Online change-point detection for a transient change

Jack Noonan

We consider a popular online change-point problem of detecting a transient change in distributions of independent random variables. For this change-point problem, several change-point procedures are formulated and some advanced results for a particular procedure are surveyed. Some new approximations for the average run length to false alarm are offered and the power of these procedures for detecting a transient change in mean of a sequence of normal random variables is compared.

AMS 2000 subject classifications: Primary 60G50, 60G35; secondary 60G70, 94C12, 93E20.
Keywords and phrases: Change-point detection, Statistical quality control, Boundary crossing probabilities.

1. INTRODUCTION

The subject of change-point detection (or statistical quality control) is devoted to monitoring and detecting changes in the structure of a time series. This paper considers a popular online change-point problem of detecting a change in distribution of a sequence of independent random variables. Online change-point problems are concerned with monitoring the structure of a random process(es) whose observations arrive sequentially. For these problems, any good monitoring procedure should reliably alert the user to unexpected changes as soon as possible or with highest probability, subject to a tolerance on false alarms. In this paper, we will assume the distributions before and after a change-point have explicit probability density functions. This is a common assumption in the field, see for example [1, 2].

Let $y_1, y_2, \ldots$ be a sequence of independent random variables arriving sequentially. The purpose of this paper is to discuss tests for the hypothesis that y_i $(i = 1, 2 \ldots)$ are identically distributed with some probability density function (pdf) $f(y)$ against the alternative that at some unknown change point $0 \le \nu < \infty$, the random variables $y_1, y_2, \ldots, y_\nu$ and $y_{\nu+l+1}, y_{\nu+l+2}, \ldots$ are identically distributed with density $f(y)$ and $y_{\nu+1}, y_{\nu+2}, \ldots, y_{\nu+l}$ are identically distributed with pdf $g(y)$ such that $g(y) \neq f(y)$. Here, l is length of the change-point period (signal) and can be known or unknown. Under a standard hypothesis testing framework, the null hypothesis is $\mathbb{H}_\infty\colon\ \nu = \infty$ and hence the pdf $f(y)$ is the density of y_i for all $i = 1, 2, \ldots$. The alternative hypothesis is $\mathbb{H}_\nu\colon 0 \le \nu < l \le \infty$ and therefore

$$\mathbb{H}_\nu\colon \begin{cases} y_i \text{ have density } f(y) & \text{if } i \le \nu \text{ or } i > \nu + l \\ y_i \text{ have density } g(y) & \text{if } \nu < i \le \nu + l \end{cases}$$

with $i = 1, 2, \ldots$. Under $\mathbb{H}_\nu$, the arrival time of the signal is $\nu + 1$ (it is unknown). Most classical results assume f and g are known completely; by this, we mean no nuisance parameters are present in the distributions. In later sections, we will briefly discuss tests designed to approach the change-point problem in the presence of nuisance parameters.

A thorough introduction to the field of online (quickest) change-point detection mainly for the case of $l = \infty$ can be found in, for example, [3, 1, 4, 5]. Some of the most popular online change-point algorithms used in practice are Shewhart's $\bar{X}$-chart [6], the CUSUM algorithm [7], the Shiryaev-Roberts procedure [8, 9] and the Exponentially Weighted Moving Average (EWMA) chart [10]. The case of $l = \infty$, and hence when a change in distribution occurs it does so permanently, is by far the most popular scenario considered in the change-point literature; a number of influential papers are [11, 12, 13, 14, 15, 16]. The CUSUM and Shiryaev Roberts procedures benefit with their simplicity and proven optimality under suitable optimality criteria; these two procedures will be the focus of discussion for the case $l = \infty$. The case of finite l, and hence when a change occurs it does so temporarily, has seen considerable attention in the past, see [17, 18, 19, 20, 21, 22, 23, 24]. More recently it has been the focus of attention in the papers of [25, 26, 27]. Examples of areas where detecting a transient change in distributions is extremely important can be found in radar and sonar [28, 29, 30], nondestructive testing [31], and medicine [32]. Non-parametric online change-point detection methods have also become very popular [33, 34]. For the state of the art techniques for multiple change-point detection, see [35, 36, 37, 38]. For sequential change-point detection in high-dimensional time series, a likelihood ratio approach can be found in [39, 40].

This survey is organised as follows. In Section 2, we survey results for $l = \infty$ and discuss known optimality results for the CUSUM and Shiryaev-Roberts procedures. This section contains well known classical results but is included to introduce the reader to change-point concepts that will be used when considering the transient change-point problem. In Section 3, we assume $l < \infty$ and discuss a number of online tests for transient changes; the likelihood ratio test providing the inspiration behind all tests. In this section, we compare procedures when applied for detecting a temporary change in mean of a sequence of Gaussian random variables. We also apply tests for monitoring stability of components used in the Oil and Gas industry.

Throughout this survey we shall use the notation $\Pr_\infty$ and $\mathbb{E}_\infty$ to denote probability and expectation under $\mathbb{H}_\infty$. Under the alternative $\mathbb{H}_\nu$, we shall use the notation $\Pr_\nu$ and $\mathbb{E}_\nu$ to denote probability and expectation assuming the change-point occurs at $\nu < \infty$.

2. PERMANENT CHANGE IN DISTRIBUTIONS

In this section, we assume $l = \infty$; if a change occurs, it does so permanently. Suppose $y_1, y_2, \ldots, y_n$ have been sampled. The likelihood ratio for testing $\mathbb{H}_\infty$ against $\mathbb{H}_\nu$ is

$$\Lambda_{\nu,n} = \prod_{i=\nu+1}^{n} \frac{g(y_i)}{f(y_i)}$$

assuming $\nu < n$, otherwise $\Lambda_{\nu,n} = 1$.

2.1 The CUSUM and Shiryaev-Roberts procedures

By maximising the statistic $\Lambda_{\nu,n}$ over all possible locations of ν, we obtain the cumulative sum (CUSUM) statistic

$$V_n := \max_{0 \le \nu \le n-1} \Lambda_{\nu,n}, \quad n \ge 1\,. \tag{2.1}$$

The CUSUM stopping rule (when to alert the user to a potential change-point) is

$$\tau_V(H) := \inf\{n \ge 1 : V_n > H\}\,. \tag{2.2}$$

An appealing property of statistic (2.1) is the recursive property

$$V_n = \max\{V_{n-1}, 1\} \cdot \frac{g(y_n)}{f(y_n)}, \quad V_0 = 1\,.$$

The threshold H in $\tau_V(H)$ is chosen on the users tolerance to false alarm risk. Page [7] and Lorden [11] measured false alarm risk through the Average Run Length to false alarm (ARL). The ARL criterion corresponds to choosing H such that $\mathbb{E}_\infty \tau_V(H) = C$, where C is a pre-defined value chosen by the user but is typically large. How to compute $\mathbb{E}_\infty \tau_V(H)$ will be discussed later in this section.

The famous CUSUM chart of Page [7] introduces a reflective barrier at zero. This procedures is defined as:

$$\begin{aligned} P_n &:= \max\left\{P_{n-1} + \log\frac{g(y_n)}{f(y_n)}, 0\right\}, \\ P_0 &= 0\,. \end{aligned} \tag{2.3}$$

The statistics (2.3) and $\log V_n$ are equivalent on the positive half plane and hence the rule

$$\tau_P(\log(H)) = \inf\{n \ge 1 : P_n > \log H\}\,,$$

and $\tau_V(H)$ are equivalent for $H > 1$. The stopping rule τ_V is more general than τ_P as thresholds $H \le 1$ are permissable. An approximation for $\mathbb{E}_\infty \tau_P(H)$ for general distributions f and g was derived in [41]. Let $I_f := -\mathbb{E}_\infty(\log(g(y_1)/f(y_1)))$ and $I_g = \mathbb{E}_0(\log(g(y_1)/f(y_1)))$ (to compute I_g we assume the change-point occurs at time zero). Then

$$\mathbb{E}_\infty \tau_P(H) \simeq \frac{e^H}{I_g \zeta^2} - \frac{H}{I_f} - \frac{1}{I_g \zeta}\,. \tag{2.4}$$

Here the constant ζ is called the limiting exponential overshoot. Let $Z_n = \sum_{i=1}^n \log(g(y_i)/f(y_i))$ be a random walk process. For a non-negative barrier a, define the stopping rule $\tau_a := \inf\{n \ge 1 : Z_n > a\}$ and define the excess over the barrier by $\kappa_a := Z_{\tau_a} - a$. Then $\zeta := \lim_{a\to\infty} \mathbb{E}_0[e^{-\kappa_a}]$. It was shown in [42, Ch. VIII] that

$$\zeta = \frac{1}{I_g} \exp\left\{-\sum_{k=1}^{\infty} \frac{1}{k}[\Pr_\infty(Z_k > 0) + \Pr_0(Z_k \le 0)]\right\}.$$

The approximation (2.4) seems extremely accurate. For example, suppose pre-change observations are i.i.d. $N(0,1)$ random variables and post-change observations are i.i.d. $N(A,1)$ for some known $A > 0$. We have

$$\begin{aligned} f(y) &= \frac{1}{\sqrt{2\pi}} \exp(-y^2/2), \\ g(y) &= \frac{1}{\sqrt{2\pi}} \exp(-(y-A)^2/2)\,. \end{aligned} \tag{2.5}$$

For $A = 1$, Monte Carlo simulations with $100{,}000$ iterations provide $\mathbb{E}_\infty \tau_P(4.39) = 500$. Application of the approximation in (2.4) provides 498. The draw back of the approximation in (2.4) is that ζ requires expensive numerical evaluation.

To construct the Shiryaev-Roberts (SR) procedure, define the generalised Bayesian detection statistic as:

$$R_n := \sum_{\nu=0}^{n} \Lambda_{\nu,n}\,. \tag{2.6}$$

Then the SR test is:

$$\tau_R(H) := \inf\{n \ge 1 : R_n > H\}\,,$$

where H is the solution of $\mathbb{E}_\infty \tau_R(H) = C$ for some predetermined C. The SR statistic (2.6) satisfies the following recurrence:

$$R_n = (1 + R_{n-1}) \cdot \frac{g(y_n)}{f(y_n)}, \quad n \ge 1,\ R_0 = 0\,.$$

2.2 Evaluating ARL for CUSUM and SR tests

Explicit expressions for $\mathbb{E}_\infty\tau_V(H)$ and $\mathbb{E}_\infty\tau_R(H)$ are not known. However, they can be numerically obtained by numerically solving particular Fredholm integral equations as proved in [43]. Here it was shown that $\mathbb{E}_\infty\tau_V(H)$ and $\mathbb{E}_\infty\tau_R(H)$ can be computed by a unified approach for general Markov statistics. Set $H>0$. For a sufficiently smooth positive valued function ξ and $s\in[0,H]$, let

$$S_n = \xi(S_{n-1})\cdot\frac{g(y_n)}{f(y_n)}, \quad n\geq 1,\ S_0 = s\in[0,H],$$

be a Markov detection statistic with stopping rule

$$\tau_S(H) := \inf\{n\geq 1 : S_n > H\}.$$

Let $\phi(s)=\mathbb{E}_\infty(\tau_S(H))$ be the ARL (note the dependence on $S_0=s$) and set $F(x)=\Pr_\infty(g(y_1)/f(y_1)\leq x)$. Then $\phi(s)$ is the solution of the following Fredholm integral equation:

$$\phi(s) = 1+\int_0^H \phi(x)\left[\frac{d}{dx}F\left(\frac{x}{\xi(s)}\right)\right]dx\,. \tag{2.7}$$

For the CUSUM and SR procedures we have $\xi(s)=\max(1,s)$ and $\xi(s)=1+s$, respectively. To solve this integral equation, see [43].

