i, \quad \hat{d} = i \quad \text{if} \quad \hat{T} = \hat{T}_i, \tag{4}$$

where

$$\hat{T} = \inf \left\{ n \geq 1 : \hat{\ell}_n \geq \max_{\substack{0 \leq j \leq N \\ j \neq i}} [\ell_n^j + a_{ji}] \right\}, \quad a_{ij} = \log A_{ij}, \quad i = 0, 1, \dots, N; \tag{5}$$

$$\hat{\ell}_n = \sum_{k=1}^n \log p_{\boldsymbol{\theta}_n^*}(X_k | \mathbf{X}_1^{k-1}), \quad \ell_n^i = \sup_{\boldsymbol{\theta} \in \Theta_i} \sum_{k=1}^n \log p_{\boldsymbol{\theta}}(X_k | \mathbf{X}_1^{k-1}).$$

2.1.2. Near Optimality of the GSLRT

In the following, we will write $\hat{\alpha}_{ij}(\boldsymbol{\theta}) = \alpha_{ij}(\hat{\delta}, \boldsymbol{\theta})$ and $\hat{\beta}(\boldsymbol{\theta}) = \beta_i(\hat{\delta}, \boldsymbol{\theta})$ for the probabilities of errors of the MGSRLT.

The developed asymptotic hypothesis testing theory is based on the SLLN and rates of convergence in the strong law for the LLR processes, specifically by strengthening the strong law into the r -quick version.

Definition 1. Let P be a probability measure and E the corresponding expectation. For $r > 0$, the random variable Y_n is said to converge P - r -quickly to a constant q if $EL_\varepsilon^r < \infty$ for all $\varepsilon > 0$, where $L_\varepsilon = \sup \{n : |Y_n - q| > \varepsilon\}$ ($\sup \emptyset = 0$).

Note that $P(L_\varepsilon < \infty) = 1$ for all $\varepsilon > 0$ is equivalent to the P -a.s. convergence of Y_n to q .

Define the LLR process

$$\lambda_n(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}}) = \log \frac{dP_{\boldsymbol{\theta}}^n}{dP_{\tilde{\boldsymbol{\theta}}}^n} = \sum_{k=1}^n \log \frac{p_{\boldsymbol{\theta}}(X_k | \mathbf{X}_1^{k-1})}{p_{\tilde{\boldsymbol{\theta}}}(X_k | \mathbf{X}_1^{k-1})}$$

and assume that there exist positive and finite numbers $I(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}})$ such that

$$\frac{1}{n} \lambda_n(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}}) \xrightarrow[n \rightarrow \infty]{P_{\boldsymbol{\theta}} - r - \text{quickly}} I(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}}) \quad \text{for all } \boldsymbol{\theta}, \tilde{\boldsymbol{\theta}} \in \Theta, \boldsymbol{\theta} \neq \tilde{\boldsymbol{\theta}}. \quad (6)$$

In addition, we certainly need some conditions on the behavior of the MLE $\boldsymbol{\theta}_n^*$ for large n , which should converge to the true value $\boldsymbol{\theta}$ in a proper way. To this end, we require the following condition on the generalized LR process:

$$\frac{1}{n} \log \hat{\Lambda}_n(\tilde{\boldsymbol{\theta}}) \xrightarrow[n \rightarrow \infty]{P_{\boldsymbol{\theta}} - r - \text{quickly}} I(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}}) \quad \text{for all } \boldsymbol{\theta}, \tilde{\boldsymbol{\theta}} \in \Theta, \boldsymbol{\theta} \neq \tilde{\boldsymbol{\theta}}, \quad (7)$$

so that the normalized by n LLR tuned to the true parameter value and its estimate converge to the same constants. In certain cases, but not always, conditions (6) and (7) imply the following conditions

$$\frac{1}{n} \log \hat{\Lambda}_n^i \xrightarrow[n \rightarrow \infty]{P_{\boldsymbol{\theta}} - r - \text{quickly}} I_i(\boldsymbol{\theta}) \quad \text{for all } \boldsymbol{\theta} \in \Theta \setminus \Theta_i, i = 0, 1, \dots, N, \quad (8)$$

where $I_i(\boldsymbol{\theta}) = \inf_{\tilde{\boldsymbol{\theta}} \in \Theta_i} I(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}})$ (the minimal “distance” from $\boldsymbol{\theta}$ to the set Θ_i) is assumed to be positive for all i . Write $\alpha_{\max} = \max_{i,j} \alpha_{ij}$ and $\beta_{\max} = \max_i \beta_i$ and define

$$J_i(\boldsymbol{\theta}) = \min_{\substack{0 \leq j \leq N \\ j \neq i}} [I_j(\boldsymbol{\theta})/c_j] \quad \text{for } \boldsymbol{\theta} \in \Theta_i, \quad J(\boldsymbol{\theta}) = \max_{0 \leq i \leq N} J_i(\boldsymbol{\theta}) \quad \text{for } \boldsymbol{\theta} \in I_{\text{in}}, \quad (9)$$

and

$$\begin{aligned} J_i^*(\boldsymbol{\theta}) &= \min_{\substack{0 \leq j \leq N \\ j \neq i}} [I_j(\boldsymbol{\theta})/c_j] \quad \text{for } \boldsymbol{\theta} \in \Theta_i, \\ J^*(\boldsymbol{\theta}) &= \max_{0 \leq i \leq N} \min_{\substack{0 \leq j \leq N \\ j \neq i}} [I_j(\boldsymbol{\theta})/c_j] = \max_{0 \leq i \leq N} J_i^*(\boldsymbol{\theta}) \quad \text{for } \boldsymbol{\theta} \in I_{\text{in}}, \end{aligned} \quad (10)$$

where

$$c_{ij} = \lim_{\alpha_{\max} \rightarrow 0} |\log \alpha_{ij}| / |\log \alpha_{\max}|, \quad c_i = \lim_{\beta_{\max} \rightarrow 0} |\log \beta_i| / |\log \beta_{\max}|.$$

Theorem 2 below establishes uniform asymptotic optimality of the MGSLRT in the general non-iid case with respect to moments of the stopping time distribution. The proof is based on the technique developed by Tartakovsky (1998) for multiple simple hypotheses. It includes a two-step procedure: first we obtain the asymptotic lower bounds for moments of the stopping time distribution $\inf_{\delta \in \mathbb{C}([\alpha_{ij}])} \mathbb{E}_{\theta}[T]^m$, $\theta \in \Theta_i$, $m > 0$, $i = 0, 1, \dots, N$, and then we show that these lower bounds are attained for the MGSLRT. The asymptotic lower bounds are given in the following theorem.

Theorem 1 (Asymptotic Lower Bounds). *Assume that there are positive and finite numbers $I(\theta, \tilde{\theta})$ such that*

$$\frac{1}{n} \lambda_n(\theta, \tilde{\theta}) \xrightarrow[t \rightarrow \infty]{P_{\theta}-a.s.} I(\theta, \tilde{\theta}) \quad \text{for all } \theta, \tilde{\theta} \in \Theta, \theta \neq \tilde{\theta}. \quad (11)$$

Let $I_i(\theta) = \inf_{\tilde{\theta} \in \Theta_i} I(\theta, \tilde{\theta})$ and suppose $\min_{0 \leq i \leq N} I_i(\theta) > 0$. Then, for all $\theta \in \Theta$ and $0 < \varepsilon < 1$,

$$\begin{aligned} \inf_{\delta \in \mathbb{C}([\alpha_{ij}])} P_{\theta} \{T > \varepsilon A_{\theta}([\alpha_{ij}])\} &\rightarrow 1 \quad \text{as } \alpha_{\max} \rightarrow 0, \\ \inf_{\delta \in \mathbb{C}(\beta)} P_{\theta} \{T > \varepsilon A_{\theta}(\beta)\} &\rightarrow 1 \quad \text{as } \beta_{\max} \rightarrow 0, \end{aligned} \quad (12)$$

and therefore, for all $m > 0$ and $\theta \in \Theta$,

$$\begin{aligned} \inf_{\delta \in \mathbb{C}([\alpha_{ij}])} \mathbb{E}_{\theta} T^m &\geq [A_{\theta}([\alpha_{ij}])]^m (1 + o(1)) \quad \text{as } \alpha_{\max} \rightarrow 0, \\ \inf_{\delta \in \mathbb{C}(\beta)} \mathbb{E}_{\theta} T^m &\geq [A_{\theta}(\beta)]^m (1 + o(1)) \quad \text{as } \beta_{\max} \rightarrow 0, \end{aligned} \quad (13)$$

where

$$A_{\theta}([\alpha_{ij}]) = \begin{cases} |\log \alpha_{\max}| / J_i(\theta) & \text{for } \theta \in \Theta_i \text{ and } i = 0, 1, \dots, N \\ |\log \alpha_{\max}| / J(\theta) & \text{for } \theta \in I_{\text{in}}. \end{cases}$$

and

$$A_{\theta}(\beta) = \begin{cases} |\log \beta_{\max}| / J_i^*(\theta) & \text{for } \theta \in \Theta_i \text{ and } i = 0, 1, \dots, N \\ |\log \beta_{\max}| / J^*(\theta) & \text{for } \theta \in I_{\text{in}}. \end{cases}$$

Next, strengthening the SLLN (11) into the r -quick version it can be shown that the lower bounds (13) are attained by the MGSLRT if the thresholds are selected appropriately. The following theorem spells out details.

Theorem 2 (MGSLRT Asymptotic Optimality). *Assume that r -quick convergence conditions (6) and (8) are satisfied.*

(i) *If the thresholds A_{ij} are so selected that $\sup_{\theta \in \Theta_i} \hat{\alpha}_{ij}(\theta) \leq \alpha_{ij}$ and $\log A_{ij} \sim \log(1/\alpha_{ij})$, then for $m \leq r$ as $\alpha_{\max} \rightarrow 0$*

$$\inf_{\delta \in \mathbb{C}([\alpha_{ij}])} \mathbb{E}_{\theta} T^m \sim \mathbb{E}_{\theta} [T^*]^m \sim \begin{cases} [|\log \alpha_{\max}| / J_i(\theta)]^m & \text{for all } \theta \in \Theta_i \text{ and } i = 0, 1, \dots, N \\ [|\log \alpha_{\max}| / J(\theta)]^m & \text{for all } \theta \in I_{\text{in}}, \end{cases} \quad (14)$$

where the functions $J_i(\theta)$, $J(\theta)$ are defined as in (9).

(ii) If the thresholds $A_{ij} = A_i$ are so selected that $\sup_{\theta \in \Theta_i} \hat{\beta}_i(\theta) \leq \beta_i$ and $\log A_i \sim \log(1/\beta_i)$, then for $m \leq r$ as $\beta_{\max} \rightarrow 0$

$$\inf_{\delta \in \mathbb{C}(\beta)} E_{\theta} T^m \sim E_{\theta} [T^*]^m \sim \begin{cases} [\log \beta_{\max} / J_i^*(\theta)]^m & \text{for all } \theta \in \Theta_i \text{ and } i = 0, 1, \dots, N \\ [\log \beta_{\max} / J^*(\theta)]^m & \text{for all } \theta \in I_{\text{in}}, \end{cases} \quad (15)$$

where the functions $J_i^*(\theta)$, $J^*(\theta)$ are defined as in (10).

Consequently, the MGSLRT minimizes asymptotically the moments of the sample size up to order r uniformly for all $\theta \in \Theta$ in the classes of tests $\mathbb{C}([\alpha_{ij}])$ and $\mathbb{C}(\beta)$.

Remark 1. One of the most important issues is to obtain upper bounds and approximations for error probabilities of the MGSLRT. However, we do not know how to upper-bound the error probabilities of the MGSLRT. The reason is that the statistics $\hat{\Lambda}_n^i$ are not likelihood ratios anymore so that the change-of-measure argument (Wald's likelihood ratio identity) cannot be applied. Some asymptotic approximations still can be obtained in the iid case for ℓ -dimensional exponential families using large and moderate deviations:

$$\sup_{\theta \in \Theta_i} P_{\theta}(\hat{d} = j) = \frac{(\log A_{ji})^{\ell/2}}{A_{ji}} + O(1) \quad \text{as } \min_{ij} A_{ij} \rightarrow \infty \quad (16)$$

(cf. Chan and Lai (2000); Lorden (1977)). In the general non-iid case this is still an open problem.

Remark 2. The assertions of Theorem 2 remain true if the normalization by n in (8) is replaced with the normalization by $\psi(n)$, where $\psi(t)$ is an increasing function, $\psi(\infty) = \infty$, in which case $[\log \alpha_{\max} / J_i(\theta)]^m$ in (14) should be replaced with $\Psi([\log \alpha_{\max} / J_i(\theta)]^m)$, where Ψ is inverse to ψ , and similarly in (15).

2.2. Detection of Structured and Unstructured Patterns in Multiple Data Streams

Rapid signal detection in multistream data or multichannel systems is widely applicable. For example, in the medical sphere, decision-makers must quickly detect an epidemic present in only a fraction of hospitals and other sources of data Chang (2003); Sonesson and Bock (2003); Tsui et al. (2012). In environmental monitoring where a large number of sensors cover a given area, decision-makers seek to detect an anomalous behavior, such as the presence of hazardous materials or intruders, that only a fraction of sensors typically capture Fienberg and Shmueli (2005); Rolka et al. (2007). In military defense applications, there is a need to detect an unknown number of targets in noisy observations obtained by radars, sonars or optical sensors that are typically multichannel in range, velocity and space Bakut et al. (1963); Tartakovsky and Brown (2008). In cyber security, there is a need to rapidly detect and localize malicious activity, such as distributed denial-of-service attacks, typically in multiple data streams Szor (2005); Tartakovsky (2014); Tartakovsky et al. (2006a,b). In genomic applications, there is a need to determine intervals of copy number variations, which are short and sparse, in multiple DNA sequences Siegmund (2013).

