

Article

Quick and Complete Convergence in the Law of Large Numbers with Applications to Statistics

Alexander G. Tartakovsky 

AGT StatConsult, 71 Cypress Way, Rolling Hills Estates, CA 90274, USA; alexg.tartakovsky@gmail.com;
Tel.: +1-310-292-7847

Abstract: In the first part of this article, we discuss and generalize the complete convergence introduced by Hsu and Robbins in 1947 to the r -complete convergence introduced by Tartakovsky in 1998. We also establish its relation to the r -quick convergence first introduced by Strassen in 1967 and extensively studied by Lai. Our work is motivated by various statistical problems, mostly in sequential analysis. As we show in the second part, generalizing and studying these convergence modes is important not only in probability theory but also to solve challenging statistical problems in hypothesis testing and changepoint detection for general stochastic non-i.i.d. models.

Keywords: complete convergence; r -quick convergence; sequential analysis; hypothesis testing; changepoint detection.

MSC: 60F15; 60G35; 60G40; 60J05; 62L10; 62C10; 62C20; 62F03; 62H15; 62M02; 62P30



Citation: Tartakovsky, A.G. Quick and Complete Convergence in the LLN with Applications to Statistics. *Mathematics* **2023**, *11*, 2687. <https://doi.org/10.3390/math11122687>

Academic Editors: Irina Shevtsova and Victor Korolev

Received: 11 May 2023

Revised: 6 June 2023

Accepted: 7 June 2023

Published: 13 June 2023



Copyright: © 2023 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In [1], Hsu and Robbins introduced the notion of complete convergence which is stronger than almost sure (a.s.) convergence. Hsu and Robbins used this notion to discuss certain aspects of the law of large numbers (LLN). In particular, let $X_1, X_2, \dots$ be independent and identically distributed (i.i.d.) random variables with the common mean $\mu = E[X_1]$. Hsu and Robbins proved that, while in Kolmogorov's strong law of large numbers (SLLN), only the first moment condition is needed for the sample mean $n^{-1} \sum_{t=1}^n X_t$ to converge to μ as $n \rightarrow \infty$, the complete version of the SLLN requires the second-moment condition $E|X_1|^2 < \infty$ (finiteness of variance). Later, Baum and Katz [2], working on the rate of convergence in the LLN established that the second-moment condition is not only necessary but also sufficient for complete convergence. Strassen [3] introduced another mode of convergence, the r -quick convergence. When $r = 1$, these two modes of convergence are closely related. In the case of i.i.d. random variables and the sample mean $n^{-1} \sum_{t=1}^n X_t$, they are identical. This fact and certain statistical applications motivated Tartakovsky [4] (see also Tartakovsky [5] and Tartakovsky et al. [6]) to introduce a natural generalization of complete convergence—the r -complete convergence, which turns out to be identical to the r -quick convergence in the i.i.d. case.

The goal of this overview paper is to discuss the importance of quick and complete convergence concepts for several challenging statistical applications. These modes of convergence are discussed in detail in the first part of this paper. Statistical applications, which constitute the second part of this paper, include such fields as sequential hypothesis testing and changepoint detection in general non-i.i.d. stochastic models when observations can be dependent and highly non-stationary. Specifically, in the second part, we first address near optimality of Wald's sequential probability ratio test (SPRT) for testing two hypotheses regarding the distributions of non-i.i.d. data. We discuss Lai's results in his fundamental paper [7], which was the first publication that used the r -quick convergence of the log-likelihood ratio processes to establish the asymptotic optimality of the SPRT as probabilities

of errors go to zero. We then go on to tackle the much more difficult multi-decision problem of testing multiple hypotheses and show that certain multi-hypothesis sequential tests asymptotically minimize moments of the stopping time distribution up to the order r when properly normalized log-likelihood ratio processes between hypotheses converge r -quickly or r -completely to finite positive numbers. These results can be established based on the former works of the author (see, e.g., Tartakovsky [4,5] and Tartakovsky et al. [6]). The second challenging application is the quickest change detection when it is necessary to detect a change that occurs at an unknown point in time as rapidly as possible. We show, using the works of the author (see, e.g., [5,6] and the references therein), that certain popular changepoint detection procedures such as CUSUM, Shiryaev, and Shiryaev–Roberts procedures are asymptotically optimal as the false alarm rate is low when the normalized log-likelihood ratio processes converge r -completely to finite numbers.

The rest of the paper is organized as follows. Section 2 discusses pure probabilistic issues related to r -complete convergence and r -quick convergence. Section 3 explores statistical applications in sequential hypothesis testing and changepoint detection. Section 4 outlines sufficient conditions for the r -complete convergence for Markov and hidden Markov models, which is needed to establish the optimality properties of sequential hypothesis tests and changepoint detection procedures. Section 5 provides a final discussion and concludes the paper.

2. Modes of Convergence and the Law of Large Numbers

We begin by listing some standard definitions in probability theory. Let $(\Omega, \mathcal{F})$ be a measurable space, i.e., Ω is a set of elementary events ω and $\mathcal{F}$ is a sigma-algebra (a system of subsets of Ω satisfying standard conditions). A probability space is a triple $(\Omega, \mathcal{F}, P)$, where P is a probability measure (completely additive measure normalized to 1) defined on the sets from the sigma-algebra $\mathcal{F}$. More specifically, by Kolmogorov's axioms, probability P satisfies: $P(\mathcal{A}) \geq 0$ for any $\mathcal{A} \in \mathcal{F}$; $P(\Omega) = 1$; and $P(\cup_{i=1}^{\infty} \mathcal{A}_i) = \sum_{i=1}^{\infty} P(\mathcal{A}_i)$ for $\mathcal{A}_i \in \mathcal{F}$, $\mathcal{A}_i \cap \mathcal{A}_j = \emptyset$, $i \neq j$, where $\emptyset$ is an empty set.

A function $X = X(\omega)$ defined on $(\Omega, \mathcal{F})$ with values in $\mathcal{X}$ is called a random variable if it is $\mathcal{F}$ -measurable, i.e., $\{\omega : X(\omega) \in B\}$ belongs to the sigma-algebra $\mathcal{F}$. The function $F(x) = P(\omega : X(\omega) \leq x)$ is the distribution function of X . It is also referred to as a cumulative distribution function (cdf). The real-valued random variables $X_1, X_2, \dots$ are independent if the events $\{X_1 \leq x_1\}, \{X_2 \leq x_2\}, \dots$ are independent for every sequence $x_1, x_2, \dots$ of real numbers. In what follows, we shall deal with real-valued random variables unless specified otherwise.

2.1. Standard Modes of Convergence

Let X be a random variable and let $\{X_n\}_{n \in \mathbb{Z}_+}$ ($\mathbb{Z}_+ = \{0, 1, 2, \dots\}$) be a sequence of random variables, both defined on the probability space $(\Omega, \mathcal{F}, P)$. We now give several standard definitions and results related to the law of large numbers.

Convergence in Distribution (Weak Convergence). Let $F_n(x) = P(\omega : X_n \leq x)$ be the cdf of X_n and let $F(x) = P(\omega : X \leq x)$ be the cdf of X . We say that the sequence $\{X_n\}_{n \in \mathbb{Z}_+}$ converges to X in distribution (or in law or weakly) as $n \rightarrow \infty$ and write $X_n \xrightarrow[n \rightarrow \infty]{\text{law}} X$ if

$$\lim_{n \rightarrow \infty} F_n(x) = F(x)$$

at all continuity points of $F(x)$.

Convergence in Probability. We say that the sequence $\{X_n\}_{n \in \mathbb{Z}_+}$ converges to X in probability as $n \rightarrow \infty$ and write $X_n \xrightarrow[n \rightarrow \infty]{P} X$ if

$$\lim_{n \rightarrow \infty} P(|X_n - X| > \varepsilon) = 0 \quad \text{for every } \varepsilon > 0.$$

Almost Sure Convergence. We say that the sequence $\{X_n\}_{n \in \mathbb{Z}_+}$ converges to X almost surely (a.s.) or with probability 1 (w.p. 1) as $n \rightarrow \infty$ under probability measure P and write $X_n \xrightarrow[n \rightarrow \infty]{P\text{-a.s.}} X$ if

$$P\left(\omega : \lim_{n \rightarrow \infty} X_n = X\right) = 1. \quad (1)$$

It is easily seen that (1) is equivalent to the condition

$$\lim_{n \rightarrow \infty} P\left(\omega : \sum_{t=n}^{\infty} |X_t - X| > \varepsilon\right) = 0 \quad \text{for every } \varepsilon > 0,$$

and that the a.s. convergence implies convergence in probability, and the convergence in probability implies convergence in distribution, while the converse statements are not generally true.

The following double implications that establish necessary and sufficient conditions (i.e., equivalences) for the a.s. convergence are useful:

$$X_n \xrightarrow[n \rightarrow \infty]{\text{a.s.}} X \iff P\left(\sup_{t \geq n} |X_t - X| > \varepsilon\right) \xrightarrow[n \rightarrow \infty]{} 0 \quad \text{for all } \varepsilon > 0. \quad (2)$$

The following result is often useful.

Lemma 1. Let $f(t)$ be a non-negative increasing function, $\lim_{t \rightarrow \infty} f(t) = \infty$. If

$$\frac{X_n}{f(n)} \xrightarrow[n \rightarrow \infty]{P\text{-a.s.}} 0,$$

then

$$\lim_{n \rightarrow \infty} P\left(\frac{1}{f(n)} \max_{0 \leq t \leq n} X_t > \varepsilon\right) = 0 \quad \text{for every } \varepsilon > 0. \quad (3)$$

Proof. For any $\varepsilon > 0$, $n_0 > 0$ and $n > n_0$, we have

$$\begin{aligned} P\left(\frac{1}{f(n)} \max_{0 \leq t \leq n} X_t > \varepsilon\right) &\leq P\left(\frac{1}{f(n)} \max_{0 \leq t \leq n_0} X_t > \varepsilon\right) + P\left(\frac{1}{f(n)} \max_{n_0 < t \leq n} X_t > \varepsilon\right) \\ &\leq P\left(\frac{1}{f(n)} \max_{0 \leq t \leq n_0} X_t > \varepsilon\right) + P\left(\sup_{t > n_0} \frac{X_t}{f(t)} > \varepsilon\right). \end{aligned}$$

Letting $n \rightarrow \infty$ and taking into account that

$$\lim_{n \rightarrow \infty} P\left(\frac{1}{f(n)} \max_{0 \leq t \leq n_0} X_t > \varepsilon\right) = 0,$$

we obtain

$$\limsup_{n \rightarrow \infty} P\left(\frac{1}{f(n)} \max_{0 \leq t \leq n} X_t > \varepsilon\right) \leq P\left(\sup_{t > n_0} \frac{X_t}{f(t)} > \varepsilon\right).$$

Since n_0 can be arbitrarily large, we can let $n_0 \rightarrow \infty$ and since, by assumption, $X_n/f(n) \xrightarrow[n \rightarrow \infty]{\text{a.s.}} 0$, it follows from (2) that the upper bound approaches 0 as $n_0 \rightarrow \infty$. This completes the proof. $\square$

Random Walk. Let $X_0, X_1, X_2, \dots$ be i.i.d. random variables with the mean $E[X_n] = \mu$ for $n \geq 1$ and the initial condition $X_0 = x$. Then, $S_n = \sum_{t=0}^n X_t$ is called a random walk with the mean $x + \mu n$.

In what follows, in the case where $X_1, X_2, \dots$ are i.i.d. random variables and $S_n = \sum_{t=0}^n X_t$, we prefer to formulate the results in terms of the random walk $\{S_n\}_{n \in \mathbb{Z}_+}$ (typically but not necessarily $S_0 = 0$).

We now recall the two strong law of large numbers (SLLN). Write $S_n = X_0 + X_1 + \dots + X_n$ for the partial sum ($X_0 = S_0 = 0$), so that $\{S_n\}_{n \in \mathbb{Z}_+}$ is a random walk with an initial condition of zero as long as $X_1, X_2, \dots$ are i.i.d. with mean μ .

Kolmogorov's SLLN. Let $\{S_n\}_{n \in \mathbb{Z}_+}$ be a random walk under probability measure P . If $E[S_1]$ exists, then the sample mean S_n/n converges to the mean value $E[S_1]$ w.p. 1, i.e.,

$$n^{-1}S_n \xrightarrow[n \rightarrow \infty]{P\text{-a.s.}} E[S_1]. \quad (4)$$

Conversely, if $n^{-1}S_n \xrightarrow[n \rightarrow \infty]{P\text{-a.s.}} \mu$, where $|\mu| < \infty$, then $E[S_1] = \mu$.

Marcinkiewicz–Zygmund's SLLN. Let $\{S_n\}_{n \in \mathbb{Z}_+}$ be a zero-mean random walk under probability measure P . The two following statements are equivalent:

- (i) $E|S_1|^p < \infty$ for $0 < p < 2$;
- (ii) $n^{-1/p}S_n \xrightarrow[n \rightarrow \infty]{P\text{-a.s.}} 0$.

2.2. Complete and r -Complete Convergence

We begin with discussing the issue of rates of convergence in the LLN.

Rates of Convergence. Let $\{X_n\}_{n \in \mathbb{Z}_+}$ be a sequence of random variables and assume that X_n converges to 0 w.p. 1 as $n \rightarrow \infty$. The question asks what the rate of convergence is. In other words, we are concerned with the speed at which the tail probability $P(|X_n| > \varepsilon)$ decays to zero. This question can be answered by analyzing the behavior of the sums

$$\Sigma(r, \varepsilon) := \sum_{n=1}^{\infty} n^{r-1} P(|X_n| > \varepsilon) \quad \text{for some } r > 0 \text{ and all } \varepsilon > 0.$$

More specifically, if $\Sigma(r, \varepsilon)$ is finite for every $\varepsilon > 0$, then the tail probability $P(|X_n| > \varepsilon)$ decays with a rate faster than $1/n^r$, so that $n^r P(|X_n| > \varepsilon) \rightarrow 0$ for all $\varepsilon > 0$ as $n \rightarrow \infty$.

To answer this question, we now consider modes of convergence that strengthen the almost sure convergence and therefore help determine the rate of convergence in the SLLN. Historically, this issue was first addressed in 1947 by Hsu and Robbins [1], who introduced the new mode of convergence which they called *complete convergence*.

Complete Convergence. The sequence $\{X_n\}_{n \in \mathbb{Z}_+}$ converges to 0 *completely* if

$$\lim_{n \rightarrow \infty} \sum_{i=n}^{\infty} P(|X_i| > \varepsilon) = 0 \quad \text{for every } \varepsilon > 0, \quad (5)$$

which is equivalent to

$$\Sigma(1, \varepsilon) = \sum_{n=1}^{\infty} P(|X_n| > \varepsilon) < \infty \quad \text{for every } \varepsilon > 0$$

Let $\{S_n\}_{n \in \mathbb{Z}_+}$ be a random walk with a mean of $E[S_n] = \mu n$. Kolmogorov's SLLN (4) implies that the sample mean S_n/n converges to μ w.p. 1. Hsu and Robbins [1] proved that, under the same assumptions (i.e., under the only first-moment condition $E|S_1| < \infty$) the sequence $\{n^{-1}S_n\}_{n \geq 1}$ does not need to completely converge to μ , but it will do so under the further second-moment condition $E|S_1|^2 < \infty$. Thus, the finiteness of variance is a sufficient condition for complete convergence in the SLLN. They conjectured that the second-moment condition is not only sufficient but also necessary for complete convergence. Thus, it follows from these results that, if the variance is finite, then the rate of convergence in Kolmogorov's SLLN is $\lim_{n \rightarrow \infty} n P(|S_n/n - \mu| > \varepsilon) = 0$ for all $\varepsilon > 0$.

In 1965, Baum and Katz [2] made a further step towards this issue. In particular, the following result follows from Theorem 3 in [2] for the zero-mean random walk $\{S_n\}_{n \in \mathbb{Z}_+}$.