Approximations for ARL of the CUSUM and SR procedures have been specifically developed for the problem of detecting the change in mean of normal random variables. Here we operate under (2.5). To approximate ARL for both the CUSUM and SR procedures or to narrow the domain of search and more efficiently numerically solve the Fredholm equation (2.7), one could use the following simple approximations developed in [44] and [45] respectively:

$$\mathbb{E}_\infty\tau_V(H) \simeq 2H/(A\kappa^2(A)), \tag{2.8}$$

$$\mathbb{E}_\infty\tau_R(H) \simeq H/\kappa(A), \tag{2.9}$$

where

$$\kappa(A) = \frac{2}{A^2}\exp\left\{-2\sum_{\nu=1}^{\infty}\frac{1}{\nu}\Phi\left(-\frac{A}{2}\sqrt{\nu}\right)\right\}$$

$$\text{and } \Phi(x)=\int_{-\infty}^{x} f(y)dy\,.$$

The approximations in (2.8) and (2.9) are very accurate, particularly for large H. In Table 1, one can observe the high accuracy of approximation (2.8) for different thresholds H. In fact, (2.9) is remarkably accurate and frequently leads to exact values of ARL. The only slight inconvenience of both approximations is the numerical evaluation required to compute $\kappa(A)$. This quantity is frequently approximated, see [42, Ch. IV], with $\kappa(A)\simeq\exp(-\rho\cdot A)$, where the constant ρ is defined later in (3.17) but can be approximated to three decimal places by $\rho\simeq 0.583$. Using this approximation for κ in (2.8) and (2.9) still results in excellent approximations and can be recommended. In this table, $\mathbb{E}_\infty\tau_V(H)$ has been approximated Monte Carlo simulations with $100,000$ repetitions.

2.3 Optimality criteria

Denote by $\Delta(C)$ the set of all stopping times of change-point procedures with ARL of at least C. More precisely, $\Delta(C):=\{\tau:\mathbb{E}_\infty\tau\geq C\},\ \ C>1$, where $\tau=\tau(H)$ is a stopping time for a sequential change-point procedure. A common criterion for comparing change-point procedures when $l=\infty$ is the supremum Average Delay to Detection (ADD) introduced by Pollak [13]. Define $ADD_\nu(\tau):=\mathbb{E}_\nu(\tau-\nu|\tau>\nu)$. Then

$$SADD(\tau) := \sup_{0\leq\nu<\infty} ADD_\nu(\tau)\,. \tag{2.10}$$

An optimal change-point procedure would satisfy $SADD(\tau_{opt}) = \inf_{\tau\in\Delta(C)} SADD(\tau)$ for all $C>1$. Finding an optimal procedure for this criterion is very difficult, where in general only asymptotic optimality as $C\to\infty$ (low false alarm rate) is known [13]. Another popular criterion is the worst-case minimax scenario of Lorden [11] defined as

$$\mathcal{L}(\tau) := \sup_{\nu\geq 0}\operatorname{ess\,sup}\mathbb{E}_\nu[(\tau-\nu)^+|y_1,y_2,\ldots,y_\nu]\,. \tag{2.11}$$

In other words, the conditional ADD is first maximized over all possible trajectories of observations up to the change-point and then over the change-point. We refer the reader to Section 6.3.3 of [1] for further discussions regarding this criterion. Asymptotic optimality (as $C\to\infty$) of the CUSUM chart of Page was proved in [11]. It was subsequently proved in [12] that the CUSUM chart of Page is in fact optimal under this criterion for every $C>1$. We refer the reader to Section 6.3.3 of [1] for further discussions regarding this criterion.

The SR procedure is optimal for every $C>1$ under the Stationary Average Delay to Detection (STADD) criterion. The STADD criterion rewards detection procedures that detect the change as quickly as possible, at the expense of raising many false alarms (using a repeated application of the same stopping rule). Formally, the STADD criterion is defined as follows. Let $\tau_1,\tau_2\ldots$ be a sequence of independent copies of the stopping time τ. Let $T_j=\tau_1+\tau_2+\ldots+\tau_j$ be the time the j^{th} alarm is raised. Let $I_\nu=\min\{j>1:T_j>\nu\}$; this is the index of the first alarm which is not false after $I_\nu-1$ false alarms. Then

$$STADD(\tau) := \lim_{\nu\to\infty}\mathbb{E}_\nu[T_{I_\nu}-\nu]\,.$$

It was proved in [46] that the STADD criterion is equivalent to the Relative Integral Average Detection Delay (RI-

H	9.32	17.33	80.65	159.35	788.00
$\mathbb{E}_\infty \tau_V(H)$	50	100	500	1000	5000
Approximation (2.8)	59	110	513	1014	5018
(2.8) with $\kappa(A) \simeq \exp(-\rho \cdot A)$	60	111	517	1023	5058

Table 1. Approximations for $\mathbb{E}_\infty \tau_V(H)$ with $A = 1$.

ADD) measure (see [43]):

$$RIADD(\tau) := \frac{\sum_{\nu=0}^{\infty} \mathbb{E}_\nu[(\tau-\nu)^+]}{\mathbb{E}_\infty[\tau]}\,.$$

It is discussed in [43] for both CUSUM and the Shiryaev–Roberts procedure Lorden's essential supremum measure (2.11) and Pollak's supremum measure $SADD$ defined in (2.10) are attained at $\nu = 0$, that is:

$$\mathcal{L}(\tau_V(H)) = SADD(\tau_V(H)) = \mathbb{E}_0 \tau_V(H)\,,$$
$$\mathcal{L}(\tau_R(H)) = SADD(\tau_R(H)) = \mathbb{E}_0 \tau_R(H)\,.$$

Similarly to the computation of $\mathbb{E}_\infty \tau_V(H)$ and $\mathbb{E}_\infty \tau_R(H)$, to obtain $\mathbb{E}_0 \tau_V(H)$ and $\mathbb{E}_0 \tau_R(H)$ one can numerically solve a Fredholm equation. Instead of setting $\phi(s) = \mathbb{E}_\infty(\tau(H))$, let $\phi(s) = \mathbb{E}_0(\tau(H))$. Also set $F(x) = \Pr_0(g(y_1)/f(y_1) \leq x)$. Then from [43], $\phi(s)$ is the solution of the Fredholm integral equation given in (2.7). The computation of STADD requires solving a slightly more difficult integral equation, see [43] for more discussions.

For the Gaussian example considered in (2.5), the findings of [43] indicate that for small values of A, say $A = 0.01$, the CUSUM noticeably outperforms the SR procedure under Lordens criterion. Vice versa, the SR procedure noticeably outperforms CUSUM under the STADD framework. When the change in A becomes large, say $A = 1$, the benefits a procedure has over the other diminishes.

3. TRANSIENT CHANGE IN DISTRIBUTIONS

In this section, we assume $1 \leq l < \infty$ and therefore study procedures aimed at detecting a transient change in distributions. Suppose $y_1, y_2, \ldots, y_n$ have been sampled. The log likelihood ratio for testing $\mathbb{H}_\infty$ against $\mathbb{H}_\nu$ is

$$\Gamma_{\nu,\nu+l} := \log \Lambda_{\nu,\nu+l} = \sum_{i=\nu+1}^{\nu+l} \log \frac{g(y_i)}{f(y_i)}\,. \tag{3.1}$$

3.1 A collection of procedures

For l unknown, the log likelihood ratio statistic is obtained by maximising (3.1) over all possible change point locations ν and transient change lengths:

$$K_n := \max_{0 \leq \nu < \nu+l \leq n} \Gamma_{\nu,\nu+l}\,, \tag{3.2}$$

with the stopping rule

$$\tau_K(H) := \inf\{n \geq 1 : K_n > H\}\,.$$

Note that in (3.2), we are maximising over l too. If there are no nuisance parameters present in f and g that require estimation, the statistic (3.2) satisfies the recursive property:

$$K_n = \max\{K_{n-1}, \max_{0 \leq \nu \leq n-1} \Gamma_{\nu,n}\}, \quad K_0 = 0\,. \tag{3.3}$$

For large n, the statistic (3.2) is very expensive to compute despite the recursive property given in (3.3). This is because in $\max_{0 \leq \nu \leq n-1} \Gamma_{\nu,n}$, one has to maximise over all possible change-point locations which is expensive for large n. For offline change-point problems, this large computational expense may be an inconvenience but it is not a fundamental problem as time is often not an issue. However, for online procedures that require calculations in real time, the statistic K_n is not practical. The assumption of no prior knowledge about the transient change length is unlikely. One can imagine that some knowledge about the length of transient change is likely, for example it may be bounded $l_0 \leq l \leq l_1$. From here on, this assumption will be made. The log likelihood ratio statistic is:

$$Z_n = Z_n(l_0, l_1) := \max_{\substack{0 \leq \nu < \nu+l \leq n \\ l_0 \leq l \leq l_1}} \Gamma_{\nu,\nu+l}\,, \tag{3.4}$$

with the stopping rule

$$\tau_Z(H) := \inf\{n \geq l_1 : Z_n(l_0, l_1) > H\}\,. \tag{3.5}$$

If no nuisance parameters require estimation, the statistic Z_n satisfies the following recursive property:

$$Z_n = \max\left\{Z_{n-1}, \max_{n-l_1 \leq \nu \leq n-l_0} \Gamma_{\nu,n}\right\}, \qquad Z_{l_1} = \max_{\substack{0 \leq \nu < \nu+l \leq l_1 \\ l_0 \leq l \leq l_1}} \Gamma_{\nu,\nu+l}\,. \tag{3.6}$$

This is much easier to compute than (3.2) for n large. If we make the additional assumption that l is known exactly and is completely contained within the sample of size n, i.e. $\nu + l \leq n$, then the MOSUM statistic is obtained by maximising (3.1) over all valid change-point locations ν:

$$M_n := \max_{0 \leq \nu \leq n-l} \Gamma_{\nu,\nu+l}\,. \tag{3.7}$$

The MOSUM statistic can be obtained by setting $l_0 = l_1 = l$ in (3.4). For this reason, the statistic Z_n can be called the generalised MOSUM procedure.

The stopping rule associated with MOSUM procedure is:

$$\tau_M(H) := \inf\{n \geq l : M_n > H\}\,.$$

In what follows, we define the MOSUM test for a general window length L, with L a fixed positive integer. The reason for doing so is we will be interested in studying quantities like the loss of power, when incorrect information is provided for the true l. Results for the likelihood ratio test can still be obtained by setting $L = l$. Define the moving sums

$$S_{n,L} := S_{n,L,L} = \sum_{j=n+1}^{n+L} \log \frac{g(y_i)}{f(y_i)} \quad (n = 0, 1, \ldots)\,.$$

Then the stopping rule $\tau_M(H)$ for a given window length L can be expressed as

$$\begin{aligned} \tau_M(H) \;&=\; \tau_{S,L}(H) + L, \;\text{ where} \\ \tau_{S,L}(H) \;&:=\; \inf\{n \geq 0 : S_{n,L} > H\} \end{aligned} \tag{3.8}$$

and therefore $\mathbb{E}_\infty \tau_M(H) = \mathbb{E}_\infty \tau_{S,L}(H) + L$. The moving sum $S_{n,L}$ provided the motivation for the MOSUM name.