Motivated by these and other applications, we consider a general sequential detection problem where observations are acquired *sequentially* in a number of data streams. The goal is to *quickly* detect the presence of a signal while controlling the probabilities of false alarms (type-I error) and missed detection (type-II error) below user-specified levels. Two scenarios are of particular interest for applications. The first is when a single signal with an unknown location is distributed

over a relatively small number of channels. For example, this may be the case when detecting an extended target with an unknown location in a sequence of images produced by a very high-resolution sensor. We call this the “structured” case, since there is a certain geometrical structure we can know at least approximately. A different, completely “unstructured” scenario is when an unknown number of “point” signals affect the channels. For example, in many target detection applications, an unknown number of point targets appear in different channels (or data streams), and it is unknown in which channels the signals will appear Tartakovsky (2013c). The multistream sequential detection problem is well-studied only in the case of a single point signal present in one (unknown) data stream Tartakovsky et al. (2003a). However, as mentioned above, in many applications, a signal (or signals) can affect multiple data streams (e.g., when detecting an unknown number of targets in multichannel sensor systems). In fact, the affected subset could be completely unknown (unknown number of signals), or known partially (e.g., knowing its size or an upper bound on its size such as a known maximal number of signals that can appear).

Our goal is to develop a general asymptotic optimality theory *without assuming iid observations in the channels*. Assuming a very general non-iid model, we focus on two multichannel sequential tests, the Generalized Sequential Likelihood Ratio Test (G-SLRT) and the Mixture Sequential Likelihood Ratio Test (M-SLRT), which are based on the maximum and average likelihood ratio over all possibly affected subsets respectively. We impose minimal conditions on the structure of the observations in channels, postulating only a certain asymptotic stability of the corresponding log-likelihood ratio statistics. Specifically, we assume that the suitably normalized log-likelihood ratios in channels almost surely converge to positive and finite numbers, which can be viewed as local limiting Kullback–Leibler information numbers. We additionally show that if the local log-likelihood ratios also have independent increments, both the G-SLRT and the M-SLRT minimize asymptotically not only the expected sample size but also every moment of the sample size distribution as the probabilities of errors vanish. Thus, we extend a result previously shown only in the case of i.i.d. observations and in the special case of a single affected stream Tartakovsky et al. (2003a). In the general case where the local log-likelihood ratios do not have independent increments, we require a certain rate of convergence in the Strong Law of Large Numbers, which is expressed in the form of r -complete convergence (cf. (Tartakovsky et al., 2014b, Ch 2)). Under this condition, we prove that both the G-SLRT and the M-SLRT asymptotically minimize the first r moments of the sample size distribution. The r -complete convergence condition is a relaxation of the r -quick convergence condition used in Tartakovsky et al. (2003a) (in the special case of detecting a single signal in a multichannel system). However, its main advantage is that it is much easier to verify in practice. Finally, we show that both the G-SLRT and the M-SLRT are computationally feasible, even with a large number of channels, when we have an upper and a lower bound on the number of signals, a general set-up that includes cases of complete ignorance as well as cases where the size of the affected subset is known.

Suppose that observations are sequentially acquired over time in N distinct sources (data streams, channels, sensors). We denote the observations in the k^{th} data stream as $X^k := \{X_n^k\}_{n \geq 1}$, $k = 1, \dots, N$. For every k , we assume that either $P^k = P_0^k$ or $P^k = P_1^k$, where P^k is the “true” distribution of X^k and P_1^k and P_0^k are two locally equivalent probability measures on the canonical space of X^k , i.e., $P_1^k \ll P_0^k$ and $P_0^k \ll P_1^k$ when both probability measures are restricted to $\mathcal{F}_t^k = \sigma(X_s^k; 0 \leq s \leq n)$ for some $n \geq 0$. We denote by Λ_n^k the Radon-Nikodým derivative

(likelihood ratio) of P_1^k versus P_0^k given $\mathcal{F}_n^k$ and by Z_n^k the corresponding LLR, i.e.,

$$\Lambda_n^k = \left. \frac{dP_1^k}{dP_0^k} \right|_{\mathcal{F}_n^k} \quad \text{and} \quad Z_n^k = \log \Lambda_n^k.$$

One possible and useful interpretation is that there is “noise” in source k under P_0^k and “signal” and noise otherwise (object/target appearance in noise). Alternatively, one may think about P_0^k as a probability measure corresponding to a “normal” scenario, while P_1^k corresponds to an “abnormal” scenario when the k -th data stream is affected by some event (malicious/unusual activity/behavior in social networks, bio-chemical threat appearance, attacks in computer networks, etc.). We want to test the global null hypothesis $H_0 : P^k = P_0^k, 1 \leq k \leq N$, according to which there is only noise in all data streams, against the alternative that a signal is present in a subset of data streams that belongs to a class $\mathcal{P}$. Thus, the alternative hypothesis takes the form $H_1 := \cup_{\mathcal{A} \in \mathcal{P}} H_1^{\mathcal{A}}$, where the distribution of X^k under $H_1^{\mathcal{A}}$ is

$$P^k = \begin{cases} P_0^k & \text{when } k \notin \mathcal{A} \\ P_1^k & \text{when } k \in \mathcal{A} \end{cases}.$$

Assuming that the observations from different data streams are mutually independent, which will be our standing assumption from now on, the distribution of $X = (X^1, \dots, X^K)$ under H_0 is described by the product measure $P_0 = P_0^1 \times \dots \times P_0^N$. On the other hand, the distribution of X when signal is present in subset $\mathcal{A}$ takes the form

$$P^{\mathcal{A}} = \prod_{k \in \mathcal{A}} P_1^k \times \prod_{k \notin \mathcal{A}} P_0^k.$$

Equivalently, for any given n and subset $\mathcal{A} \in \mathcal{P}$, we have:

$$\Lambda_n^{\mathcal{A}} = \left. \frac{dP^{\mathcal{A}}}{dP_0} \right|_{\mathcal{F}_n} = \prod_{k \in \mathcal{A}} \Lambda_n^k.$$

The goal is to find a pair $\delta = (T, d)$ that consists of an $\{\mathcal{F}_n\}$ -stopping time T and an $\mathcal{F}_T$ -measurable random variable d taking values in $\{0, 1\}$, so that H_i is selected on $\{d = i, T < \infty\}$, $i = 0, 1$, where $\{\mathcal{F}_n\}$ is the filtration generated by all sources of observations, i.e.,

$$\mathcal{F}_n = \bigvee_{1 \leq k \leq N} \mathcal{F}_n^k = \sigma(X_s^k; 0 \leq s \leq n, 1 \leq k \leq N).$$

Specifically, the goal is to find a sequential test that (a) controls type-I and type-II error probabilities below α and β , respectively, i.e., belongs to the class of tests

$$\mathbb{C}_{\alpha, \beta}(\mathcal{P}) = \{\delta : P_0(d = 1) \leq \alpha \text{ and } \sup_{\mathcal{A} \in \mathcal{P}} P^{\mathcal{A}}(d = 0) \leq \beta\},$$

and (b) it is asymptotically optimal as $\alpha, \beta \rightarrow 0$ in the sense that it attains

$$\inf_{(\tau, d) \in \mathbb{C}_{\alpha, \beta}(\mathcal{P})} E_0 T \quad \text{and} \quad \inf_{\delta \in \mathbb{C}_{\alpha, \beta}(\mathcal{P})} E^{\mathcal{A}} T \quad \forall \mathcal{A} \in \mathcal{P}.$$

More generally, we are interested in establishing conditions under which a specific sequential test $\delta_0 = (T_0, d_0)$ is first-order asymptotically optimal with respect to higher moments of the stopping time distribution, i.e., for all $0 < m \leq r$ and some $r > 1$

$$\lim_{\alpha, \beta \rightarrow 0} \frac{\inf_{\delta \in \mathbb{C}_{\alpha, \beta}(\mathcal{P})} \mathbb{E}_0 T^m}{\mathbb{E}_0 T_0^m} = 1 \quad \text{and} \quad \lim_{\alpha, \beta \rightarrow 0} \frac{\inf_{\delta \in \mathbb{C}_{\alpha, \beta}(\mathcal{P})} \mathbb{E}^{\mathcal{A}} T^m}{\mathbb{E}^{\mathcal{A}} T_0^m} = 1 \quad \forall \mathcal{A} \in \mathcal{P}.$$

Of course, the answer to this question depends heavily on the class of alternatives $\mathcal{P}$. We will only assume that there is a lower bound ($\underline{m} \geq 1$) and an upper bound ($\overline{m} \leq N$) on the cardinality of the subset of affected data streams, i.e.,

$$\mathcal{P} = \{\mathcal{A} : \underline{m} \leq |\mathcal{A}| \leq \overline{m}\}. \quad (17)$$

This sequential testing problem is well understood when the signal can be present in at most one data stream ($\overline{m} = 1$). Specifically, in this case, the optimality of the GSLRT was established by Tartakovsky et al. (2003b) under general conditions on the underlying distributions.

In this project, we propose the GSLRT and the Weighted SLRT (WSLRT) that are feasible for a large number of data streams on one hand and asymptotically optimal on the other hand. In addition, error probabilities of these tests can be explicitly controlled.

2.2.1. Asymptotic Optimality of the G-SLRT

We begin with establishing lower bounds for moments of the stopping time distribution. Recall that we consider very general non-iid models for the observations $(X_n^k)_{n \geq 1}$ in “channels,” so the LLR processes Z_n^k , $k = 1, \dots, N$ have no particular structure. However, to obtain some meaningful results certain assumptions have to be made. We formulate these assumptions in the form of a certain stability of the behavior of the LLRs for large n . Specifically, in the following we suppose that there are positive and finite numbers I_0^k and I_1^k such that the normalized LLRs $n^{-1}Z_n^k$, $k = 1, \dots, N$ converge in probability to $-I_0^k$ under $\mathbb{P}_0^k$ and to I_1^k under $\mathbb{P}_1^k$,

$$\frac{1}{n} Z_n^k \xrightarrow[t \rightarrow \infty]{\mathbb{P}_0^k} -I_0^k, \quad \frac{1}{n} Z_n^k \xrightarrow[t \rightarrow \infty]{\mathbb{P}_1^k} I_1^k, \quad k = 1, \dots, N, \quad (18)$$

in which case also

$$\frac{1}{n} Z_n^{\mathcal{A}} \xrightarrow[t \rightarrow \infty]{\mathbb{P}_0} -I_0^{\mathcal{A}}, \quad \frac{1}{n} Z_n^{\mathcal{A}} \xrightarrow[t \rightarrow \infty]{\mathbb{P}_1^{\mathcal{A}}} I_1^{\mathcal{A}},$$

where

$$I_0^{\mathcal{A}} = \sum_{k \in \mathcal{A}} I_0^k \quad \text{and} \quad I_1^{\mathcal{A}} = \sum_{k \in \mathcal{A}} I_1^k. \quad (19)$$

The following theorem establishes asymptotic lower bounds for all positive moments of the stopping time distribution in the class $\mathbb{C}_{\alpha, \beta}(\mathcal{P})$. We write $\alpha_{\max} = \max(\alpha, \beta)$.

Theorem 3. Assume there exist positive and finite numbers I_0^k and I_1^k such that, for all $\varepsilon > 0$ and $k = 1, \dots, N$,

$$\begin{aligned} \lim_{M \rightarrow \infty} \mathbb{P}_1^k \left\{ \frac{1}{M} \max_{1 \leq n \leq M} Z_n^k \geq (1 + \varepsilon) I_1^k \right\} &= 1, \\ \lim_{M \rightarrow \infty} \mathbb{P}_0^k \left\{ \frac{1}{M} \max_{1 \leq n \leq M} (-Z_n^k) \geq (1 + \varepsilon) I_0^k \right\} &= 1. \end{aligned} \quad (20)$$

Then, for all $m > 0$,

$$\begin{aligned} \liminf_{\alpha_{\max} \rightarrow 0} \frac{\inf_{\delta \in \mathbb{C}_{\alpha, \beta}(\mathcal{P})} \mathbb{E}_0 T^m}{|\log \beta|^m} &\geq \left(\frac{1}{\min_{\mathcal{A} \in \mathcal{P}} I_0^{\mathcal{A}}} \right)^m, \\ \liminf_{\alpha_{\max} \rightarrow 0} \frac{\inf_{\delta \in \mathbb{C}_{\alpha, \beta}(\mathcal{P})} \mathbb{E}^{\mathcal{A}} T^m}{|\log \alpha|^m} &\geq \left(\frac{1}{I_1^{\mathcal{A}}} \right)^m. \end{aligned} \quad (21)$$

When $\mathcal{P} = \{\mathcal{A}\}$, i.e., there is no uncertainty regarding the subset of streams in which the signal may be present, the asymptotic lower bounds (21) are attained by the Sequential Probability Ratio Test (SPRT),

$$\tau_{a,b}^{\mathcal{A}} = \inf\{n \geq 1 : Z_n^{\mathcal{A}} \notin (-a, b)\}, \quad d^{\mathcal{A}} = \begin{cases} 1 & \text{when } Z_{\tau^{\mathcal{A}}}^{\mathcal{A}} \geq b \\ 0 & \text{when } Z_{\tau^{\mathcal{A}}}^{\mathcal{A}} \leq -a \end{cases}, \quad (22)$$

under r -quick convergence conditions for the LLRs, which can be deduced from Lai (1981); Tartakovsky (1998); Tartakovsky et al. (2014a). To be specific, for $\varepsilon > 0$, introduce the last entree times

$$L_0^k(\varepsilon) = \sup\{n \geq 1 : |n^{-1} Z_n^k + I_0^k| > \varepsilon\} \quad \text{and} \quad L_1^k(\varepsilon) = \sup\{n \geq 1 : |n^{-1} Z_n^k - I_1^k| > \varepsilon\}$$

($\sup\{\emptyset\} = 0$) and assume that for some $r > 0$,

$$\mathbb{E}_0^k[L_0^k(\varepsilon)]^r < \infty \quad \text{and} \quad \mathbb{E}_1^k[L_1^k(\varepsilon)]^r < \infty, \quad k = 1, \dots, N. \quad (23)$$

According to Definition 1, conditions (23) mean that the normalized LLRs $n^{-1} Z_n^k$, $k = 1, \dots, N$ converge to $-I_0^k$ and I_1^k r -quickly under $\mathbb{P}_0^k$ and $\mathbb{P}_1^k$, respectively.

