Performance Analysis of Wald’s SPRT with Independent but Non-Stationary Log-Likelihood Ratios

Yu Liu and X. Rong Li

Department of Electrical Engineering
University of New Orleans
New Orleans, LA 70148, U.S.A.
Email: {lyu2, xli}@uno.edu

Abstract—The characteristics and behavior of Wald’s sequential probability ratio test are revealed by two important functions — operating characteristic (OC) and average sample number (ASN). These two functions have been studied extensively under the assumption of independent and identically distributed (i.i.d.) log-likelihood ratios, which is too stringent for many applications. This paper relaxes the requirement of identical distribution. Two inductive equations governing the OC and ASN are developed. Unfortunately, they have non-unique solutions in the general case. They do have unique solutions in two special cases: (a) the log-likelihood ratios converge in distributions and (b) the log-likelihood ratios have periodic distributions. Numerical solutions for these two special cases are obtained. They are compared with the results of Monte Carlo simulations because existing methods for this problem setting are lacking.

Keywords: sequential probability ratio test, operating characteristic function, average sample number, non-stationary.

I. INTRODUCTION

The well-known sequential probability ratio test (SPRT) proposed by Wald [1], [2] is a powerful tool of sequential analysis and is widely used in medicine, statistics, social science and engineering, such as clinical test, quality control, radar signal processing. Many versions of generalizations of the SPRT (referred to as GSPRT [3]) have been proposed [4]–[6] to improve the performance further for more complicated practical applications or to meet the requirements for more general settings than the simplest case of binary simple hypothesis testing with i.i.d. observations. Their behavior was studied in [3], [7]. Further, SPRT also forms the foundation of some change detection techniques, e.g., the celebrated Page’s cumulative sum (CUSUM) test [8].

A comprehensive survey of sequential analysis as well as the challenges was provided by Lai [9] and the subsequent comments and discussions by the reviewers.

The SPRT is generally applicable to sequential testing problems. Theoretically speaking, it requires neither the hypotheses to be simple nor the observations to be i.i.d., provided the likelihood ratios can be computed sequentially. However, analysis of SPRT’s properties and behavior becomes much more complicated with these assumptions relaxed.

The performance of the SPRT has been studied extensively. Its optimality for binary simple hypothesis testing with i.i.d. observations was proved in [10]–[12] — the expected sample sizes under both hypotheses are simultaneously minimized among all the tests that do not exceed the given type I and type II error rates. This optimality is remarkable and usually not achievable elsewhere. But when it comes to the composite hypothesis problem, this “miracle” is gone and analysis of the optimality becomes much more difficult (see [7], [9], [13], [14] and the references therein). Also, the optimality properties of the SPRT without the assumption of the i.i.d. observations have been studied, but only a few asymptotic results for some special cases are available (see [3], [15]–[18]).

Two important functions — operating characteristic (OC) and average sample number (ASN) — characterize the behavior of the SPRT. Unlike the optimality problem, the existing methods proposed to evaluate these two functions are almost all based on the assumption of an i.i.d. process. In this case, it has been known [8], [19], [20] for a long time that these two functions satisfy the Fredholm integral equations of the second kind (FIESK) [21], [22]. Analytical solutions can not be obtained except for some special cases — for example, Vardeman and Ray [23] provided an exact solution to the average run length (ARL) function of Page’s CUSUM test when the observations are i.i.d. and exponentially distributed. It is interesting since the ARL satisfies a similar FIESK as OC and ASN do. In general, one has to resort to some numerical techniques or to approximate the solutions under some assumptions. Wald [2] approximated the solution of OC and ASN by omitting the overshoot when the test statistic crosses one of the boundaries. An iterative method was proposed by Page [24] to numerically solve the FIESK, and the convergence property of this method has been examined by Kemp [25]. Besides, extensive studies have been done to compute the OC and ASN for the truncated SPRT with some specific processes (e.g., Poisson and Binomial) [26]–[29]. If the log-likelihood ratios are i.i.d. and Gaussian distributed, Goel and Wu [30] proposed a method to convert the FIESK to a system of linear algebraic equations (SLAE) by approximating the integrals with the Gaussian quadrature. It can be actually applied to the non-Gaussian case provided the Gaussian quadrature can work well. Mikhailova, et al. [31]

Research supported in part by ARO through grant W911NF-08-1-0409 and ONR-DEPSCoR through grant N00014-09-1-1169.
considered a problem with a discontinuous core of FIESK, and the solution becomes much more complicated. Other methods, such as finite element analysis [32]–[34], can also be employed to find the numerical solutions of the FIESK. Further, the bounds of the OC and ASN were calculated in [2], [20], [25], [35] (see the references therein).