For the transient change-point problem, the false alarm risk can be measured through ARL. However, this is not the only approach taken in the change-point literature. In [18] and [47] , the false alarm risk is measured through:

$$\sup_{k \geq 1} \Pr_\infty(k \leq \tau < k + m_\alpha) \leq \alpha\,,$$

where τ is a stopping rule, α is your false alarm tolerance (type 1 error) and $\liminf m_\alpha / |\log(\alpha)| > I_g^{-1}$ but $\log m_\alpha = o(|\log \alpha|)$ as $\alpha \to 0$; recall $I_g = \mathbb{E}_0(\log(g(y_1)/f(y_1)))$. Another alternative to the usual ARL constraint has been proposed in [44, 48]. Here, the suggested criterion is

$$\sup_{k \geq 1} \Pr_\infty(\tau < k + m_\alpha | \tau \geq l) \leq \alpha\,.$$

From here on in, we measure false alarm risk through ARL $\mathbb{E}_\infty \tau$ and simply refer to [18, 47, 44, 48] for more discussions on other approaches. The majority of research has focused on detecting transient changes in the mean of a sequence of Gaussian random variables. The next section is devoted solely to this problem.

3.2 Detecting a transient change in Gaussian random variables

Consider the problem of detecting the change in mean of normal random variables. Suppose pre-change observations are i.i.d. $N(\mu, 1)$ random variables and post-change observations are i.i.d. $N(\mu + A, 1)$ for some $A > 0$. The values of μ, l and A may be known or unknown, with μ and A playing the roles of nuisance parameters if unknown. We have

$$\begin{aligned} f(y) \;&=\; \frac{1}{\sqrt{2\pi}} \exp(-(y-\mu)^2/2)\,, \\ g(y) \;&=\; \frac{1}{\sqrt{2\pi}} \exp(-(y-\mu-A)^2/2)\,. \end{aligned} \tag{3.9}$$

The offline version of this change-point problem is devoted to testing for change-points in a sample of fixed length and has seen significant attention in the past, see [49, 50, 51, 52, 53]. An excellent survey of several statistics aimed at addressing the offline problem can be found in [2]. Despite the fact Z_n defined in (3.4) is a generalisation of M_n given in (3.7), we will initially discuss recent results for M_n. These results will provide inspiration for addressing the much more complicated problems associated with Z_n.

3.2.1 The MOSUM statistic

The MOSUM stopping rule given in (3.8) specialised for this Gaussian example is

$$\begin{aligned} \tau'_M(H') &= \tau'_{S,L}(H') + L\,, \\ \tau'_{S,L}(H') &= \inf\{n \geq 0 : S'_{n,L} > H'\}\,, \\ S'_{n,L} &= A \sum_{j=n+1}^{n+L} (y_j - \mu - A/2)\;. \end{aligned}$$

Knowledge of A is required to set the ARL constraint $\mathbb{E}_\infty \tau'_M(H') = C$. However, if we consider the stopping rule:

$$\begin{aligned} \tau_M(H) &= \tau_{S,L}(H) + L\,, \\ \tau_{S,L}(H) &= \inf\{n \geq 0 : S_{n,L} > H\}\,, \\ S_{n,L} &= \sum_{j=n+1}^{n+L} y_j\;, \end{aligned} \tag{3.10}$$

then one can show that $\tau'_M(H') \overset{d}{=} \tau_M(H'/A + \mu L + AL/2)$. As a result, the stopping rules $\tau'_M(H')$ and $\tau_M(H)$ are equal in distribution provided that $\mathbb{E}_\infty \tau'_M(H') = \mathbb{E}_\infty \tau_M(H) = C$. The stopping rule $\tau_M(H)$ has the benefit of not requiring knowledge of A to set the ARL constraint. We will refer to the stopping rule $\tau_M(H)$ as the MOSUM test in this Gaussian setting.

The problem of approximating $\mathbb{E}_\infty \tau_{S,L}(H)$ assuming μ is known was considered in [26]. Here will recall the main steps in the construction. Define

$$h = \frac{H - \mu L}{\sqrt{L}} \;\text{ so that }\; H = \mu L + h\sqrt{L}$$

and consider the standardised versions of $S_{n,L}$:

$$\xi_{n,L} := \frac{S_{n,L} - \mathbb{E}_\infty S_{n,L}}{\sqrt{\mathrm{Var}_\infty(S_{n,L})}} = \frac{S_{n,L} - \mu L}{\sqrt{L}}, n = 0, 1, \dots .$$

Then the stopping time $\tau_{S,L}(H)$ is equivalent to the stopping time

$$\tau_\xi(h) := \inf\{n \geq 0 : \xi_{n,L} \geq h\} \tag{3.11}$$

and hence $\mathbb{E}_\infty \tau_\xi(h) = \mathbb{E}_\infty \tau_{S,L}(H)$.

For any integer $M \geq 0$, the discrete time process $\xi_{0,L}, \xi_{1,L}, \dots, \xi_{M,L}$ is approximated by a continuous time analogue $S(t)$ on $[0, T = M/L]$. The process $S(t)$ is a zero mean, stationary Gaussian process with correlation function $R(t) = \max\{0, 1 - |t|\}$. The ARL $\mathbb{E}_\infty \tau_\xi(h)$ then has the continuous-time approximation

$$\mathbb{E}_\infty \tau_\xi(h) \cong -L \int_0^\infty s \, dF_h(s) , \tag{3.12}$$

where $F_h(T) := \mathrm{P}_\infty(S(t) < h \text{ for all } t \in [0, T])$.

Explicit formulas for the probability $F_h(T)$ with $T \leq 1$ were first derived in [54]. Here it was shown

$$F_h(T) = \int_{-\infty}^h \Phi\left(\frac{h(Z+1) - x(-Z+1)}{2\sqrt{Z}}\right) \varphi(x) dx$$
$$- \frac{2\sqrt{Z}}{Z+1} \varphi(h) \left[h\sqrt{Z}\, \Phi(h\sqrt{Z}) + \frac{1}{\sqrt{2\pi}} (\sqrt{2\pi}\varphi(h))^Z \right] .$$

For $T = 1$ this reduces to

$$F_h(1) = \Phi^2(h) - \varphi(h)\left[h\Phi(h) + \varphi(h)\right] . \tag{3.13}$$

For $T > 1$, formulae for $F_h(T)$ were first derived in [55]; these expressions take different forms depending on whether or not T is integer. The result of [55, p.949] states that if $T = n$ is a positive integer then

$$F_h(n) = \int_{-\infty}^h \int_{D_x} \det[\varphi(y_i - y_{j+1} + h)]_{i,j=0}^n \, dy_2 \dots dy_{n+1} dx , \tag{3.14}$$

where $y_0 = 0, y_1 = h - x$, $D_x = \{y_2, \dots, y_{n+1} \mid h - x < y_2 < y_3 < \dots < y_{n+1}\}$. For non-integer $T \geq 1$, the exact formula for $F_h(T)$ is even more complex (the integral has the dimension $\lceil 2T \rceil + 1$) see [55, p.950]. For $T = 2$, (3.14) yields

$$\begin{aligned} F_h(2) = {} & \Phi^3(h) - 2h\varphi(h)\Phi^2(h) \\ & + \frac{h^2 - 3 + \sqrt{\pi}h}{2} \varphi^2(h)\Phi(h) + \frac{h + \sqrt{\pi}}{2} \varphi^3(h) \\ & + \int_0^\infty \Phi(h - y)[\varphi(h + y)\Phi(h - y) \\ & - \sqrt{\pi}\varphi^2(h)\Phi(\sqrt{2}y)\,]dy . \end{aligned} \tag{3.15}$$

The complicated nature of these expressions for $F_h(T)$ made them impractical for the use in the ARL approximation (3.12). One simple yet still very accurate approximation has the form (see [56]):

$$F_h(T) \simeq F_h(2) \left[\theta(h)\right]^{T-2} , \tag{3.16}$$

where $\theta(h) = F_h(2)/F_h(1)$ and the probabilities $F_h(1)$ and $F_h(2)$ are given in (3.13) and (3.15) respectively. Here, $\varphi(x)$ and $\Phi(x)$ are the standard normal density and distribution functions respectively. The approximation given in (3.16) applied to (3.12) results in the following continuous-time ARL approximation:

$$\mathbb{E}_\infty \tau_\xi(h) \quad \simeq \quad -\frac{L \cdot F_h(2)}{\theta(h)^2 \log(\theta(h))} .$$

This approximation was then corrected in [26, Section 7] for discrete time to improve results for small L. This amounted to correcting the probabilities $F_h(1)$ and $F_h(2)$ for discrete time; this was performed by specialising results of D. Siegmund; primarily on expected overshoot a discrete time normal random walk has the first time it crosses a threshold. From [42, p. 225], this expected overshoot was computed as

$$\rho := -\int_0^\infty \frac{1}{\pi\lambda^2} \log\{2(1 - \exp(-\lambda^2/2))/\lambda^2\} \, d\lambda \simeq 0.582597 . \tag{3.17}$$

Define the probability

$$F_h(M; L) := \Pr\left(\max_{n=0,1,\dots,M} \xi_{n,L} < h\right) . \tag{3.18}$$

From [26, p. 18]:

$$\mathbb{E}_\infty \tau_{S,L}(H) = \mathbb{E}_\infty \tau_\xi(h) \simeq -\frac{L \cdot F_h(2L; L)}{\theta_L(h)^2 \log(\theta_L(h))} \quad \text{with } \theta_L(h) = \frac{F_h(2L; L)}{F_h(L; L)} , \tag{3.19}$$

where, for $h_L := h + \omega_L$ with $\omega_L = \sqrt{2}\rho/\sqrt{L}$, the probabilities $F_h(L; L)$ and $F_h(2L; L)$ can be approximated by:

$$F_h(L; L) \simeq \Phi(h)\Phi(h_L) - \varphi(h_L)[h\Phi(h) + \varphi(h)] , \tag{3.20}$$

$$\begin{aligned} F_h(2L; L) \simeq {} & \frac{\varphi^2(h_L)}{2}[(h^2 - 1 + \sqrt{\pi}h)\Phi\,(h) \\ & + \quad (h + \sqrt{\pi})\varphi\,(h)] \\ & - \quad \varphi\,(h_L)\,\Phi\,(h_L)\,[(h + h_L)\,\Phi\,(h) + \varphi\,(h)] \\ & + \quad \Phi\,(h)\,\Phi^2(h_L) \end{aligned}$$

Table 2. Approximations for $\mathbb{E}_\infty \tau_\xi(h)$ with $L = 10$.

h	2	2.25	2.5	2.75	3	3.25	3.5
(3.19)	126	217	395	759	1551	3375	7837
$\mathbb{E}_\infty \tau_\xi(h)$	127	218	396	757	1550	3344	7721

$$+ \int_0^\infty \Phi(h-y)[\varphi(h_L+y)\Phi(h_L-y) - \sqrt{\pi}\varphi^2(h_L)\Phi(\sqrt{2}y)]dy\,. \tag{3.21}$$

Only a one-dimensional integral has to be numerically evaluated for approximating $F_h(2L;L)$. Tables 2 and Tables 3 demonstrate that (3.19) using (3.20) and (3.21) is extremely accurate.

For approximating the boundary-crossing probability $F_h(M;L)$ for all M, the discrete time corrected form of (3.16) suggests using the approximation

$$F_h(M;L) \simeq F_h(2L;L)\,[\theta_L(h)]^{M/L-2}\,. \tag{3.22}$$

One could then approximate $F_h(2L;L)$ and $\theta_L(h)$ using (3.20) and (3.21); the high accuracy of the resulting approximation was comprehensively studied in [26].