Theorem 1. Let $r > 0$ and $\alpha > 1/2$. If $\{S_n\}_{n \in \mathbb{Z}_+}$ is a zero-mean random walk, then the following statements are equivalent:

$$\begin{aligned} \mathbb{E}[|S_1|^{(r+1)/\alpha}] < \infty &\iff \sum_{n=1}^{\infty} n^{r-1} \mathbb{P}\{n^{-\alpha}|S_n| > \varepsilon\} < \infty \text{ for all } \varepsilon > 0 \\ &\iff \sum_{n=1}^{\infty} n^{r-1} \mathbb{P}\left\{\sup_{k \geq n} \frac{1}{k^\alpha} |S_k| > \varepsilon\right\} < \infty \text{ for all } \varepsilon > 0. \end{aligned} \quad (6)$$

Setting $r = 1$ and $\alpha = 1$ in (6), we obtain the following equivalence

$$\mathbb{E}[|S_1|^2] < \infty \iff \sum_{n=1}^{\infty} \mathbb{P}\{|n^{-1}S_n| > \varepsilon\} < \infty \text{ for all } \varepsilon > 0,$$

which shows that the conjecture of Hsu and Robbins is correct—the second-moment condition $\mathbb{E}|S_1|^2 < \infty$ is both necessary and sufficient for complete convergence

$$n^{-1}S_n \xrightarrow[n \rightarrow \infty]{\text{P-completely}} 0.$$

Furthermore, if for some $r > 0$, the $(r+1)$ -th moment is finite, $\mathbb{E}|S_1|^{r+1} < \infty$, then the rate of convergence in the SLLN is $\lim_{n \rightarrow \infty} n^r \mathbb{P}(|n^{-1}S_n| > \varepsilon) = 0$ for all $\varepsilon > 0$.

Previous results suggest that it is reasonable to generalize the notion of complete convergence into the following mode of convergence that we will refer to as *r-complete convergence*, which is also related to the so-called *r-quick convergence* that we will discuss later on (see Section 2.3).

Definition 1 (*r-Complete Convergence*). Let $r > 0$. We say that the sequence of random variables $\{X_n\}_{n \in \mathbb{Z}_+}$ converges to X as $n \rightarrow \infty$ *r-completely* under probability measure $\mathbb{P}$ and write $X_n \xrightarrow[n \rightarrow \infty]{\text{P-r-completely}} X$ if

$$\Sigma(r, \varepsilon) := \sum_{n=1}^{\infty} n^{r-1} \mathbb{P}(|X_n - X| > \varepsilon) < \infty \text{ for every } \varepsilon > 0. \quad (7)$$

Note that the a.s. convergence of $\{X_n\}$ to X can be equivalently written as

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\sum_{t=n}^{\infty} |X_t - X| > \varepsilon\right) = 0 \text{ for every } \varepsilon > 0,$$

so that the *r-complete convergence* with $r \geq 1$ implies the a.s. convergence, but the converse is not true in general.

Suppose that X_n converges a.s. to X . If $\Sigma(r, \varepsilon)$ is finite for every $\varepsilon > 0$, then

$$\lim_{n \rightarrow \infty} \sum_{t=n}^{\infty} t^{r-1} \mathbb{P}(|X_t - X| > \varepsilon) = 0 \text{ for every } \varepsilon > 0$$

and probability $\mathbb{P}(|X_n - X| > \varepsilon)$ goes to 0 as $n \rightarrow \infty$ with the rate faster than $1/n^r$. Hence, as already mentioned above, the *r-complete convergence* allows one to determine the rate of convergence of X_n to X , i.e., to answer the question of how fast the tail probability $\mathbb{P}(|X_n - X| > \varepsilon)$ decays to zero.

The following result provides a very useful implication of complete convergence.

Theorem 2. Let $\{X_n\}_{n \in \mathbb{Z}_+}$ and $\{Y_n\}_{n \in \mathbb{Z}_+}$ be two arbitrary, possibly dependent sequences of random variables. Assume that there are positive and finite numbers μ_1 and μ_2 such that

$$\sum_{n=1}^{\infty} \mathbb{P}\left(\left|\frac{1}{n}X_n - \mu_1\right| > \varepsilon\right) < \infty \quad \text{for every } \varepsilon > 0 \quad (8)$$

and

$$\sum_{n=1}^{\infty} \mathbb{P}\left(\left|\frac{1}{n}Y_n - \mu_2\right| > \varepsilon\right) < \infty \quad \text{for every } \varepsilon > 0, \quad (9)$$

i.e., $n^{-1}X_n \xrightarrow[n \rightarrow \infty]{\text{P-completely}} \mu_1$ and $n^{-1}Y_n \xrightarrow[n \rightarrow \infty]{\text{P-completely}} \mu_2$. If $\mu_1 \geq \mu_2$, then for any random time T

$$\mathbb{P}(X_T < b, Y_{T+1} \geq b(1 + \delta)) \longrightarrow 0 \quad \text{as } b \rightarrow \infty \quad \text{for any } \delta > 0. \quad (10)$$

Proof. Fix $\delta > 0$, $c \in (0, \delta)$ and let $N_b = \lceil (1 + c)b/\mu_2 \rceil$ be the smallest integer that is larger than or equal to $(1 + c)b/\mu_2$. Observe that

$$\begin{aligned} \mathbb{P}(X_T < b, Y_{T+1} \geq b(1 + \delta)) &\leq \mathbb{P}(X_T \leq b, T \geq N_b) + \mathbb{P}(Y_{T+1} \geq (1 + \delta)b, T < N_b) \\ &\leq \mathbb{P}(X_T \leq b, T \geq N_b) + \mathbb{P}\left(\max_{1 \leq n \leq N_b} Y_n \geq (1 + \delta)b\right). \end{aligned}$$

Thus, to prove (10), it suffices to show that the two terms on the right-hand side go to 0 as $b \rightarrow \infty$.

For the first term, we notice that, for any $n \geq N_b$,

$$\frac{b}{n} \leq \frac{b}{N_b} \leq \frac{\mu_2}{1 + c} \leq \frac{\mu_1}{1 + c} < \mu_1,$$

so that

$$\begin{aligned} \mathbb{P}(X_T \leq b, T \geq N_b) &= \sum_{n=N_b}^{\infty} \mathbb{P}(X_n \leq b, T = n) \leq \sum_{n=N_b}^{\infty} \mathbb{P}\left(\frac{X_n}{n} \leq \frac{b}{n}\right) \\ &\leq \sum_{n=N_b}^{\infty} \mathbb{P}\left(\frac{X_n}{n} \leq \frac{\mu_1}{1 + c}\right) = \sum_{n=N_b}^{\infty} \mathbb{P}\left(\frac{X_n}{n} - \mu_1 \leq -\frac{c}{1 + c}\mu_1\right). \end{aligned}$$

Since $N_b \rightarrow \infty$ as $b \rightarrow \infty$, the upper bound goes to 0 as $b \rightarrow \infty$ due to condition (8).

Next, since $c \in (0, \delta)$, there exists $\varepsilon' > 0$ such that

$$\frac{(1 + \delta)b}{N_b} = \frac{(1 + \delta)b}{\lceil b(1 + c)/\mu_2 \rceil} \geq (1 + \varepsilon')\mu_2.$$

As a result,

$$\mathbb{P}\left(\max_{1 \leq n \leq N_b} Y_n \geq (1 + \delta)b\right) \leq \mathbb{P}\left(\frac{1}{N_b} \max_{1 \leq n \leq N_b} Y_n \geq (1 + \varepsilon')\mu_2\right),$$

where the upper bound goes to 0 as $b \rightarrow \infty$ by condition (9) (see Lemma 1). $\square$

Remark 1. The proof suggests that the assertion (10) of Theorem 2 holds under the following one-sided conditions

$$\mathbb{P}\left(n^{-1} \max_{1 \leq s \leq n} Y_s - \mu_2 > \varepsilon\right) \xrightarrow[n \rightarrow \infty]{} 0, \quad \sum_{n=1}^{\infty} \mathbb{P}\left(n^{-1}X_n - \mu_1 < -\varepsilon\right) < \infty.$$

Complete convergence conditions (8) and (9) guarantee both these conditions.

Remark 2. Theorem 2 can be applied to the overshoot problem. Indeed, if $X_n = Y_n = Z_n$ and the random time T is the first time n when Z_n exceeds the level b , $T = \inf\{n \geq 1 : Z_n > b\}$, then Theorem 2 shows that the relative excess of boundary crossing (overshoot) $(Z_T - b)/b$ converges to 0 in probability as $b \rightarrow \infty$ when Z_n/n completely converges as $n \rightarrow \infty$ to a positive number μ .

2.3. r -Quick Convergence

In 1967, Strassen [3] introduced the notion of r -quick limit points of a sequence of random variables. The r -quick convergence has been further addressed by Lai [7,8], Chow and Lai [9], Fuh and Zhang [10], and Tartakovsky [4,5] (see certain details in Section 2.4).

We define r -quick convergence in a way suitable for this paper. Let $\{X_n\}_{n \in \mathbb{Z}_+}$ be a sequence of real-valued random variables and let X be a random variable defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Definition 2 (r -Quick Convergence). Let $r > 0$ and for $\varepsilon > 0$, let

$$L_\varepsilon = \sup\{n \geq 1 : |X_n - X| > \varepsilon\} \quad (\sup\{\emptyset\} = 0)$$

be the last entry time of X_n in the region $(X + \varepsilon, \infty) \cup (-\infty, X - \varepsilon)$. We say that the sequence $\{X_n\}_{n \in \mathbb{Z}_+}$ converges to X r -quickly as $n \rightarrow \infty$ under the probability measure $\mathbb{P}$ and write $X_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}-r\text{-quickly}} X$ if and only if

$$\mathbb{E}[L_\varepsilon^r] < \infty \quad \text{for every } \varepsilon > 0, \quad (11)$$

where $\mathbb{E}$ is the operator of expectation under probability $\mathbb{P}$.

This definition can be generalized to random variables $X, \{X_n\}_{n \in \mathbb{Z}_+}$ taking values in a metric space $(\mathcal{X}, d)$ with distance d : $X_n \xrightarrow[n \rightarrow \infty]{r\text{-quickly}} X$ if

$$\mathbb{E}[(\sup\{n \geq 1 : d(X, X_n) > \varepsilon\})^r] < \infty \quad \text{for every } \varepsilon > 0.$$

Note that the a.s. convergence $X_n \rightarrow \mu$ ($|\mu| < \infty$) as $n \rightarrow \infty$ to a constant μ can be expressed as $\mathbb{P}(L_\varepsilon(\mu) < \infty) = 1$, where $L_\varepsilon(\mu) = \sup\{n \geq 1 : |X_n - \mu| > \varepsilon\}$. Therefore, the r -quick convergence implies the convergence w.p. 1 but not conversely.

Also, in general, r -quick convergence is stronger than r -complete convergence. Specifically, the following lemma shows that

$$\max_{1 \leq i \leq n} X_i \xrightarrow[n \rightarrow \infty]{r\text{-completely}} \mu \implies X_n \xrightarrow[n \rightarrow \infty]{r\text{-quickly}} \mu \implies X_n \xrightarrow[n \rightarrow \infty]{r\text{-completely}} \mu. \quad (12)$$

Lemma 2. Let $\{X_n\}_{n \in \mathbb{Z}_+}$ be a sequence of random variables. Let $f(t)$ be a non-negative increasing function, $f(0) = 0$, $\lim_{t \rightarrow \infty} f(t) = +\infty$, and let for $\varepsilon > 0$

$$L_\varepsilon(f) = \sup\{n \geq 1 : |X_n| > \varepsilon f(n)\} \quad (\sup\{\emptyset\} = 0)$$

be the last time that X_n leaves the interval $[-\varepsilon f(n), +\varepsilon f(n)]$.

(i) For any $r > 0$ and any $\varepsilon > 0$, the following inequalities hold:

$$r \sum_{n=1}^{\infty} n^{r-1} \mathbb{P}\{|X_n| \geq \varepsilon f(n)\} \leq \mathbb{E}[L_\varepsilon(f)^r] \leq r \sum_{n=1}^{\infty} n^{r-1} \mathbb{P}\left\{\sup_{t \geq n} \frac{|X_t|}{f(t)} \geq \varepsilon\right\}. \quad (13)$$

Therefore,

$$\sum_{n=1}^{\infty} n^{r-1} \mathbb{P}\left\{\sup_{t \geq n} \frac{|X_t|}{f(t)} \geq \varepsilon\right\} < \infty \quad \text{for all } \varepsilon > 0 \implies X_n \xrightarrow[n \rightarrow \infty]{r\text{-quickly}} 0.$$

(ii) If $f(t)$ is a power function, $f(t) = t^\gamma$, $\gamma > 0$, then the finiteness of

$$\sum_{n=1}^{\infty} n^{r-1} \mathbb{P} \left\{ \max_{1 \leq t \leq n} X_t \geq \varepsilon n^\gamma \right\}$$

for some $r > 0$ and every $\varepsilon > 0$ implies the r -quick convergence of X_n to 0:

$$\left\{ \sum_{n=1}^{\infty} n^{r-1} \mathbb{P} \left(\max_{1 \leq t \leq n} X_t \geq \varepsilon n^\gamma \right) < \infty \forall \varepsilon > 0 \right\} \implies \{E[L_\varepsilon(\gamma)^r] < \infty \forall \varepsilon > 0\}, \quad (14)$$

where $L_\varepsilon(\gamma) = \sup\{n \geq 1 : |X_n| > \varepsilon n^\gamma\}$.

Proof. Proof of (i). Obviously,

$$\mathbb{P}\{|X_n| \geq \varepsilon f(n)\} \leq \mathbb{P}\{L_\varepsilon(f) \geq n\} \leq \mathbb{P}\left\{\sup_{t \geq n} \frac{1}{f(t)} |X_t| \geq \varepsilon\right\}$$

from which the inequalities (13) follow immediately.

Proof of (ii). Write $M_u = \max_{1 \leq n \leq \lceil u \rceil} |X_n|$, where $\lceil u \rceil$ is a smallest integer greater or equal to u . We have the following chain of inequalities and equalities:

$$\begin{aligned} E[L_{2\varepsilon}(\gamma)^r] &\leq r \int_0^\infty t^{r-1} \mathbb{P} \left\{ \sup_{u \geq t} u^{-\gamma} |X_u| \geq 2\varepsilon \right\} dt \\ &\leq r \int_0^\infty t^{r-1} \mathbb{P} \left\{ \sup_{u \geq t} [|X_u| - \varepsilon u^\gamma] \geq \varepsilon t^\gamma \right\} dt \\ &\leq r \int_0^\infty t^{r-1} \mathbb{P} \left\{ \sup_{u > 0} [|X_u| - \varepsilon u^\gamma] \geq \varepsilon t^\gamma \right\} dt \\ &\leq r \sum_{n=1}^{\infty} \int_0^\infty t^{r-1} \mathbb{P} \left\{ \sup_{(2^{n-1}-1)t^\gamma < u^\gamma \leq (2^n-1)t^\gamma} [|X_u| - \varepsilon u^\gamma] \geq \varepsilon t^\gamma \right\} dt \\ &\leq r \sum_{n=1}^{\infty} \int_0^\infty t^{r-1} \mathbb{P} \left\{ \sup_{u^\gamma \leq 2^n t^\gamma} |X_u| \geq 2^{n-1} \varepsilon t^\gamma \right\} dt \\ &= r \sum_{n=1}^{\infty} \int_0^\infty t^{r-1} \mathbb{P} \{ M_{2^{n/\gamma}u} \geq 2^{n-1} \varepsilon t^\gamma \} dt \\ &= r \left[\sum_{n=1}^{\infty} 2^{-n/\gamma} \right] \int_0^\infty u^{r-1} \mathbb{P} \{ M_u \geq (\varepsilon/2) u^\gamma \} du. \end{aligned}$$

It follows that

$$\begin{aligned} E[L_{2\varepsilon}(\gamma)^r] &\leq r(2^{1/\gamma} - 1)^{-1} \int_0^\infty u^{r-1} \mathbb{P} \{ M_u \geq (\varepsilon/2) u^\gamma \} du \leq \\ &\leq r(2^{1/\gamma} - 1)^{-1} \sum_{n=1}^{\infty} n^{r-1} \mathbb{P} \left\{ \max_{1 \leq t \leq n} X_n \geq \varepsilon n^\gamma \right\} \end{aligned}$$

which yields the implication (14) and completes the proof. $\square$

The following theorem shows that, in the i.i.d. case, the implications in (12) become equivalences.