Obviously, conditions (23) imply the corresponding r -quick convergence of $n^{-1} Z_n^{\mathcal{A}}$:

$$\mathbb{E}_0[L_0^{\mathcal{A}}(\varepsilon)]^r < \infty \quad \text{and} \quad \mathbb{E}^{\mathcal{A}}[L_1^{\mathcal{A}}(\varepsilon)]^r < \infty, \quad (24)$$

where $L_0^{\mathcal{A}}(\varepsilon) = \sup\{n \geq 1 : |n^{-1} Z_n^{\mathcal{A}} + I_0^{\mathcal{A}}| > \varepsilon\}$ and $L_1^{\mathcal{A}}(\varepsilon) = \sup\{t \geq 1 : |n^{-1} Z_n^{\mathcal{A}} - I_1^{\mathcal{A}}| > \varepsilon\}$.

If the thresholds b and a are selected so that $(\tau^{\mathcal{A}}, d^{\mathcal{A}}) \in \mathbb{C}_{\alpha, \beta}(\mathcal{A})$ and $b \sim |\log \alpha|$, $a \sim |\log \beta|$, in particular $b = |\log \alpha|$ and $a = |\log \beta|$, then using (Tartakovsky et al., 2014a, Theorem 3.4.2) yields, for all $0 < m \leq r$ as $\alpha_{\max} \rightarrow 0$,

$$\begin{aligned} \inf_{\delta \in \mathbb{C}_{\alpha, \beta}(\mathcal{A})} \mathbb{E}_0[\tau]^m &\sim \mathbb{E}_0[\tau^{\mathcal{A}}]^m \sim \left(\frac{|\log \beta|}{I_0^{\mathcal{A}}} \right)^m, \\ \inf_{\delta \in \mathbb{C}_{\alpha, \beta}(\mathcal{A})} \mathbb{E}^{\mathcal{A}}[\tau]^m &\sim \mathbb{E}^{\mathcal{A}}[\tau^{\mathcal{A}}]^m \sim \left(\frac{|\log \alpha|}{I_1^{\mathcal{A}}} \right)^m. \end{aligned} \quad (25)$$

When $\mathcal{P}$ is not a singleton, it is natural to apply a generalized likelihood ratio approach and consider the G-SLRT $\hat{\delta}_{a,b} = (\hat{T}_{a,b}, \hat{d})$ given by

$$\begin{aligned} \hat{T}_{a,b} &= \inf\left\{n \geq 1 : \max_{\mathcal{A} \in \mathcal{P}} Z_n^{\mathcal{A}} \notin (-a, b)\right\}, \\ \hat{d} &= \begin{cases} 1 & \text{when } \max_{\mathcal{A} \in \mathcal{P}} Z_{\hat{\tau}}^{\mathcal{A}} \geq b \\ 0 & \text{when } \max_{\mathcal{A} \in \mathcal{P}} Z_{\hat{\tau}}^{\mathcal{A}} \leq -a \end{cases}. \end{aligned} \quad (26)$$

This test was considered by Tartakovsky et al. (2003b) where its asymptotic optimality was established in the special case that signal can be present in only a single data stream, i.e., $\mathcal{P} = \{\mathcal{A} : |\mathcal{A}| = 1\}$. Theorem 4 below is a generalization of this result for an arbitrary class of alternatives $\mathcal{P}$.

The following lemma gives upper bounds on the error probabilities of the G-SLRT, which suggest threshold values that guarantee the target error probabilities. This lemma does not require any assumptions on the local distributions. Let $|\mathcal{P}| = C_N$ denote the cardinality of class $\mathcal{P}$, i.e., the number of possible alternatives in $\mathcal{P}$. Note that $|\mathcal{P}|$ takes its maximum value when there is no prior information regarding the subset of affected channels ($\overline{\mathcal{P}}_N$), in which case $|\mathcal{P}| = 2^N - 1$.

Lemma 1. *For any thresholds $a, b > 0$,*

$$\mathbf{P}_0(\hat{d} = 1) \leq |\mathcal{P}| e^{-b} \quad \text{and} \quad \max_{\mathcal{A} \in \mathcal{P}} \mathbf{P}^{\mathcal{A}}(\hat{d} = 0) \leq e^{-a}. \quad (27)$$

Therefore, for any target error probabilities $\alpha, \beta \in (0, 1)$, we can guarantee that $(\hat{\tau}, \hat{d}) \in \mathbb{C}_{\alpha, \beta}(\mathcal{P})$ when thresholds are selected as

$$b = |\log(\alpha/|\mathcal{P}|)| \quad \text{and} \quad a = |\log \beta|. \quad (28)$$

Theorem 4. *Let the thresholds b and a in the GSLRT (26) be chosen so that $\hat{\delta}_{a,b} \in \mathbb{C}_{\alpha, \beta}(\mathcal{P})$ and $b \sim |\log \alpha|$, $a \sim |\log \beta|$ as $\alpha_{\max} \rightarrow 0$, in particular $b = |\log \alpha/|\mathcal{P}||$ and $a = |\log \beta|$. If, for some $r > 0$, the conditions (23) hold, i.e.,*

$$\frac{1}{n} Z_n^k \xrightarrow[t \rightarrow \infty]{\mathbf{P}_1^k - r - \text{quickly}} I_1^k \quad \text{and} \quad \frac{1}{n} Z_n^k \xrightarrow[t \rightarrow \infty]{\mathbf{P}_0^k - r - \text{quickly}} -I_0^k, \quad k = 1, \dots, N, \quad (29)$$

then, for any class of alternatives $\mathcal{P}$ and all $0 < m \leq r$ as $\alpha_{\max} \rightarrow 0$,

$$\mathbf{E}_0 \hat{T}^m \sim \left(\frac{|\log \beta|}{\min_{\mathcal{A} \in \mathcal{P}} I_0^{\mathcal{A}}} \right)^m \sim \inf_{\delta \in \mathbb{C}_{\alpha, \beta}(\mathcal{P})} \mathbf{E}_0 T^m, \quad (30)$$

and for every $\mathcal{A} \in \mathcal{P}$,

$$\mathbf{E}^{\mathcal{A}} \hat{T}^m \sim \left(\frac{|\log \alpha|}{I_1^{\mathcal{A}}} \right)^m \sim \inf_{\delta \in \mathbb{C}_{\alpha, \beta}(\mathcal{P})} \mathbf{E}^{\mathcal{A}} T^m. \quad (31)$$

Definition 2. Let $r > 0$. We say that the sequence $(Y_n)_{n \geq 1}$ converges r -completely under probability measure $\mathbf{P}$ to a constant q as $n \rightarrow \infty$ and write

$$Y_n \xrightarrow[n \rightarrow \infty]{\mathbf{P} - r - \text{completely}} q$$

if

$$\sum_{n=1}^{\infty} n^{r-1} \mathbf{P}(|Y_n - q| > \varepsilon) < \infty \quad \text{for all } \varepsilon > 0.$$

This condition turns out to be weaker than the corresponding r -quick convergence (in general), and more importantly it is easier to check the complete convergence condition than r -quick condition. Therefore, as a next step, it is natural to replace conditions (29) with the corresponding r -complete convergence conditions for the LLRs:

$$\frac{1}{n}Z_n^k \xrightarrow[n \rightarrow \infty]{\mathbb{P}_1^{k-r}\text{-completely}} I_1^k \quad \text{and} \quad \frac{1}{n}Z_n^k \xrightarrow[n \rightarrow \infty]{\mathbb{P}_0^{k-r}\text{-completely}} -I_0^k, \quad k = 1, \dots, K, \quad (32)$$

i.e., that for all $\varepsilon > 0$ and all $k = 1, \dots, N$,

$$\sum_{n=1}^{\infty} n^{r-1} \mathbb{P}_1^k \left(\left| \frac{1}{n}Z_n^k - I_1^k \right| > \varepsilon \right) < \infty, \quad \sum_{n=1}^{\infty} n^{r-1} \mathbb{P}_0^k \left(\left| \frac{1}{n}Z_n^k + I_0^k \right| > \varepsilon \right) < \infty. \quad (33)$$

The following theorem spells out details.

Theorem 5. *Let the thresholds b and a in the GSLRT (26) be chosen so that $\hat{\delta}_{a,b} \in \mathbb{C}_{\alpha,\beta}(\mathcal{P})$ and $b \sim |\log \alpha|$, $a \sim |\log \beta|$ as $\alpha_{\max} \rightarrow 0$, in particular $b = |\log \alpha|/|\mathcal{P}|$ and $a = |\log \beta|$. If, for some $r > 0$, the r -complete convergence conditions (32) hold, then, for any class of alternatives $\mathcal{P}$ and all $0 < m \leq r$ as $\alpha_{\max} \rightarrow 0$,*

$$\mathbb{E}_0 \hat{T}^m \sim \left(\frac{|\log \beta|}{\min_{\mathcal{A} \in \mathcal{P}} I_0^{\mathcal{A}}} \right)^m \sim \inf_{\delta \in \mathbb{C}_{\alpha,\beta}(\mathcal{P})} \mathbb{E}_0 T^m, \quad (34)$$

and for every $\mathcal{A} \in \mathcal{P}$,

$$\mathbb{E}^{\mathcal{A}} \hat{T}^m \sim \left(\frac{|\log \alpha|}{I_1^{\mathcal{A}}} \right)^m \sim \inf_{\delta \in \mathbb{C}_{\alpha,\beta}(\mathcal{P})} \mathbb{E}^{\mathcal{A}} T^m. \quad (35)$$

We now consider a special case where the LLR increments $\ell_n^k = Z_n^k - Z_{n-1}^k$, $n \geq 1$ in the k^{th} channel are independent, *but not necessarily identically distributed*, random variables, and show that the asymptotic optimality properties (34)–(35) hold true for any positive integer m , as long as only the SLLN holds, i.e., as long as the almost sure convergence conditions

$$\frac{1}{n}Z_n^k \xrightarrow[n \rightarrow \infty]{\mathbb{P}_1^{k-\text{a.s.}}} I_1^k \quad \text{and} \quad \frac{1}{n}Z_n^k \xrightarrow[n \rightarrow \infty]{\mathbb{P}_0^{k-r-\text{a.s.}}} -I_0^k, \quad k = 1, \dots, K, \quad (36)$$

are satisfied. To this end, we need the following renewal theorem.

Lemma 2. *Let $\xi^k := (\xi_t^k)_{t \geq 1}$, $1 \leq k \leq N$ be (possibly dependent) sequences of random variables on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let $\mathbb{E}$ be the corresponding expectation. Define the stopping time*

$$\nu(b) := \inf \left\{ t \geq 1 : \min_{1 \leq k \leq N} S_t^k > b \right\}; \quad S_t^k := \sum_{u=1}^t \xi_u^k.$$

Suppose that for every $1 \leq k \leq N$ there is a positive constant μ_k such that $S_t^k/t \xrightarrow{\text{a.s.}} \mu_k$. Then, as $b \rightarrow \infty$ we have

$$\frac{\nu(b)}{b} \xrightarrow{\text{a.s.}} \left(\min_{1 \leq k \leq N} \mu_k \right)^{-1}.$$

Moreover, the convergence holds in $\mathcal{L}^r$ for every $r > 0$, if each ξ^k is a sequence of independent random variables and there is a $\lambda \in (0, 1)$ such that

$$\sup_{t \geq 1} \mathbb{E} \left[\exp \{ \lambda (\xi_t^k)^- \} \right] < \infty. \quad (37)$$

The following theorem establishes a stronger asymptotic optimality property for the G-SLRT in the case of LLRs with independent increments.

Theorem 6. *Let $\mathcal{P}$ be an arbitrary class of possibly affected subsets of channels and suppose that the thresholds in the G-SLRT are selected so that $\hat{\delta}_{a,b} \in \mathbb{C}_{\alpha,\beta}(\mathcal{P})$ and $b \sim |\log \alpha|$, $a \sim |\log \beta|$ as $\alpha_{\max} \rightarrow 0$, in particular $b = |\log \alpha|/|\mathcal{P}|$ and $a = |\log \beta|$. If the LLR increments, $\{\ell_n^k\}_{n \geq 1}$, are independent over time under $\mathbf{P}_0^k$ and $\mathbf{P}_1^k$ for every $1 \leq k \leq N$, then the asymptotic optimality properties (34)–(35) hold true for any $m \geq 1$, as long as the almost sure convergence conditions (36) hold.*

2.2.2. Asymptotic Optimality of the M-SLRT

In this section, we propose an alternative sequential test that is based on averaging, instead of maximizing, the likelihood ratios that correspond to the different hypotheses. We show that it has the same asymptotic optimality properties and similar feasibility as the G-SLRT.

Let $\mathcal{P}$ be an arbitrary class, $\{p_{\mathcal{A}}\}_{\mathcal{A} \in \mathcal{P}}$ an arbitrary family of positive numbers that add up to 1 (weights) and consider the probability measure

$$\bar{\mathbf{P}} := \sum_{\mathcal{A} \in \mathcal{P}} p_{\mathcal{A}} \mathbf{P}^{\mathcal{A}}. \quad (38)$$

Then the Radon-Nikodým derivative of $\bar{\mathbf{P}}$ versus $\mathbf{P}_0$ given $\mathcal{F}_n$ is

$$\bar{\Lambda}_n := \frac{d\bar{\mathbf{P}}}{d\mathbf{P}_0} \Big|_{\mathcal{F}_n} = \sum_{\mathcal{A} \in \mathcal{P}} p_{\mathcal{A}} \Lambda_n^{\mathcal{A}} = \sum_{n=1}^N \sum_{\mathcal{A} \in \mathcal{P} \cap \mathcal{P}_n} p_{\mathcal{A}} \Lambda_n^{\mathcal{A}}. \quad (39)$$

If we replace the GLRo statistic $\hat{Z}_n = \max_{\mathcal{A} \in \mathcal{P}} Z_n^{\mathcal{A}}$ in (26) by the logarithm of the mixture likelihood ratio, $\bar{Z}_n := \log \bar{\Lambda}_n$, then we obtain the following sequential test:

$$\bar{\tau} = \inf \{n \geq 1 : \bar{Z}_n \notin (-a, b)\}, \quad \bar{d} := \begin{cases} 1 & \text{when } \bar{Z}_{\bar{\tau}} \geq b \\ 0 & \text{when } \bar{Z}_{\bar{\tau}} \leq -a \end{cases}, \quad (40)$$

to which we refer as the Mixture Sequential Likelihood Ratio Test (M-SLRT).