3.2.2 MOSUM and scan statistics

We remark that the moving sum process $S_{n,L}$ forms the basis of a number of change-point detection algorithms and the stopping rule $\tau_M(H)$ is used even when the r.v.'s y_j are not necessarily Gaussian, see for example [57, 58, 59, 60]; the moving sum $S_{n,L}$ is sometimes called a 'scan statistic'. In [57, 58, 60] and [p.45, p.187][59] the authors consider the application of scan statistics for either monitoring a Poisson process or monitoring a sequence of i.i.d. Bernoulli trials (among others). In [57, 58, 59], approximations of a similar form to (3.22) are made, where $F_h(2L;L)$ and $\theta_L(h)$ can be computed exactly in certain non-Gaussian settings. The robustness of approximation (3.19) and formulas (3.20) and (3.21) to non-normality was studied in Sections 6.2 and 6.3 of [26]. Here it was demonstrated with non-Gaussian y_j and potentially non-uniform weights associated with each y_j, that these formulas remain very accurate for suitably large L.

3.2.3 The stopping rule $\tau_Z(H)$

Here, we assume l is not known exactly but can be bounded between l_0 and l_1. We will initially assume μ and A are known. The stopping rule given in (3.5) specialised for this Gaussian example is tantamount to:

$$\tau_Z(H) = \inf\left\{ n \geq l_1 : \max_{\substack{0\leq\nu<\nu+l\leq n\\ l_0\leq l\leq l_1}} Z_n(l_0,l_1) \right\},$$

$$\text{with } Z_n = Z_n(l_0,l_1) = A\sum_{j=\nu+1}^{\nu+l}\left(y_j-\mu-\frac{A}{2}\right). \tag{3.23}$$

Using the recursive property outlined in (3.6), for $n > l_1$ the statistic Z_n satisfies:

$$Z_n = \max\{Z_{n-1}, S_{n,l_0,l_1}\}\,, \text{ with}$$
$$S_{n,l_0,l_1} := \max_{n-l_1\leq\nu\leq n-l_0} A\sum_{j=\nu+1}^{n}\left(y_j-\mu-\frac{A}{2}\right),$$
$$Z_{l_1} = \max_{\substack{0\leq\nu<\nu+l\leq l_1\\ l_0\leq l\leq l_1}} A\sum_{j=\nu+1}^{\nu+l}\left(y_j-\mu-\frac{A}{2}\right).$$

The short memory of the MOSUM statistic is paramount to the form of the approximation given in (3.22). This short memory is also present within the generalised moving sum statistic Z_n if one considers its recursive definition above. After the initialising value Z_{l_1}, Z_n essentially becomes a moving sum process given by S_{n,l_0,l_1}. The process $\{S_{n,l_0,l_1}\}$ exhibits a short memory, where dependence between two values is lost after l_1 observations i.e. S_{n,l_0,l_1} and S_{n+l_1,l_0,l_1} are independent for all n. This suggests that stationary behaviour of the combined process $(Z_{l_1}, \{S_{n,l_0,l_1}\})$, under the condition of not crossing the barrier H, should be achieved quickly. One would then anticipate that the form of approximations (3.19) and (3.22) would also be suitable when applied to Z_n.

For $M \geq 0$, introduce the probability:

$$F_{l_0,l_1}(H,M) := \Pr_\infty(Z_{M+l_1} < H) = \Pr_\infty\{Z_{l_1} < H, S_{j,l_0,l_1} < H \;\; \forall\, j = l_1+1,\dots l_1+M\}\,.$$

Then one would anticipate the following approximation to be accurate:

$$F_{l_0,l_1}(H,M) \simeq F_{l_0,l_1}(H,2l_1)\,[\theta_{l_1}(H)]^{M/l_1-2} \quad \text{with } \theta_{l_1}(H) = \frac{F_{l_0,l_1}(H,2l_1)}{F_{l_0,l_1}(H,l_1)}\,, \tag{3.24}$$

$$\mathbb{E}_\infty\tau_Z(H) \simeq l_1 - \frac{l_1 F_{l_0,l_1}(H,2l_1)}{[\theta_{l_1}(H)]^2\log(\theta_{l_1}(H))}\,. \tag{3.25}$$

Table 3. Approximations for $\mathbb{E}_\infty \tau_\xi(h)$ with $L = 50$.

h	2	2.25	2.5	2.75	3	3.25	3.5
(3.19)	471	791	1392	2587	5099	10695	23918
$\mathbb{E}_\infty \tau_\xi(h)$	472	792	1397	2588	5085	10749	24131

Unfortunately, the probability $F_{l_0,l_1}(H;M)$ is complex and to the author's knowledge no formula or approximations are known. The probabilities $F_{l_0,l_1}(H;2l_1)$ and $F_{l_0,l_1}(H;l_1)$ can be approximated via Monte Carlo; this is not too cumbersome as at most $3l_1$ random variables need to be simulated at each iteration. As commonly $\mathbb{E}_\infty \tau_Z(H) = C$ with C large, the right tail of the distribution of the random variable $\max\{Z_{l_1}, \max_{j=0\dots M} S_{j,l_0,l_1}\}$ is of the most interest. Large deviation theory, see [51, 53, 61], could be used to approximate the right tail of this distribution, however numerical results indicate approximations of these kind would not be accurate enough for general l_0 and l_1 (those that are not astronomically large). If the prior knowledge that $1 \le l \le l_1$ is known, and an explicit formula to approximate $F_{1,l_1}(H;M)$ or $\mathbb{E}_\infty \tau_Z(H)$ is desired, the following heuristic argument could be used. Using inspiration from [49], a continuous time analogue of the probability $F_{1,l_1}(H;M)$ that allows for the application of existing large deviation results is:

$$\Pr\left\{ \max_{\substack{0\le s<t\le M+l_1 \\ 0\le t-s\le l_1}} \left[W(t) - W(s) - \frac{A}{2}(t-s) \right] < \frac{H}{A} \right\}, \tag{3.26}$$

where $W(t)$, $0 \le t < \infty$, is the standard Brownian motion process. Ideally, a large deviation approximation for (3.26) should be computed explicitly. However, for $M = l_1$, $M = 2l_1$ and A large, say $A \ge 1$, simulation studies indicate that the additional maximisation constraint in (3.26) of $0 < t - s < l_1$ has very little influence on this probability. If this constraint is ignored, the following large deviation result of [49] can be applied.

Lemma 1. *Suppose $\gamma > 0$, $m \to \infty$ and $u \to \infty$ such that $m\gamma u^{-1}$ is some fixed number in $(1,\infty)$. Then*

$$\Pr\left\{ \max_{0\le s<t\le m} [W(t) - W(s) - \gamma(t-s)] > u \right\} = [2\gamma(m\gamma - u) + 3 + o(1)] \exp(-2\gamma u).$$

To subsequently correct this result for discrete time, it is recommended in [49] to increase the barrier H by 2ρ, where ρ is defined in (3.17).

This results in the approximations

$$F_{1,l_1}(H,l_1) \simeq 1 - (A(Al_1 - H/A - 2\rho) + 3)\exp\{-A(H/A + 2\rho)\}$$

$$F_{1,l_1}(H,2l_1) \simeq 1 - (A(3Al_1/2 - H/A - 2\rho) + 3)\exp\{-A(H/A + 2\rho)\}.$$

As a result, using the approximations given in (3.24) and (3.25):

$$F_{1,l_1}(H,M) \simeq 1 - (A(3Al_1/2 - H/A - 2\rho) + 3) \times \exp\{-A(H/A + 2\rho)\} \left[\hat{\theta}_{l_1}(H)\right]^{M/l_1 - 2}, \tag{3.27}$$

with

$$\hat{\theta}_{l_1}(H) = \frac{1 - (A(3Al_1/2 - H/A - 2\rho) + 3)\exp\{-A(H/A + 2\rho)\}}{1 - (A(Al_1 - H/A - 2\rho) + 3)\exp\{-A(H/A + 2\rho)\}}.$$

Also

$$\mathbb{E}_\infty \tau_Z(H) \simeq l_1 - \frac{l_1[1 - (A(3Al_1/2 - H/A - 2\rho) + 3)\exp\{-A(H/A + 2\rho)\}]}{[\hat{\theta}_{l_1}(H)]^2 \log(\hat{\theta}_{l_1}(H))}. \tag{3.28}$$

The accuracy of the approximation in (3.24) is demonstrated in Figures 1-2 for different l_0, l_1, M and A as a function of H. In this approximation, $F_{l_0,l_1}(H,2l_1)$ and $F_{l_0,l_1}(H,l_1)$ have been approximated using Monte Carlo simulations with 100,000 repetitions. In these figures, the probability $F_{l_0,l_1}(H;M)$ is depicted with a thick dashed black line and is obtained from simulations. The approximation in (3.24) is depicted with a solid blue line. From these figures the high accuracy of approximation (3.24) is clearly demonstrated. In Figures 3-6, we assess the accuracy of the approximation in (3.27). In these figures, for $A = 1$ and various M, the probability $F_{1,l_1}(H;M)$ is depicted with a thick dashed black line whereas the approximation provided in (3.27) is shown with a solid red line. The number present on the figure is used to show the value of l_1 used. From these figures, we see for large H the approximation in (3.27) is adequate. In Tables 4-5 the accuracy of the approximations provided in (3.25) and (3.28) are assessed for different H. We see the approximation in (3.25) is extremely accurate for all H. For large H, the approximation in (3.28) is fairly accurate and has the benefit of explicit evaluation. For small H and small A, the accuracy of (3.28) should deteriorate.

Table 4. Approximations for $\mathbb{E}_\infty \tau_Z(H)$ with $l_0 = 25, l_1 = 50, A = 1$.

H	-5	-4.5	-4	-3.5	-3	-2.5	-2
(3.25)	126	145	166	196	228	268	319
$\mathbb{E}_\infty \tau_Z(H)$	127	144	167	194	229	272	323

Table 5. Approximations for $\mathbb{E}_\infty \tau_Z(H)$ with $l_0 = 1, l_1 = 10, A = 1$.

H	2	2.25	2.5	2.75	3	3.25	3.5
(3.28)	30	42	59	81	111	148	195
(3.25)	41	53	70	91	120	156	205
$\mathbb{E}_\infty \tau_Z(H)$	41	54	70	91	120	157	207

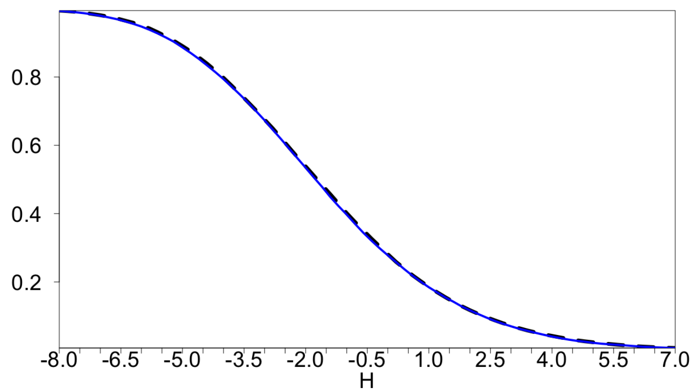

Figure 1: Empirical probabilities of reaching the barrier H (dashed black) and approximation (3.24) (solid blue): $A = 1$, $M/l_1 = 4$ with $l_0 = 25$ and $l_1 = 50$.