However, all these studies of OC and ASN are based on the i.i.d. assumption. The stationarity of an i.i.d. process simplifies the analysis of OC and ASN greatly under which the statistical properties of the SPRT are the same regardless of the start time of the test (provided the initial values are the same). In this paper, we consider the case of an independent but non-stationary process, meaning that the log-likelihood ratios need not be identically distributed. This scenario is frequently encountered in applications. For example, if the SPRT is implemented based on the innovations of the Kalman filter [36], the log-likelihood ratios are independent or weakly coupled, but their distributions at different time instants are different in general. Or the parameter under test may be time varying, rendering the log-likelihood ratios not identically distributed. Once the log-likelihood ratios are not stationary, the OC and ASN in general do not obey FIESK any more, which invalidate the existing methods based on solving FIESK.

In this paper, we attack this problem in two steps: First, following the derivation of FIESK for the i.i.d. case, two inductive integral equations governing the OC and ASN respectively for the non-stationary case are obtained. They can be viewed as a generalization of the FIESK. The governing equations are in an inductive form since without the stationarity of the log-likelihood ratios, the OC and ASN are changing w.r.t. the start time of the test — the statistical properties of the SPRT with different start times are related but not identical. Unfortunately, the uniqueness of solution can not be guaranteed in general, rendering numerical solution difficult. By extending the SLAE method, these two inductive integral equations can be approximated by a system of linear equations. But it is under-determined and there are infinite many solutions. This also reveals the uniqueness problem of the solution. However, the theoretical values of these two governing equations can not be ignored since they form the foundation to analyze the OC and ASN and help understand the behavior of the SPRT. Second, for two frequently encountered special cases — (a) the log-likelihood ratios converge in distribution and (b) they have periodic distributions — additional conditions (equations) can be imposed, rendering a unique numerical solution, which can be easily obtained by solving a system of linear equations. As explained in Sec. IV, these two special cases are not uncommon in applications and their solutions are simple. No existing method known to the authors can be employed to compare in this problem setting, so our solutions for these two special cases are inspected and verified by Monte Carlo (MC) simulations.

This paper is organized as follows. First, Wald’s SPRT is reviewed in Sec. II. The OC and ASN with independent but non-stationary log-likelihood ratios are analyzed in Sec. III, and numerical solutions for two special cases are developed in Sec. IV. In Sec. V, two illustrative examples are provided and our solutions are compared with the results of MC simulation. Conclusions are made in Sec. VI.

II. OVERVIEW OF WALD’S SPRT

For a binary hypothesis testing problem,

\[ \begin{align*}
H_0 & : \theta \in \Theta_0 \\
H_1 & : \theta \in \Theta_1
\end{align*} \]

where \( \theta \) is the parameter under test, if the observations \( x_k \) are collected sequentially, then the SPRT computes the cumulative sum \( S_k \) of the log-likelihood ratios and the decision is made when enough data are collected:

- Declare \( H_1 \) if
  \[ S_k \geq B \approx \ln \frac{1 - \beta}{\alpha} \]
- Declare \( H_0 \) if
  \[ S_k \leq A \approx \ln \frac{\beta}{1 - \alpha} \]
- Else, continue
  \[ S_{k+1} = S_k + s_{k+1} \] (1)

where

\[ s_k = \ln \left( \frac{f(x_k|H_1)}{f(x_k|H_0)} \right) \]

is the log-likelihood ratio at time \( k \), \( A \) and \( B \) are the lower and upper thresholds depending on the type I error probability \( \alpha \) and type II error probability \( \beta \) desired by the users, \( f(x_k|H_i) \) is the likelihood function. Wald’s approximations of \( A \) and \( B \) are widely used in practical applications.

A. Operating Characteristic and Average Sample Number

If \( s_k \) are i.i.d. (conditioned on parameter \( \theta \)), extensive results for the OC and ASN are available. Denote the probability density function (PDF) of \( s_k \) as \( f_\theta(s) \), where the subscript \( \theta \) denotes the ground truth of the underlying parameter. This notation is used throughout this paper. Note that the ground truth of \( \theta \) need not be in either \( \Theta_0 \) or \( \Theta_1 \). First, define

\[ \tau_k = \min \{ t : S_t \leq A \text{ or } S_t \geq B, t \geq k \} \] (2)

as the stopping time variable. The subscript \( k \) denotes the start time of the test. Because of the stationarity for the i.i.d. case, it does not matter when the SPRT starts (i.e., its statistical properties do not change w.r.t. \( k \)). Without loss of generality, we assume the test starts at time \( k = 0 \).