Theorem 3. Let $\{S_n\}_{n \in \mathbb{Z}_+}$ be a zero-mean random walk. The following statements are equivalent

$$E|S_1|^{r+1} < \infty \iff n^{-1} S_n \xrightarrow[n \rightarrow \infty]{r\text{-completely}} 0, \quad (15)$$

$$E|S_1|^{r+1} < \infty \iff n^{-1}S_n \xrightarrow[n \rightarrow \infty]{r\text{-quickly}} 0, \quad (16)$$

$$E|S_1|^{r+1} < \infty \iff \sum_{n=1}^{\infty} n^{r-1} P\left\{\sup_{t \geq n} \frac{1}{t} |S_t| > \varepsilon\right\} < \infty \quad \text{for all } \varepsilon > 0. \quad (17)$$

Proof. By Theorem 1, in the i.i.d. case,

$$E|S_1|^{r+1} < \infty \iff \sum_{n=1}^{\infty} n^{r-1} P\left(\frac{1}{n} |S_n| > \varepsilon\right) < \infty \quad \forall \varepsilon > 0 \quad (18)$$

and

$$E|S_1|^{r+1} < \infty \iff \sum_{n=1}^{\infty} n^{r-1} P\left(\sup_{t \geq n} \frac{1}{t} |S_t| > \varepsilon\right) < \infty \quad \forall \varepsilon > 0, \quad (19)$$

so that assertion (15) follows from (18) and (17) from (19).

Next, let

$$L_\varepsilon = \sup\{n \geq 1 : |S_n| \geq n\varepsilon\} \quad (\sup \emptyset = 0).$$

By Lemma 2(i),

$$E[L_\varepsilon^r] \leq r \sum_{n=1}^{\infty} n^{r-1} P\left\{\sup_{t \geq n} (|S_t|/t) \geq \varepsilon\right\} \quad \forall \varepsilon > 0, \quad (20)$$

which, along with (19), implies (16). $\square$

2.4. Further Remarks on r -Complete Convergence, r -Quick Convergence, and Rates of Convergence in SLLN

Let $\{S_n\}_{n \in \mathbb{Z}_+}$ be a random walk. Without loss of generality, let $S_0 = 0$ and $E[S_1] = 0$.

1. Strassen [3] proved, in particular, that if $f(n) = (2n \log n)^{1/2}$ in Lemma 2, then for $r > 0$

$$\limsup_{n \rightarrow \infty} \frac{S_n}{\sqrt{2n \log n}} = \sqrt{r E[S_1^2]} \quad r\text{-quickly} \quad (21)$$

whenever $E|S_1|^p < \infty$ for $p > (2r + 1)$. He also proved the functional form of the law of the iterated logarithm.

2. Lai [8] improved this result, showing that Strassen's moment condition $E|S_1|^p < \infty$ for $p > (2r + 1)$ can be relaxed. Specifically, he showed that a weaker condition

$$E\left[|S_1|^{2(r+1)} (\log^+ |S_1| + 1)^{-(r+1)}\right] < \infty \quad \text{for } r > 0 \quad (22)$$

is the best one can do (i.e., both necessary and sufficient):

$$E\left[|S_1|^{2(r+1)} (\log^+ |S_1| + 1)^{-(r+1)}\right] < \infty \iff \limsup_{n \rightarrow \infty} \frac{S_n}{\sqrt{2n \log n}} < \infty \quad r\text{-quickly},$$

in which case equality (21) holds.

Note, however, that for $r = 0$, in terms of the a.s. convergence,

$$E\left[|S_1|^2\right] < \infty \iff \limsup_{n \rightarrow \infty} \frac{S_n}{\sqrt{2n \log \log n}} = \sqrt{E[|S_1|^2]} \quad \text{a.s.}$$

but under condition (22) for all $r > 0$

$$\limsup_{n \rightarrow \infty} \frac{S_n}{\sqrt{2n \log \log n}} = \infty \quad r\text{-quickly}.$$

3. Let $\alpha > 1/2$ and $r > 0$. Chow and Lai [9] established the following one-sided inequality for tail probabilities:

$$\sum_{n=1}^{\infty} n^{r-1} \mathbb{P} \left(\max_{1 \leq t \leq n} S_t \geq n^{\alpha} \right) \leq C_{r,\alpha} \left\{ \mathbb{E} \left[(S_1^+)^{(r+1)/\alpha} \right] + \left(\mathbb{E}[S_1^2] \right)^{r/(2\alpha-1)} \right\} \quad (23)$$

whenever $\mathbb{E}|S_1|^2 < \infty$. Under the same hypotheses, this one-sided inequality implies the two-sided one:

$$\sum_{n=1}^{\infty} n^{r-1} \mathbb{P} \left(\max_{1 \leq t \leq n} |S_t| \geq n^{\alpha} \right) \leq C_{r,\alpha} \left\{ \mathbb{E} \left[|S_1|^{(r+1)/\alpha} \right] + \left(\mathbb{E}[S_1^2] \right)^{r/(2\alpha-1)} \right\}. \quad (24)$$

The upper bound in (24) turns out to be sharp since the lower bound also holds:

$$\sum_{n=1}^{\infty} n^{r-1} \mathbb{P} \left(\max_{1 \leq t \leq n} |S_t| \geq n^{\alpha} \right) \geq 1 + B_{r,\alpha} \left\{ \mathbb{E} \left[|S_1|^{(r+1)/\alpha} \right] + \left(\mathbb{E}[S_1^2] \right)^{r/(2\alpha-1)} \right\}.$$

Here, the constants $C_{r,\alpha}$ and $B_{r,\alpha}$ are universal only depending on r, α .

The results of Chow and Lai [9] provide one-sided analogues of the results of Baum and Katz [2] as well as extend their results. Indeed, the one-sided inequality (23) implies that the following statements are equivalent for the zero-mean random walk $\{S_n\}_{n \in \mathbb{N}}$:

- (i) $\mathbb{E}[(S_1^+)^{(r+1)/\alpha}] < \infty$;
- (ii) $\sum_{n=1}^{\infty} n^{r-1} \mathbb{P}(n^{-\alpha} S_n \geq \varepsilon) < \infty$ for all $\varepsilon > 0$;
- (iii) $\sum_{n=1}^{\infty} n^{r-1} \mathbb{P} \left(\sup_{k \geq n} k^{-\alpha} S_k \geq \varepsilon \right) < \infty$ for all $\varepsilon > 0$,

where $\alpha > 1/2$.

Clearly, the two-sided inequality (24) yields the assertions of Theorem 1.

4. The Marcinkiewicz–Zygmund SLLN states that, for $\alpha > 1/2$, the following implications hold:

$$\mathbb{E}|S_1|^{1/\alpha} < \infty \iff n^{-\alpha} S_n \xrightarrow[n \rightarrow \infty]{\text{a.s.}} 0. \quad (25)$$

The strengthened r -quick equivalent of this SLLN is: for any $r > 0$ and $\alpha > 1/2$, the following statements are equivalent,

$$\begin{aligned} \mathbb{E}[|S_1|^{(r+1)/\alpha}] < \infty &\iff \sum_{i=1}^{\infty} n^{r-1} \mathbb{P} \left\{ \frac{1}{n^{\alpha}} |S_n| > \varepsilon \right\} < \infty \text{ for all } \varepsilon > 0 \\ &\iff \sum_{n=1}^{\infty} n^{r-1} \mathbb{P} \left\{ \sup_{k \geq n} \frac{1}{k^{\alpha}} |S_k| > \varepsilon \right\} < \infty \text{ for all } \varepsilon > 0 \\ &\iff n^{-\alpha} S_n \xrightarrow[n \rightarrow \infty]{r\text{-quickly}} 0. \end{aligned} \quad (26)$$

Implications (26) follow from Theorem 1, Theorem 3 and inequality (24). The proof is almost obvious and omitted.

3. Applications of r -Complete and r -Quick Convergences in Statistics

In this section, we outline certain statistical applications which show the usefulness of r -complete and r -quick versions of the SLLN.

3.1. Sequential Hypothesis Testing

We begin by formulating the following multi-hypothesis testing problem for a general non-i.i.d. stochastic model. Let $(\Omega, \mathcal{F}, \mathcal{F}_n, \mathbb{P})$, $n \in \mathbb{Z}_+ = \{0, 1, 2, \dots\}$ be a filtered probability space with standard assumptions about the monotonicity of the sub- σ -algebras $\mathcal{F}_n$. The sub- σ -algebra $\mathcal{F}_n = \sigma(\mathbf{X}^n)$ of $\mathcal{F}$ is assumed to be generated by the sequence $\mathbf{X}^n = \{X_t, 1 \leq t \leq n\}$ observed up to time n , which is defined on the space $(\Omega, \mathcal{F})$. The hypotheses are $H_i : \mathbb{P} = \mathbb{P}_i$, $i = 0, 1, \dots, N$, where $\mathbb{P}_0, \mathbb{P}_1, \dots, \mathbb{P}_N$ are given probability measures assumed to be locally mutually absolutely continuous, i.e., their restrictions $\mathbb{P}_i^n$ and $\mathbb{P}_j^n$ to $\mathcal{F}_n$ are

equivalent for all $1 \leq n < \infty$ and all $i, j = 0, 1, \dots, N, i \neq j$. Let Q^n be a restriction to $\mathcal{F}_n$ of a σ -finite measure Q on $(\Omega, \mathcal{F})$. Under P_i , the sample $\mathbf{X}^n = (X_1, \dots, X_n)$ has a joint density $p_{i,n}(\mathbf{X}^n)$ with respect to the dominating measure Q^n for all $n \in \mathbb{N}$, which can be written as

$$p_{i,n}(\mathbf{X}^n) = \prod_{t=1}^n f_{i,t}(X_t | \mathbf{X}^{t-1}), \quad (27)$$

where $f_{i,n}(X_n | \mathbf{X}^{n-1})$, $n \geq 1$ are corresponding conditional densities.

For $n \in \mathbb{N}$, define the likelihood ratio (LR) process between the hypotheses H_i and H_j

$$\Lambda_{ij}(n) = \frac{dP_i^n}{dP_j^n}(\mathbf{X}^n) = \frac{p_{i,n}(\mathbf{X}^n)}{p_{j,n}(\mathbf{X}^n)} = \prod_{t=1}^n \frac{f_{i,t}(X_t | \mathbf{X}^{t-1})}{f_{j,t}(X_t | \mathbf{X}^{t-1})}$$

and the log-likelihood ratio (LLR) process

$$\lambda_{ij}(n) = \log \Lambda_{ij}(n) = \sum_{t=1}^n \log \left[\frac{f_{i,t}(X_t | \mathbf{X}^{t-1})}{f_{j,t}(X_t | \mathbf{X}^{t-1})} \right].$$

A multi-hypothesis sequential test is a pair $\delta = (d, T)$, where T is a stopping time with respect to the filtration $\{\mathcal{F}_n\}_{n \in \mathbb{Z}_+}$ and $d = d(\mathbf{X}^T)$ is an $\mathcal{F}_T$ -measurable terminal decision function with values in the set $\{0, 1, \dots, N\}$. Specifically, $d = i$ means that the hypothesis H_i is accepted upon stopping, i.e., $\{d = i\} = \{T < \infty, \delta \text{ accepts } H_i\}$. Let $\alpha_{ij}(\delta) = P_i(d = j)$, $i \neq j$, $i, j = 0, 1, \dots, N$ denote the error probabilities of the test δ , i.e., the probabilities of accepting the hypothesis H_j when H_i is true.

Introduce the class of tests with probabilities of errors $\alpha_{ij}(\delta)$ that do not exceed the prespecified numbers $0 < \alpha_{ij} < 1$:

$$\mathbb{C}(\alpha) = \{\delta : \alpha_{ij}(\delta) \leq \alpha_{ij} \text{ for } i, j = 0, 1, \dots, N, i \neq j\}, \quad (28)$$

where $\alpha = (\alpha_{ij})$ is a matrix of given error probabilities that are positive numbers less than 1.

Let E_i denote the expectation under the hypothesis H_i (i.e., under the measure P_i). The goal of a statistician is to find a sequential test that would minimize the expected sample sizes $E_i[T]$ for all hypotheses H_i , $i = 0, 1, \dots, N$ at least approximately, say asymptotically for small probabilities of errors, i.e., as $\alpha_{ij} \rightarrow 0$.

3.1.1. Asymptotic Optimality of Wald's SPRT

First, assume that $N = 1$, i.e., that we are dealing with two hypotheses H_0 and H_1 . In the mid-1940s, Wald [11,12] introduced the *sequential probability ratio test* (SPRT) for the sequence of i.i.d. observations $X_1, X_2, \dots$, in which case $f_{i,t}(X_t | \mathbf{X}^{t-1}) = f_i(X_t)$ in (27) and the LR $\Lambda_{1,0}(n) = \Lambda_n$ is

$$\Lambda_n = \prod_{t=1}^n \frac{f_1(X_t)}{f_0(X_t)}.$$

After n observations have been made, Wald's SPRT prescribes for each $n \geq 1$:

$$\begin{aligned} &\text{stop and accept } H_1 && \text{if } \Lambda_n \geq A_1; \\ &\text{stop and accept } H_0 && \text{if } \Lambda_n \leq A_0; \\ &\text{continue sampling} && \text{if } A_0 < \Lambda_n < A_1, \end{aligned}$$

where $0 < A_0 < 1 < A_1$ are two thresholds.

Let $Z_t = \log[f_1(X_t)/f_0(X_t)]$ be the LLR for the observation X_t , so the LLR for the sample $\mathbf{X}^n$ is the sum

$$\lambda_{10}(n) = \lambda_n = \sum_{t=1}^n Z_t, \quad n = 1, 2, \dots$$

Let $a_0 = -\log A_0 > 0$ and $a_1 = \log A_1 > 0$. The SPRT $\delta_*(a_0, a_1) = (d_*, T_*)$ can be represented in the form

$$T_*(a_0, a_1) = \inf\{n \geq 1 : \lambda_n \notin (-a_0, a_1)\}, \quad d_*(a_0, a_1) = \begin{cases} 1 & \text{if } \lambda_{T_*} \geq a_1 \\ 0 & \text{if } \lambda_{T_*} \leq -a_0. \end{cases} \quad (29)$$

In the case of two hypotheses, the class of tests (28) is of the form

$$\mathbb{C}(\alpha_0, \alpha_1) = \{\delta : \alpha_0(\delta) \leq \alpha_0 \text{ and } \alpha_1(\delta) \leq \alpha_1\}.$$

That is, it includes hypothesis tests with upper bounds α_0 and α_1 on the probabilities of errors of Type 1 (false positive) $\alpha_0(\delta) = \alpha_{0,1}(\delta)$ and Type 2 (false negative) $\alpha_1(\delta) = \alpha_{1,0}(\delta)$, respectively.

Wald's SPRT has an extraordinary optimality property: it minimizes both expected sample sizes $E_0[T]$ and $E_1[T]$ in the class of sequential (and non-sequential) tests $\mathbb{C}(\alpha_0, \alpha_1)$ with given error probabilities as long as the observations are i.i.d. under both hypotheses. More specifically, Wald and Wolfowitz [13] proved, using a Bayesian approach, that if $\alpha_0 + \alpha_1 < 1$ and thresholds $-a_0$ and a_1 can be selected in such a way that $\alpha_0(\delta_*) = \alpha_0$ and $\alpha_1(\delta_*) = \alpha_1$, then the SPRT δ_* is strictly optimal in class $\mathbb{C}(\alpha_0, \alpha_1)$. A rigorous proof of this fundamental result is tedious and involves several delicate technical details. Alternative proofs can be found in [14–18].