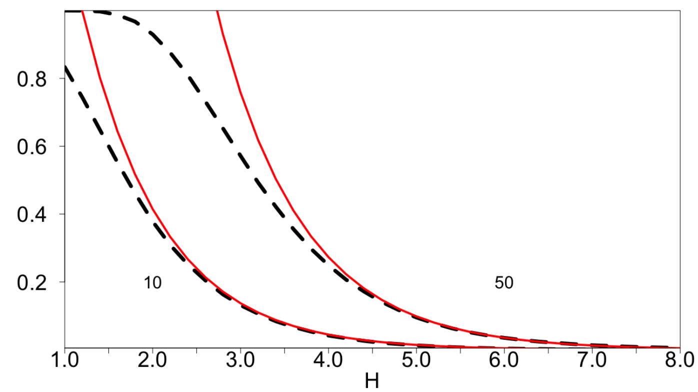

Figure 3: Empirical probabilities of reaching the barrier H (dashed black) and corresponding versions of approximation (3.27) (solid red): $A = 1$, $m/l_1 = 1$ with (a) $l_1 = 10$ and (b) $l_1 = 50$.

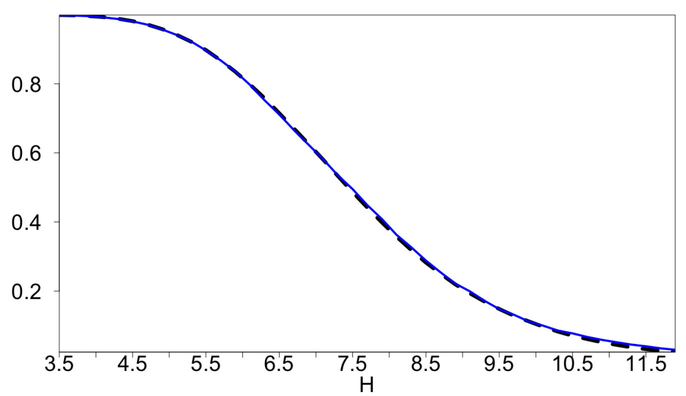

Figure 2: Empirical probabilities of reaching the barrier H (dashed black) and approximation (3.24) (solid blue): $A = 0.5$, $M/l_1 = 25$ with $l_0 = 10$ and $l_1 = 20$.

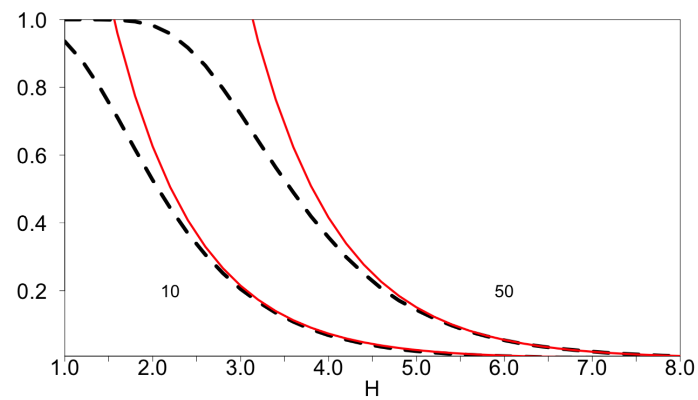

Figure 4: Empirical probabilities of reaching the barrier H (dashed black) and corresponding versions of approximation (3.27) (solid red): $A = 1$, $m/l_1 = 2$ with $l_1 = 10$ and (b) $l_1 = 50$.

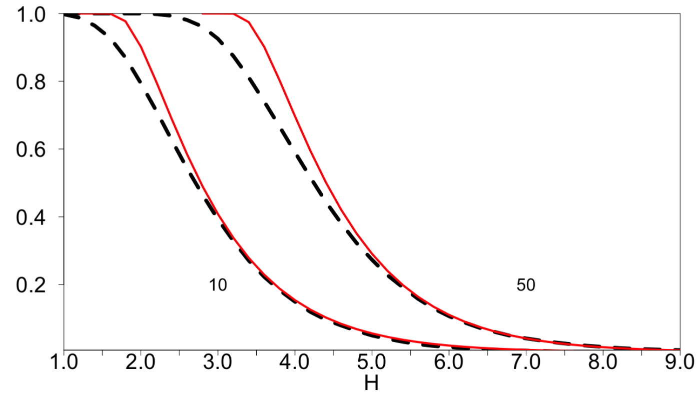

Figure 5: Empirical probabilities of reaching the barrier H (dashed black) and corresponding versions of approximation (3.27) (solid red): $A = 1$, $m/l_1 = 10$ with $l_1 = 10$ and (b) $l_1 = 50$.

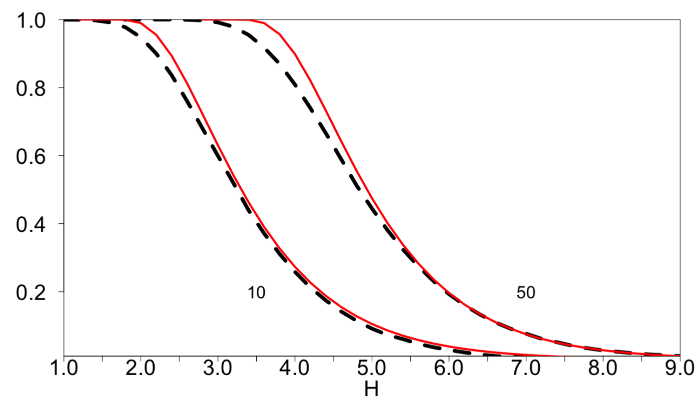

Figure 6: Empirical probabilities of reaching the barrier H (dashed black) and corresponding versions of approximation (3.27) (solid red): $A = 1$, $m/l_1 = 5$ with (a) $l_1 = 10$ and (b) $l_1 = 50$.

3.2.4 The presence of nuisance parameters

Here we briefly consider statistics aimed at detecting a transient change when certain nuisance parameters require estimation. The brevity of this discussion is because (in the authors opinion) in practice for online change-point problems, the behaviour of the time series under the null hypothesis of no change-point is often observed for a lengthy period of time. This allows for the accurate estimation of certain nuisance parameters and they can therefore be assumed known. The counterargument to this is that small trend changes or small undetectable changes may contaminate estimates. For the situation of small trend changes, the use of singular spectrum analysis (SSA) could be used to extract the trend allowing for the study of only the residuals, see Section 3.5 where an approach similar to this is discussed. Many of the following statistics appear in some form in [2] when addressing the offline change-point problem, and a number of approximations for the false alarm error are provided. The log likelihood ratio given in (3.1), where f and g are given in (3.9), is

$$\Gamma_{\nu,\nu+l} = A \sum_{j=\nu+1}^{\nu+l} \left(y_j - \mu - \frac{A}{2} \right) .$$

Using motivation from [50], if μ is unknown, l is unknown but bounded $l_0 \leq l \leq l_1$ and A is known, then one can replace μ with its maximum likelihood estimator under H_∞; $\hat{\mu} := \sum_{i=1}^{n} y_i/n$ to obtain:

$$Z_n^1 := \max_{\substack{0 \leq \nu < \nu+l \leq n \\ l_0 \leq l \leq l_1}} A \sum_{j=\nu+1}^{\nu+l} \left(y_j - \hat{\mu} - \frac{A}{2} \right) .$$

In [51], μ was replaced with its average over the null and alternative hypotheses to obtain the true likelihood ratio statistic:

$$Z_n^2 := \max_{\substack{0 \leq \nu < \nu+l \leq n \\ l_0 \leq l \leq l_1}} A \sum_{j=\nu+1}^{\nu+l} \left(y_j - \hat{\mu} - \frac{A}{2} \left(1 - \frac{l}{n} \right) \right) .$$

If μ and A are both unknown, the square root of the log likelihood ratio statistic is:

$$Z_n^3 := \max_{\substack{0 \leq \nu < \nu+l \leq n \\ l_0 \leq l \leq l_1}} \frac{\left(\sum_{j=\nu+1}^{\nu+l} y_j \right) - l\hat{\mu}}{\sqrt{l(1 - \frac{l}{n})}} .$$

The statistic Z_n^3 is also studied in [62, p. 497] for detecting a transient change in mean of random variables in an offline setting. Here the statistic is formulated without likelihood arguments and is therefore used when y_j are not necessarily Gaussian. In [62], the authors view the change-point statistic as a discretizations of some Holder norms or semi-norms allowing them to obtain limiting distributions under the null hypothesis of no change in mean.

Instead of testing for the existence of a single transient change in the mean, the problem of detecting multiple changes in the means of i.i.d. random variables has been studied in [63]. Here, a MOSUM-like statistic is used for detecting any possible number of changes in a sample of fixed length n and the values of the mean after the change-point do not necessarily have to be known. This can be seen as a significant generalisation of the problems considered in this paper and in [2], if one considers only the offline setting. The statistic studied in [63] is proportional to the following quantity:

$$\max_{L\le\nu\le n-L} |T_{\nu,n}(L)| \,, \tag{3.29}$$

with

$$T_{\nu,n}(L) = \frac{1}{\sqrt{2L}}\left(\sum_{i=\nu+1}^{\nu+L} y_i - \sum_{i=\nu-L+1}^{\nu} y_i\right).$$

The statistic $T_{\nu,n}(L)$ has a simple interpretation of comparing at every time point $L \le \nu \le n-L$ the mean of the subsample $y_{\nu-L+1},\dots,y_\nu$ with the mean of the subsample $y_{\nu+1},\dots,y_{\nu+L}$. Naturally, a large difference between the two means (the sign is irrelevant because of the absolute value in (3.29)) would indicate a change at this point. As mentioned in [63], at a point ν this statistic is similar to the likelihood ratio statistic for the sample $y_{\nu+1},\dots,y_{\nu+L}$ at the potential change-point ν. The asymptotic behaviour of a normalised form of the statistic in (3.29) as $n\to\infty$ is given Theorem 2.1 of [63] (here we are ignoring a number of technical details) and is shown to follow a Gumbel extreme value distribution.

For the statistics considered in this section, it is not obvious how one can translate the offline change-point results of [63, 49, 50, 62, 51, 52, 53] to address the online change-point problem in the presence of nuisance parameters.

3.3 Optimality criteria

For online detection of transient changes, optimality criteria like (2.10) and (2.11) do not have much meaning as the change in distributions is not permanent (signal can be missed). Instead, optimality involving the maximisation of the probability of detection under a constraint on the false alarm risk is more applicable, see [64, 65]. One could use a worst-case criterion of the form:

$$\inf_\nu \mathrm{P}_\nu\{\tau(H)-\nu < T \,|\, \tau(H) > \nu\}\,, \tag{3.30}$$

where $T > 1$ is the maximum length of time after the change-point occurs that it must be detected; this is problem specific and is therefore chosen by the user. By imposing the condition of a long run with no false alarms, another possible criterion is

$$\lim_{\nu\to\infty} \mathrm{P}_\nu\{\tau(H)-\nu < T \,|\, \tau(H) > \nu\}\,. \tag{3.31}$$

Using ARL as the measure of false alarm risk, a stopping rule $\tau\in\Delta(C)$ is then optimal for a given C if it maximises (3.30) or (3.31); recall $\Delta(C) = \{\tau : E_\infty\tau \ge C\},\ \ C>1$.

3.3.1 MOSUM procedure

For the MOSUM procedure given in (3.10), the quantity (3.31) was the focus of study in [25] and built on the continuous time results of [27]. For $T = l+L$, the quantity (3.31) is equivalent to:

$$\mathcal{P}_S(H,A,L) \;:=\; \lim_{\nu\to\infty} \mathrm{P}_\nu\{S_{n,L} > H \text{ for some } n \in [\nu'+1, \nu+l-1] \,|\, \tau_{S,L}(H) > \nu'\}\,, \tag{3.32}$$

with $\nu' := \nu - L$.

Formally, we require $\nu\to\infty$ in (3.32). This is to ensure that the sequence of moving sums $\{S_{n,L}\}_n$ reaches the stationary behaviour under the null hypothesis and given that we have not crossed the threshold H. However, as discussed [25, 27], this stationary regime is reached very quickly and in all approximations it is enough to only require $\nu \ge 2L$.