The OC is defined as the probability that the test statistic finally drops below \( A \) as a function of the test initial value \( s \) (i.e., \( S_0 = s \)) and the ground truth \( \theta \):

\[ P_b(s) \triangleq P_b\{S_{\tau_0} \leq A|S_0 = s\} = 1 - P_b\{S_{\tau_0} \geq B|S_0 = s\} \]

Notice that \( P_b(\bullet) \) denotes the OC while the \( P_b(\bullet) \) denotes the probability of an event — the reader should be able to distinguish from the context without any ambiguity. The
second equality of Eq. (3) holds if and only if the SPRT terminates within finite steps almost surely:

\[ P_\theta \{ \tau_0 < \infty \} = 1 \]

Wald [2] has proved it under very mild conditions. One of the sufficient conditions [35] is

\[ P_\theta \{ s_k = 0 \} < 1, \forall k \]  

(4)

Define

\[ N_\theta (s) \triangleq E_\theta [ \tau_0 | S_0 = s ] \]

as the average sample number (ASN) [2], which is also a function of \( s \) and \( \theta \). In this paper, we do not consider the case that the ASN does not exist.

B. Governing Equations and Solutions with i.i.d. Log-Likelihood Ratios

It is known [8], [19], [20] that \( P_\theta (s) \) and \( N_\theta (s) \) satisfy the following Fredholm integral equations of the second kind (FIESK) [21], [22].

\[ P_\theta (s) = \int_{-\infty}^{A-s} f_\theta (x) \, dx + \int_{A}^{B} P_\theta (x) f_\theta (x-s) \, dx \]  

(5)

\[ N_\theta (s) = 1 + \int_{A}^{B} N_\theta (x) f_\theta (x-s) \, dx \]  

(6)

The existence and uniqueness of the solution for general FIESK are guaranteed under mild conditions, given in [37], [38]. In general, people have to resort to numerical approximation to the solutions.

III. THE OC AND ASN WITH INDEPENDENT BUT NON-STATIONARY LOG-LIKELIHOOD RATIOS

If the sequence \( \{ s_k \} \) is independent but not stationary, it does matter when the SPRT starts. So we define

\[ P_\theta^k (s) \triangleq P_\theta \{ S_{\tau_k} \leq A | S_k = s \} \]

as the OC for this case with the superscript \( k \) explicitly indicating the start time of the test. Obviously, this definition degenerates to Eq. (3) if \( \{ s_k \} \) is i.i.d. Similarly, define

\[ N_\theta^k (s) \triangleq E_\theta [ \tau_k - k | S_k = s ] \]

as the ASN. Note that the start time \( k \) is subtracted since we only consider how many future samples are needed on average.

Then \( P_\theta^k (s) \) and \( N_\theta^k (s) \) are governed by the following inductive integral equations,

\[ P_\theta^k (s) = \int_{-\infty}^{A-s} f_\theta^{k+1} (x) \, dx + \int_{A}^{B} P_\theta^{k+1} (x) f_\theta^{k+1} (x-s) \, dx \]  

(7)

\[ N_\theta^k (s) = 1 + \int_{A}^{B} N_\theta^{k+1} (x) f_\theta^{k+1} (x-s) \, dx \]  

(8)

where \( f_\theta^k (\bullet) \) denotes the PDF of \( s_k \), and the non-stationarity is explicitly revealed by the superscript \( k \). The idea of the derivation follows similarly as in [19], [20]. Since by definition \( P_\theta^k (s) \) is the probability of the event that the test statistic crosses the lower bound \( A \) with the start time \( k \) and initial value \( s \), this event can be partitioned to two mutually exclusive events:

\[ \Lambda_1 = \{ S_{k+1} = S_k + s_{k+1} = s + s_{k+1} \leq A \} \]

\[ \Lambda_2 = \{ A < S_{k+1} < B \} \cap \{ S_{\tau_{k+1}} \leq A | S_{k+1} \} \]

The first is the event that the test statistic crosses \( A \) at time \( k+1 \) while the second is that \( S_{k+1} \) is between the two bounds (i.e., the SPRT does not stop at time \( k+1 \)) and crosses \( A \) after time \( k+1 \). This second event can be viewed as the test starts at time \( k+1 \) with the initial value of \( S_{k+1} \). It is clear that the probability of the first event equals

\[ P_\theta \{ \Lambda_1 \} = \int_{-\infty}^{\Lambda-s} f_\theta^{k+1} (x) \, dx \]