Regardless of the strict optimality of SPRT which holds if and only if thresholds are selected so that the probabilities of errors of SPRT are exactly equal to the prescribed values α_0, α_1 , which is usually impossible, suppose that thresholds a_0 and a_1 are so selected that

$$a_0 \sim \log(1/\alpha_1) \quad \text{and} \quad a_1 \sim \log(1/\alpha_0) \quad \text{as } \alpha_{\max} \rightarrow 0. \quad (30)$$

Then

$$E_1[T_*] \sim \frac{|\log \alpha_0|}{I_1}, \quad E_0[T_*] \sim \frac{|\log \alpha_1|}{I_0} \quad \text{as } \alpha_{\max} \rightarrow 0, \quad (31)$$

where $I_1 = E_1[Z_1]$ and $I_0 = E_0[-Z_1]$ are Kullback–Leibler (K-L) information numbers so that the following asymptotic lower bounds for expected sample sizes are attained by SPRT:

$$\inf_{\delta \in \mathbb{C}(\alpha_0, \alpha_1)} E_1[T] \geq \frac{|\log \alpha_0|}{I_1} + o(1), \quad \inf_{\delta \in \mathbb{C}(\alpha_0, \alpha_1)} E_0[T] \geq \frac{|\log \alpha_1|}{I_0} + o(1) \quad \text{as } \alpha_{\max} \rightarrow 0$$

(cf. [6]). Hereafter, $\alpha_{\max} = \max(\alpha_0, \alpha_1)$.

The following inequalities for the error probabilities of the SPRT hold in the most general non-i.i.d. case

$$\alpha_1(\delta_*) \leq \exp\{-a_0\}[1 - \alpha_0(\delta_*)], \quad \alpha_0(\delta_*) \leq \exp\{-a_1\}[1 - \alpha_1(\delta_*)]. \quad (32)$$

These bounds can be used to guarantee asymptotic relations (30).

In the i.i.d. case, by the SLLN, the LLR λ_n has the following stability property

$$n^{-1}\lambda_n \xrightarrow[n \rightarrow \infty]{P_1\text{-a.s.}} I_1, \quad n^{-1}(-\lambda_n) \xrightarrow[n \rightarrow \infty]{P_0\text{-a.s.}} I_0. \quad (33)$$

This allows one to conjecture that, if in the general non-i.i.d. case, the LLR is also stable in the sense that the almost sure convergence conditions (33) are satisfied with some positive and finite numbers I_1 and I_0 , then the asymptotic formulas (31) still hold. In the general case, these numbers represent the local K-L information in the sense that often (while not always) $I_1 = \lim_{n \rightarrow \infty} n^{-1}E_1[\lambda_n]$ and $I_0 = \lim_{n \rightarrow \infty} n^{-1}E_0[-\lambda_n]$. Note, however, that in the general non-i.i.d. case, the SLLN does not even guarantee the finiteness of the expected sample sizes $E_i[T_*]$ of the SPRT, so some additional conditions are needed, such as a certain rate of convergence in the strong law, e.g., complete or quick convergence.

In 1981, Lai [7] was the first to prove the asymptotic optimality of Wald's SPRT in a general non-i.i.d. case as $\alpha_{\max} = \max(\alpha_0, \alpha_1) \rightarrow 0$. While the motivation was the near optimality of invariant SPRTs with respect to nuisance parameters, Lai proved a more general result using the r -quick convergence concept.

Specifically, for $0 < I_0 < \infty$ and $0 < I_1 < \infty$, define

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

($\sup \{\emptyset\} = 0$) and suppose that $E_i[L_i(\varepsilon)^r] < \infty$ ($i = 0, 1$) for some $r > 0$ and every $\varepsilon > 0$, i.e., that the normalized LLR converges r -quickly to I_1 under P_1 and to $-I_0$ under P_0 :

$$n^{-1}\lambda_n \xrightarrow[n \rightarrow \infty]{P_1-r\text{-quickly}} I_1 \quad \text{and} \quad n^{-1}\lambda_n \xrightarrow[n \rightarrow \infty]{P_0-r\text{-quickly}} -I_0. \quad (34)$$

Strengthening the a.s. convergence (33) into the r -quick version (34), Lai [7] established the first-order asymptotic optimality of Wald's SPRT for moments of the stopping time distribution up to order r : If thresholds $a_1(\alpha_0, \alpha_1)$ and $-a_0(\alpha_0, \alpha_1)$ in the SPRT are so selected that $\delta_*(\alpha_0, \alpha_1) \in \mathbb{C}(\alpha_0, \alpha_1)$ and asymptotics (30) hold, then as $\alpha_{\max} \rightarrow 0$,

$$\inf_{\delta \in \mathbb{C}(\alpha_0, \alpha_1)} E_1[T^r] \sim \left(\frac{|\log \alpha_0|}{I_1} \right)^r \sim E_1[T_*^r],$$

$$\inf_{\delta \in \mathbb{C}(\alpha_0, \alpha_1)} E_0[T^r] \sim \left(\frac{|\log \alpha_1|}{I_0} \right)^r \sim E_0[T_*^r]. \quad (35)$$

Wald's ideas have been generalized in many publications to construct sequential tests of composite hypotheses with nuisance parameters when these hypotheses can be reduced to simple ones by the principle of invariance. If M_n is the maximal invariant statistic and $p_i(M_n)$ is the density of this statistic under hypothesis H_i , then the invariant SPRT is defined as in (29) with the LLR $\lambda_n = \log[p_1(M_n)/p_0(M_n)]$. However, even if the observations $X_1, X_2, \dots$ are i.i.d. the invariant LLR statistic λ_n is not a random walk anymore and Wald's methods cannot be applied directly. Lai [7] has applied the asymptotic optimality property (35) of Wald's SPRT in the non-i.i.d. case to investigate the optimality properties of several classical invariant SPRTs such as the sequential t -test, the sequential T^2 -test, and Savage's rank-order test.

In the sequel, we will call the case where the a.s. convergence in the non-i.i.d. model (33) holds with the rate $1/n$ *asymptotically stationary*. Assume now that (33) is generalized to

$$\lambda_n / \psi(n) \xrightarrow[n \rightarrow \infty]{P_1\text{-a.s.}} I_1, \quad (-\lambda_n) / \psi(n) \xrightarrow[n \rightarrow \infty]{P_0\text{-a.s.}} I_0, \quad (36)$$

where $\psi(t)$ is a positive increasing function. If $\psi(t)$ is not linear, then this case will be referred to as *asymptotically non-stationary*.

A simple example where this generalization is needed is testing H_0 versus H_1 regarding the mean of the normal distribution:

$$X_n = i S_n + \xi_n, \quad n \in \mathbb{Z}_+, \quad i = 0, 1,$$

where $\{\xi_n\}_{n \geq 1}$ is a zero-mean i.i.d. standard Gaussian sequence $\mathcal{N}(0, 1)$ and $S_n = \sum_{j=0}^k c_j n^j$ is a polynomial of order $k \geq 1$. Then,

$$\lambda_n = \sum_{t=1}^n S_t X_t - \frac{1}{2} \sum_{t=1}^n S_t^2,$$

$E_1[\lambda_n] = -E_0[\lambda_n] = \frac{1}{2} \sum_{t=1}^n S_t^2 \sim c_k^2 n^{2k}$ for a large n , so $\psi(n) = n^{2k}$ and $I_1 = I_0 = c_k^2/2$ in (36). This example is of interest for certain practical applications, in particular, for the recognition of ballistic objects and satellites.

Tartakovsky et al. ([6] Section 3.4) generalized Lai's results for the asymptotically non-stationary case. Write $\Psi(t)$ for the inverse function of $\psi(t)$.

Theorem 4 (SPRT asymptotic optimality). *Let $r \geq 1$. Assume that there exist finite positive numbers I_0 and I_1 and an increasing non-negative function $\psi(t)$ such that the r -quick convergence conditions*

$$\frac{\lambda_n}{\psi(n)} \xrightarrow[n \rightarrow \infty]{P_1-r\text{-quickly}} I_1, \quad \frac{-\lambda_n}{\psi(n)} \xrightarrow[n \rightarrow \infty]{P_0-r\text{-quickly}} I_0$$

hold. If thresholds $-a_0(\alpha_0, \alpha_1)$ and $a_1(\alpha_0, \alpha_1)$ are selected so that $\delta_(a_0, a_1) \in \mathbb{C}(\alpha_0, \alpha_1)$ and $a_0 \sim |\log \alpha_1|$ and $a_1 \sim |\log \alpha_0|$, then, as $\alpha_{\max} \rightarrow 0$,*

$$\begin{aligned} \inf_{\delta \in \mathbb{C}(\alpha_0, \alpha_1)} E_1[T^r] &\sim \left[\Psi\left(\frac{|\log \alpha_0|}{I_1}\right) \right]^r \sim E_1[T_*^r], \\ \inf_{\delta \in \mathbb{C}(\alpha_0, \alpha_1)} E_0[T^r] &\sim \left[\Psi\left(\frac{|\log \alpha_1|}{I_0}\right) \right]^r \sim E_0[T_*^r]. \end{aligned} \quad (37)$$

This theorem implies that the SPRT asymptotically minimizes the moments of the stopping time distribution up to order r .

The proof of this theorem is performed in two steps which are related to our previous discussion of the rates of convergence in Section 2. The first step is to obtain the asymptotic lower bounds in class $\mathbb{C}(\alpha_0, \alpha_1)$:

$$\liminf_{\alpha_{\max} \rightarrow 0} \frac{\inf_{\delta \in \mathbb{C}(\alpha_0, \alpha_1)} E_1[T^r]}{[\Psi(|\log \alpha_0|/I_1)]^r} \geq 1, \quad \liminf_{\alpha_{\max} \rightarrow 0} \frac{\inf_{\delta \in \mathbb{C}(\alpha_0, \alpha_1)} E_0[T^r]}{[\Psi(|\log \alpha_1|/I_0)]^r} \geq 1.$$

These bounds hold whenever the following right-tail conditions for the LLR are satisfied:

$$\begin{aligned} \lim_{M \rightarrow \infty} P_1 \left\{ \frac{1}{\psi(M)} \max_{1 \leq n \leq M} \lambda_n \geq (1 + \varepsilon) I_1 \right\} &= 1, \\ \lim_{M \rightarrow \infty} P_0 \left\{ \frac{1}{\psi(M)} \max_{1 \leq n \leq M} (-\lambda_n) \geq (1 + \varepsilon) I_0 \right\} &= 1. \end{aligned}$$

Note that, by Lemma 1, these conditions are satisfied when the SLLN (36) holds so that the almost sure convergence (36) is sufficient. However, as we already mentioned, the SLLN for the LLR is not sufficient to guarantee even the finiteness of the SPRT stopping time.

The second step is to show that the lower bounds are attained by the SPRT. To do so, it suffices to impose the following additional left-tail conditions:

$$\sum_{n=1}^{\infty} n^{r-1} P_1 \{ \lambda_n \leq (I_1 - \varepsilon) \psi(n) \} < \infty, \quad \sum_{n=1}^{\infty} n^{r-1} P_0 \{ -\lambda_n \leq (I_0 - \varepsilon) \psi(n) \} < \infty$$

for all $0 < \varepsilon < \min(I_0, I_1)$. Since both right-tail and left-tail conditions hold if the LLR converges r -completely to I_i ,

$$\sum_{n=1}^{\infty} n^{r-1} P_1 \left\{ \left| \frac{\lambda_n}{\psi(n)} - I_1 \right| \geq \varepsilon \right\} < \infty, \quad \sum_{n=1}^{\infty} n^{r-1} P_0 \left\{ \left| \frac{\lambda_n}{\psi(n)} + I_0 \right| \geq \varepsilon \right\} < \infty,$$

and since r -quick convergence implies r -complete convergence (see (12)), we conclude that the assertions (37) hold.

Remark 3. *In the i.i.d. case, Wald's approach allows us to establish asymptotic equalities (37) with $I_1 = E_1[\lambda_1]$ and $I_0 = -E_0[\lambda_1]$ being K-L information numbers under the only condition of finiteness I_i . However, Wald's approach breaks down in the non-i.i.d. case. Certain generalizations in the case of independent but non-identically and substantially non-stationary observations, extending Wald's ideas, were considered in [19–21]. Theorem 4 covers all these non-stationary models.*

Fellouris and Tartakovsky [22] extended previous results on the asymptotic optimality of the SPRT to the case of the multistream hypothesis testing problem when the observations are sequentially acquired in multiple data streams (or channels or sources). The problem is to test the null hypothesis H_0 that none of the N streams are affected against the composite hypothesis H_B that a subset $B \subset \{1, \dots, N\}$ is affected. Write P_B and E_B for the distribution of observations and expectation under hypothesis H_B . Let $\mathcal{P}$ denote a class of subsets of $\{1, \dots, N\}$ that incorporates prior information which is available regarding the subset of affected streams, e.g., not more than $K < N$ streams can be affected. (In many practical problems, K is substantially smaller than the total number of streams N , which can be very large.)

Two sequential tests were studied in [22]—the generalized sequential likelihood ratio test and the mixture sequential likelihood ratio test. It has been shown that both tests are first-order asymptotically optimal, minimizing the moments of the sample size $E_0[T^r]$ and $E_B[T^r]$ for all $B \in \mathcal{P}$ up to order r as $\max(\alpha_0, \alpha_1) \rightarrow 0$ in the class of tests

$$\mathbb{C}_{\mathcal{P}}(\alpha_0, \alpha_1) = \left\{ \delta : P_0(d=1) \leq \alpha_0 \text{ and } \max_{B \in \mathcal{P}} P_B(d=0) \leq \alpha_1 \right\}, \quad 0 < \alpha_i < 1.$$

The proof is essentially based on the concept of r -complete convergence of LLR with the rate $1/n$. See also Chapter 1 in [5].

3.1.2. Asymptotic Optimality of the Multi-hypothesis SPRT

We now return to the multi-hypothesis model with $N > 1$ that we started to discuss at the beginning of this section (see (27) and (28)). The problem of the sequential testing of many hypotheses is substantially more difficult than that of testing two hypotheses. For multiple-decision testing problems, it is usually very difficult, if even possible, to obtain optimal solutions. Finding an optimal non-Bayesian test in the class of tests (28) that minimizes expected sample sizes $E_i[T]$ for all hypotheses H_i , $i = 0, 1, \dots, N$ is not manageable even in the i.i.d. case. For this reason, a substantial part of the development of sequential multi-hypothesis testing in the 20th century has been directed towards the study of certain combinations of one-sided sequential probability ratio tests when observations are i.i.d. (see, e.g., [23–28]).

We will focus on the following first-order asymptotic criterion: Find a multi-hypothesis test $\delta_*(\alpha) = (d_*(\alpha), T_*(\alpha))$ such that, for some $r \geq 1$,

$$\lim_{\alpha_{\max} \rightarrow 0} \frac{\inf_{\delta \in \mathbb{C}(\alpha)} E_i[T^r]}{E_i[T_*(\alpha)^r]} = 1 \quad \text{for all } i = 0, 1, \dots, N, \quad (38)$$

where $\alpha_{\max} = \max_{0 \leq i, j \leq N, i \neq j} \alpha_{ij}$.

In 1998, Tartakovsky [4] was the first who considered the sequential multiple hypothesis testing problems for general non-i.i.d. stochastic models following Lai's idea of exploiting the r -quick convergence in the SLLN for two hypotheses. The results were obtained for both discrete and continuous-time scenarios and for the asymptotically non-stationary case where the LLR processes between hypotheses converge to finite numbers with the rate $1/\psi(t)$. Two multi-hypothesis tests were investigated: (1) the *rejecting* test, which rejects the hypotheses one by one, and the last hypothesis, which is not rejected, is accepted; and (2) the *matrix accepting* test that accepts a hypothesis for which all component SPRTs that involve this hypothesis vote for accepting it.

We now proceed with introducing this accepting test which we will refer to as the *matrix SPRT* (MSPRT). In the present article, we do not consider the continuous-time scenarios. Those who are interested in continuous time are referred to [4,6,19,21,29].

Write $\mathcal{N} = \{0, 1, \dots, N\}$. For a threshold matrix $(A_{ij})_{i,j \in \mathcal{N}}$, with $A_{ij} > 0$ and the A_{ii} being immaterial (say 0), define the matrix SPRT $\delta_*^N = (T_*^N, d_*^N)$, built on $(N+1)N/2$ one-sided SPRTs between the hypotheses H_i and H_j , as follows:

$$\text{Stop at the first } n \geq 1 \text{ such that, for some } i, \Lambda_{ij}(n) \geq A_{ji} \text{ for all } j \neq i, \quad (39)$$

and accept the unique H_i that satisfies these inequalities. Note that, for $N = 1$, the MSPRT coincides with Wald's SPRT.