In the following lemma we show how to select the thresholds in order to guarantee the desired error control for the M-SLRT.

Lemma 3. *For any positive thresholds a and b we have*

$$\mathbf{P}_0(\bar{d} = 1) \leq e^{-b} \quad \text{and} \quad \max_{\mathcal{A} \in \mathcal{P}} \mathbf{P}^{\mathcal{A}}(\bar{d} = 0) \leq \left(\min_{\mathcal{A} \in \mathcal{P}} p_{\mathcal{A}} \right)^{-1} e^{-a}. \quad (41)$$

Therefore, for any $\alpha, \beta \in (0, 1)$, $(\bar{\tau}, \bar{d}) \in \mathbb{C}_{\alpha,\beta}(\mathcal{P})$ when the thresholds are selected as follows:

$$b = |\log \alpha| \quad \text{and} \quad a = |\log \beta| - \min_{\mathcal{A} \in \mathcal{P}} (\log p_{\mathcal{A}}). \quad (42)$$

The following theorem shows that the M-SLRT has exactly the same asymptotic optimality properties as the G-SLRT.

Theorem 7. Consider an arbitrary class of possibly affected subsets, $\mathcal{P}$, and suppose that the thresholds of the M-SLRT are selected so that $\bar{\delta}_{a,b} \in \mathbb{C}_{\alpha,\beta}(\mathcal{P})$ and $b \sim |\log \alpha|$, $a \sim |\log \beta|$ as $\alpha_{\max} \rightarrow 0$, in particular according to (42). If r -complete convergence conditions (32) hold, then for all $1 \leq m \leq r$ we have as $\alpha_{\max} \rightarrow 0$:

$$\mathbb{E}_0[\bar{\tau}^m] \sim \left(\frac{|\log \beta|}{\min_{\mathcal{A} \in \mathcal{P}} I_0^{\mathcal{A}}} \right)^m \sim \inf_{(\tau,d) \in \mathbb{C}_{\alpha,\beta}(\mathcal{P})} \mathbb{E}_0[\tau^m], \quad (43)$$

$$\mathbb{E}^{\mathcal{A}}[\bar{\tau}^m] \sim \left(\frac{|\log \alpha|}{I_1^{\mathcal{A}}} \right)^m \sim \inf_{(\tau,d) \in \mathbb{C}_{\alpha,\beta}(\mathcal{P})} \mathbb{E}^{\mathcal{A}}[\tau^m] \quad \text{for every } \mathcal{A} \in \mathcal{P}. \quad (44)$$

Moreover, if the LLRs Z_t^k have independent increments, then the asymptotic relationships (43)–(44) hold for every $m > 0$ as long as the almost sure convergence conditions (36) are satisfied.

2.2.3. Feasibility

The implementation of the G-SLRT requires computing at each time t the generalized log-likelihood ratio statistic

$$\hat{Z}_n = \max_{\mathcal{A} \in \mathcal{P}} Z_n^{\mathcal{A}} = \max_{\mathcal{A} \in \mathcal{P}} \sum_{k \in \mathcal{A}} Z_n^k.$$

A direct computation of each $Z_n^{\mathcal{A}}$ for every $\mathcal{A} \in \mathcal{P}$ can be a very computationally expensive task when the cardinality of class $\mathcal{P}$, $|\mathcal{P}|$, is very large. However, the computation of $\hat{Z}_n$ is very easy for a class $\mathcal{P}$ of the form $\mathcal{P}_{\underline{m}, \bar{m}}$, which contains all subsets of size at least $\underline{m}$ and at most $\bar{m}$. In order to see this, let us use the following notation for the order statistics: $Z_n^{(1)} \geq \dots \geq Z_n^{(N)}$, i.e., $Z_n^{(1)}$ is the top local LLR statistic and $Z_n^{(N)}$ is the smallest LLR at time n .

When the size of the affected subset is known in advance, i.e., $\underline{m} = \bar{m} = m$, we have

$$\hat{Z}_n = \sum_{k=1}^m Z_n^{(k)}. \quad (45)$$

Indeed, for any $\mathcal{A} \in \mathcal{P}_m$ we have $Z_n^{\mathcal{A}} \leq \sum_{k=1}^m Z_n^{(k)}$. Therefore, $\hat{Z}_n \leq \sum_{k=1}^m Z_n^{(k)}$, and the upper bound is attained by the subset which consists of the m channels with the highest LLR values at time n .

In the more general case that $\underline{m} < \bar{m}$ we have

$$\hat{Z}_n = \sum_{k=1}^{\underline{m}} Z_n^{(k)} + \sum_{k=\underline{m}+1}^{\bar{m}} (Z_n^{(k)})^+,$$

and the G-SLRT takes the following form:

$$\begin{aligned} \hat{\tau} &= \inf \left\{ n \geq 1 : \sum_{k=1}^{\bar{m}} (Z_n^{(k)})^+ \geq b \quad \text{or} \quad \sum_{k=1}^{\underline{m}} Z_n^{(k)} \leq -a \right\} \\ \hat{d} &= \begin{cases} 1 & \text{when } \sum_{k=1}^{\bar{m}} (Z_{\hat{\tau}}^{(k)})^+ \geq b \\ 0 & \text{when } \sum_{k=1}^{\underline{m}} Z_{\hat{\tau}}^{(k)} \leq -a \end{cases}. \end{aligned} \quad (46)$$

Indeed, for any $\mathcal{A} \in \mathcal{P}_{\underline{m}, \overline{m}}$ we have

$$Z_n^{\mathcal{A}} \leq \sum_{k=1}^{\underline{m}} Z_n^{(k)} + \sum_{k=\underline{m}+1}^{\overline{m}} (Z_n^{(k)})^+,$$

and the upper bound is attained by the subset which consists of the $\underline{m}$ channels with the top $\underline{m}$ LLRs and the next (if any) top $\overline{m} - \underline{m}$ channels that have positive LLRs.

Similarly to the G-SLRT, the M-SLRT is computationally feasible even when N is large. Indeed, the mixture likelihood ratio takes the form

$$\overline{\Lambda}_n = C(\mathcal{P}) \sum_{m=1}^N \sum_{\mathcal{A} \in \mathcal{P} \cap \mathcal{P}_m} \prod_{k \in \mathcal{A}} (p_k \Lambda_n^k).$$

When in particular there is an upper and a lower bound on the size of the affected subset, i.e., $\mathcal{P} = \mathcal{P}_{\underline{m}, \overline{m}}$ for some $1 \leq \underline{m} \leq \overline{m} \leq N$, the mixture likelihood ratio statistic takes the form

$$\overline{\Lambda}_N = C(\mathcal{P}) \sum_{m=\underline{m}}^{\overline{m}} \sum_{\mathcal{A} \in \mathcal{P}_m} \prod_{k \in \mathcal{A}} (p_k \Lambda_n^k) \quad (47)$$

and its computational complexity is polynomial in the number of channels, N . However, in the special case of complete uncertainty ($\underline{m} = 1, \overline{m} = N$), the M-SLRT requires only $O(N)$ operations. Indeed, if we set for simplicity $p_k = p$ and $\pi = p/(1+p)$, then the mixture likelihood ratio in (47) admits the following representation for the class $\mathcal{P} = \overline{\mathcal{P}}_N$:

$$\overline{\Lambda}_n = C(\mathcal{P}) [(1 - \pi)^{-N} \tilde{\Lambda}_n - 1] \quad (48)$$

where the statistic $\tilde{\Lambda}_n$ is defined as follows:

$$\tilde{\Lambda}_n = \prod_{k=1}^N (1 - \pi + \pi \Lambda_n^k). \quad (49)$$

Note that the statistic $\tilde{\Lambda}_n$ has an appealing statistical interpretation, as it is the likelihood ratio that corresponds to the case that each channel belongs to the affected subset with probability $\pi \in (0, 1)$. It is possible to use $\tilde{\Lambda}_n$ as the detection statistic and incorporate prior information by an appropriate selection of π . For instance, if we know the exact size of the affected subset, say $\mathcal{P} = \mathcal{P}_m$, we may set $\pi = m/N$, whereas if we know that at most m channels may be affected, i.e., $\mathcal{P} = \overline{\mathcal{P}}_m$, then we may set $\pi = m/(2N)$.

2.3. Asymptotic Bayesian Theory of Quickest Changepoint Detection

The problem of rapid detection of abrupt changes in a state of a process or a system arises in a variety of applications from engineering problems (e.g., navigation integrity monitoring Basseville and Nikiforov (1993); Tartakovsky et al. (2014b)), military applications (e.g., target detection and tracking in heavy clutter Tartakovsky et al. (2014b)) to cyber security (e.g., quick detection of attacks in computer networks Kent (2000); Tartakovsky (2013b); Tartakovsky et al. (2006a,b, 2014b)). In the present project, we are interested in a sequential setting assuming that as long as

the behavior of the observation process is consistent with a “normal” (initial in-control) state, we allow the process to continue. If the state changes, then we need to detect this event as rapidly as possible while controlling for the risk of false alarms. In other words, we are interested in designing the quickest change-point detection procedure that optimizes the tradeoff between a measure of detection delay and a measure of the frequency of false alarms.

In the beginning of the 1960s, Shiryaev (1963) developed a Bayesian sequential changepoint detection (quickest disorder detection) theory in the iid case assuming that the observations are independent and identically distributed (iid) according to a distribution F pre-change and another distribution G post-change and with the prior distribution of the change point being geometric. In particular, Shiryaev (1963) proved that the detection procedure that is based on thresholding the posterior probability of the change being active before the current time is strictly optimal, minimizing the average delay to detection in the class of procedures with a given probability of false alarm. Tartakovsky and Veeravalli (2005) generalized Shiryaev’s theory for the non-iid case that covers very general discrete-time non-iid stochastic models and a wide class of prior distributions that include distributions with both exponential tails and heavy tails. In particular, it was proved that the Shiryaev detection procedure is asymptotically optimal – it minimizes the average delay to detection as well as higher moments of the detection delay as the probability of a false alarm vanishes. Baron and Tartakovsky (2006) developed an asymptotic Bayesian theory for general continuous-time stochastic processes.

The key assumption in general asymptotic theories developed in Baron and Tartakovsky (2006); Tartakovsky and Veeravalli (2005) is a certain stability property of the log-likelihood ratio process between the “change” and “no-change” hypotheses, which was expressed in the form of the strong law of large numbers with a positive and finite number and its strengthened r -quick version. However, it is not easy (and in fact can be quite difficult) to verify r -quick convergence in particular applications and examples. For this reason, it was conjectured in Baron and Tartakovsky (2006); Tartakovsky and Veeravalli (2005) that essentially the same asymptotic results may be obtained under a weaker r -complete version of the strong law of large numbers for the log-likelihood ratio. In fact, in most examples provided in Baron and Tartakovsky (2006); Tartakovsky and Veeravalli (2005) and in the recent book by Tartakovsky et al. (2014b), verification of the r -quick convergence is replaced by verification of the r -complete convergence. Our main goal in this project is to confirm this conjecture, proving that the Shiryaev changepoint detection procedure is asymptotically optimal under the r -complete convergence condition for the suitably normalized log-likelihood ratio process.

In the following, we deal only with discrete time $t = n \in \mathbb{Z}_+ = \{0, 1, 2, \dots\}$. The continuous time case $t \in \mathbb{R}_+ = [0, \infty)$ is more “delicate” and will be considered elsewhere. Having said that, let $(\Omega, \mathcal{F}, \mathcal{F}_n, P)$, $n \in \mathbb{Z}_+$ be a filtered probability space, where the sub- σ -algebra $\mathcal{F}_n = \sigma(\mathbf{X}^n)$ of $\mathcal{F}$ is assumed to be generated by the process $\mathbf{X}^n = \{X_t\}_{1 \leq t \leq n}$ observed up to time n . Let P_0 and P_∞ be two probability measures defined on this space, which are assumed to be mutually locally absolutely continuous, so that the restrictions of these measures P_0^n and P_∞^n to the sigma-algebras $\mathcal{F}_n$ are mutually absolutely continuous for all $n \geq 1$.

We are interested in the following changepoint problem. In a “normal” mode, the observed process X_n follows the measure P_∞ , and at an unknown time ν ($\nu \geq 0$) something happens and X_n follows the measure P_0 . The goal is to detect the change as soon as possible after it occurs, subject to a constraint on the risk of false alarms. The exact optimality criteria will be specified in Section 2.3.2.

2.3.1. A General Changepoint Model

Let $p_j(\mathbf{X}^n)$, $j = \infty, 0$ denote densities of P_j^n (with respect to some non-degenerate σ -finite measure), where $\mathbf{X}^n = (X_1, \dots, X_n)$ is the sample of size n . For a fixed $\nu \in \mathbb{Z}_+$, the change induces a probability measure P_ν (correspondingly density $p_\nu(\mathbf{X}^n) = p(\mathbf{X}^n|\nu)$), which is a combination of the pre- and post-change densities:

$$p_\nu(\mathbf{X}^n) = p_\infty(\mathbf{X}^\nu) \cdot p_0(\mathbf{X}_{\nu+1}^n|\mathbf{X}^\nu) = \prod_{i=1}^{\nu} p_\infty(X_i|\mathbf{X}^{i-1}) \cdot \prod_{i=\nu+1}^n p_0(X_i|\mathbf{X}^{i-1}), \quad (50)$$

where $\mathbf{X}_m^n = (X_m, \dots, X_n)$ and $p_j(X_n|\mathbf{X}^{n-1})$ is the conditional density of X_n given $\mathbf{X}^{n-1}$. In the sequel we assume that ν is the serial number of the last pre-change observation. Note that in general the conditional densities $p_0(X_i|\mathbf{X}^{i-1})$, $i = \nu+1, \nu+2, \dots$ may depend on the changepoint ν , i.e., $p_0(X_i|\mathbf{X}^{i-1}) = p_0^{(\nu)}(X_i|\mathbf{X}^{i-1})$ for $i > \nu$. Certainly the densities $p_j(X_i|\mathbf{X}^{i-1}) = p_{j,i}(X_i|\mathbf{X}^{i-1})$, $j = 0, \infty$ may depend on i .