The probability of the second event equals

\[ P_\theta \{ \Lambda_2 \} = \int_{A}^{B} P_\theta \{ S_{\tau_{k+1}} \leq A | S_{k+1} = x \} P_\theta \{ S_{k+1} = x \} \, dx \]

\[ = \int_{A}^{B} P_\theta \{ S_{\tau_{k+1}} \leq A \mid S_{k+1} = x \} f_\theta^{k+1} (x-s) \, dx \]

\[ = \int_{A}^{B} P_\theta^{k+1} (x) f_\theta^{k+1} (x-s) \, dx \]

Hence Eq. (7) follows. The derivation for \( N_\theta^k (s) \) follows similarly. First, define two mutually exclusive events

\[ \Gamma_1 = \{ S_{k+1} \leq A \} \cup \{ S_{k+1} \geq B \} \]

\[ \Gamma_2 = \{ A < S_{k+1} < B \} \]

It is clear that

\[ P_\theta \{ \Gamma_1 \} = 1 - P_\theta \{ \Gamma_2 \} \]

\[ P_\theta \{ \Gamma_2 \} = \int_{A}^{B} f_\theta^{k+1} (x-s) \, dx \]

so,

\[ N_\theta^k (s) = N_\theta^k (s|\Gamma_1) P_\theta \{ \Gamma_1 \} + N_\theta^k (s|\Gamma_2) P_\theta \{ \Gamma_2 \} \]

\[ = 1 \cdot (1 - P_\theta \{ \Gamma_2 \}) \]

\[ + \left( 1 + \int_{A}^{B} N_\theta^{k+1} (x) P_\theta \{ S_{k+1} = x | \Gamma_2 \} \, dx \right) P_\theta \{ \Gamma_2 \} \]

\[ = 1 \cdot (1 - P_\theta \{ \Gamma_2 \}) \]

\[ + \left( 1 + \int_{A}^{B} N_\theta^{k+1} (x) P_\theta \{ S_{k+1} = x, \Gamma_2 \} \, dx \right) P_\theta \{ \Gamma_2 \} \]

\[ = 1 \cdot (1 - P_\theta \{ \Gamma_2 \}) \]

\[ + \left( 1 + \int_{A}^{B} N_\theta^{k+1} (x) f_\theta^{k+1} (x-s) \, dx \right) P_\theta \{ \Gamma_2 \} \]

\[ = 1 + \int_{A}^{B} N_\theta^{k+1} (x) f_\theta^{k+1} (x-s) \, dx \]

This is Eq. (8).

Unlike Eqs. (5)–(6), even the numerical solutions of Eqs. (7)–(8) are difficult to obtain (if not impossible) in general. There are two major difficulties. First, although these two
equations are in inductive forms (w.r.t. time $k$), it is virtually impossible to implement the induction, be it forward or backward, since no initial value (this is actually exactly what we want to have) is available for the induction. Second, if $P^k_\theta(s)$ and $N^k_\theta(s)$ are viewed as functions with two variables $k$ and $s$, then the solutions of these two equations are not unique. It is clear that for the i.i.d. case, Eqs. (7)–(8) degenerate to Eqs. (5)–(6) since the start time of the test has no impact on OC and ASN. In the next section, we try to solve these two equations numerically for two special cases.

IV. NUMERICAL SOLUTIONS FOR SOME SPECIAL CASES

If it is further assumed that the sequence $\{s_k\}$ converges in distribution or $s_k$ have periodic distributions, additional equations can be imposed to Eqs. (7)–(8), resulting in unique numerical solutions.

A. System of Linear Algebraic Equations (SLAE) Method

The SLAE method was proposed [30] to solve Eqs. (5) and (6) numerically under the Gaussian assumption. It approximates the integral term in the Fredholm equation by Gaussian quadrature to form a system of linear equations. Numerical values of $P_\theta(s)$ and $N_\theta(s)$ on every quadrature point are calculated by solving this linear equation system. Although the SLAE method was developed under the Gaussian assumption, actually it is generally applicable provided the integral term in Eqs. (5)–(6) can be well approximated by Gaussian quadrature. Further, the fact that many density functions can be well approximated by Gaussian mixtures of only a few components also broadens its applicable situations.