In the following, we omit the superscript N in $\delta_*^N = (T_*^N, d_*^N)$ for brevity. Obviously, with $a_{ji} = \log A_{ji}$, the MSPRT in (39) can be written as

$$T_* = \inf\{n \geq 1 : \lambda_{ij}(n) \geq a_{ji} \text{ for all } j \neq i \text{ and some } i\}, \quad (40)$$

$$d_* = i \text{ for which (40) holds.} \quad (41)$$

Introducing the Markov accepting times for the hypotheses H_i as

$$T_i = \inf\left\{n \geq 1 : \lambda_{i0}(n) \geq \max_{\substack{1 \leq j \leq N \\ j \neq i}} [\lambda_{j0}(n) + a_{ji}]\right\}, \quad i = 0, 1, \dots, N, \quad (42)$$

the test in (40), (41) can be also written in the following form:

$$T_* = \min_{0 \leq j \leq N} T_j, \quad d_* = i \text{ if } T_* = T_i. \quad (43)$$

Thus, in the MSPRT, each component SPRT is extended until, for some $i \in \mathcal{N}$, all N SPRTs involving H_i accept H_i .

Using Wald's likelihood ratio identity, it is easily shown that $\alpha_{ij}(\delta_*) \leq \exp(-a_{ij})$ for $i, j \in \mathcal{N}$, $i \neq j$, so selecting $a_{ji} = |\log \alpha_{ji}|$ implies that $\delta_* \in \mathbb{C}(\alpha)$. These inequalities are similar to Wald's ones in the binary hypothesis case and are very imprecise. In his ingenious paper, Lorden [27] showed that, with a very sophisticated design that includes the accurate estimation of thresholds accounting for overshoots, the MSPRT is nearly optimal in the third-order sense, i.e., it minimizes the expected sample sizes for all hypotheses up to an additive disappearing term: $\inf_{\delta \in \mathbb{C}(\alpha)} E_i[T] = E_i[T_*] + o(1)$ as $\alpha_{\max} \rightarrow 0$. This result only holds for i.i.d. models with the finite second moment $E_i[\lambda_{ij}(1)^2] < \infty$. In the non-i.i.d. case (and even in the i.i.d. case for higher moments $r > 1$), there is no way to obtain such a result, so we focus on the first-order optimality (38).

The following theorem establishes asymptotic operating characteristics and the optimality of MSPRT under the r -quick convergence of $\lambda_{ij}(n)/\psi(n)$ to finite K-L-type numbers I_{ij} , where $\psi(n)$ is a positive increasing function, $\psi(\infty) = \infty$.

Theorem 5 (MSPRT asymptotic optimality [4]). Let $r \geq 1$. Assume that there exist finite positive numbers I_{ij} , $i, j = 0, 1, \dots, N$, $i \neq j$ and an increasing non-negative function $\psi(t)$ such that, for some $r > 0$,

$$\frac{\lambda_{ij}(n)}{\psi(n)} \xrightarrow[n \rightarrow \infty]{P_i-r\text{-quickly}} I_{ij} \text{ for all } i, j = 0, 1, \dots, N, i \neq j. \quad (44)$$

Then, the following assertions are true.

(i) For $i = 0, 1, \dots, N$,

$$E_i[T_*^r] \sim \left[\Psi \left(\max_{\substack{0 \leq j \leq N \\ j \neq i}} \frac{a_{ji}}{I_{ij}} \right) \right]^r \text{ as } \min_{j,i} a_{ji} \rightarrow \infty. \quad (45)$$

- (ii) If the thresholds are so selected that $\alpha_{ij}(\delta^*) \leq \alpha_{ij}$ and $a_{ji} \sim |\log \alpha_{ji}|$, particularly as $a_{ji} = |\log \alpha_{ji}|$, then for all $i = 0, 1, \dots, N$

$$\inf_{\delta \in \mathbb{C}(\alpha)} E_i[T^r] \sim \left[\Psi \left(\max_{\substack{0 \leq j \leq N \\ j \neq i}} \frac{|\log \alpha_{ji}|}{I_{ij}} \right) \right]^r \sim E_i[T_*^r] \quad \text{as } \alpha_{\max} \rightarrow 0. \quad (46)$$

Assertion (ii) implies that the MSPRT asymptotically minimizes the moments of the stopping time distribution up to order r for all hypotheses $H_0, H_1, \dots, H_N$ in the class of tests $\mathbb{C}(\alpha)$.

Remark 4. Both assertions of Theorem 5 are correct under the r -complete convergence

$$\frac{\lambda_{ij}(n)}{\psi(n)} \xrightarrow[n \rightarrow \infty]{P_i-r\text{-complete}} I_{ij} \quad \text{for all } i, j = 0, 1, \dots, N, i \neq j,$$

i.e., when

$$\sum_{n=1}^{\infty} n^{r-1} P_i \left\{ \left| \frac{1}{\psi(n)} \lambda_{ij}(n) - I_{ij} \right| > \varepsilon \right\} < \infty \quad \text{for all } \varepsilon > 0.$$

While this statement has not been proven anywhere to date, it can be easily proven using the methods developed for multistream hypothesis testing and changepoint detection ([5] Ch 1, Ch 6).

Remark 5. As shown in the example given in Section 3.4.3 of [6], the r -quick convergence conditions in Theorem 5 (or corresponding r -complete convergence conditions for LLR processes) cannot be generally relaxed into the almost sure convergence

$$\frac{\lambda_{ij}(n)}{\psi(n)} \xrightarrow[n \rightarrow \infty]{P_i-a.s.} I_{ij} \quad \text{for all } i, j = 0, 1, \dots, N, i \neq j. \quad (47)$$

However, the following weak asymptotic optimality result holds for the MSPRT under the a.s. convergence: if the a.s. convergence (47) holds with the power function $\psi(t) = t^k$, $k > 0$, then, for every $0 < \varepsilon < 1$,

$$\inf_{\delta \in \mathbb{C}(\alpha)} P_i(T > \varepsilon T_*) \rightarrow 1 \quad \text{as } \alpha_{\max} \rightarrow 0 \quad \text{for all } i = 0, 1, \dots, N \quad (48)$$

whenever thresholds a_{ji} are selected as in Theorem 5 (ii).

Note that several interesting statistical and practical applications of these results to invariant sequential testing and multisample slippage scenarios are discussed in Sections 4.5 and 4.6 of Tartakovsky et al. [6] (see Mosteller [30] and Ferguson [16] for terminology regarding multisample slippage problems).

3.2. Sequential Changepoint Detection

Sequential (or quickest) changepoint detection is an important subfield of sequential analysis. The observations are made one at a time and as long as their behavior suggests that the process of interest is in control (i.e., in a normal state), the process is allowed to continue. If the state is believed to have lost control, the goal is to detect the change in distribution as rapidly as possible. Quickest change detection problems have an enormous number of important applications, e.g., object detection in noise and clutter, industrial quality control, environment surveillance, failure detection, navigation, seismology, computer network security, genomics, and epidemiology (see, e.g., [31–40]). Many challenging application areas are discussed in the books by Tartakovsky, Nikiforov, and Basseville ([6] Ch 11) and Tartakovsky ([5] Ch 8).

3.2.1. Changepoint Models

The probability distribution of the observations $\mathbf{X} = \{X_n\}_{n \in \mathbb{N}}$ is subject to a change at an unknown point in time $\nu \in \{0, 1, 2, \dots\} = \mathbb{Z}_+$ so that $X_1, \dots, X_\nu$ are generated by one stochastic model and $X_{\nu+1}, X_{\nu+2}, \dots$ are generated by another model. A sequential detection rule is a stopping time T for an observed sequence $\{X_n\}_{n \geq 1}$, i.e., T is an integer-valued random variable such that the event $\{T = n\}$ belongs to the sigma-algebra $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$ generated by observations $X_1, \dots, X_n$.

Let P_∞ denote the probability measure corresponding to the sequence of observations $\{X_n\}_{n \geq 1}$ when there is never a change ($\nu = \infty$) and, for $k = 0, 1, 2, \dots$, let P_k denote the measure corresponding to the sequence $\{X_n\}_{n \geq 1}$ when $\nu = k < \infty$. We denote the hypothesis that the change never occurs by $H_\infty : \nu = \infty$ and we denote the hypothesis that the change occurs at time $0 \leq k < \infty$ by $H_k : \nu = k$.

First consider a general non-i.i.d. model assuming that the observations may have a very general stochastic structure. Specifically, if we let, as before, $\mathbf{X}^n = (X_1, \dots, X_n)$ denote the sample of size n , then when $\nu = \infty$ (there is no change), the conditional density of X_n given $\mathbf{X}^{n-1}$ is $g_n(X_n|\mathbf{X}^{n-1})$ for all $n \geq 1$ and when $\nu = k < \infty$, then the conditional density is $g_n(X_n|\mathbf{X}^{n-1})$ for $n \leq k$ and $f_n(X_n|\mathbf{X}^{n-1})$ for $n > k$. Thus, for the general non-i.i.d. changepoint model, the joint density $p(\mathbf{X}^n|H_k)$ under hypothesis H_k can be written as follows

$$p(\mathbf{X}^n|H_k) = \begin{cases} \prod_{t=1}^n g_t(X_t|\mathbf{X}^{t-1}) & \text{for } \nu = k \geq n, \\ \prod_{t=1}^k g_t(X_t|\mathbf{X}^{t-1}) \times \prod_{t=k+1}^n f_t(X_t|\mathbf{X}^{t-1}) & \text{for } \nu = k < n, \end{cases} \quad (49)$$

where $g_n(X_n|\mathbf{X}^{n-1})$ is the pre-change conditional density and $f_n(X_n|\mathbf{X}^{n-1})$ is the post-change conditional density which may depend on ν , $f_n(X_n|\mathbf{X}^{n-1}) = f_n^{(\nu)}(X_n|\mathbf{X}^{n-1})$, but we will omit the superscript ν for brevity.

The classical changepoint detection problem deals with the i.i.d. case where there is a sequence of observations $X_1, X_2, \dots$ that are identically distributed with a probability density function (pdf) $g(x)$ for $n \leq \nu$ and with a pdf $f(x)$ for $n > \nu$. That is, in the i.i.d. case, the joint density of the vector $\mathbf{X}^n = (X_1, \dots, X_n)$ under hypothesis H_k has the form

$$p(\mathbf{X}^n|H_k) = \begin{cases} \prod_{t=1}^n g(X_t) & \text{for } \nu = k \geq n, \\ \prod_{t=1}^k g(X_t) \times \prod_{t=k+1}^n f(X_t) & \text{for } \nu = k < n. \end{cases} \quad (50)$$

Note that, as discussed in [5,6], in applications, there are two different kinds of changes—additive and non-additive. Additive changes lead to a change in the mean value of the sequence of observations. Non-additive changes are typically produced by a change in variance or covariance, i.e., these are spectral changes.

We now proceed by discussing the models for the change point ν . The change point ν may be considered either as an unknown deterministic number or as a random variable. If the change point is treated as a random variable, then the model has to be supplied with the *prior distribution* of the change point. There may be several changepoint mechanisms, and, as a result, a random variable ν may be dependent on or independent of the observations. In particular, Moustakides [41] assumed that ν can be a $\{\mathcal{F}_n\}$ -adapted stopping time. In this article, we will not discuss Moustakides's concept by allowing the prior distribution to depend on some additional information available to "Nature" (see [5] for a detailed discussion); rather, when considering a Bayesian approach, we will assume that the prior distribution of the unknown change point is independent of the observations.

3.2.2. Popular Changepoint Detection Procedures

Before formulating the criteria of optimality in the next subsection, we begin by defining the three most popular and common change detection procedures, which are

either optimal or nearly optimal in different settings. To define these procedures, we need to introduce the partial likelihood ratio and the corresponding log-likelihood ratio

$$\text{LR}_t = \frac{f_t(X_t|\mathbf{X}^{t-1})}{g_t(X_t|\mathbf{X}^{t-1})}, \quad Z_t = \log \frac{f_t(X_t|\mathbf{X}^{t-1})}{g_t(X_t|\mathbf{X}^{t-1})}, \quad t = 1, 2, \dots$$

It is worth iterating that, for general non-i.i.d. models, the post-change density often depends on the point of change, $f_t(X_t|\mathbf{X}^{t-1}) = f_t^{(\nu)}(X_t|\mathbf{X}^{t-1})$, so in general $\text{LR}_t = \text{LR}_t^{(\nu)}$ and $Z_t = Z_t^{(\nu)}$ also depend on the change point ν . However, this is not the case for the i.i.d. model (50).

The CUSUM Procedure

We now introduce the *Cumulative Sum* (CUSUM) algorithm, which was first proposed by Page [42] for the i.i.d. model (50). Recall that we consider the changepoint detection problem as a problem of testing two hypotheses: H_ν that the change occurs at a fixed-point $0 \leq \nu < \infty$ against the alternative H_∞ that the change never occurs. The LR between these hypotheses is $\Lambda_n^\nu = \prod_{t=\nu+1}^n \text{LR}_t$ for $\nu < n$ and 1 for $\nu \geq n$. Since the hypothesis H_ν is composite, we may apply the generalized likelihood ratio (GLR) approach maximizing the LR Λ_n^ν over ν to obtain the GLR statistic

$$V_n = \max_{0 \leq \nu < n} \prod_{t=\nu+1}^n \text{LR}_t, \quad n \geq 1.$$

It is easy to verify that this statistic follows the recursion

$$V_n = \max\{1, V_{n-1}\} \text{LR}_n, \quad n \geq 1, \quad V_0 = 1 \quad (51)$$

as long as the partial LR LR_n does not depend on the change point, i.e., the post-change conditional density $f_n(X_n|\mathbf{X}^{n-1})$ does not depend on ν . This is always the case for i.i.d. models (50) when $f_n(X_n|\mathbf{X}^{n-1}) = f(X_n)$. However, as we already mentioned, for non-i.i.d. models, $f_n(X_n|\mathbf{X}^{n-1}) = f_n^{(\nu)}(X_n|\mathbf{X}^{n-1})$ often depends on the change point ν , so $\text{LR}_n = \text{LR}_n^{(\nu)}$, in which case the recursion (51) does not hold.

The logarithmic version of V_n , $W_n = \log V_n$, is related to Page's CUSUM statistic G_n introduced by Page [42] in the i.i.d. case as $G_n = \max(0, W_n)$. The statistic G_n can also be obtained via the GLR approach by maximizing the LLR $\lambda_n^\nu = \log \Lambda_n^\nu$ over $0 \leq \nu < \infty$. However, since the hypotheses H_∞ and H_ν are indistinguishable for $\nu \geq n$, the maximization over $\nu \geq n$ does not make very much sense. Note also that, in contrast to Page's CUSUM statistic G_n , the statistic W_n may take values smaller than 0, so the CUSUM procedure

$$\text{T}_{\text{CS}} = \inf\{n \geq 1 : W_n \geq a\} \quad (52)$$

makes sense even for negative values of the threshold a . Thus, it is more general than Page's CUSUM. Note the recursions

$$W_n = W_{n-1}^+ + Z_n, \quad n \geq 1, \quad W_0 = 0 \quad (53)$$

and

$$G_n = (G_{n-1} + Z_n)^+, \quad n \geq 1, \quad G_0 = 0$$

in cases where $Z_n = \log[f_n(X_n|\mathbf{X}^{n-1})/g_n(X_n|\mathbf{X}^{n-1})]$ does not depend on ν .

Shiryaev's Procedure

In the i.i.d. case and for the zero-modified geometric prior distribution of the change point, Shiryaev [43] introduced the change detection procedure that prescribes the thresholding of the posterior probability $P(\nu < n | \mathbf{X}^n)$. Introducing the statistic

$$S_n^\pi = \frac{P(\nu < n | \mathbf{X}^n)}{1 - P(\nu < n | \mathbf{X}^n)}$$

one can write the stopping time of the Shiryaev procedure in the general non-i.i.d. case and for an arbitrary prior π as

$$T_{SH} = \inf\{n \geq 1 : S_n^\pi \geq A\}, \quad (54)$$

where A ($A > 0$) is a threshold controlling for the false alarm risk. The statistic S_n^π can be written as

$$\begin{aligned} S_n^\pi &= \frac{1}{P(\nu \geq n)} \sum_{k=0}^{n-1} \pi_k \Lambda_n^k \\ &= \frac{1}{P(\nu \geq n)} \sum_{k=0}^{n-1} \pi_k \prod_{t=k+1}^n LR_t, \quad n \geq 1, \quad S_0^\pi = 0, \end{aligned} \quad (55)$$

where the product $\prod_{t=j}^i LR_t = 1$ for $j < i$.