The reasoning behind the choice $T = l+L$ is as follows. Assume $\mathbb{H}_\nu$ with $\nu<\infty$, and that ν is suitably large. If the barrier H is reached for any sum $S_{n,L}$ with $n\le\nu'$ then, since there are no parts of the signal in the sums $S_{0,L},\dots S_{\nu',L}$, we classify the event of reaching the barrier as a false alarm. Each one of the sums $S_{\nu'+1,L},\dots,S_{\nu+l-1,L}$ has mean larger than $L\mu$ as it contains at least a part of the signal. Reaching the barrier H by any of these sums will be classified as a correct detection of the signal. If neither of these sums reaches H, then we say that we failed to detect the signal and further events when $S_{n,L}\ge H$ with $n\ge\nu+l$ will again be classified as false alarms. In Figure 7 we display the values $\mathbb{E}_\nu S_{n,L}$ as a function of n.

Define the function

$$Q(n;A,L,\nu') := \begin{cases} 0 & \text{for } n\le\nu' \text{ or } n\ge\nu+l\\ A(n-\nu') & \text{for } \nu'<n\le\nu'+\min(l,L)\\ A\min(l,L) & \text{for } \nu'+\min(l,L)<n\le\nu'+\max(l,L)\\ A(L+l+\nu'-n) & \text{for } \nu'+\max(l,L)<n\le\nu+l-1\,. \end{cases}$$

Then Figure 7 is also a plot of $\mu L + Q(n;A,L,\nu')$. By subtracting $\mathbb{E}_\nu S_{n,L}$ from the threshold H and standardising the random variables $S_{n,L}$ the power of the test given in (3.32) can be expressed in terms of probability under $\mathbb{H}_\infty$:

$$\mathcal{P}_\xi(h,A,L) := \lim_{\nu\to\infty} \mathrm{Pr}_\infty\left\{\xi_{n,L} > h - \frac{Q(n;A,L,\nu')}{\sigma\sqrt{L}} \text{ for some } n\in[\nu'+1,\nu+l-1] \,\middle|\, \tau_\xi(h)>\nu'\right\},$$

where $\mathcal{P}_S(H,A,L) = \mathcal{P}_\xi(h,A,L)$. Recall the relation $H = \mu L + h\sqrt{L}$. To approximate $\mathcal{P}_\xi(h,A,L)$, the approach taken in [25] was similar to the approach taken to approximate ARL in Section 3.2.1. The approach is as follows. We firstly approximate the problem in the continuous-time setting and compute probabilities for the Gaussian process $S(t)$. Then, use the results of D. Siegmund to correct the continuous time probability for discrete time. Fix $\gamma = A\sqrt{L}$, $\kappa = \nu'/L$,

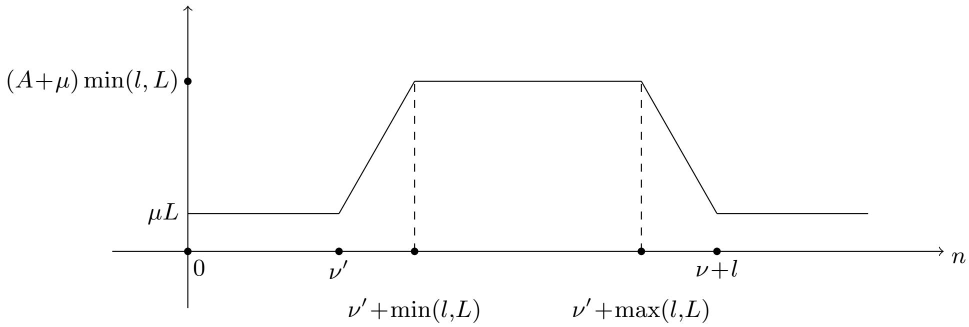

Figure 7: $\mathbb{E}_\nu S_{n,L}$ as a function of n

$\lambda = l/L$ and define the function

$$Q(t;\gamma,\kappa,\lambda) = \begin{cases} 0 \\ \gamma(t-\kappa) \\ \gamma\min(1,\lambda) \\ \gamma(1+\lambda+\kappa-t) \end{cases}$$

$$\begin{array}{c} \text{for } t \le \kappa \text{ or } t \ge \kappa+1+\lambda. \\ \text{for } \kappa < t \le \kappa+\min(1,\lambda) \\ \text{for } \kappa+\min(1,\gamma) < t \le \kappa+\max(1,\lambda) \\ \text{for } \kappa+\max(1,\lambda) < t \le \kappa+1+\lambda\,. \end{array}$$

The diffusion approximation for the power of the test is

$$\begin{aligned} \mathcal{P}(h,A) :=& \\ &\lim_{\kappa\to\infty} \Pr{}_\infty\{S(t) > h - Q(t;\gamma,\kappa,\lambda) \\ &\text{for some } t \in [\kappa,\kappa+1+\lambda] \,|\, \tilde\tau(h) > \kappa\}\,, \end{aligned} \tag{3.33}$$

where $\tilde\tau(h) = \inf\{t > 0 : S(t) > h\}$. We refer to Lemma 4.1 in [25] for more details about this approach. That is, by assuming $L \to \infty$, we make the approximation

$$\mathcal{P}_\xi(h,A,L) \cong \mathcal{P}(h,A)\,.$$

The complexity of computation of the diffusion approximation $\mathcal{P}(h,A)$ and its discrete-time corrected version depends on the choice of L in comparison to l. Here, we will only consider the scenario of $\lambda = l/L = 1$ which corresponds to the case of l known at the MOSUM construction stage. The two other cases of $\lambda > 1$ and $\lambda < 1$ are studied in [25].

For $\lambda = 1$, the diffusion approximation for $\mathcal{P}_\xi(h,A,L)$ given in (3.33) reduces to

$$\begin{aligned} \mathcal{P}(h,A) = \lim_{\kappa\to\infty} \mathrm{P}_\infty\{S(t) \ge h - Q(t;\gamma,\kappa) \\ \text{for some } t\in[\kappa,\kappa+2] \big|\, \tilde\tau(h) > \kappa\}\,, \end{aligned} \tag{3.34}$$

where $Q(t;\gamma,\kappa) = \gamma\max\{0, 1-|t-(\kappa+1)|\}$. The barrier $h - Q(t;\gamma,\kappa)$ is depicted in Figure 8.

The probability (3.34) was considered in [27], where approximations accurate to more than 4 decimal places were developed. Define the following two conditional probabilities:

$$\begin{aligned} F_{h,0}(1|x) :=& \mathrm{P}_\infty(S(t) < h \text{ for all } t \in [0,1] \\ &\mid S(0) = x)\,, \\ F_{h,0,-\gamma,\gamma}(3|x) :=& \mathrm{P}_\infty(S(t) < B(t;h,0,-\gamma,\gamma) \\ &\text{for all } t \in [0,3] \mid S(0) = x)\,, \end{aligned}$$

where the barrier $B(t;h,0,-\gamma,\gamma)$ is defined as

$$B(t;h,0,-\gamma,\gamma) = \begin{cases} h, & 0 \le t \le 1 \\ h-\gamma(t-1), & 1 < t \le 2 \\ h-\gamma+\gamma(t-2), & 2 < t \le 3 \\ 0 & \text{otherwise}\,, \end{cases}$$

and is depicted in Figure 9. From [27] we obtain

$$\mathcal{P}(h,A) \cong 1 - \frac{F_{h,0,-\gamma,\gamma}(3|0)}{F_{h,0}(1|0)}\,, \tag{3.35}$$

where

$$F_{h,0}(1\,|\,x) = \Phi(h) - \exp\big(-(h^2-x^2)/2\big)\,\Phi(x) \tag{3.36}$$

and

$$\begin{aligned} F_{h,0,-\gamma,\gamma}(3\,|\,x) =& \frac{e^{\gamma^2/2}}{\varphi(x)} \int_{-x-h}^{\infty}\int_{x_2-h+\gamma}^{\infty} e^{-\gamma(x_3-x_2)} \\ &\times \det\begin{bmatrix} \varphi(x) & \varphi(-x_2-h) & \varphi(-x_3-2h+\gamma) & \Phi(-x_3-2h+\gamma) \\ \varphi(h) & \varphi(-x-x_2) & \varphi(-x-x_3-h+\gamma) & \Phi(-x-x_3-h+\gamma) \\ \varphi(x_2+2h+x) & \varphi(h) & \varphi(x_2-x_3+\gamma) & \Phi(x_2-x_3+\gamma) \\ \varphi(x_3+3h-\gamma+x) & \varphi(x_3+2h-\gamma-x_2) & \varphi(h) & \Phi(h) \end{bmatrix} dx_3 dx_2\,. \end{aligned}$$

To compute the approximation (3.35) one needs to numerically evaluate a two-dimensional integral which is a routine problem for modern computers.

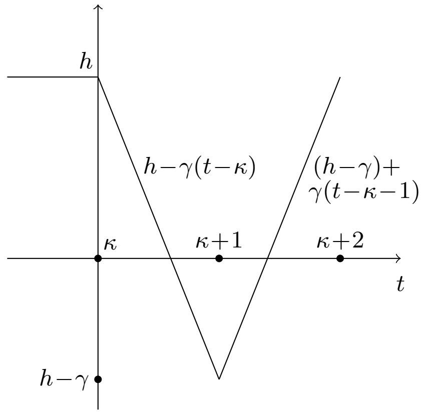

Figure 8: Barrier $h - Q(t;\gamma,\kappa)$ for $\lambda = 1$.

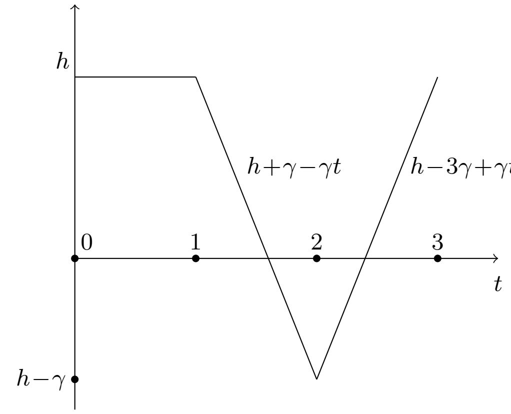

Figure 9: Barrier $B(t;h,0,-\gamma,\gamma)$.

Correcting approximation (3.35) for discrete time can be performed in the same manner as correcting the ARL approximations in Section 3.2.1. This results in the approximation

$$\mathcal{P}_\xi(h,A,L) \cong 1 - \frac{F_{h_L,0,-\gamma,\gamma}(3|0)}{F_{h_L,0}(1|0)}\,,$$

$$\text{where } h_L := h + \omega_L\,. \tag{3.37}$$

The quantity $\omega_L = \sqrt{2}\rho/\sqrt{L}$ corresponds to the specialised discrete time correction of D. Siegmund, see [25].

In Figures 10-11, the thicker black dashed line corresponds to the empirical values of the BCP $\mathcal{P}_\xi(h,A,L)$ computed from 100,000 simulations with different values of L and γ, where $\mu = 0$ and $\sigma = 1$. The solid red line corresponds to the approximation in (3.37). The dot-dashed blue line corresponds to the diffusion approximation given in (3.35). The axis are: the x-axis shows the value of γ. The y-axis denotes the probabilities of reaching the barrier. The graphs, therefore, show the empirical probabilities of $\mathcal{P}_\xi(h,A,L)$ and values of approximation (3.37).

From Figures 10-11, we see that approximation (3.37) is very accurate even for a very small $L = 5$. We also see the significance of the discrete-time correction; whilst the diffusion approximation provides sensible results should you compare it with $L = 100$, for $L = 5$ the diffusion approximation is very far off.