The SLAE method can be generalized to solve Eqs. (7)–(8). Replacing the integral term with $n$-point Gaussian quadrature yields

$$\int_A^B P^{k+1}_\theta(x) f^{k+1}_\theta(x-s) \, dx \approx \sum_{i=1}^n \omega_i P^{k+1}_\theta(z_i) f^{k+1}_\theta(z_i - s)$$

where $\omega_i$ and $z_i$ are the weights and points of the Gaussian quadrature respectively. Applying to Eq. (7) this approximation yields

$$P^k_\theta(s) = F^{k+1}_\theta(A-s) + \sum_{i=1}^n \omega_i P^{k+1}_\theta(z_i) f^{k+1}_\theta(z_i - s)$$

where $F^k_\theta(\bullet)$ is the cumulative distribution function (CDF) of $s_k$. Let $s = z_1, \ldots, z_n$. Then a system of linear equations is obtained

$$P^k_\theta(z_j) = \sum_{i=1}^n \omega_i P^{k+1}_\theta(z_i) f^{k+1}_\theta(z_i - z_j) = f^{k+1}_\theta(A-z_j)$$

for $j = 1, 2, \ldots, n$ which has the equivalent matrix form

$$\begin{bmatrix} I & M_{k+1} \\ P_{k+1} \end{bmatrix} P_k = \Phi_{k+1}$$

where

$$M_{k+1} = \left( M^{(ij)}_{k+1} \right) \quad \text{and} \quad M^{(ij)}_{k+1} = -\omega_i f^{k+1}_\theta(z_i - z_j)$$

$$\Phi_{k+1} = \left[ P^{k+1}_\theta(A-z_1), \ldots, P^{k+1}_\theta(A-z_n) \right]^T$$

$$P_k = \left[ P^{k}_\theta(z_1), P^{k}_\theta(z_2), \ldots, P^{k}_\theta(z_n) \right]^T$$

Eq. (9) has $n$ equations but $2n$ unknowns, so it is under-determined and there are infinite many solutions. Combining Eq. (9) for different time $k$ yields

$$\begin{bmatrix} I & M_{k+1} \\ I & M_{k+2} \\ \vdots & \vdots \\ I & M_{k+m} \end{bmatrix} \begin{bmatrix} P_k \\ P_{k+1} \\ \vdots \\ P_{k+m} \end{bmatrix} = \begin{bmatrix} \Phi_{k+1} \\ \Phi_{k+2} \\ \vdots \\ \Phi_{k+m} \end{bmatrix}$$

(10)

The equation for ASN can be derived similarly

$$\begin{bmatrix} I & M_{k+1} \\ I & M_{k+2} \\ \vdots & \vdots \\ I & M_{k+m} \end{bmatrix} \begin{bmatrix} N_k \\ N_{k+1} \\ \vdots \\ N_{k+m} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

(11)

where

- $m$ is a positive integer;
- $N_k = \left[ N^{k}_\theta(z_1), N^{k}_\theta(z_2), \ldots, N^{k}_\theta(z_n) \right]^T$;
- $1 = n \times 1$ vector of all 1’s.

Eqs. (10)–(11) are still under-determined since there are only $nm$ equations but $n(m+1)$ unknowns. Additional $n$ equations are needed to have unique solutions. We consider two special cases.

B. With Sequence $\{s_k\}$ Converging in Distribution

First, we consider a special case that the random sequence $\{s_k\}$ converges to $\bar{s}$ in distribution [39] and the PDF of $s_k$ are continuous for all $k$; that is, for every point $s$ of $F^k_\theta(s)$,

$$\lim_{k \to \infty} F^k_\theta(s) = F(s)$$

where $F(s)$ is the CDF of $\bar{s}$. This case is not uncommon — e.g., if the Kalman filter is applied to a linear time-invariant system, the innovations are converging in distribution (provided the standard Kalman filter assumptions are satisfied and the filter itself is converging). In this case, the sequence $\{s_k\}$ asymptotically becomes i.i.d. and thus there exists a positive integer $M$ such that

$$P^k_\theta(s) \approx P^{k+1}_\theta(s) \approx \cdots \approx P^{\infty}_\theta(s) \quad \forall k > M$$

$$N^k_\theta(s) \approx N^{k+1}_\theta(s) \approx \cdots \approx N^{\infty}_\theta(s) \quad \forall k > M$$