Often (following Shiryaev's assumptions), it is supposed that the change point ν is distributed according to the geometric distribution $\text{Geometric}(q)$

$$P(\nu = k) = q(1 - q)^k \quad \text{for } k = 0, 1, 2, \dots, \quad (56)$$

where $q \in (0, 1)$.

If LR_n does not depend on the change point ν and the prior distribution is geometric (56), then the statistic $\tilde{S}_n^q = S_n^\pi / q$ can be rewritten in the recursive form

$$\tilde{S}_n^q = \left(1 + \tilde{S}_{n-1}^q\right) \frac{LR_n}{1 - q}, \quad n \geq 1, \quad \tilde{S}_0^q = 0. \quad (57)$$

However, as mentioned above, this may not be the case for non-i.i.d. models, since LR_n often depends on ν .

Shiryaev–Roberts Procedure

The generalized Shiryaev–Roberts (SR) change detection procedure is based on the thresholding of the generalized SR statistic

$$R_n^{r_0} = r_0 \Lambda_n^0 + \sum_{k=0}^{n-1} \Lambda_n^k = r_0 \prod_{t=1}^n LR_t + \sum_{k=0}^{n-1} \prod_{t=k+1}^n LR_t, \quad n \geq 1, \quad (58)$$

with a non-negative head-start $R_0 = r_0$, $r_0 \geq 0$, i.e., the stopping time of the SR procedure is given by

$$T_{SR}^{r_0} = \inf\{n \geq 1 : R_n^{r_0} \geq A\}, \quad A > 0. \quad (59)$$

This procedure is usually referred to as the SR- r detection procedure in contrast to the standard SR procedure $T_{SR} \equiv T_{SR}^{r_0}$, $r_0 = 0$ that starts with a zero initial condition $r_0 = 0$. In the i.i.d. case (50), this modification of the SR procedure was introduced and studied in detail in [44,45].

If LR_n does not depend on the change point ν , then the SR- r detection statistic satisfies the recursion

$$R_n^{r_0} = (1 + R_{n-1}^{r_0})LR_n, \quad n \geq 1, \quad R_0^{r_0} = r_0.$$

Note that, as the parameter of the geometric prior distribution $q \rightarrow 0$, the Shiryaev statistic $\tilde{S}_n^q$ converges to the SR statistic $R_n^{r_0=0}$.

3.2.3. Optimality Criteria

The goal of online change detection is to detect the change with the smallest delay controlling for a false alarm rate at a given level. Tartakovsky et al. [6] suggested several changepoint problem settings, including Bayesian, minimax, and uniform (pointwise) approaches.

Let E_k denote the expectation with respect to measure P_k when the change occurs at $\nu = k < \infty$ and E_∞ with respect to P_∞ when there is no change.

In 1954, Page [42] suggested measuring the risk due to a false alarm by the mean time to false alarm $E_\infty[T]$ and the risk associated with a true change detection by the mean time to detection $E_0[T]$ when the change occurs at the very beginning. He called these performance characteristics the *average run length* (ARL). Page also introduced the now most famous change detection procedure—the CUSUM procedure (see (52) with W_n replaced by G_n)—and analyzed it using these operating characteristics in the i.i.d. case.

While the false alarm rate can be reasonably measure by the ARL to false alarm

$$\text{ARL2FA}(T) = E_\infty[T],$$

as Figure 1 suggests, the risk due to a true change detection can be reasonably measured by the conditional expected delay to detection

$$\text{CEDD}_\nu(T) = E_\nu[T - \nu | T > \nu], \quad \nu = 0, 1, 2, \dots$$

for any possible change point $\nu \in \mathbb{Z}_+ = \{0, 1, 2, \dots\}$ but not necessarily by the ARL to detection $E_0[T] \equiv \text{CEDD}_0(T)$. A good detection procedure has to guarantee small values of the expected detection delay $\text{CEDD}_\nu(T)$ for all change points $\nu \in \mathbb{Z}_+$ when $\text{ARL2FA}(T)$ is set at a certain level. However, if the false alarm risk is measured in terms of the ARL to false alarm, i.e., it is required that $\text{ARL2FA}(T) \geq \gamma$ for some $\gamma \geq 1$, then a procedure that minimizes the conditional expected delay to detection $\text{CEDD}_\nu(T)$ uniformly over all ν does not exist. For this reason, we must resort to different optimality criteria, e.g., to Bayesian and minimax criteria.

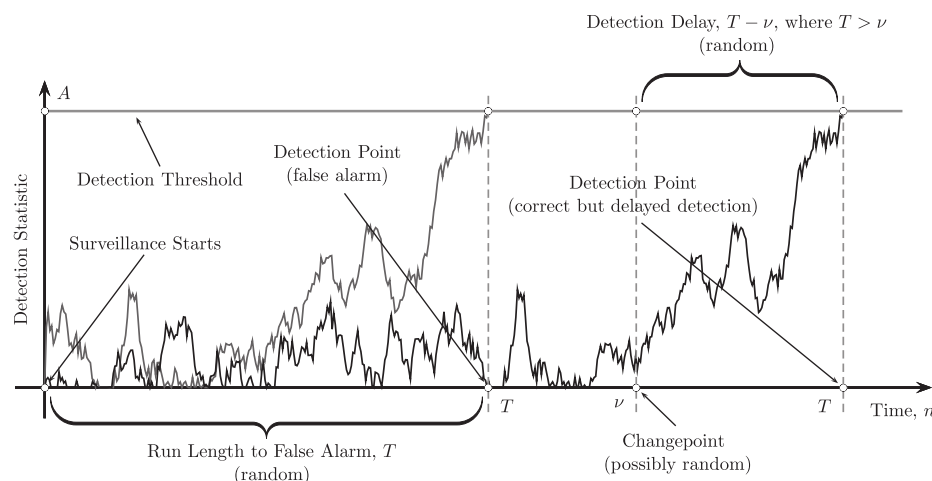


Figure 1. Illustration of a single-run sequential changepoint detection. Two possibilities in the detection process: false alarm (left) and correct detection (right).

Minimax Changepoint Optimization Criteria

There are two popular minimax criteria. The first one was introduced by Lorden [46]:

$$\inf_T \sup_{\nu \in \mathbb{Z}_+} \text{ess sup } E_\nu[T - \nu | T > \nu, \mathcal{F}_\nu] \quad \text{subject to } \text{ARL2FA}(T) \geq \gamma.$$

This requires minimizing the conditional expected delay to detection $E_\nu[T - \nu \mid T > \nu, \mathcal{F}_\nu]$ in the worst-case scenario with respect to both the change point ν and the trajectory $(X_1, \dots, X_\nu)$ of the observed process in the class of detection procedures

$$\mathbb{C}_{\text{ARL}}(\gamma) = \{T : \text{ARL2FA}(T) \geq \gamma\}, \quad \gamma \geq 1,$$

for which the ARL to false alarm exceeds the prespecified value $\gamma \in [1, \infty)$. Let $\text{ESADD}(T) = \sup_{\nu \geq 0} \text{ess sup } E_\nu[T - \nu \mid T > \nu, \mathcal{F}_\nu]$ denote Lorden's speed detection measure. Under Lorden's minimax approach, the goal is to find a stopping time $T_{\text{opt}} \in \mathbb{C}_{\text{ARL}}(\gamma)$ such that

$$\text{ESADD}(T_{\text{opt}}) = \inf_{T \in \mathbb{C}_{\text{ARL}}(\gamma)} \text{ESADD}(T) \quad \text{for any } \gamma \geq 1.$$

In the classical i.i.d. scenario (50), Lorden [46] proved that the CUSUM detection procedure (52) is asymptotically first-order minimax optimal as $\gamma \rightarrow \infty$, i.e.,

$$\inf_{T \in \mathbb{C}_{\text{ARL}}(\gamma)} \text{ESADD}(T) = \text{ESADD}(T_{\text{CS}})(1 + o(1)), \quad \gamma \rightarrow \infty.$$

Later on, Moustakides [47], using optimal stopping theory, in his ingenious paper, established the exact optimality of CUSUM for any ARL to the false alarm $\gamma \geq 1$.

Another popular, less pessimistic minimax criterion is from Pollak [48]:

$$\inf_T \sup_{\nu \in \mathbb{Z}_+} \text{CEDD}_\nu(T) \quad \text{subject to } \text{ARL2FA}(T) \geq \gamma,$$

which requires minimizing the conditional expected delay to detection $\text{CEDD}_\nu(T) = E_\nu[T - \nu \mid T > \nu]$ in the worst-case scenario with respect to the change point ν in class $\mathbb{C}_{\text{ARL}}(\gamma)$. Under Pollak's minimax approach, the goal is to find a stopping time $T_{\text{opt}} \in \mathbb{C}_{\text{ARL}}(\gamma)$ such that

$$\sup_{\nu \in \mathbb{Z}_+} \text{CEDD}_\nu(T_{\text{opt}}) = \inf_{T \in \mathbb{C}_{\text{ARL}}(\gamma)} \sup_{\nu \in \mathbb{Z}_+} \text{CEDD}_\nu(T) \quad \text{for any } \gamma \geq 1.$$

For the i.i.d. model (50), Pollak [48] showed that the modified SR detection procedure that starts from the quasi-stationary distribution of the SR statistic (i.e., the head-start r_0 in the SR- r procedure is a specific random variable) is third-order asymptotically optimal as $\gamma \rightarrow \infty$, i.e., the best one can attain up to an additive term $o(1)$:

$$\inf_{T \in \mathbb{C}_{\text{ARL}}(\gamma)} \sup_{\nu \in \mathbb{Z}_+} \text{CEDD}_\nu(T) = \sup_{\nu \in \mathbb{Z}_+} \text{CEDD}_\nu(T_{\text{SR}}^{r_0}) + o(1), \quad \gamma \rightarrow \infty,$$

where $o(1) \rightarrow 0$ as $\gamma \rightarrow \infty$. Later, Tartakovsky et al. [49] proved that this is also true for the SR- r procedure (59) that starts from the fixed but specially designed point $r_0 = r_0(\gamma)$ that depends on γ , which was first introduced and thoroughly studied by Moustakides et al. [44]. See also Polunchenko and Tartakovsky [50] on the exact optimality of the SR- r procedure.

Bayesian Changepoint Optimization Criterion

In Bayesian problems, the point of change ν is treated as random with a prior distribution $\pi_k = P(\nu = k)$, $k \in \mathbb{Z}_+$. Define the probability measure on the Borel σ -algebra $\mathcal{B}$ in $\mathbb{R}^\infty \times \mathbb{Z}_+$ as

$$P^\pi(\mathcal{A} \times \mathcal{K}) = \sum_{k \in \mathcal{K}} \pi_k P_k(\mathcal{A}), \quad \mathcal{A} \in \mathcal{B}(\mathbb{R}^\infty), \quad \mathcal{K} \in \mathbb{Z}_+.$$

Under measure P^π , the change point ν has a distribution $\pi = \{\pi_k\}$ and the model for the observations is given in (49).

From the Bayesian point of view, it is reasonable to measure the false alarm risk with the *weighted probability of false alarm* (PFA) defined as

$$\text{PFA}^\pi(T) := \mathbb{P}^\pi(T \leq \nu) = \sum_{k=0}^{\infty} \pi_k \mathbb{P}_k(T \leq k) = \sum_{k=0}^{\infty} \pi_k \mathbb{P}_\infty(T \leq k). \quad (60)$$

The last equality follows from the fact that $\mathbb{P}_k(T \leq k) = \mathbb{P}_\infty(T \leq k)$ because the event $\{T \leq k\}$ depends on the first k observations which under measure $\mathbb{P}_k$ correspond to the no-change hypothesis H_∞ . Thus, for $\alpha \in (0, 1)$, introduce the class of changepoint detection procedures

$$\mathbb{C}_\pi(\alpha) = \{T : \text{PFA}^\pi(T) \leq \alpha\} \quad (61)$$

for which the weighted PFA does not exceed a prescribed level α .

Let $\mathbb{E}^\pi$ denote the expectation with respect to the measure $\mathbb{P}^\pi$.

Shiryaev [18,43] introduced the Bayesian optimality criterion

$$\inf_{T \in \mathbb{C}_\pi(\alpha)} \mathbb{E}^\pi[(T - \nu)^+],$$

which is equivalent to minimizing the conditional expected detection delay $\text{EDD}^\pi(T) = \mathbb{E}^\pi[T - \nu | T > \nu]$

$$\inf_T \text{EDD}^\pi(T) \quad \text{subject to } \text{PFA}^\pi(T) \leq \alpha.$$

Under the Bayesian approach, the goal is to find a stopping time $T_{\text{opt}} \in \mathbb{C}_\pi(\alpha)$ such that

$$\text{EDD}^\pi(T_{\text{opt}}) = \inf_{T \in \mathbb{C}_\pi(\alpha)} \text{EDD}^\pi(T) \quad \text{for any } \alpha \in (0, 1). \quad (62)$$

For the i.i.d. model (50) and for the geometric prior distribution $\text{Geometric}(q)$ of the changepoint ν (see (56)), this problem was solved by Shiryaev [18,43]. Shiryaev [18,43,51] proved that the detection procedure given by the stopping time $T_{\text{SH}}(A)$ defined in (54) is strictly optimal in class $\mathbb{C}_\pi(\alpha)$ if $A = A_\alpha$ in (54) can be selected in such a way that $\text{PFA}^\pi(T_{\text{SH}}(A_\alpha)) = \alpha$, that is

$$\inf_{T \in \mathbb{C}_\pi(\alpha)} \text{EDD}^\pi(T) = \text{EDD}^\pi(T_{\text{SH}}(A_\alpha)) \quad \text{for any } \alpha \in (0, 1).$$

Uniform Pointwise Optimality Criterion

In many applications, the most reasonable optimality criterion is the pointwise uniform criterion of minimizing the conditional expected detection delay $\text{CEDD}_\nu(T) = \mathbb{E}_\nu[T - \nu | T \geq \nu]$ for all $\nu \in \mathbb{Z}_+$ when the false alarm risk is fixed at a certain level. However, as we already mentioned, if it is required that $\text{ARL2FA}(T) \geq \gamma$ for some $\gamma \geq 1$, then a procedure that minimizes $\text{CEDD}_\nu(T)$ for all ν does not exist. More importantly, as discussed in ([5] Section 2.3), the requirement of having large values of the $\text{ARL2FA}(T)$ generally does not guarantee small values of the maximal local probability of false alarm $\text{MLPFA}(T) = \sup_{\ell \geq 0} \mathbb{P}_\infty(T \leq \ell + m | T > \ell)$ in a time window of a length $m \geq 1$, while the opposite is always true (see Lemmas 2.1–2.2 in [5]). Hence, the constraint $\text{MLPFA}(T) \leq \beta$ is more stringent than $\text{ARL2FA}(T) \geq \gamma$.

Another reason for considering the MLPFA constraint instead of the ARL to false alarm constraint is that the latter one makes sense if and only if the $\mathbb{P}_\infty$ -distribution of stopping times are geometric or at least close to geometric, which is often the case for many popular detection procedures such as CUSUM and SR in the i.i.d. case. However, for general non-i.i.d. models, this is not necessarily true (see [5,52] for a detailed discussion).