3.4 Comparison of tests

In this section, we compare the power of the MOSUM test in (3.10) against the generalised MOSUM statistic (3.23) and the CUSUM test given in (2.2) specialised for this Gaussian example when used to detect a transient change. The CUSUM statistic in (2.1) can be expressed as:

$$V_n = \max_{0\le\nu\le n-1} \prod_{j=\nu+1}^{n} \exp\left(\frac{(y_j-\mu)^2 - (y_j-\mu-A)^2}{2}\right),$$

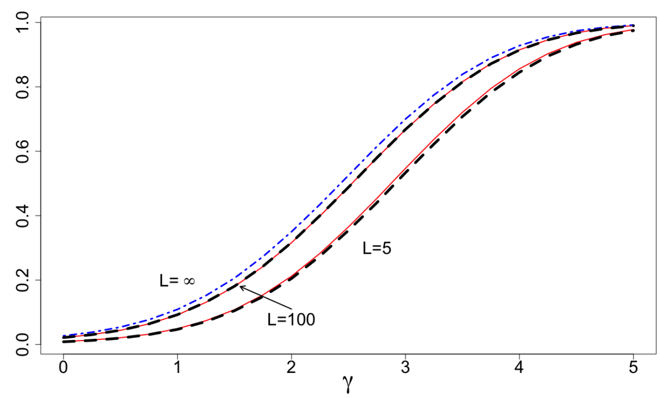

Figure 10: Empirical probabilities of $\mathcal{P}_\xi(h,A,L)$ (thick dashed black) and its approximations (solid red and solid blue) for $h = 3$.

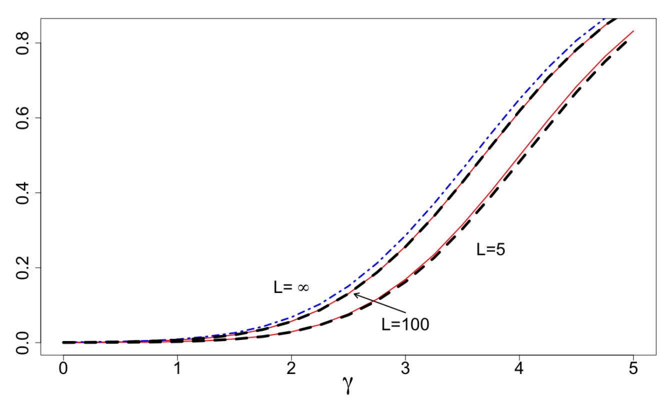

Figure 11: Empirical probabilities of $\mathcal{P}_\xi(h,A,L)$ (thick dashed black) and its approximations (solid red and solid blue) for $h = 4$.

with the CUSUM test being

$$\tau_V(H) = \inf\{n \geq 1 : V_n > H\}\,.$$

(The choice of H will be discussed shortly.) Secondly, but also simultaneously, we compare the power of the MOSUM test as $\lambda = l/L$ varies in $[0.5, 2]$; the purpose is to demonstrate when the generalised MOSUM statistic becomes beneficial when the exact value of l is unknown and we make a potentially poor guess in the MOSUM test. This corresponds to a reasonable choice of $l/l_0 = 2$ and $l/l_1 = 0.5$. Here, we shall consider the power criterion given in (3.31) and set $T = 2l$. That is, we want to detect the presence of the change point within $2l - 1$ after its occurence. For the MOSUM test, the power is then

$$\begin{aligned}&\mathcal{P}_S(H_1, A, L) := \\ &\lim_{\nu\to\infty} \Pr_\nu\{S_{n,L} > H_1 \text{ for some} \\ &n \in [\nu - L + 1, \nu - L + 2l - 1] \,|\, \tau_{S,L}(H_1) > \nu - L\}\,.\end{aligned}$$

For the generalised MOSUM test, the power is

$$\begin{aligned}&\mathcal{P}_Z(H_2, A, l_0, l_1) := \\ &\lim_{\nu\to\infty} \Pr_\nu\{Z_n(l_0, l_1) > H_2 \text{ for some} \\ &n \in [\nu + 1, \nu + 2l - 1] \,|\, \tau_Z(H_2) > \nu\}\,.\end{aligned}$$

The power of the CUSUM test for the transient change considered in then equivalent to

$$\begin{aligned}&\mathcal{P}_V(H_3, A) := \\ &\lim_{\nu\to\infty} \Pr_\nu\{V_n > H_3 \text{ for some} \\ &n \in [\nu + 1, \nu + 2l - 1] \,|\, \tau_V(H_3) > \nu\}\,.\end{aligned}$$

To compare the three tests, the thresholds H_1, H_2 and H_3 have been set such that $\mathbb{E}_\infty \tau_M(H_1) = \mathbb{E}_\infty \tau_Z(H_2) = \mathbb{E}_\infty \tau_V(H_3) = 500$. Determination of H_1 for MOSUM has been computed using the accurate approximation in (3.19). For the generalised MOSUM procedure, H_2 is found via Monte Carlo simulations with $50,000$ repetitions. Determination of H_3 for CUSUM was obtained using tabulated values given in [43, p. 3237].

In the first example shown in Figure 12, we have set $A = 1$ and $l = 10$. For the MOSUM test, we consider values of $L \in [5, 20]$ to ensure $\lambda \in [0.5, 2]$. For each λ, the values of $\mathcal{P}_S(H_1, A, L)$ can be accurately approximated using the results of [25] or via Monte Carlo methods and are displayed with a solid black line. The dashed orange line depicts $\mathcal{P}_Z(H_2, A, 5, 20)$ which corresponds to prior knowledge that l is between $[5, 20]$. The shorter dashed blue line corresponds to $\mathcal{P}_V(H_3, A)$ which has been obtained via Monte Carlo simulations. In Figure 13, we set $A = 0.5$ and $l = 20$. For the MOSUM procedure, we consider values of $L \in [10, 40]$ to ensure $\lambda \in [0.5, 2]$. In this figure, the dashed orange line depicts $\mathcal{P}_Z(H_2, A, 10, 40)$ which corresponds to prior knowledge that l is between $[10, 40]$. The shorter dashed blue line corresponds to $\mathcal{P}_V(H_3, A)$ obtained via Monte Carlo simulations. In all Monte Carlo simulations, we have used 50,000 repetitions.

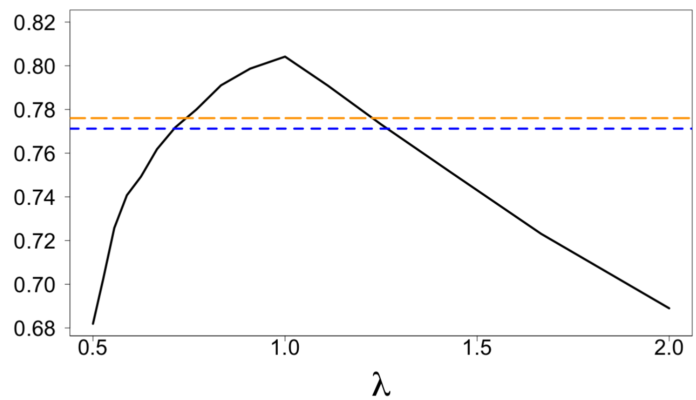

Figure 12: Power of three tests with $A = 1$ and $l = 10$ and ARL= 500.

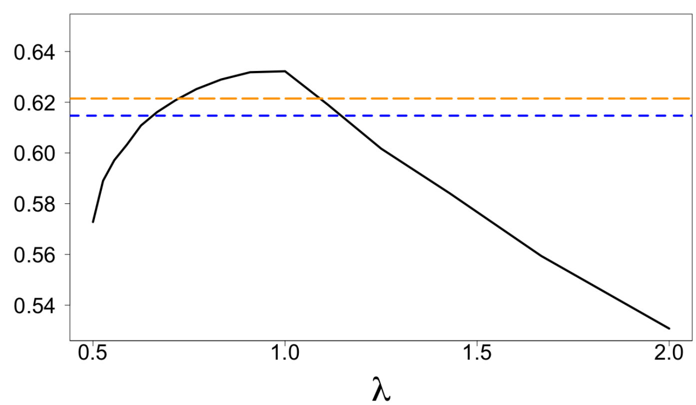

Figure 13: Power of three tests with $A = 0.5$ and $l = 20$ and ARL= 500.

From Figures 12-13, one can observe the advantage of knowing l since the largest value of $\mathcal{P}_S(H_1, A, L)$ is the largest power of all three tests and is obtained for $\lambda = l/L = 1$. In these figures, the values of $\lambda = l/L$ such that $\mathcal{P}_S(H_1, A, L)$ exceeds $\mathcal{P}_Z(H_2, A, 5, 20)$ (Figures 12) and $\mathcal{P}_Z(H_2, A, 10, 40)$ (Figures 13) shows the freedom in the choice of L such that when l is unknown, you still benefit over only assuming l is bounded (similarly for CUSUM case when considering the dashed blue line). From these figures it is clear that unless you are very fortunate in choosing L close to l for the MOSUM test, you should use the generalised MOSUM test if A is known. Unfortunately, there are no convenient analytic results for this test. Moreover, both the generalised MOSUM procedure and CUSUM procedures require the additional knowledge of A; this is not true for MOSUM. For the choice of parameters considered in both examples, the additional knowledge of a transient change leads to obvious benefits in power; those is seen by comparing the generalised MOSUM orange lines with the blue CUSUM lines. Of course, $\mathcal{P}_Z(H_2, A, l_0, l_1) \to \mathcal{P}_S(H_1, A, l)$ as $l_0, l_1 \to l$.

3.5 An application to real world data

Hydrostatic pressure testing is important safety precaution for the Oil and Gas industry, see [66]. Pressure testing is performed to confirm a pressure containing system is structurally sound and not leaking. Tests are performed by increasing the pressure in the system, expanding the pressure body, until the pressure reaches a pre-defined value typically equal to or larger than the body rated design pressure, then holding it there for a long enough time period to confirm there are no leaks, until eventually releasing the pressure. When performing tests offshore on floating Vessel/Drilling Rigs (Rig) this is complicated by the Rig's movement due to the ocean waves, which introduce nearly sinusoidal fluctuations in pressure. Many of these tests are performed in real time and in parallel. Locating automatically when a test has been performed is essential for pressure analysis to determine if a leak is present and this is not obvious when noise is large. Typical example data is shown in Figure 14. When performing pressure tests, the hold periods can differ in length and amplitudes (pressure).