Hence

$$P_k \approx P_{k+1} \approx \cdots \approx \hat{P}$$

$$N_k \approx N_{k+1} \approx \cdots \approx \hat{N}$$

(12)

where $\hat{P}$ and $\hat{N}$ are the solutions of Eqs. (5)–(6) respectively with $f_\theta(s) = f_\theta(\bar{s})$. This can be intuitively understood: If $M$ is large enough, for $k > M$ the distributions of $s_k$ become almost identical (i.e., approximately i.i.d.), and then the impact of the test start time on the statistical properties of the SPRT can be ignored, which validates Eq. (12). Therefore, any existing
techniques for the i.i.d. case can be applied to obtain \( \hat{P} \) and \( \hat{N} \). Then, we choose the \( m \) in Eqs. (10)–(11) such that

\[
k + m = M
\]

Plugging Eq. (12) into Eqs. (10) and (11) respectively yields the following two \( n(m + 1) \times n(m + 1) \) equation systems for OC

\[
\begin{bmatrix}
I & M_{k+1} & I & M_{k+2} \\
& & \ddots & \ddots \\
& & I & M_{k+m} \\
\end{bmatrix}
\begin{bmatrix}
P_k \\
P_{k+1} \\
\vdots \\
P_{k+m} \\
\end{bmatrix}
= 
\begin{bmatrix}
\Phi_{k+1} \\
\Phi_{k+2} \\
\vdots \\
\Phi_{k+m} \\
\end{bmatrix}
\]
(13)

and for ASN

\[
\begin{bmatrix}
I & M_{k+1} & I & M_{k+2} \\
& & \ddots & \ddots \\
& & I & M_{k+m} \\
\end{bmatrix}
\begin{bmatrix}
N_k \\
N_{k+1} \\
\vdots \\
N_{k+m} \\
\end{bmatrix}
= 
\begin{bmatrix}
1 \\
1 \\
\vdots \\
1 \\
\end{bmatrix}
\]
(14)

Note that the coefficient matrices in Eq. (13) and (14) are always invertible. By Gaussian elimination, the solutions are

\[
P_{k+i} = X_{k+i+1} + \sum_{l=i+1}^{m} (-1)^{l-i} \left( \prod_{j=i+1}^{l} M_{k+j} \right) X_{k+l+1}
\]

\[
N_{k+i} = Y_{k+i+1} + \sum_{l=i+1}^{m} (-1)^{l-i} \left( \prod_{j=i+1}^{l} M_{k+j} \right) Y_{k+l+1}
\]

for \( i = 0, 1, \ldots, m \), where

\[
\begin{bmatrix}
X_{k+1} \\
X_{k+2} \\
\vdots \\
X_{k+m} \\
X_{k+m+1} \\
\end{bmatrix}
= 
\begin{bmatrix}
\Phi_{k+1} \\
\Phi_{k+2} \\
\vdots \\
\Phi_{k+m} \\
\hat{P} \\
\end{bmatrix}
\]

\[
\begin{bmatrix}
Y_{k+1} \\
Y_{k+2} \\
\vdots \\
Y_{k+m} \\
Y_{k+m+1} \\
\end{bmatrix}
= 
\begin{bmatrix}
1 \\
1 \\
\vdots \\
1 \\
\hat{N} \\
\end{bmatrix}
\]

Here and in the sequel the summation is defined satisfying

\[
\sum_{i=a}^{b} (\bullet)_i \equiv 0, \quad \text{if} \quad a > b
\]

C. With Periodic PDF of \( s_k \)

If the distributions of \( s_k \) are periodic with period \( T \) (a positive integer), i.e.,

\[
f^B_\theta(s) = f^{k+T}_\theta(s)
\]

then

\[
P^B_\theta(s) = P^{k+T}_\theta(s)
\]

\[
N^B_\theta(s) = N^{k+T}_\theta(s)
\]

and hence

\[
P_k = P_{k+T}
\]

\[
N_k = N_{k+T}
\]
(15)

It is easily understood since the SPRT starting at time \( k \) or \( k + T \) are statistically equivalent, provided the initial values are the same. Inserting Eq. (15) into Eqs. (10) and (11) respectively yields the following two equation systems for OC

\[
\begin{bmatrix}
I & M_{k+1} & I & M_{k+2} \\
& & \ddots & \ddots \\
& & I & M_{k+T} \\
\end{bmatrix}
\begin{bmatrix}
P_k \\
P_{k+1} \\
\vdots \\
P_{k+T} \\
\end{bmatrix}
= 
\begin{bmatrix}
\Phi_{k+1} \\
\Phi_{k+2} \\
\vdots \\
\Phi_{k+T} \\
\end{bmatrix}
(16)
\]

and for ASN

\[
\begin{bmatrix}
I & M_{k+1} & M_{k+2} \\
& & \ddots & \ddots \\
& & I & M_{k+T} \\
\end{bmatrix}
\begin{bmatrix}
N_k \\
N_{k+1} \\
\vdots \\
N_{k+T} \\
\end{bmatrix}
= 
\begin{bmatrix}
1 \\
1 \\
\vdots \\
1 \\
\end{bmatrix}
(17)
\]