For these reasons, introduce the most stringent class of change detection procedures for which the MLPFA(T) is upper-bounded by the prespecified level $\beta \in (0, 1)$:

$$\mathbb{C}_{\text{PFA}}(m, \beta) = \left\{ T : \sup_{\ell \geq 0} \mathbb{P}_{\infty}(T \leq \ell + m | T > \ell) \leq \beta \right\}. \quad (63)$$

The goal is to find a stopping time $T_{\text{opt}} \in \mathbb{C}_{\text{PFA}}(m, \beta)$ such that

$$\text{CEDD}_{\nu}(T_{\text{opt}}) = \inf_{T \in \mathbb{C}_{\text{PFA}}(m, \beta)} \text{CEDD}_{\nu}(T) \quad \text{for all } \nu \in \mathbb{Z}_+ \text{ and any } 0 < \beta < 1. \quad (64)$$

3.2.4. Asymptotic Optimality for General Non-i.i.d. Models via r -Quick and r -Complete Convergence

Complete Convergence and General Bayesian Changepoint Detection Theory

First consider the Bayesian problem assuming that the change point ν is a random variable independent of the observations with a prior distribution $\pi = \{\pi_k\}$. Unfortunately, in the general non-i.i.d. case and for an arbitrary prior π , the Bayesian optimization problem (62) is intractable for arbitrary values of PFA $\alpha \in (0, 1)$. For this reason, we will consider the first-order asymptotic problem assuming that the given PFA α approaches zero. To be specific, the goal is to design such a detection procedure T^* that asymptotically minimizes the expected detection delay $\text{EDD}^{\pi}(T)$ to first order as $\alpha \rightarrow 0$:

$$\inf_{T \in \mathbb{C}_{\pi}(\alpha)} \text{EDD}^{\pi}(T) = \text{EDD}^{\pi}(T^*)(1 + o(1)) \quad \text{as } \alpha \rightarrow 0, \quad (65)$$

where $o(1) \rightarrow 0$ as $\alpha \rightarrow 0$. It turns out that, in the asymptotic setting, it is also possible to find a procedure that minimizes the conditional expected detection delay $\text{EDD}_k(T) = \mathbb{E}_k[T - k | T > k]$ uniformly for all possible values of the change point $\nu = k \in \mathbb{Z}_+$, i.e.,

$$\lim_{\alpha \rightarrow 0} \frac{\inf_{T \in \mathbb{C}_{\pi}(\alpha)} \text{EDD}_k(T)}{\text{EDD}_k(T^*)} = 1 \quad \text{for all } k \in \mathbb{Z}_+. \quad (66)$$

Furthermore, asymptotic optimality results can also be established for higher moments of the detection delay of the order of $r > 1$

$$\mathbb{E}_k[(T - k)^r | T > k] \quad \text{and} \quad \mathbb{E}^{\pi}[(T - \nu)^r | T > \nu].$$

Since the Shiryaev procedure $T_{\text{SH}}(A)$, which was defined in (54), (55), is optimal for the i.i.d. model and Geometric(q) prior, it is reasonable to assume that it is asymptotically optimal for the more general prior and the non-i.i.d model. However, to study asymptotic optimality, we need certain constraints imposed on the prior distribution and on the asymptotic behavior of the decision statistics as the sample size increases, i.e., on the general stochastic model (49).

Assume that the prior distribution $\{\pi_k\}$ is fully supported, i.e., $\pi_k > 0$ for all $k \in \mathbb{Z}_+$ and $\pi_{\infty} = 0$ and that the following condition holds:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \left| \log \sum_{k=n+1}^{\infty} \pi_k \right| = \mu \quad \text{for some } 0 \leq \mu < \infty. \quad (67)$$

Obviously, if $\mu > 0$, then the prior π has an exponential right tail (e.g., the geometric distribution Geometric(q), in which case $\mu = |\log(1 - q)|$). If $\mu = 0$, then it has a heavier tail than an exponential tail. In this case, we will refer to it as a heavy-tailed distribution.

Define the LLR of the hypotheses H_k and H_{∞}

$$\lambda_n^k = \log \frac{dP_k^n}{dP_{\infty}^n} = \sum_{t=k+1}^n \frac{f_t(X_t | \mathbf{X}^t)}{g_t(X_t | \mathbf{X}^t)}, \quad n > k$$

($\lambda_n^k = 0$ for $n \leq k$). To obtain asymptotic optimality results, the general non-i.i.d. model for observations is restricted to the case that the normalized LLR $n^{-1}\lambda_{k+n}^k$ obeys the SLLN as $n \rightarrow \infty$ with a finite and positive number I under the probability measure P_k and its r -complete strengthened version

$$\sum_{n=1}^{\infty} n^{r-1} \sup_{k \in \mathbb{Z}_+} P_k \left\{ |n^{-1}\lambda_{k+n}^k - I| > \varepsilon \right\} < \infty \quad \text{for every } \varepsilon > 0. \quad (68)$$

It follows from Lemma 7.2.1 in [6] that, for any $A > 0$,

$$\text{PFA}^\pi(\text{T}_{\text{SH}}(A)) \leq (1 + A)^{-1},$$

so that $\text{T}_{\text{SH}}(A_\alpha) \in \mathbb{C}_\pi(\alpha)$ if $A = A_\alpha = (1 - \alpha)/\alpha$.

The following theorem that can be deduced from Theorem 3.7 in [5] shows that the Shiryaev detection procedure is asymptotically optimal if the normalized LLR $n^{-1}\lambda_{k+n}^k$ converges r -completely to a positive and finite number I and the prior distribution satisfies condition (67).

Theorem 6. Suppose that the prior distribution $\pi = \{\pi_k\}_{k \in \mathbb{Z}_+}$ of the change point satisfies condition (67) with some $0 \leq \mu < \infty$. Assume that there exists some number $0 < I < \infty$ such that the LLR process $n^{-1}\lambda_{k+n}^k$ converges to I uniformly r -completely as $n \rightarrow \infty$ under P_k , i.e., condition (68) holds for some $r \geq 1$. If threshold $A = A_\alpha$ in the Shiryaev procedure is so selected that $\text{PFA}^\pi(\text{T}_{\text{SH}}(A_\alpha)) \leq \alpha$ and $\log A_\alpha \sim |\log \alpha|$ as $\alpha \rightarrow 0$, e.g., as $A = (1 - \alpha)/\alpha$, then as $\alpha \rightarrow 0$

$$\inf_{T \in \mathbb{C}_\pi(\alpha)} E_k[(T - k)^r | T > k] \sim \left(\frac{|\log \alpha|}{I + \mu} \right)^r \sim E_k[(\text{T}_{\text{SH}} - k)^r | \text{T}_{\text{SH}} > k] \quad \text{for all } k \in \mathbb{Z}_+$$

and

$$\inf_{T \in \mathbb{C}_\pi(\alpha)} E^\pi[(T - \nu)^r | T > \nu] \sim \left(\frac{|\log \alpha|}{I + \mu} \right)^r \sim E^\pi[(\text{T}_{\text{SH}} - \nu)^r | \text{T}_{\text{SH}} > \nu].$$

Therefore, the Shiryaev procedure $\text{T}_{\text{SH}}(A_\alpha)$ is first-order asymptotically optimal as $\alpha \rightarrow 0$ in class $\mathbb{C}_\pi(\alpha)$, minimizing the moments of the detection delay up to order r whenever the r -complete version of the SLLN (68) holds for the LLR process.

For $r = 1$, the assertions of this theorem imply the asymptotic optimality of the Shiryaev procedure for the expected detection delays (65) and (66) as well as asymptotic approximations for the expected detection delays.

Remark 6. The results of Theorem 6 can be generalized to the asymptotically non-stationary case where $\lambda_{k+n}^k/\psi(n)$ converges to I uniformly r -completely as $n \rightarrow \infty$ under P_k with a non-linear function $\psi(n)$ similarly to the hypothesis testing problem discussed in Section 3.1. See also the recent paper [53] for the minimax change detection problem with independent but substantially non-stationary post-change observations.

It is also interesting to see how two other most popular changepoint detection procedures—the SR and CUSUM—perform in the Bayesian context.

Consider the SR procedure defined by (58), (59). By Lemma 3.4 (p. 100) in [5],

$$\text{PFA}^\pi(\text{T}_{\text{SR}}^{r_0}(A)) \leq \frac{r_0 \sum_{k=1}^{\infty} \pi_k + \sum_{k=1}^{\infty} k \pi_k}{A} \quad \text{for every } A > 0,$$

and therefore, setting $A = A_\alpha = \alpha^{-1}(r_0 + \sum_{k=1}^{\infty} k\pi_k)$ implies that $T_{SR}^{r_0}(A_\alpha) \in \mathbb{C}_\pi(\alpha)$. If threshold $A = A_\alpha$ in the SR procedure is so selected that $\text{PFA}_\pi(T_{SR}^{r_0}(A_\alpha)) \leq \alpha$ and $\log A_\alpha \sim |\log \alpha|$ as $\alpha \rightarrow 0$, e.g., as $A_\alpha = \alpha^{-1}(r_0 + \sum_{k=1}^{\infty} k\pi_k)$, then as $\alpha \rightarrow 0$

$$E_k[(T_{SR}^{r_0} - k)^r | T_{SR}^{r_0} > k] \sim \left(\frac{|\log \alpha|}{I}\right)^r \quad \text{for all } k \in \mathbb{Z}_+ \quad (69)$$

and

$$E^\pi[(T_{SR}^{r_0} - \nu)^r | T_{SR}^{r_0} > \nu] \sim \left(\frac{|\log \alpha|}{I}\right)^r \quad (70)$$

whenever the uniform r -complete convergence condition (68) holds. Therefore, the SR procedure $T_{SR}^{r_0}(A_\alpha)$ is first-order asymptotically optimal as $\alpha \rightarrow 0$ in class $\mathbb{C}_\pi(\alpha)$, minimizing the moments of the detection delay up to order r , when the prior distribution π is heavy-tailed (i.e., when $\mu = 0$) and the r -complete version of the SLLN holds. In the case where $\mu > 0$ (i.e., the prior distribution has an exponential tail), the SR procedure is not optimal. This can be expected since it uses the improper uniform prior in the detection statistic.

The same asymptotic results (69), (70) are true for the CUSUM procedure $T_{CS}(a)$ defined in (52) if threshold $a = a_\alpha$ is so selected that $\text{PFA}_\pi(T_{CS}(a_\alpha)) \leq \alpha$ and $a_\alpha \sim |\log \alpha|$ as $\alpha \rightarrow 0$ and the uniform r -complete convergence condition (68) holds.

Hence, the r -complete convergence of the LLR process is the sufficient condition for the uniform asymptotic optimality of several popular change detection procedures in class $\mathbb{C}_\pi(\alpha)$.

Complete Convergence and General Non-Bayesian Changepoint Detection Theory

Consider the non-Bayesian problem where the change point ν is an unknown deterministic number. We focus on the most interesting for a variety of applications uniform optimality criterion (64) that requires minimizing the conditional expected delay to detection $\text{CEDD}_\nu(T) = E_\nu[T - \nu | T > \nu]$ for all values of the change point $\nu \in \mathbb{Z}_+$ in the class of change detection procedures $\mathbb{C}_{\text{PFA}}(m, \beta)$ defined in (63). Recall that this class includes change detection procedures with the maximal local probability of false alarm in the time window m ,

$$\text{MLPFA}(T) = \sup_{\ell \geq 0} P_\infty(T \leq \ell + m | T > \ell),$$

which does not exceed the prescribed value $\beta \in (0, 1)$. However, the exact solution to this challenging problem is unknown even in the i.i.d. case.

Instead consider the following asymptotic problem assuming that the given MLPFA β goes to zero: find a change detection procedure T^* which asymptotically minimizes the expected detection delay $E_\nu[T - \nu | T > \nu]$ to the first order as $\beta \rightarrow 0$. That is, the goal is to design such a detection procedure T^* that

$$\inf_{T \in \mathbb{C}_{\text{PFA}}(m, \beta)} E_\nu[T - \nu | T > \nu] = E_\nu[T^* - \nu | T^* > \nu](1 + o(1)) \quad \text{for all } \nu \in \mathbb{Z}_+ \text{ as } \beta \rightarrow 0.$$

More generally, we may focus on the asymptotic problem of minimizing the moments of the detection delay of order $r \geq 1$:

$$\inf_{T \in \mathbb{C}_{\text{PFA}}(m, \beta)} E_\nu[(T - \nu)^r | T > \nu] = E_\nu[(T^* - \nu)^r | T^* > \nu](1 + o(1)) \quad \text{for all } \nu \in \mathbb{Z}_+ \text{ as } \beta \rightarrow 0.$$

To solve this problem, we need to assume that the window length $m = m_\beta$ is a function of the MLPFA constraint β and that m_β goes to infinity as $\beta \rightarrow 0$ with a certain appropriate rate. Using [54], the following results can be established.

Consider the SR procedure defined by (58), (59) with $r_0 = 0$, in which case write $T_{SR}^{r_0}(A) = T_{SR}(A)$. Let $r \geq 1$ and assume that the r -complete version of the SLLN holds with some number $0 < I < \infty$, i.e., $n^{-1}\lambda_{\nu+n}^\nu$ converges to I uniformly r -completely as

$n \rightarrow \infty$ under P_ν . If $m_\beta = O(|\log \beta|^2)$ as $\beta \rightarrow \infty$ and threshold $A = A_\beta$ in the SR procedure is so selected that $\text{MLPFA}(\text{T}_{\text{SR}}(A_\beta)) \leq \beta$ and $\log A_\beta \sim |\log \beta|$ as $\beta \rightarrow 0$, e.g., as defined in [54], then as $\beta \rightarrow 0$

$$\inf_{T \in \mathbb{C}_{\text{PFA}}(m_\beta, \beta)} E_\nu[(T - \nu)^r | T > \nu] \sim \left(\frac{|\log \beta|}{I} \right)^r \sim E_\nu[(\text{T}_{\text{SR}} - \nu)^r | \text{T}_{\text{SR}} > \nu] \quad \text{for all } \nu \in \mathbb{Z}_+.$$

A similar result also holds for the CUSUM procedure $\text{T}_{\text{CS}}(a)$ if threshold $a = a_\beta$ is selected so that $\text{MLPFA}(\text{T}_{\text{CS}}(a_\beta)) \leq \beta$ and $a_\beta \sim |\log \beta|$ as $\beta \rightarrow 0$ and the r -complete version of the SLLN holds for the normalized LLR $n^{-1} \lambda_{\nu+n}^\nu$ as $n \rightarrow \infty$.

Hence, the r -complete convergence of the LLR process is the sufficient condition for the uniform asymptotic optimality of SR and CUSUM change detection procedures with respect to the moments of the detection delay of order r in class $\mathbb{C}_{\text{PFA}}(m_\beta, \beta)$.

4. Quick and Complete Convergence for Markov and Hidden Markov Models

Usually, in particular problems, the verification of the SLLN for the LLR process is relatively easy. However, in practice, verifying the strengthened r -complete or r -quick versions of the SLLN, i.e., checking condition (68), can cause some difficulty. Many interesting examples where this verification was performed can be found in [5,6]. However, it is interesting to find sufficient conditions for the r -complete convergence for a relatively large class of stochastic models.

In this section, we outline this issue for Markov and hidden Markov models based on the results obtained by Pergamenchikov and Tartakovsky [54] for ergodic Markov processes and by Fuh and Tartakovsky [55] for hidden Markov models (HMM). See also Tartakovsky ([5] Ch 3).

Let $\{X_n\}_{n \in \mathbb{Z}_+}$ be a time-homogeneous Markov process with values in a measurable space $(\mathcal{X}, \mathcal{B})$ with the transition probability $P(x, A)$ with density $p(y|x)$. Let E_x denote the expectation with respect to this probability. Assume that this process is geometrically ergodic, i.e., there exist positive constants $0 < R < \infty$, $\kappa > 0$, and probability measure π on $(\mathcal{X}, \mathcal{B})$ and the Lyapunov $\mathcal{X} \rightarrow [1, \infty)$ function V with $\pi(V) < \infty$ such that

$$\sup_{n \in \mathbb{Z}_+} e^{\kappa n} \sup_{0 < \psi \leq V} \sup_x \frac{1}{V(x)} |E_x[\psi(X_n)] - \pi(\psi)| \leq R.$$

In the change detection problem, the sequence $\{X_n\}_{n \in \mathbb{Z}_+}$ is a Markov process, such that $\{X_n\}_{1 \leq n \leq \nu}$ is a homogeneous process with the transition density $g(y|x)$ and $\{X_n\}_{n > \nu}$ is homogeneous positive ergodic with the transition density $f(y|x)$ and the ergodic (stationary) distribution π . In this case, the LLR process λ_n^k can be represented as

$$\lambda_n^k = \sum_{t=k+1}^n G(X_t, X_{t-1}), \quad n > k,$$

where $G(y, x) = \log[f(y|x)/g(y|x)]$.