In a particular *iid case*, addressed in detail in the past the observations are independent and identically distributed (iid) with density $f_\infty(x)$ in the normal (pre-change) mode and with another density $f_0(x)$ in the abnormal (post-change) mode, i.e., in this case, (50) holds with $p_\infty(X_i|\mathbf{X}^{i-1}) = f_\infty(X_i)$ and $p_0(X_i|\mathbf{X}^{i-1}) = f_0(X_i)$.

We are interested in a Bayesian setting where the change point ν is assumed to be a random variable independent of the observations with prior probability distribution $\Pi_n = P(\nu \leq n)$, $n \in \mathbb{Z}_+$. We also write $\pi_k = P(\nu = k)$ for the probability on non-negative integers, $k = 0, 1, 2, \dots$. Formally, we allow the change point ν to take negative values too, but the detailed distribution for $k < 0$ is not important. The only value we need is the cumulative probability $q = P(\nu < 0)$. The probability $P(\nu \leq 0) = q + \pi_0$ is the probability of the “atom” associated with the event that the change already took place before the observations became available.

In the past, the typical choice for the prior distribution was (zero modified) geometric distribution,

$$P(\nu < 0) = q \quad \text{and} \quad P(\nu = k) = (1 - q)\rho(1 - \rho)^k \quad \text{for } k = 0, 1, 2, \dots, \quad (51)$$

where $0 \leq q < 1$, $0 < \rho < 1$.

In the rest of the paper, we consider an arbitrary prior distribution that belongs to the class of distributions that satisfy the following condition:

C. For some $0 \leq \mu < \infty$,

$$\lim_{n \rightarrow \infty} \frac{|\log(1 - \Pi_n)|}{n} = \mu. \quad (52)$$

In the case that $\mu = 0$, we assume in addition that for some $r \geq 1$

$$\sum_{k=0}^{\infty} \pi_k |\log \pi_k|^r < \infty. \quad (53)$$

If $\mu > 0$, then the prior distribution has an exponential right tail. Such distributions, as geometric and discrete versions of gamma and logistic distributions, i.e., models with bounded hazard rate, belong to this class. In this case, condition (53) holds automatically. If $\mu = 0$, then the distribution has a heavy tail, i.e., such a distribution belongs to the model with a vanishing hazard rate. However, we cannot allow this distribution to have a tail that is too heavy, which is guaranteed by condition (53).

2.3.2. Optimality Criteria

Any sequential detection procedure is a stopping time T for the observed process $\{X_n\}_{n \in \mathbb{Z}_+}$, i.e., T is an extended random variable, such that the event $\{T = n\}$ belongs to the sigma-algebra $\mathcal{F}_n$. A false alarm is raised whenever $T \leq \nu$. A good detection procedure should guarantee a small delay to detection $T - \nu$ provided that there is no false alarm, while the rate (or risk) of false alarms should be kept at a given, usually low level.

Let P_k and E_k denote the probability and the corresponding expectation when the change occurs at time $\nu = k \in \mathbb{Z}_+$. In what follows, P^π denotes the probability measure on the Borel sigma-algebra in $\mathbb{R}^\infty \times \mathbb{N}$ defined as $P^\pi(\mathcal{A} \times J) = \sum_{k \in J} \pi_k P_k(\mathcal{A})$ for $\mathcal{A} \in \mathcal{B}(\mathbb{R}^\infty)$, $J \subseteq \mathbb{N}$ and E^π denotes the expectation with respect to P^π .

In a Bayesian setting, the risk associated with the delay to detection is usually measured by the average delay to detection

$$E^\pi(T - \nu | T > \nu) = \frac{\sum_{k=0}^{\infty} \pi_k E_k(T - k | T > k) P_\infty(T > k)}{1 - \text{PFA}(T)} \quad (54)$$

and the risk associated with a false alarm by the weighted probability of false alarm (PFA) defined as

$$\text{PFA}(T) = P^\pi(T \leq \nu) = \sum_{k=1}^{\infty} \pi_k P_\infty(T \leq k). \quad (55)$$

In (54) and (55) we use the fact that $P_k(T > k) = P_\infty(T > k)$ and $P_k(T \leq k) = P_\infty(T \leq k)$ for $k \in \mathbb{Z}_+$ and that $P_\infty(T \leq 0) = 0$.

For $0 < \alpha < 1$, let $\mathbb{C}_\alpha = \{T : \text{PFA}(T) \leq \alpha\}$ be a class of detection procedures for which the weighted probability of false alarm does not exceed the predefined level α . In a Bayesian setting, the goal is to find an optimal procedure that minimizes in the class $\mathbb{C}_\alpha$ the average delay to detection, i.e.,

$$\text{find } T_{\text{opt}} \in \mathbb{C}_\alpha \text{ such that } E^\pi(T_{\text{opt}} - \nu | T_{\text{opt}} > \nu) = \inf_{T \in \mathbb{C}_\alpha} E^\pi(T - \nu | T > \nu).$$

However, except for the iid case, the solution of this problem is not tractable. For this reason, we address the asymptotic problem of minimizing the average detection delay as α approaches zero. For practical purposes, it is also interesting to consider the problem of minimizing higher moments of the detection delay $E^\pi[(T - \nu)^m | T > \nu]$ for some $m \geq 1$, i.e., to find a first-order asymptotically optimal detection procedure $T_o \in \mathbb{C}_\alpha$ that satisfies

$$\lim_{\alpha \rightarrow 0} \frac{E^\pi[(T_o - \nu)^m | T_o > \nu]}{\inf_{T \in \mathbb{C}_\alpha} E^\pi[(T - \nu)^m | T > \nu]} = 1. \quad (56)$$

2.3.3. Change Detection Procedures

Let “ $H_k : \nu = k$ ” and “ $H_\infty : \nu = \infty$ ” be the hypotheses that the change occurs at the point $0 \leq k < \infty$ and that the change never happens, respectively. Then, using (50), we obtain that the likelihood ratio (LR) between these hypotheses when the sample $\mathbf{X}^n = (X_1, \dots, X_n)$ is observed is

$$\frac{dP_k^n}{dP_\infty^n} = \prod_{i=k+1}^n \frac{p_0(X_i | \mathbf{X}^{i-1})}{p_\infty(X_i | \mathbf{X}^{i-1})}, \quad k < n.$$

Write $\mathcal{L}_i = p_0(X_i|\mathbf{X}^{i-1})/p_\infty(X_i|\mathbf{X}^{i-1})$ and introduce the normalized average (weighted) LR

$$\Lambda_n = \frac{1}{P(\nu \geq n)} \left(q \prod_{i=1}^n \mathcal{L}_i + \sum_{k=0}^{n-1} \pi_k \prod_{i=k+1}^n \mathcal{L}_i \right), \quad n \in \mathbb{Z}_+.$$

Note that $\Lambda_0 = q/(1-q)$. Let $g_n = P(\nu < n|\mathbf{X}^n)$ stand for the posterior probability of the change being in effect up to time n . Shiryaev (1963) proved that, in the iid case, the detection procedure $T_a = \inf \{n : g_n \geq a\}$ is strictly optimal for every $0 < \alpha < 1$ – it minimizes the average detection delay $E^\pi(T - \nu | T > \nu)$ if $a = a_\alpha$ is selected so that $\text{PFA}(T_a) = \alpha$ and the prior distribution is geometric. We refer to this procedure as the Shiryaev detection procedure in the general non-iid case too. We now show that $\Lambda_n = g_n/(1 - g_n)$, so that the Shiryaev procedure can be written as

$$T_A = \inf \{n \geq 1 : \Lambda_n \geq A\}, \quad A > 0. \quad (57)$$

Indeed, $g_n = \sum_{k=-\infty}^{n-1} P(\nu = k|\mathbf{X}^n)$, where

$$\begin{aligned} P(\nu = k|\mathbf{X}^n) &= \frac{\pi_k \prod_{j=1}^k p_\infty(X_j|\mathbf{X}^{j-1}) \prod_{i=k+1}^n p_0(X_i|\mathbf{X}^{i-1})}{\sum_{k=-\infty}^{\infty} \pi_k \prod_{j=1}^k p_\infty(X_j|\mathbf{X}^{j-1}) \prod_{i=k+1}^n p_0(X_i|\mathbf{X}^{i-1})} \\ &= \frac{\pi_k \prod_{i=k+1}^n \mathcal{L}_i}{q \prod_{i=1}^n \mathcal{L}_i + \sum_{k=0}^{n-1} \pi_k \prod_{i=k+1}^n \mathcal{L}_i + P(\nu \geq n)}, \end{aligned}$$

and we obtain

$$g_n = \frac{q \prod_{i=1}^n \mathcal{L}_i + \sum_{k=0}^{n-1} \pi_k \prod_{i=k+1}^n \mathcal{L}_i}{q \prod_{i=1}^n \mathcal{L}_i + \sum_{k=0}^{n-1} \pi_k \prod_{i=k+1}^n \mathcal{L}_i + P(\nu \geq n)}.$$

Therefore,

$$\frac{g_n}{1 - g_n} = \frac{1}{P(\nu \geq n)} \left(q \prod_{i=1}^n \mathcal{L}_i + \sum_{k=1}^{n-1} \pi_k \prod_{i=k+1}^n \mathcal{L}_i \right) = \Lambda_n.$$

In particular, in the popular case of zero modified geometric prior (51), the statistic Λ_n is

$$\Lambda_n = \frac{q}{1-q} \prod_{i=1}^n \left(\frac{\mathcal{L}_i}{1-\rho} \right) + \rho \sum_{k=1}^n \prod_{i=k}^n \left(\frac{\mathcal{L}_i}{1-\rho} \right). \quad (58)$$

In the following, to avoid triviality, we assume that $A > q/(1-q)$, since otherwise $T_A = 0$ with probability 1.

By Lemma 7.2.1 in Tartakovsky et al. (2014b),

$$\text{PFA}(T_A) \leq 1/(1+A) \quad \text{for every } A > q/(1-q), \quad (59)$$

and therefore setting $A = A_\alpha = (1-\alpha)/\alpha$ guarantees that $T_A \in \mathbb{C}_\alpha$.

Another popular change detection procedure is the Shiryaev–Roberts (SR) procedure (due to Shiryaev (1963) and Roberts (1966)) given by the stopping time

$$\tilde{T}_B = \inf \{n \geq 1 : R_n \geq B\}, \quad B > 0, \quad (60)$$

where the statistic R_n , the SR statistic, is given by

$$R_n = \sum_{k=1}^n \prod_{i=k}^n \mathcal{L}_i, \quad n \geq 0 \quad (R_0 = 0). \quad (61)$$

The statistic R_n can be viewed as a limit of the statistic Λ_n/ρ as $\rho \rightarrow 0$ when the prior distribution of the change point is geometric (51) with $q = 0$. Indeed, see (58).

2.3.4. r -Quick Convergence Versus r -Complete Convergence

Introduce the LLRs

$$Z_i = \log \frac{p_0(X_i | \mathbf{X}^{i-1})}{p_\infty(X_i | \mathbf{X}^{i-1})}, \quad \lambda_{k+n}^k = \log \frac{d\mathbf{P}_k^{k+n}}{d\mathbf{P}_\infty^{k+n}} = \sum_{i=k+1}^{k+n} Z_i, \quad n \geq 1.$$

We need the following two definitions.

Definition 3. Let $r > 0$. For $k = 0, 1, 2, \dots$, we say that the normalized LLR $n^{-1}\lambda_{k+n}^k$ converges r -quickly to a constant I as $n \rightarrow \infty$ under probability $\mathbf{P}_k$ if $\mathbf{E}_k[L_k(\varepsilon)]^r < \infty$ for all $\varepsilon > 0$, where $L_k(\varepsilon) = \sup \{n \geq 1 : |n^{-1}\lambda_{k+n}^k - I| > \varepsilon\}$ ($\sup\{\emptyset\} = 0$) is the last time when $n^{-1}\lambda_{k+n}^k$ leaves the interval $[I - \varepsilon, I + \varepsilon]$.

Definition 4. Let $r > 0$. For $k = 0, 1, 2, \dots$, we say that the normalized LLR $n^{-1}\lambda_{k+n}^k$ converges r -completely to a constant I as $n \rightarrow \infty$ under probability $\mathbf{P}_k$ if for all $\varepsilon > 0$,

$$\sum_{n=1}^{\infty} n^{r-1} \mathbf{P}_k \{ |n^{-1}\lambda_{k+n}^k - I| > \varepsilon \} < \infty. \quad (62)$$

(For $r = 1$ this mode of convergence was introduced by Hsu and Robbins (1947).)

Note first that in general r -quick convergence is a stronger property than r -complete convergence. See Lemma 2.4.1 in Tartakovsky et al. (2014b). More importantly, checking r -quick convergence in applications is often much more difficult than checking r -complete convergence.

In the discrete time case, Tartakovsky and Veeravalli (2005) developed a general asymptotic Bayesian theory of changepoint detection assuming that the LLR obeys the strong law of large numbers (SLLN) with some positive and finite constant I , i.e.,

$$\frac{1}{n} \lambda_{k+n}^k \xrightarrow[n \rightarrow \infty]{\mathbf{P}_k\text{-a.s.}} I \quad \text{for all } k \in \mathbb{Z}_+, \quad (63)$$

with a certain rate of convergence expressed via the r -quick convergence, specifically assuming in addition that for some $r \geq 1$

$$\sum_{k=0}^{\infty} \pi_k \mathbf{E}_k[L_k(\varepsilon)]^r < \infty. \quad (64)$$

A similar development was performed by Baron and Tartakovsky (2006) in continuous time, assuming that

$$\int_0^\infty \mathbf{E}_u[L_u(\varepsilon)]^r d\Pi_u < \infty.$$

However, as we already mentioned, verification of the latter r -quick convergence condition in particular examples is not an easy task.