By Gaussian elimination, the solutions are

\[
P_{k+i} = U_{k+i+1} + \sum_{l=i+1}^{T} (-1)^{l-i} \left( \prod_{j=i+1}^{l} M_{k+j} \right) U_{k+l+1}
\]

\[
N_{k+i} = V_{k+i+1} + \sum_{l=i+1}^{T} (-1)^{l-i} \left( \prod_{j=i+1}^{l} M_{k+j} \right) V_{k+l+1}
\]

for \( i = 0, 1, \ldots, T \), where

\[
\begin{bmatrix}
U_{k+1} \\
U_{k+2} \\
\vdots \\
U_{k+T} \\
U_{k+T+1} \\
\end{bmatrix}
= 
\begin{bmatrix}
\Phi_{k+1} \\
\Phi_{k+2} \\
\vdots \\
\Phi_{k+T} \\
V_{k+T+1} \\
\end{bmatrix}
\]

\[
\begin{bmatrix}
V_{k+1} \\
V_{k+2} \\
\vdots \\
V_{k+T} \\
V_{k+T+1} \\
\end{bmatrix}
= 
\begin{bmatrix}
1 \\
1 \\
\vdots \\
1 \\
V \\
\end{bmatrix}
\]

and

\[
\hat{U} = Z^{-1} \left[ \Phi_{k+1} + \sum_{l=1}^{T-1} (-1)^{l} \left( \prod_{j=1}^{l} M_{k+j} \right) \Phi_{k+l+1} \right]
\]

\[
\hat{V} = Z^{-1} \left( I + \sum_{l=1}^{T-1} (-1)^{l} \prod_{j=1}^{l} M_{k+j} \right) 1
\]

\[
Z = \left( I + (-1)^{T} \prod_{j=1}^{T} M_{k+j} \right)
\]

provided \( Z \) is invertible. It is clear that when \( T = 1 \), Eqs. (16)–(17) degenerate to the i.i.d. case.

V. ILLUSTRATIVE EXAMPLES

Two numerical examples are provided in this section. Since the ground truth of the OC and ASN are unknown, we have to resort to the Monte Carlo (MC) method. In both of the following two cases, results of MC simulation were obtained from 10,000 runs.
A. Case 1: \( \{ s_k \} \) Converges in Distribution

If the log-likelihood ratio \( s_k \) has the Gaussian distribution

\[
f^{k}_\theta(s) = N(s; \mu_k, \sigma^2), \quad k = 1, 2, \ldots
given \theta = \frac{c}{\Gamma} - k
\]

\( \mu_k = \theta + ce^{-k} \)

where \( \theta \) is the ground truth of the underlying parameter, \( c \) a constant, and \( \sigma^2 \) the variance. It is easy to verify that \( \{ s_k \} \) converges in distribution:

\[
\lim_{k \to +\infty} \Phi_k(s) = \Phi(s)
\]

where \( \Phi(s) \) is the Gaussian CDF with mean \( \theta \) and variance \( \sigma^2 \). Two groups of parameters were simulated and the results were compared with the solutions of our algorithm (see the parameters in Table I and the simulation results in Fig. 1). \( A \) and \( B \) were computed by Wald’s approximation with both type I and type II error rates setting to 0.01. \( m \) is the parameter in Eqs. (13)-(14) and \( n \) is the number of Gaussian quadrature points.

It can be observed from Fig. 1 that \( P^k_\theta(s) \) is monotonically decreasing w.r.t. \( s \), as expected. This should happen because of the definition of OC. Since the distribution of \( s_k \) converges very fast (the term \( ce^{-k} \) diminishes exponentially), the differences of OC \( P^k_\theta(s) \) and ASN \( N^k_\theta(s) \) curves with different start time \( k \) are almost unobservable for \( k > 10 \). Further, as the mean of \( s_k \) is converging to \( \theta \) exponentially, for \( \theta > 0 \) the mean of the cumulative sum \( S_k \) is increasing, rendering \( S_k \) less likely to drop below the lower threshold \( A \). For \( \theta < 0 \), the mean of \( S_k \) will finally decrease, increasing the chance that \( S_k \) drops below \( A \). Hence, \( P^k_\theta(s) \) for group one is much smaller than for group two. For \( N^k_\theta(s) \), when the initial value \( s \) of the test is close to the bounds, it is more likely that the test statistic crosses the bounds in fewer steps. It makes sense
Figure 2. Results for Case 2, where the PDF of $s_k$ is periodic in $k$. The left and right columns correspond to parameters in the first and the second groups in Table II respectively. Again, the MC simulated results (blue solid line with “+”) agree very well with the SLAE solutions (red dashed line), which further verifies the solutions of our methods.