Define

$$I = \int_{\mathcal{X}} \left\{ \int_{\mathcal{X}} G(y, x) f(y|x) dy \right\} \pi(dx).$$

Under a set of quite sophisticated sufficient conditions, the LLR λ_{k+n}^n/n converges to I as $n \rightarrow \infty$ r -completely (cf. [54]). We omit the details and only mention that the main condition is the finiteness of $(r+1)$ -th moment of the LLR increment, $E_0[(G(X_1, X_0))^{r+1}] < \infty$.

Now consider the HMM with finite state space. Then again, as in the pure Markov case, the main condition for the r -complete convergence of λ_{k+n}^n/n to I , where I is specified in Fuh and Tartakovsky [55], is $E_0[(\lambda_1^0)^{r+1}] < \infty$. Further details can be found in [55].

Similar results for Markov and hidden Markov models hold for the hypothesis testing problem considered in Section 3.1. Specifically, if in the Markov case we assume that the

observed Markov process $\{X_n\}_{n \in \mathbb{Z}_+}$ is a time-homogeneous geometrically ergodic with a transition density $f_i(y|x)$ under hypothesis H_i ($i = 0, 1, \dots, N$) and invariant distribution π_i , then the LLR processes are

$$\lambda_{ij}(n) = \sum_{t=1}^n G_{ij}(X_t, X_{t-1}), \quad i, j = 0, 1, \dots, N, i \neq j,$$

where $G_{ij}(y, x) = \log[f_i(y|x)/f_j(y|x)]$. If $E_i[(G_{ij}(X_1, X_0))^{r+1}] < \infty$, then the LLR $n^{-1}\lambda_{ij}(n)$ converges r -completely to a finite number

$$I_{ij} = \int_{\mathcal{X}} \left\{ \int_{\mathcal{X}} G_{ij}(y, x) f_i(y|x) dy \right\} \pi_i(dx).$$

5. Discussion and Conclusions

The purpose of this article is to provide an overview of two modes of convergence in the LLN— r -quick and r -complete convergences. These strengthened versions of the SLLN are often neglected in the theory of probability. In the first part of this paper (Section 2), we discussed in detail these two modes of convergence and corresponding strengthened versions of the SLLN. The main motivation was the fact that both r -quick and r -complete versions of the SLLN can be effectively used for establishing near optimality results in sequential analysis, in particular, in sequential hypothesis testing and quickest changepoint detection problems for very general stochastic models of dependent and non-stationary observations. These models are not limited to Markov and hidden Markov models. The results presented in the second part of this paper (Section 3) show that the constraints imposed on the models for observations can be formulated in terms of either the r -quick or r -complete convergence of properly normalized log-likelihood ratios between hypotheses to finite numbers, which can be interpreted as local Kullback–Leibler information numbers. This is natural and can be intuitively expected since optimal or nearly optimal decision-making rules are typically based on a combination of log-likelihood ratios. Therefore, if one is interested in the asymptotic optimality properties of decision-making rules, the asymptotic behavior of log-likelihood ratios as the sample size goes to infinity not only matters but provides the main contribution.

The results presented in this article allow us to conclude that the strengthened r -quick and r -complete versions of the SLLN are useful tools for many statistical problems for general non-i.i.d. stochastic models. In particular, r -quick and r -complete convergences for log-likelihood ratio processes are sufficient for the near optimality of sequential hypothesis tests and changepoint detection procedures for models with dependent and non-identically distributed observations. Such non-i.i.d. models are typical for modern large-scale information and physical systems that produce big data in numerous practical applications. Readers interested in specific applications may find detailed discussions in [4–7,21,22,33,35,37,53–58].

Funding: This article received no external funding.

Data Availability Statement: No real data were used in this research.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Hsu, P.L.; Robbins, H. Complete convergence and the law of large numbers. *Proc. Natl. Acad. Sci. USA* **1947**, *33*, 25–31. [[CrossRef](#)] [[PubMed](#)]
2. Baum, L.E.; Katz, M. Convergence rates in the law of large numbers. *Trans. Am. Math. Soc.* **1965**, *120*, 108–123. [[CrossRef](#)]
3. Strassen, V. Almost sure behavior of sums of independent random variables and martingales. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, San Diego, CA, USA, 21 June–18 July 1965 and 27 December 1965–7 January 1966*; Le Cam, L.M., Neyman, J., Eds.; Vol. 2: Contributions to Probability Theory. Part 1; University of California Press: Berkeley, CA, USA, 1967; pp. 315–343.

4. Tartakovsky, A.G. Asymptotic optimality of certain multihypothesis sequential tests: Non-i.i.d. case. *Stat. Inference Stoch. Process.* **1998**, *1*, 265–295. [\[CrossRef\]](#)
5. Tartakovsky, A.G. *Sequential Change Detection and Hypothesis Testing: General Non-i.i.d. Stochastic Models and Asymptotically Optimal Rules*; Monographs on Statistics and Applied Probability 165; Chapman & Hall/CRC Press, Taylor & Francis Group: Boca Raton, FL, USA; London, UK; New York, NY, USA, 2020.
6. Tartakovsky, A.G.; Nikiforov, I.V.; Basseville, M. *Sequential Analysis: Hypothesis Testing and Changepoint Detection*; Monographs on Statistics and Applied Probability 136; Chapman & Hall/CRC Press, Taylor & Francis Group: Boca Raton, FL, USA; London, UK; New York, NY, USA, 2015.
7. Lai, T.L. Asymptotic optimality of invariant sequential probability ratio tests. *Ann. Stat.* **1981**, *9*, 318–333. [\[CrossRef\]](#)
8. Lai, T.L. On r -quick convergence and a conjecture of Strassen. *Ann. Probab.* **1976**, *4*, 612–627. [\[CrossRef\]](#)
9. Chow, Y.S.; Lai, T.L. Some one-sided theorems on the tail distribution of sample sums with applications to the last time and largest excess of boundary crossings. *Trans. Am. Math. Soc.* **1975**, *208*, 51–72. [\[CrossRef\]](#)
10. Fuh, C.D.; Zhang, C.H. Poisson equation, moment inequalities and quick convergence for Markov random walks. *Stoch. Process. Their Appl.* **2000**, *87*, 53–67. [\[CrossRef\]](#)
11. Wald, A. Sequential tests of statistical hypotheses. *Ann. Math. Stat.* **1945**, *16*, 117–186. [\[CrossRef\]](#)
12. Wald, A. *Sequential Analysis*; John Wiley & Sons, Inc.: New York, NY, USA, 1947.
13. Wald, A.; Wolfowitz, J. Optimum character of the sequential probability ratio test. *Ann. Math. Stat.* **1948**, *19*, 326–339. [\[CrossRef\]](#)
14. Burkholder, D.L.; Wijsman, R.A. Optimum properties and admissibility of sequential tests. *Ann. Math. Stat.* **1963**, *34*, 1–17. [\[CrossRef\]](#)
15. Matthes, T.K. On the optimality of sequential probability ratio tests. *Ann. Math. Stat.* **1963**, *34*, 18–21. [\[CrossRef\]](#)
16. Ferguson, T.S. *Mathematical Statistics: A Decision Theoretic Approach*; Probability and Mathematical Statistics; Academic Press: Cambridge, MA, USA, 1967.
17. Lehmann, E.L. *Testing Statistical Hypotheses*; John Wiley & Sons, Inc.: New York, NY, USA, 1968.
18. Shiryaev, A.N. *Optimal Stopping Rules*; Series on Stochastic Modelling and Applied Probability; Springer: New York, NY, USA, 1978; Volume 8.
19. Golubev, G.K.; Khas'minskii, R.Z. Sequential testing for several signals in Gaussian white noise. *Theory Probab. Appl.* **1984**, *28*, 573–584. [\[CrossRef\]](#)
20. Tartakovsky, A.G. Asymptotically optimal sequential tests for nonhomogeneous processes. *Seq. Anal.* **1998**, *17*, 33–62. [\[CrossRef\]](#)
21. Verdenskaya, N.V.; Tartakovskii, A.G. Asymptotically optimal sequential testing of multiple hypotheses for nonhomogeneous Gaussian processes in an asymmetric situation. *Theory Probab. Appl.* **1991**, *36*, 536–547. [\[CrossRef\]](#)
22. Fellouris, G.; Tartakovsky, A.G. Multichannel sequential detection—Part I: Non-i.i.d. data. *IEEE Trans. Inf. Theory* **2017**, *63*, 4551–4571. [\[CrossRef\]](#)
23. Armitage, P. Sequential analysis with more than two alternative hypotheses, and its relation to discriminant function analysis. *J. R. Stat. Soc.-Ser. Methodol.* **1950**, *12*, 137–144. [\[CrossRef\]](#)
24. Chernoff, H. Sequential design of experiments. *Ann. Math. Stat.* **1959**, *30*, 755–770. [\[CrossRef\]](#)
25. Kiefer, J.; Sacks, J. Asymptotically optimal sequential inference and design. *Ann. Math. Stat.* **1963**, *34*, 705–750. [\[CrossRef\]](#)
26. Lorden, G. Integrated risk of asymptotically Bayes sequential tests. *Ann. Math. Stat.* **1967**, *38*, 1399–1422. [\[CrossRef\]](#)
27. Lorden, G. Nearly-optimal sequential tests for finitely many parameter values. *Ann. Stat.* **1977**, *5*, 1–21. [\[CrossRef\]](#)
28. Pavlov, I.V. Sequential procedure of testing composite hypotheses with applications to the Kiefer-Weiss problem. *Theory Probab. Appl.* **1990**, *35*, 280–292. [\[CrossRef\]](#)
29. Baron, M.; Tartakovsky, A.G. Asymptotic optimality of change-point detection schemes in general continuous-time models. *Seq. Anal.* **2006**, *25*, 257–296. [\[CrossRef\]](#)
30. Mosteller, F. A k -sample slippage test for an extreme population. *Ann. Math. Stat.* **1948**, *19*, 58–65. [\[CrossRef\]](#)
31. Bakut, P.A.; Bolshakov, I.A.; Gerasimov, B.M.; Kuriksha, A.A.; Repin, V.G.; Tartakovsky, G.P.; Shirokov, V.V. *Statistical Radar Theory*; Tartakovsky, G.P., Ed.; Sovetskoe Radio: Moscow, Russia, 1963; Volume 1. (In Russian)
32. Basseville, M.; Nikiforov, I.V. *Detection of Abrupt Changes—Theory and Application*; Information and System Sciences Series; Prentice-Hall, Inc.: Englewood Cliffs, NJ, USA, 1993.
33. Jeske, D.R.; Steven, N.T.; Tartakovsky, A.G.; Wilson, J.D. Statistical methods for network surveillance. *Appl. Stoch. Model. Bus. Ind.* **2018**, *34*, 425–445. [\[CrossRef\]](#)
34. Jeske, D.R.; Steven, N.T.; Wilson, J.D.; Tartakovsky, A.G. Statistical network surveillance. In *Wiley StatsRef: Statistics Reference Online*; Wiley: New York, NY, USA, 2018; pp. 1–12. [\[CrossRef\]](#)
35. Tartakovsky, A.G.; Brown, J. Adaptive spatial-temporal filtering methods for clutter removal and target tracking. *IEEE Trans. Aerosp. Electron. Syst.* **2008**, *44*, 1522–1537. [\[CrossRef\]](#)
36. Szor, P. *The Art of Computer Virus Research and Defense*; Addison-Wesley Professional: Upper Saddle River, NJ, USA, 2005.
37. Tartakovsky, A.G. Rapid detection of attacks in computer networks by quickest changepoint detection methods. In *Data Analysis for Network Cyber-Security*; Adams, N., Heard, N., Eds.; Imperial College Press: London, UK, 2014; pp. 33–70.
38. Tartakovsky, A.G.; Rozovskii, B.L.; Blažek, R.B.; Kim, H. Detection of intrusions in information systems by sequential change-point methods. *Stat. Methodol.* **2006**, *3*, 252–293. [\[CrossRef\]](#)

39. Tartakovsky, A.G.; Rozovskii, B.L.; Blažek, R.B.; Kim, H. A novel approach to detection of intrusions in computer networks via adaptive sequential and batch-sequential change-point detection methods. *IEEE Trans. Signal Process.* **2006**, *54*, 3372–3382. [\[CrossRef\]](#)
40. Siegmund, D. Change-points: From sequential detection to biology and back. *Seq. Anal.* **2013**, *32*, 2–14. [\[CrossRef\]](#)
41. Moustakides, G.V. Sequential change detection revisited. *Ann. Stat.* **2008**, *36*, 787–807. [\[CrossRef\]](#)
42. Page, E.S. Continuous inspection schemes. *Biometrika* **1954**, *41*, 100–114. [\[CrossRef\]](#)
43. Shiryaev, A.N. On optimum methods in quickest detection problems. *Theory Probab. Appl.* **1963**, *8*, 22–46. [\[CrossRef\]](#)
44. Moustakides, G.V.; Polunchenko, A.S.; Tartakovsky, A.G. A numerical approach to performance analysis of quickest change-point detection procedures. *Stat. Sin.* **2011**, *21*, 571–596. [\[CrossRef\]](#)
45. Moustakides, G.V.; Polunchenko, A.S.; Tartakovsky, A.G. Numerical comparison of CUSUM and Shiryaev–Roberts procedures for detecting changes in distributions. *Commun. Stat.-Theory Methods* **2009**, *38*, 3225–3239. [\[CrossRef\]](#)
46. Lorden, G. Procedures for reacting to a change in distribution. *Ann. Math. Stat.* **1971**, *42*, 1897–1908. [\[CrossRef\]](#)
47. Moustakides, G.V. Optimal stopping times for detecting changes in distributions. *Ann. Stat.* **1986**, *14*, 1379–1387. [\[CrossRef\]](#)
48. Pollak, M. Optimal detection of a change in distribution. *Ann. Stat.* **1985**, *13*, 206–227. [\[CrossRef\]](#)
49. Tartakovsky, A.G.; Pollak, M.; Polunchenko, A.S. Third-order asymptotic optimality of the generalized Shiryaev–Roberts changepoint detection procedures. *Theory Probab. Appl.* **2012**, *56*, 457–484. [\[CrossRef\]](#)
50. Polunchenko, A.S.; Tartakovsky, A.G. On optimality of the Shiryaev–Roberts procedure for detecting a change in distribution. *Ann. Stat.* **2010**, *38*, 3445–3457. [\[CrossRef\]](#)
51. Shiryaev, A.N. The problem of the most rapid detection of a disturbance in a stationary process. *Sov. Math.-Dokl.* **1961**, *2*, 795–799; Translation from *Doklady Akademii Nauk SSSR* **1961**, *138*, 1039–1042.
52. Tartakovsky, A.G. Discussion on “Is Average Run Length to False Alarm Always an Informative Criterion?” by Yajun Mei. *Seq. Anal.* **2008**, *27*, 396–405. [\[CrossRef\]](#)
53. Liang, Y.; Tartakovsky, A.G.; Veeravalli, V.V. Quickest change detection with non-stationary post-change observations. *IEEE Trans. Inf. Theory* **2023**, *69*, 3400–3414. [\[CrossRef\]](#)
54. Pergamenchikov, S.; Tartakovsky, A.G. Asymptotically optimal pointwise and minimax quickest change-point detection for dependent data. *Stat. Inference Stoch. Process.* **2018**, *21*, 217–259. [\[CrossRef\]](#)
55. Fuh, C.D.; Tartakovsky, A.G. Asymptotic Bayesian theory of quickest change detection for hidden Markov models. *IEEE Trans. Inf. Theory* **2019**, *65*, 511–529. [\[CrossRef\]](#)
56. Kolessa, A.; Tartakovsky, A.; Ivanov, A.; Radchenko, V. Nonlinear estimation and decision-making methods in short track identification and orbit determination problem. *IEEE Trans. Aerosp. Electron. Syst.* **2020**, *56*, 301–312. [\[CrossRef\]](#)
57. Tartakovsky, A.; Berenkov, N.; Kolessa, A.; Nikiforov, I. Optimal sequential detection of signals with unknown appearance and disappearance points in time. *IEEE Trans. Signal Process.* **2021**, *69*, 2653–2662. [\[CrossRef\]](#)
58. Pergamenchikov, S.M.; Tartakovsky, A.G.; Spivak, V.S. Minimax and pointwise sequential changepoint detection and identification for general stochastic models. *J. Multivar. Anal.* **2022**, *190*, 104977. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.