In Baron and Tartakovsky (2006); Tartakovsky and Veeravalli (2005), it was conjectured that all asymptotic results, including near optimality of the Shiryaev procedure (in the sense defined in (56)), hold if the r -quick convergence condition (64) is weakened into the r -complete convergence

$$\sum_{k=0}^{\infty} \pi_k \left[\sum_{n=1}^{\infty} n^{r-1} \mathbb{P}_k \left\{ |n^{-1} \lambda_{k+n}^k - I| > \varepsilon \right\} \right] < \infty$$

(with an obvious modification in continuous time). In the following subsections, we justify this conjecture.

2.3.5. Asymptotic Operating Characteristics and Optimality of the Shiryaev Procedure

In this subsection, we present the main results related to asymptotic optimality of the Shiryaev detection procedure in the general non-iid case as well as in the case of independent observations.

The following lemma, that establishes the asymptotic lower bounds for moments of the detection delay, will be used for proving asymptotic optimality properties.

Lemma 4. *Let T_A be the Shiryaev changepoint detection procedure defined in (57). Let, for some $\mu \geq 0$, the prior distribution of the change point satisfy condition (52). Assume that for some positive and finite I*

$$\lim_{M \rightarrow \infty} \mathbb{P}_k \left(\frac{1}{M} \max_{1 \leq n \leq M} \lambda_{k+n}^k \geq (1 + \varepsilon)I \right) = 0 \quad \text{for all } \varepsilon > 0 \text{ and all } k \in \mathbb{Z}_+. \quad (65)$$

Then, for all $m > 0$,

$$\liminf_{\alpha \rightarrow 0} \frac{\inf_{T \in \mathbb{C}_\alpha} \mathbb{E}^\pi [(T - \nu)^m | T > \nu]}{|\log \alpha|^m} \geq \frac{1}{(I + \mu)^m}. \quad (66)$$

and

$$\liminf_{A \rightarrow \infty} \frac{\mathbb{E}^\pi [(T_A - \nu)^m | T_A > \nu]}{(\log A)^m} \geq \frac{1}{(I + \mu)^m}. \quad (67)$$

Define

$$\Upsilon_{k,r}(\varepsilon) = \sum_{n=1}^{\infty} n^{r-1} \mathbb{P}_k \left(\frac{1}{n} \lambda_{k+n}^k < I - \varepsilon \right). \quad (68)$$

Recall that by (59), $\text{PFA}(T_A) \leq (1 + A)^{-1}$ for any $0 < A < q/(1 - q)$, which implies that $\text{PFA}(T_{A_\alpha}) \leq \alpha$ (i.e., $T_{A_\alpha} \in \mathbb{C}_\alpha$) for any $0 < \alpha < 1 - q$ if $A = A_\alpha = (1 - \alpha)/\alpha$.

The following theorem is the main result in the general non-iid case, which shows that the Shiryaev detection procedure is asymptotically optimal under mild conditions for the observations and prior distributions.

Theorem 8. *Let T_A be the Shiryaev changepoint detection procedure defined in (57). Let $r \geq 1$ and let the prior distribution of the change point satisfy condition (C). Assume that for some number $0 < I < \infty$ condition (65) is satisfied and that the following condition holds as well*

$$\sum_{k=0}^{\infty} \pi_k \Upsilon_{k,r}(\varepsilon) < \infty \quad \text{for all } \varepsilon > 0. \quad (69)$$

(i) Then for all $0 < m \leq r$

$$\lim_{A \rightarrow \infty} \frac{\mathbb{E}^\pi[(T_A - \nu)^m | T_A > \nu]}{(\log A)^m} = \frac{1}{(I + \mu)^m}. \quad (70)$$

(ii) If $A = A_\alpha = (1 - \alpha)/\alpha$, where $0 < \alpha < 1 - q$, then $T_{A_\alpha} \in \mathbb{C}_\alpha$ and it is asymptotically optimal as $\alpha \rightarrow 0$ in class $\mathbb{C}_\alpha$, minimizing moments of the detection delay up to order r , i.e., for all $0 < m \leq r$,

$$\lim_{\alpha \rightarrow 0} \frac{\inf_{T \in \mathbb{C}_\alpha} \mathbb{E}^\pi[(T - \nu)^m | T > \nu]}{\mathbb{E}^\pi[(T_{A_\alpha} - \nu)^m | T_{A_\alpha} > \nu]} = 1. \quad (71)$$

Also, the following first-order asymptotic approximations hold:

$$\inf_{T \in \mathbb{C}_\alpha} \mathbb{E}^\pi[(T - \nu)^m | T > \nu] \sim \mathbb{E}^\pi[(T_{A_\alpha} - \nu)^m | T_{A_\alpha} > \nu] \sim \left(\frac{|\log \alpha|}{I + \mu} \right)^m \quad \text{as } \alpha \rightarrow 0. \quad (72)$$

This assertion also holds if $A = A_\alpha$ is selected so that $\text{PFA}(T_{A_\alpha}) \leq \alpha$ and $\log A_\alpha \sim |\log \alpha|$ as $\alpha \rightarrow 0$.

Corollary 1. Let $r \geq 1$. Let the prior distribution of the change point satisfy condition (C). Assume that for some $0 < I < \infty$

$$\sum_{k=0}^{\infty} \pi_k \left[\sum_{n=1}^{\infty} n^{r-1} \mathbb{P}_k \left(\left| \frac{1}{n} \lambda_{k+n}^k - I \right| > \varepsilon \right) \right] < \infty \quad \text{for all } \varepsilon > 0. \quad (73)$$

Then (70), (71) and (72) hold true.

The above results show that the lower bounds (66) and (67) for moments of the detection delay hold whenever the LLR process λ_{k+n}^k obeys the SLLN (63), since in this case condition (65) is satisfied. However, in general, an almost sure convergence (63) is not sufficient for obtaining the upper bounds, and therefore, for asymptotic optimality of the Shiryaev procedure. In fact, this condition does not even guarantee finiteness of the average delay to detection $\mathbb{E}^\pi(T_A - \nu | T_A > \nu)$, and to obtain meaningful results we need to strengthen the SLLN into the r -complete version. On the other hand, in the iid case, where conditioned on $\nu = k$ the observations $X_1, \dots, X_k$ are iid with pre-change density $f_\infty(x)$ and $X_{k+1}, X_{k+2}, \dots$ are iid with post-change density $f_0(x)$, the situation is dramatically different. By Theorem 4 of Tartakovsky and Veeravalli (2005), the Shiryaev procedure asymptotically (as $\alpha \rightarrow 0$) minimizes all positive moments of the detection delay in class $\mathbb{C}_\alpha$ if the prior distribution is geometric and the Kullback–Leibler information number

$$\mathcal{K} = \mathbb{E}_0 \lambda_1^0 = \int \log \left(\frac{f_0(x)}{f_\infty(x)} \right) d\mu(x) \quad (74)$$

is positive and finite.

We now extend this result to the case where observations are independent, but not necessarily identically distributed, i.e., $p_\infty(X_i | \mathbf{X}^{i-1}) = f_{\infty,i}(X_i)$ and $p_0(X_i | \mathbf{X}^{i-1}) = f_{0,i}(X_i)$ in (50). More generally, we may assume that the increments Z_i of the LLR $\lambda_n^k = \sum_{i=k+1}^n Z_i$ are independent, which is always the case if the observations are independent. This slight generalization is important for certain examples with dependent observations that lead to the LLR with independent increments.

Theorem 9. Let T_A be the Shiryaev changepoint detection procedure defined in (57). Let $r \geq 1$. Assume that the LLR process $\{\lambda_{k+n}^k\}_{n \geq 1}$ has independent, not necessarily identically distributed increments under P_k , $k \in \mathbb{Z}_+$. Suppose that condition (65) holds and the following condition

$$\lim_{n \rightarrow \infty} P_k \left(\frac{1}{n} \lambda_{\ell+n}^\ell < I - \varepsilon \right) = 0 \quad \text{for all } \varepsilon > 0, \text{ all } \ell \geq k \text{ and all } k \in \mathbb{Z}_+ \quad (75)$$

is satisfied. Let the prior distribution of the change point be geometric (51) with $q = 0$. Then relations (70), (71) and (72) hold true for all $m > 0$ with $\mu = |\log(1 - \rho)|$. Therefore, the Shiryaev procedure T_{A_α} minimizes asymptotically as $\alpha \rightarrow 0$ all positive moments of the detection delay in class $\mathbb{C}_\alpha$.

The idea of relaxing the r -complete convergence condition by condition (75) is based on splitting integration, when obtaining the upper bound for the expectation $E_k[(T_A - k)^+]^r$, into a sequence of intervals (cycles) of the size $N_A \approx \log A / (I + \mu)$ and then showing that $P_k(T_A - k > \ell N_A) \leq \delta^\ell$, $\ell = 1, 2, \dots$ for some small δ under condition (75), using independence of the LLR increments.

There are many examples associated with Markov and Hidden Markov models (and even more general) that show that the developed theory is useful since the suggested r -complete convergence conditions hold. These examples may be found in Pergamenchikov and Tartakovsky (Submitted in 2016); Tartakovsky (Submitted in 2016).

2.4. Asymptotic Pointwise and Minimax Theory of Quickest Changepoint Detection

In the area of quickest detection, there are four conventional approaches to the optimum tradeoff problem: Bayesian, generalized Bayesian, multicyclic detection of changes in a stationary regime, and minimax (see Tartakovsky et al. (2014b, Ch 6)). The Bayesian problem was considered in the previous section where we developed a general Bayesian change detection theory.

By contrast, in a minimax formulation, the change point is assumed to be an unknown non-random number and the goal is to minimize the worst-case delay (with respect to the point of change) subject to a lower bound on the mean time until false alarm. Specifically, Lorden (1971) suggested the worst-worst-case average delay to detection measure

$$\text{ESADD}(\tau) = \sup_{\nu \geq 0} \text{ess sup } E_\nu(\tau - \nu | \tau > \nu, \mathcal{F}_\nu)$$

that should be minimized in the class of procedures $\mathcal{H}_\gamma = \{\tau : E_\infty \tau \geq \gamma\}$ for which the average run length (mean time) to false alarm $E_\infty \tau$ is not smaller than a given number $\gamma > 1$. Here τ is a generic change detection procedure (stopping time), E_ν stands for the operator of expectation when the change point is ν ($\nu = \infty$ corresponds to a no-change scenario) and $\mathcal{F}_\nu = \sigma(X_1, \dots, X_\nu)$ is the sigma-algebra generated by the first ν observations $X_1, \dots, X_\nu$. Lorden (1971) developed an asymptotic minimax theory of change detection (in the iid case) as $\gamma \rightarrow \infty$, proving in particular that Page's CUSUM procedure is asymptotically first-order minimax. Later Moustakides (1986) established strict optimality of CUSUM for any value of the average run length to false alarm $\gamma > 1$. In the 1980s, Pollak (1985) introduced a less pessimistic worst-case detection delay measure — maximal conditional average delay to detection,

$$\text{SAD2D}(\tau) = \sup_{\nu \geq 0} E_\nu(\tau - \nu | \tau > \nu), \quad (76)$$

and found an almost optimal procedure that minimizes $\text{SAD2D}(\tau)$ subject to the constraint on the average run length to false alarm (i.e., in the class $\mathcal{H}_\gamma$) as γ becomes large. Pollak's idea was to modify the Shiryaev–Roberts statistic by randomization of the initial condition in order to make it an equalizer. Pollak proved that the randomized Shiryaev–Roberts procedure that starts from a random point sampled from the quasi-stationary distribution of the Shiryaev–Roberts statistic is asymptotically nearly minimax within an additive vanishing term.

In the early stages the theoretical development was focused primarily on the iid case. However, in practice the observations may be non-identically distributed and dependent. A general asymptotic minimax theory of change-point detection for non-iid models was developed by Lai (1995, 1998) (see also Fuh (2003) for hidden Markov models with a finite state-space). In particular, for a low false alarm rate (large γ) the asymptotic minimaxity of the CUSUM procedure was established in Fuh (2003); Lai (1998).

In the iid case, the suitably standardized distributions of the stopping times of the CUSUM and Shiryaev–Roberts detection procedures are asymptotically exponential for large thresholds and fit well into the geometric distribution even for a moderate false alarm rate (see Pollak and Tartakovsky (2009b)). In this case, the average run length to false alarm is an appropriate measure of false alarms. However, for non-iid models the limiting distribution is not guaranteed to be exponential or even close to it. In general, we cannot even guarantee that large values of the average run length to false alarm will produce small values of the maximal local false alarm probability. Therefore, the average run length to false alarm is not appropriate in general, and instead it is more adequate to use the local conditional false alarm probability, as suggested in Tartakovsky (2005); Tartakovsky et al. (2014b). This issue is extremely important for non-iid models, as a discussion in Mei (2008); Tartakovsky (2008) shows.

In the project, we pursue two objectives. First, we introduce two novel classes of changepoint detection procedures, which, instead of imposing a lower bound on the average run length to false alarm, require more adequate upper bounds on the uniform probability of false alarm or uniform conditional probability of false alarm in the spirit of works by Lai (1998), Tartakovsky (2005) and Tartakovsky et al. (2014b). However, these classes slightly differ from those proposed in Lai (1998); Tartakovsky (2005); Tartakovsky et al. (2014b). This modification allows us to substantially relax Lai's essential supremum conditions Lai (1998), which do not hold for certain interesting practical models. In fact, our conditions are equivalent to the uniform version of the complete convergence for the log-likelihood ratio processes, i.e., they are related to the rate of convergence in the strong law of large numbers for the log-likelihood ratio between the “change” and “no-change” hypotheses. We concentrate on a minimax problem of minimizing Pollak's maximal conditional average delay to detection defined in (76) as well as on a pointwise problem of minimizing the conditional average delay to detection $E_\nu(\tau - \nu | \tau > \nu)$ for every change point $\nu \geq 0$. For the sake of completeness, we also consider the other popular risks $\sup_{\nu \geq 0} E_\nu(\tau - \nu)^+$ and $E_\nu(\tau - \nu)^+$, $\nu \geq 0$, while we strongly believe that the conditional versions $E_\nu(\tau - \nu | \tau > \nu)$ and (76) are more appropriate for most applications. We consider extremely general non-iid stochastic models for the observations, and it is our goal to find reasonable sufficient conditions for the observation models under which the Shiryaev–Roberts (or CUSUM) procedure is asymptotically optimal. To achieve the first goal we exploit the asymptotic Bayesian theory of changepoint detection developed in the previous section that offers a constructive and flexible approach for studying asymptotic efficiency of Bayesian type procedures. It turns out that a similar method can be used for the analysis of minimax risks and that the complete convergence type conditions for the log-likelihood ratio are also

sufficient in the minimax setting. These sufficient conditions as well as the main results related to asymptotic optimality of the Shiryaev–Roberts procedure in the classes of procedures with upper bounds on the weighted false alarm probability and local false alarm probabilities are given below.