for the test to take more steps if the test starts around the middle of the two bounds.

Table I
PARAMETERS FOR CASE 1

<table>
<thead>
<tr>
<th>Group</th>
<th>$\theta$</th>
<th>$\sigma^2$</th>
<th>$c$</th>
<th>$A$</th>
<th>$B$</th>
<th>$m$</th>
<th>$n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>9</td>
<td>10</td>
<td>-4.6</td>
<td>4.6</td>
<td>10</td>
<td>20</td>
</tr>
<tr>
<td>2</td>
<td>-1</td>
<td>9</td>
<td>10</td>
<td>-4.6</td>
<td>4.6</td>
<td>10</td>
<td>20</td>
</tr>
</tbody>
</table>

B. Case 2: Periodic PDF

Again, $s_k$ are assumed of the Gaussian distribution of Eq. (18), but the mean $\mu_k$ is changing periodically:

$$\mu_k = \theta + \cos\left(\frac{2\pi k}{T}\right)$$

where the period $T$ is an integer. It is obvious that $f_k^\theta(s) = f_{k+T}^\theta(s)$. Two groups of parameters, given in Table II, were simulated by the MC method and the results were compared with our solutions. The results are plotted in Fig. 2. The differences between the MC results and our numerical solutions are tiny, as expected. Some patterns of the curves — e.g., $P_k^\theta(s)$ is monotonically decreasing and $N_k^\theta(s)$ has its peak value in the middle of $A$ and $B$ — are similar as in case 1 for the same reasons. But in this case, the curve for $k$ and $k+T$ are exactly overlapped, meaning that there are only $T$ different curves.

As shown in Fig. 1 and Fig. 2, results of MC simulation support our numerical solutions favorably. The average computational cost (measured in seconds) of our algorithms and MC simulation is given in Table III, which shows our algorithm is also computationally efficient.

Table II
PARAMETERS FOR CASE 2

<table>
<thead>
<tr>
<th>Group</th>
<th>$\theta$</th>
<th>$\sigma^2$</th>
<th>$T$</th>
<th>$A$</th>
<th>$B$</th>
<th>$m$</th>
<th>$n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.6</td>
<td>4</td>
<td>9</td>
<td>-4.6</td>
<td>4.6</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>-0.6</td>
<td>4</td>
<td>9</td>
<td>-4.6</td>
<td>4.6</td>
<td>20</td>
<td></td>
</tr>
</tbody>
</table>

VI. CONCLUSIONS

In this paper we have developed two inductive equations governing the OC and ASN functions respectively when the log-likelihood ratios are independent but non-stationary. They can be viewed as a generalization of the Fredholm integral equations for the i.i.d. case. Numerical algorithms for two special cases — the sequence \( \{s_k\} \) of log-likelihood ratios converges in distribution and \( s_k \) have periodic distributions — have been proposed based on the existing SLAE method developed for solving the Fredholm integral equations. The central idea of the SLAE method is approximating the integral term in the Fredholm equations by the Gaussian quadrature. Of course, other numerical approximations may also be applied here. Identifying other useful special cases that have unique solutions are under investigation. To our knowledge, no algorithm has been proposed to compute the OC and ASN in this problem setting, and consequently our method can not be compared or verified by existing methods other than MC simulation. Two numerical examples of our solutions have been provided, and close agreement with the results of MC simulation has been found. It is worth to note that since the ARL function of the CUSUM test satisfies a similar Fredholm equation, our methods can be easily transplanted to compute the ARL function with independent but non-stationary log-likelihood ratios.

REFERENCES


Table III

<table>
<thead>
<tr>
<th>SLAE/MC simulation</th>
<th>SLAE MC simulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case 1</td>
<td>0.03s</td>
</tr>
<tr>
<td>Case 2</td>
<td>0.05s</td>
</tr>
</tbody>
</table>