The second objective is to find a method for verification of the required sufficient conditions in a number of particular, still very general, challenging models. The natural question is how one may check the proposed sufficient conditions and even whether there are more or less general models, except of course the iid case, for which these conditions hold. To this end, we focus on the class of data models for which one can exploit the method of geometric ergodicity for homogeneous Markov processes. These results can be found in Section 5 of our recently submitted paper Pergamenchtchikov and Tartakovsky (Submitted in 2016) and show that our sufficient conditions for pointwise and minimax optimality hold for homogeneous Markov ergodic processes. In Pergamenchtchikov and Tartakovsky (Submitted in 2016), these conditions are further illustrated for several examples that include autoregressive, autoregressive GARCH, and other models widely used in many applications.

2.4.1. Novel Optimality Criteria

In this project, we study the Shiryaev–Roberts (SR) procedure given by the following stopping time

$$T(h) = \inf \left\{ n \geq 1 : \sum_{k=1}^n e^{\lambda_n^{k-1}} \geq h \right\}, \quad (77)$$

where $h > 0$ is some fixed positive threshold which will be specified later. We set $\inf\{\emptyset\} = +\infty$. In the iid case, this procedure has certain interesting strict optimality properties (see Pollak and Tartakovsky (2009a) and Tartakovsky et al. (2014b)).

Our main goal is to show that the SR detection procedure $T(h)$ is nearly optimal in pointwise and minimax problems described below.

To describe these problems we introduce for any $0 < \beta < 1$, $m^* \geq 1$ and $k^* > m^*$ the following class of change detection procedures

$$\mathcal{H}^*(\beta, k^*, m^*) = \left\{ \tau : \sup_{1 \leq k \leq k^* - m^*} \mathbb{P}_\infty(\tau < k + m^* | \tau > k) \leq \beta \right\}. \quad (78)$$

Note that the probability $\mathbb{P}_\infty(\tau < k + m | \tau \geq k) = \mathbb{P}_\infty(k \leq \tau < k + m | \tau \geq k)$ is the conditional probability of false alarm in the time interval $[k, k + m - 1]$ of the length m , which we refer to as the *local conditional probability of false alarm* (LCPFA).

We consider the *conditional detection delay risk*

$$\mathcal{R}_\nu^*(\tau) = \mathbb{E}_\nu(\tau - \nu | \tau > \nu) \quad (79)$$

(compare with (76)) and the following problems: the pointwise minimization, i.e., for any $\nu \geq 0$

$$\inf_{\tau \in \mathcal{H}^*(\beta, k^*, m^*)} \mathcal{R}_\nu^*(\tau); \quad (80)$$

and the minimax optimization

$$\inf_{\tau \in \mathcal{H}^*(\beta, k^*, m^*)} \max_{0 \leq \nu \leq k^*} \mathcal{R}_\nu^*(\tau). \quad (81)$$

The parameters k^* and m^* will be specified later.

In addition, we consider a Bayesian-type problem of minimizing the risk (79) in a class of procedures with the given weighted probability of false alarm.

2.4.2. Asymptotic Optimality of the SR Procedure

We now proceed with tackling the pointwise and minimax problems (80) and (81) in the class of procedures with given LCPFA. The method of establishing asymptotic optimality of the SR procedure is based on the lower-upper bounding technique. Specifically, we first obtain asymptotic lower bounds for the risk $\mathcal{R}_\nu^*(\tau)$ in the class $\mathcal{H}^*(\beta, k^*, m^*)$, and then we show that these asymptotic lower bounds are attained for the SR procedure $T(h)$ with a certain threshold $h = h_\beta$.

We do not assume any particular model or even class of models for the observations, and as a result, there is no “structure” of the LLR process. We therefore have to impose some conditions on the behavior of the LLR process at least for large n . It is natural to assume that there exists a positive finite number I such that $\lambda_n^k/(n - k)$ converges almost surely to I under P_k , i.e.,

(A₁) Assume that there exists a number $I > 0$ such that for any $k \geq 0$

$$\frac{1}{n} \lambda_{k+n}^k \xrightarrow[n \rightarrow \infty]{P_k\text{-a.s.}} I. \quad (82)$$

This is always true for iid data models with

$$I = I(f_1, f_0) = \mathbb{E}_0 Z_1 = \int \log \left[\frac{f_1(x)}{f_0(x)} \right] f_1(x) d\mu(x)$$

being the Kullback–Leibler information number. It turns out that the a.s. convergence condition (82) is sufficient for obtaining lower bounds for all positive moments of the detection delay.

Next, for any $0 < \beta < 1$, $m^* \geq 1$ and $k^* > m^*$, define

$$\alpha_1 = \alpha_1(\beta, m^*) = \beta + (1 - \varrho_{1,\beta})^{m^*+1} \quad (83)$$

and

$$\alpha_2 = \alpha_2(\beta, k^*) = \beta(1 - \varrho_{2,\beta})^{k^*}, \quad (84)$$

where

$$\varrho_{1,\beta} = \frac{1}{1 + |\log \beta|} \quad \text{and} \quad \varrho_{2,\beta} = \frac{\varrho_{1,\beta}}{1 + |\log |\log \beta||}. \quad (85)$$

To find asymptotic lower bounds for the problems (80) and (81) in addition to condition (A₁) we impose the following condition related to the growth of the window size m^* in the LPFA:

(H₁) The size of the window m^* in (83) is a function of β , i.e. $m^* = m_\beta^*$, such that

$$\lim_{\beta \rightarrow 0} \frac{|\log \alpha_{1,\beta}|}{|\log \beta|} = 1, \quad (86)$$

where $\alpha_{1,\beta} = \alpha_1(\beta, m_\beta^*)$.

For example, we can take $m_\beta^* = 1 + \lfloor (1 + |\log \beta|)^2 \rfloor$.

The following theorem establishes asymptotic lower bounds.

Theorem 10. Assume that conditions $(\mathbf{A}_1)$ and $(\mathbf{H}_1)$ hold. Then, for any $k^* > m^*$ and $\nu \geq 0$,

$$\liminf_{\beta \rightarrow 0} \frac{1}{|\log \beta|} \inf_{\tau \in \mathcal{H}^*(\beta, k^*, m^*)} \sup_{\nu \geq 0} \mathcal{R}_\nu^*(\tau) \geq \liminf_{\beta \rightarrow 0} \frac{1}{|\log \beta|} \inf_{\tau \in \mathcal{H}^*(\beta, k^*, m^*)} \mathcal{R}_\nu^*(\tau) \geq \frac{1}{I}. \quad (87)$$

In order to study asymptotics for the average detection delay of the SR procedure and for establishing its asymptotic optimality, we impose the following constraint on the rate of convergence for

$$\tilde{\lambda}_{k,n} = \frac{1}{n} \lambda_{k+n}^k - I. \quad (88)$$

$(\mathbf{A}_2)$ Assume that $\tilde{\lambda}_{k,n}$ converges uniformly completely to 0 as $n \rightarrow \infty$, i.e., for any $\varepsilon > 0$

$$\Upsilon^*(\varepsilon) = \sum_{n=1}^{\infty} \sup_{k \geq 0} \mathbb{P}_k \left\{ |\tilde{\lambda}_{k,n}| > \varepsilon \right\} < \infty. \quad (89)$$

To establish asymptotic optimality properties of the SR procedure with respect to the risks $\mathcal{R}_\nu^*(\tau)$ (for all $\nu \geq 0$) and $\sup_{\nu \geq 0} \mathcal{R}_\nu^*(\tau)$ in the class $\mathcal{H}^*(\beta, k^*, m^*)$ we need the uniform complete convergence condition $(\mathbf{A}_2)$ as well as the following condition.

$(\mathbf{H}_2)$ Parameters k^* and m^* are functions of β , i.e. $k^* = k_\beta^*$ and $m^* = m_\beta^*$, such that

$$\lim_{\beta \rightarrow 0} \left(|\log \alpha_{3,\beta}| + k_\beta^* \log(1 - \varrho_{2,\beta}) \right) = +\infty \quad \text{and} \quad \lim_{\beta \rightarrow 0} \frac{|\log \alpha_{3,\beta}|}{|\log \beta|} = 1. \quad (90)$$

where $\alpha_{3,\beta} = \alpha_3(\beta, k_\beta^*)$.

We can take, for example, the parameters $k^* = k_\beta^*$ and $m^* = m_\beta^*$ as

$$m_\beta^* = 1 + \lfloor (1 + |\log \beta|)^2 \rfloor \quad \text{and} \quad k_\beta^* = 2m_\beta^*. \quad (91)$$

Denote by T_β^* the SR procedure $T(h_\beta^*)$ defined in (77) with the threshold h_β^* given by

$$h_\beta^* = \frac{1 - \alpha_{3,\beta}}{\varrho_{2,\beta} \alpha_{3,\beta}}. \quad (92)$$

Theorem 11. If conditions $(\mathbf{H}_1)$ and $(\mathbf{H}_2)$ hold, then, for any $0 < \beta < 1$, the SR procedure T_β^* with the threshold h_β^* given by (92) belongs to the class $\mathcal{H}^*(\beta, k^*, m^*)$. Assume that in addition condition $(\mathbf{A}_2)$ is satisfied. Then the SR procedure T_β^* is first-order asymptotically uniformly pointwise optimal and minimax in the class $\mathcal{H}^*(\beta, k^*, m^*)$, i.e.,

$$\lim_{\beta \rightarrow 0} \frac{\inf_{\tau \in \mathcal{H}^*(\beta, k^*, m^*)} \mathcal{R}_\nu^*(\tau)}{\mathcal{R}_\nu^*(T_\beta^*)} = 1 \quad \text{for all fixed } \nu \geq 0. \quad (93)$$

and

$$\lim_{\beta \rightarrow 0} \frac{\inf_{\tau \in \mathcal{H}^*(\beta, k^*, m^*)} \max_{0 \leq \nu \leq k_\beta^*} \mathcal{R}_\nu^*(\tau)}{\max_{0 \leq \nu \leq k_\beta^*} \mathcal{R}_\nu^*(T_\beta^*)} = 1. \quad (94)$$

Also, as $\beta \rightarrow 0$, the following first-order asymptotic approximations hold for the pointwise and maximal risks:

$$\mathcal{R}_\nu^*(T_\beta^*) \sim \inf_{\tau \in \mathcal{H}(\beta, k^*, m^*)} \mathcal{R}_\nu^*(\tau) \sim \frac{|\log \beta|}{I} \quad \text{for any } \nu \geq 0 \quad (95)$$

and

$$\sup_{0 \leq \nu \leq k_\beta^*} \mathcal{R}_\nu^*(T_\beta^*) \sim \inf_{\tau \in \mathcal{H}(\beta, k^*, m^*)} \sup_{0 \leq \nu \leq k_\beta^*} \mathcal{R}_\nu^*(\tau) \sim \frac{|\log \beta|}{I}. \quad (96)$$

The results of Theorem 11 can be extended to higher moments of the detection delay by strengthening the complete convergence with the uniform r -complete convergence. More specifically, the following asymptotic optimality result holds true.

Theorem 12. Assume that conditions (H_1) and (H_2) hold, and in addition, for some $r > 1$ the uniform r -complete convergence condition

$$\sum_{n=1}^{\infty} n^{r-1} \sup_{k \geq 0} P_k \left\{ \left| \tilde{\lambda}_{k,n} \right| > \varepsilon \right\} < \infty \quad \text{for all } \varepsilon > 0 \quad (97)$$

is satisfied. Then, for any $0 < \beta < 1$, the SR procedure T_β^* with the threshold h_β^* given by (92) belongs to the class $\mathcal{H}^*(\beta, k^*, m^*)$ and as $\beta \rightarrow 0$ for any $0 < \ell \leq r$

$$\begin{aligned} E_\nu \left[(T_\beta^* - \nu)^\ell | T_\beta^* > \nu \right] &\sim \inf_{\tau \in \mathcal{H}^*(\beta, k^*, m^*)} E_\nu \left[(\tau - \nu)^\ell | \tau > \nu \right] \\ &\sim \left(\frac{|\log \beta|}{I} \right)^\ell \quad \text{for all } \nu \geq 0 \end{aligned} \quad (98)$$

and

$$\begin{aligned} \max_{0 \leq \nu \leq k_\beta^*} E_\nu \left[(T_\beta^* - \nu)^\ell | T_\beta^* > \nu \right] &\sim \inf_{\tau \in \mathcal{H}^*(\beta, k^*, m^*)} \max_{0 \leq \nu \leq k_\beta^*} E_\nu \left[(\tau - \nu)^\ell | \tau > \nu \right] \\ &\sim \left(\frac{|\log \beta|}{I} \right)^\ell. \end{aligned} \quad (99)$$

Therefore, the SR procedure T_β^* is first-order asymptotically uniformly pointwise optimal and also minimax in the class $\mathcal{H}^*(\beta, k^*, m^*)$ with respect to the moments of the detection delay up to order r .

3. POTENTIAL IMPACTS

The research produces general theories of sequential hypothesis testing and quickest changepoint detection for very general non-iid stochastic models, as well as novel nearly optimal tests of composite hypotheses and changepoint detection procedures that significantly impact the effectiveness of DOD in recognizing unusual patterns of activity in heterogeneous volumes of data and automatic threat detection. We believe that our research results in practical and scalable algorithms for on-line detection and recognition of threats, in particular in cyber security applications related to rapid detection of intrusions in computer networks with very low false alarm rates.

4. SCIENTIFIC PERSONNEL SUPPORTED BY THIS PROJECT

1. Alexander Tartakovsky, Professor, Department of Statistics
2. Anthony Labarga, PhD student, Department of Statistics (fall 2014)
3. Djedjiga Belfadel, Ph.D. student, Department of Electrical and Computer Engineering (fall 2014)

5. REQUIRED NUMERICAL DATA RELATED TO THIS GRANT

(1) List of papers, books and book chapters submitted or published under ARO sponsorship during reporting period

(a) Books, book chapters and manuscripts submitted, but not published:

1. A. G. Tartakovsky, On Asymptotic Optimality in Sequential Changepoint Detection: Non-iid Case. *Bernoulli*, Submitted in 2016.
2. S. Pergamenchtchikov and A. G. Tartakovsky, Asymptotically Optimal Pointwise and Minimax Quickest Change-Point Detection for Dependent Data. *Statistical Inference for Stochastic Processes*, Submitted in 2016.
3. G. Fellouris and A.G. Tartakovsky, Multichannel Sequential Detection—Part I: Non-i.i.d. Data. *IEEE Transactions on Information Theory*, Submitted in 2016.

(b) Books, book chapters and papers published in peer-reviewed journals:

1. A.S. Polunchenko, G. Sokolov, and A.G. Tartakovsky, Optimal Design and Analysis of the Exponentially Weighted Moving Average Chart for Exponential Data, *Sri Lankan Journal of Applied Statistics*, Vol. 15, pp. 55–82, 2014.
2. A.G. Tartakovsky, Nearly Optimal Sequential Tests of Composite Hypotheses Revisited, *Proceedings of the Steklov Institute of Mathematics*, Vol. 287, pp. 268–288, 2014 (Invited Paper for Special Issue in honor of the 80th birthday of Professor Albert Shiryaev).
3. **BOOK:** A.G. Tartakovsky, I. Nikiforov, and M. Basseville, *Sequential Analysis: Hypothesis Testing and Changepoint Detection* (Monographs on Statistics and Applied Probability.) Chapman & Hall/CRC, Boca Raton, FL, 2014.
4. **BOOK CHAPTER:** A. G. Tartakovsky, *Rapid Detection of Attacks in Computer Networks by Quickest Changepoint Detection Methods* (In N. Adams and N. Heard, editors, *Data Analysis for Network Cyber-Security*), pages 33–70. Imperial College Press, London, UK, 2014.

(c) Papers published in non-peer-reviewed journals or in conference proceedings: None

(d) Papers presented at meetings, but not published in conference proceedings:

1. A.G. Tartakovsky (joint with G. Fellouris), Optimal Unstructured Sequential Detection in Multiple Channel Systems, *The 7th International Workshop on Applied Probability*, Antalya, Turkey, June 16-19, 2014 (Invited).

(2) Demographic Data for this Reporting Period:

- (a) Number of Manuscripts submitted: 7
- (b) Number of Peer Reviewed Papers: 5
- (c) Number of books and/or book chapters submitted or published: 2
- (c) Number of Non-Peer Reviewed Papers submitted during this reporting period: 0
- (d) Number of Presented but not Published Papers submitted during this reporting period: 1

(3) Demographic Data for the life of this agreement:

- (a) Number of Scientists Supported by this agreement: 3
- (b) Number of Inventions resulting from this agreement: 0
- (c) Number of PhD(s) awarded as a result of this agreement: 0
- (d) Number of Bachelor Degrees awarded as a result of this agreement: 0
- (e) Number of Patents Submitted as a result of this agreement: 0
- (f) Number of Patents Awarded as a result of this agreement: 0
- (g) Number of Grad Students supported by this agreement: 2
- (h) Number of FTE Grad Students supported by this agreement: 0
- (i) Number of Post Doctorates supported by this agreement: 0
- (j) Number of FTE Post Doctorates supported by this agreement: 0
- (k) Number of Faculty supported by this agreement: 1
- (l) Number of Other Staff supported by this agreement: 0
- (m) Number of Undergrads supported by this agreement: 0
- (n) Number of Master Degrees awarded as a result of this agreement: 0

(4) Student Metrics for graduating undergraduates funded by this agreement

- (a) Number of undergraduates funded by your agreement during this reporting period: 0
- (b) Number of undergraduate funded by your agreement, who graduated during this period: 0
- (c) Number of undergraduates funded by your agreement, who graduated during this period with a degree in a science, mathematics, engineering, or technology field: 0
- (d) Number of undergraduates funded by your agreement, who graduated during this period and will continue to pursue a graduate or Ph.D degree in a science, mathematics, engineering, or technology field: 0
- (e) Number of undergraduates funded by your agreement, who graduated during this period and intend to work for the Defense Department: 0
- (f) Number of undergraduates graduating during this period, who achieved at least a 3.5 GPA based on a scale with a maximum of a 4.0 GPA. (Convert GPAs on any other scale to be an equivalent value on a 4.0 scale.): 0
- (g) Number of undergraduates working on your agreement, who graduated during this period and were funded by a DoD funded Center of Excellence for Education, Research or Engineering: 0
- (h) Number of undergraduates funded by your agreement, who graduated during this period and will receive a scholarship or fellowship for further studies in a science, mathematics, engineering or technology field: 0

(5) Report of inventions

None.

REFERENCES

- P. A. Bakut, I. A. Bolshakov, B. M. Gerasimov, A. A. Kuriksha, V. G. Repin, G. P. Tartakovsky, and V. V. Shirokov. *Statistical Radar Theory*, volume 1 (G. P. Tartakovsky, Editor). Sovetskoe Radio, Moscow, USSR, 1963. In Russian.
- M. Baron and A. G. Tartakovsky. Asymptotic optimality of change-point detection schemes in general continuous-time models. *Sequential Analysis*, 25(3):257–296, Oct. 2006. Invited Paper in Memory of Milton Sobel.
- M. Basseville and I. V. Nikiforov. *Detection of Abrupt Changes – Theory and Application*. Information and System Sciences Series. Prentice-Hall, Inc, Englewood Cliffs, NJ, USA, 1993. ISBN 0131267809. Online.
- H. P. Chan and T. L. Lai. Asymptotic approximations for error probabilities of sequential or fixed sample size tests in exponential families. *Annals of Statistics*, 28(6):1638–1669, Dec. 2000.
- F.-K. Chang. Structural health monitoring: Promises and challenges. In *Proceedings of the 30th Annual Review of Progress in Quantitative NDE (QNDE)*, Green Bay, WI, USA. American Institute of Physics, July 2003.
- S. E. Fienberg and G. Shmueli. Statistical issues and challenges associated with rapid detection of bio-terrorist attacks. *Statistics in Medicine*, 24(4):513–529, July 2005.
- C.-D. Fuh. SPRT and CUSUM in Hidden Markov Models. *Annals of Statistics*, 31(3):942–977, June 2003.
- P. L. Hsu and H. Robbins. Complete convergence and the law of large numbers. *Proceedings of the National Academy of Sciences of the United States of America*, 33(2):25–31, Feb. 1947.
- S. Kent. On the trail of intrusions into information systems. *IEEE Spectrum*, 37(12):52–56, Dec. 2000.
- T. L. Lai. Asymptotic optimality of invariant sequential probability ratio tests. *Annals of Statistics*, 9(2):318–333, Mar. 1981.
- T. L. Lai. Sequential changepoint detection in quality control and dynamical systems (with discussion). *Journal of the Royal Statistical Society - Series B Methodology*, 57(4):613–658, 1995.
- T. L. Lai. Information bounds and quick detection of parameter changes in stochastic systems. *IEEE Transactions on Information Theory*, 44(7):2917–2929, Nov. 1998.
- G. Lorden. Procedures for reacting to a change in distribution. *Annals of Mathematical Statistics*, 42(6):1897–1908, Dec. 1971.
- G. Lorden. Nearly optimal sequential tests for exponential families. *Unpublished Manuscript*, 1977.
- Y. Mei. Is average run length to false alarm always an informative criterion? *Sequential Analysis*, 27(4):354–376, Oct. 2008.

- G. V. Moustakides. Optimal stopping times for detecting changes in distributions. *Annals of Statistics*, 14(4):1379–1387, Dec. 1986.
- S. Pergamenchtchikov and A. G. Tartakovsky. Asymptotically optimal pointwise and minimax quickest change-point detection for dependent data. *Statistical Inference for Stochastic Processes*, Submitted in 2016.
- M. Pollak. Optimal detection of a change in distribution. *Annals of Statistics*, 13(1):206–227, Mar. 1985.
- M. Pollak and A. G. Tartakovsky. Optimality properties of the Shiryaev–Roberts procedure. *Statistica Sinica*, 19(4):1729–1739, Oct. 2009a.
- M. Pollak and A. G. Tartakovsky. Asymptotic exponentiality of the distribution of first exit times for a class of Markov processes with applications to quickest change detection. *Theory of Probability and its Applications*, 53(3):430–442, 2009b.
- S. W. Roberts. A comparison of some control chart procedures. *Technometrics*, 8(3):411–430, Aug. 1966.
- H. Rolka, H. Burkom, G. F. Cooper, M. Kulldorff, D. Madigan, and W. K. Wong. Issues in applied statistics for public health bioterrorism surveillance using multiple data streams: research needs. *Statistics in Medicine*, 26(8):1834–1856, 2007.
- A. N. Shiryaev. On optimum methods in quickest detection problems. *Theory of Probability and its Applications*, 8(1):22–46, 1963.
- D. Siegmund. Change-points: From sequential detection to biology and back. *Sequential Analysis*, 32(1):2–14, Jan. 2013.
- C. Sonesson and D. Bock. A review and discussion of prospective statistical surveillance in public health. *Journal of the Royal Statistical Society A*, 166:5–21, 2003.
- P. Szor. *The Art of Computer Virus Research and Defense*. Addison-Wesley Professional, Upper Saddle River, NJ, USA, 2005.
- A. G. Tartakovsky. Asymptotic optimality of certain multihypothesis sequential tests: Non-i.i.d. case. *Statistical Inference for Stochastic Processes*, 1(3):265–295, Oct. 1998.
- A. G. Tartakovsky. Asymptotic performance of a multichart CUSUM test under false alarm probability constraint. In *Proceedings of the 44th IEEE Conference Decision and Control and European Control Conference (CDC-ECC'05), Seville, SP*, pages 320–325. IEEE, Omnipress CD-ROM, 2005. ISBN 0780395689.
- A. G. Tartakovsky. Discussion on “Is average run length to false alarm always an informative criterion?” by Yajun Mei. *Sequential Analysis*, 27(4):396–405, Oct. 2008.

- A. G. Tartakovsky. General multidecision theory: Hypothesis testing and changepoint detection with applications to homeland security. Interim Progress Report, ARO Grant W911NF-13-1-0073, Department of Mathematics, University of Southern California, Los Angeles, CA, USA, 2013a.
- A. G. Tartakovsky. Rapid detection of attacks in computer networks by quickest changepoint detection methods. In *Proceedings of the Data Analysis for Cybersecurity Workshop, Heilbronn Institute for Mathematical Research, Bristol, UK*, Mar. 2013b.
- A. G. Tartakovsky. Discussion on “Change-points: From sequential detection to biology and back” by david siegmund. *Sequential Analysis*, 32(1):36–42, Jan. 2013c.
- A. G. Tartakovsky. Rapid detection of attacks in computer networks by quickest changepoint detection methods. In N. Adams and N. Heard, editors, *Data Analysis for Network Cyber-Security*, pages 33–70. Imperial College Press, London, UK, 2014.
- A. G. Tartakovsky. On asymptotic optimality in sequential changepoint detection: Non-iid case. *Bernoulli*, Submitted, Submitted in 2016.
- A. G. Tartakovsky and J. Brown. Adaptive spatial-temporal filtering methods for clutter removal and target tracking. *IEEE Transactions on Aerospace and Electronic Systems*, 44(4):1522–1537, Oct. 2008.
- A. G. Tartakovsky and V. V. Veeravalli. General asymptotic Bayesian theory of quickest change detection. *Theory of Probability and its Applications*, 49(3):458–497, 2005.
- A. G. Tartakovsky, X. R. Li, and G. Yaralov. Sequential detection of targets in multichannel systems. *IEEE Transactions on Information Theory*, 49(2):425–445, Feb. 2003a.
- A. G. Tartakovsky, X. R. Li, and G. Yaralov. Sequential detection of targets in multichannel systems. *IEEE Transactions on Information Theory*, 49(2):425–445, Feb. 2003b.
- A. G. Tartakovsky, B. L. Rozovskii, R. B. Blažek, and H. Kim. A novel approach to detection of intrusions in computer networks via adaptive sequential and batch-sequential change-point detection methods. *IEEE Transactions on Signal Processing*, 54(9):3372–3382, Sept. 2006a.
- A. G. Tartakovsky, B. L. Rozovskii, R. B. Blažek, and H. Kim. Detection of intrusions in information systems by sequential change-point methods. *Statistical Methodology*, 3(3):252–293, July 2006b.
- A. G. Tartakovsky, I. Nikiforov, and M. Basseville. *Sequential Analysis: Hypothesis Testing and Changepoint Detection*. Monographs on Statistics and Applied Probability. Chapman & Hall/CRC, Boca Raton, FL, 2014a.
- A. G. Tartakovsky, I. V. Nikiforov, and M. Basseville. *Sequential Analysis: Hypothesis Testing and Changepoint Detection*. Monographs on Statistics and Applied Probability. Chapman & Hall/CRC Press, Boca Raton, London, New York, 2014b.
- K.-L. Tsui, S. W. Han, W. Jiang, and W. H. Woodall. A review and comparison of likelihood based charting methods. *IIE Transactions*, 44(9):724–743, Sept. 2012.