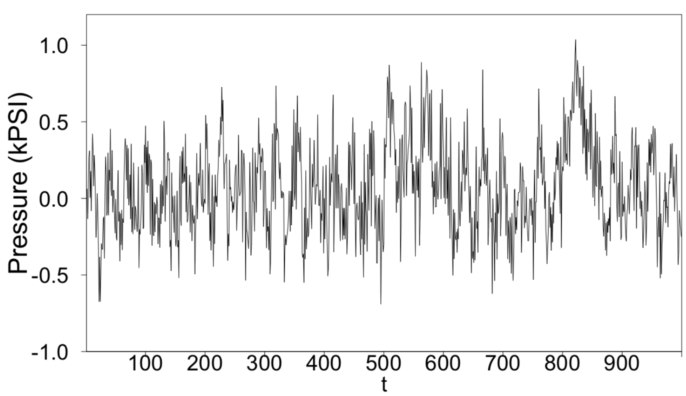

Figure 14: Typical oil pressure data

A sensible way of modelling the data under the null hypothesis of no pressure test could be $z_t = s_t + y_t$, where s_t represents the signal introduced by the wave motion and y_t can be modelled as i.i.d. $N(\mu, \sigma^2)$ and reflects the random noise that is present in the system. In most scenarios, there is significant pre-test data so s_t, μ and σ can be estimated with great accuracy and therefore assumed known. How to estimate s_t or in general how to remove all main components of a signal leaving only noise can be performed using Singular Spectrum Analysis, see [67, 68]. When a pressure test begins, this can be reflected with a change in mean of the y_t; that is, under a pressure test $\mathbb{E}y_t = \mu + A$. The value of A is often constant, but can differ between tests and is generally unknown. Each test can differ in duration but typical lengths vary between $l \in [50, 100]$ units of time. One has to detect a transient change in mean of $y_t = z_t - s_t$. The behaviour of $z_t - s_t$ is shown in Figure 15. In Figure 16, we depict the MOSUM statistic setting $L = 75$. The horizontal line in this figure corresponds to the threshold required for an ARL of 5000. The MOSUM statistic indicates the location of three performed pressure tests and has the great advantage of not requiring knowledge A when determining the ARL threshold unlike the generalised MOSUM and CUSUM procedures. A similar example is shown in Figures 17-18, where $L = 150$ has been selected; three tests have been clearly located. Note that in Figure 16 and Figure 18, the MOSUM statistic is depicted with a shift in time by L $(t \to t - L)$. This explains the early exceedance seen in Figure 18.

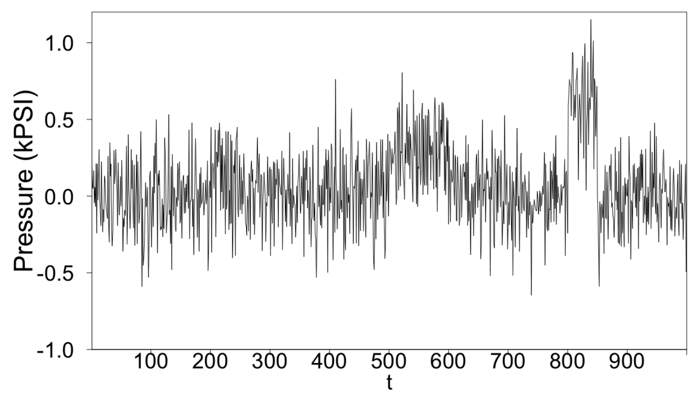

Figure 15: Behaviour of y_t.

ACKNOWLEDGEMENTS

This work was partially supported by EPSRC grant EP/M024830. The author is grateful to both reviewers for their careful reading of the manuscript.

Received 6 April 2021

REFERENCES

[1] A. Tartakovsky, I. Nikiforov, and M. Basseville. *Sequential analysis: Hypothesis testing and changepoint detection.* CRC Press, 2014.

[2] Q. Yao. Tests for change-points with epidemic alternatives. *Biometrika*, 80(1):179–191, 1993.

[3] M. Basseville and I. Nikiforov. *Detection of Abrupt Changes: Theory and Application.* Prentice-Hall Inc, 1993.

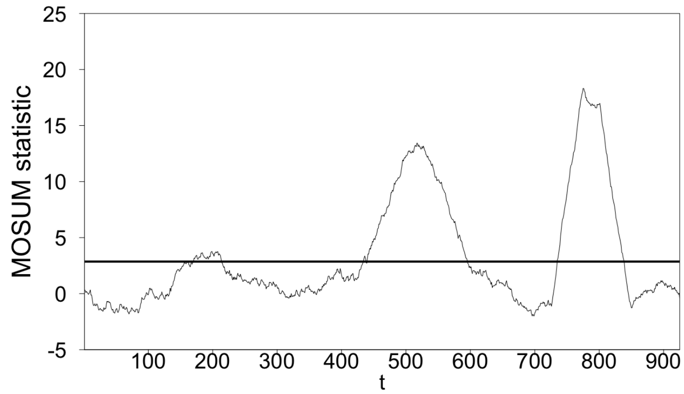

Figure 16: MOSUM statistic based on Figure 15 with $L = 50$ and ARL= 5000.

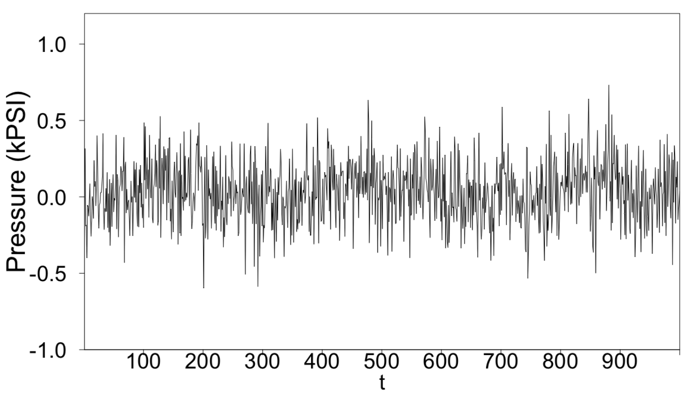

Figure 17: Behaviour of y_t.

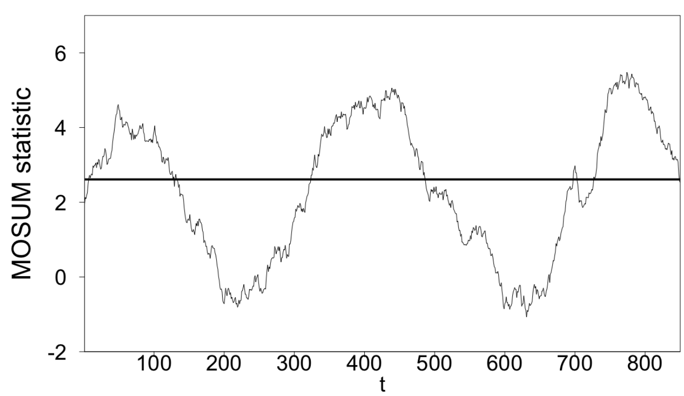

Figure 18: MOSUM statistic based on Figure 17 with $L = 150$ and ARL= 5000.

[4] A. Zhigljavsky and A. Kraskovsky. *Detection of abrupt changes of random processes in radiotechnics problems*. St. Petersburg University Press, St. Petersburg, Russia.(in Russian), 1988.

[5] V. Poor and O. Hadjiliadis. *Quickest detection*. Cambridge University Press, 2008.

[6] W. Shewhart. *Economic control of quality of manufactured product*. Macmillan And Co Ltd, London, 1931.

[7] E. Page. Continuous inspection schemes. *Biometrika*, 41(1/2):100–115, 1954.

[8] A. Shiryaev. On optimum methods in quickest detection problems. *Theory of Probability & Its Applications*, 8(1):22–46, 1963.

[9] S. Roberts. A comparison of some control chart procedures. *Technometrics*, 8(3):411–430, 1966.

[10] S. Roberts. Control chart tests based on geometric moving averages. *Technometrics*, 42(1):97–101, 2000.

[11] G. Lorden. Procedures for reacting to a change in distribution. *The Annals of Mathematical Statistics*, 42(6):1897–1908, 1971.

[12] G. Moustakides. Optimal stopping times for detecting changes in distributions. *Annals of Statistics*, 14(4):1379–1387, 1986.

[13] M. Pollak. Optimal detection of a change in distribution. *The Annals of Statistics*, pages 206–227, 1985.

[14] Y. Ritov. Decision theoretic optimality of the CUSUM procedure. *The Annals of Statistics*, pages 1464–1469, 1990.

[15] L. Gordon and M. Pollak. An efficient sequential nonparametric scheme for detecting a change of distribution. *The Annals of Statistics*, pages 763–804, 1994.

[16] T. Lai. Sequential changepoint detection in quality control and dynamical systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(4):613–644, 1995.

[17] B. Guépié, L. Fillatre, and I. Nikiforov. Detecting a suddenly arriving dynamic profile of finite duration. *IEEE Transactions on Information Theory*, 63(5):3039–3052, 2017.

[18] B. Guépié, L. Fillatre, and I. Nikiforov. Sequential detection of transient changes. *Sequential Analysis*, 31(4):528–547, 2012.

[19] T. Lai. Control charts based on weighted sums. *The Annals of Statistics*, 2(1):134–147, 1974.

[20] C. Chu, K. Hornik, and C. Kaun. MOSUM tests for parameter constancy. *Biometrika*, 82(3):603–617, 1995.

[21] P. Bauer and P. Hackl. The use of MOSUMS for quality control. *Technometrics*, 20(4):431–436, 1978.

[22] P. Bauer and P. Hackl. An extension of the MOSUM technique for quality control. *Technometrics*, 22(1):1–7, 1980.

[23] J. Glaz, V. Pozdnyakov, and S. Wallenstein. *Scan statistics: Methods and applications*. Springer Science & Business Media, 2009.

[24] J. Glaz, J. Naus, and X. Wang. Approximations and inequalities for moving sums. *Methodology and Computing in Applied Probability*, 14(3):597–616, 2012.

[25] J. Noonan and A. Zhigljavsky. Power of the MOSUM test for online detection of a transient change in mean. *Sequential Analysis*, 39(2):269–293, 2020.

[26] J. Noonan and A. Zhigljavsky. Approximations for the Boundary Crossing Probabilities of Moving Sums of Random Variables. *Methodology and Computing in Applied Probability*, 2020.

[27] A. Zhigljavsky and J. Noonan. First passage times for Slepian process with linear and piecewise linear barriers. *Extremes*, pages 1–25, 2021.

[28] M. Bell. Information theory and radar waveform design. *IEEE Transactions on Information Theory*, 39(5):1578–1597, 1993.

[29] V. Poor. *An introduction to signal detection and estimation*. Springer Science & Business Media, 2013.

[30] R. Streit and P. Willett. Detection of random transient signals via hyperparameter estimation. *IEEE transactions on signal processing*, 47(7):1823–1834, 1999.

[31] L. Schmerr. *Fundamentals of ultrasonic nondestructive evaluation*. Springer, 2016.

[32] A. Bianchi, L. Mainardi, E. Petrucci, M. Signorini, M. Mainardi, and S. Cerutti. Time-variant power spectrum analysis for the detection of transient episodes in HRV signal. *IEEE Transactions on Biomedical Engineering*, 40(2):136–144, 1993.

[33] V. Moskvina and A. Zhigljavsky. An algorithm based on Singular Spectrum Analysis for change-point detection. *Communications in Statistics—Simulation and Computation*, 32(2):319–352, 2003.

[34] E. Brodsky and B. Darkhovsky. *Nonparametric methods in change point problems*, volume 243. Springer Science & Business Media, 2013.

[35] P. Fryzlewicz. Wild binary segmentation for multiple change-point detection. *Annals of Statistics*, 42(6):2243–2281, 2014.

[36] H. Cho and P. Fryzlewicz. Multiple-change-point detection for high dimensional time series via sparsified binary segmentation. *Journal of the Royal Statistical Society: Series B: Statistical Methodology*, pages 475–507, 2015.

[37] K. Korkas and P. Fryzlewicz. Multiple change-point detection for non-stationary time series using wild binary segmentation. *Statistica Sinica*, pages 287–311, 2017.

[38] P. Fryzlewicz. Detecting possibly frequent change-points: Wild Binary Segmentation 2 and steepest-drop model selection. *Journal of the Korean Statistical Society*, 49(4):1027–1070, 2020.

[39] H. Dette and J. Gösmann. A likelihood ratio approach to sequential change point detection for a general class of parameters. *Journal of the American Statistical Association*, 115(531):1361–1377, 2020.

[40] J. Gösmann, T. Kley, and H. Dette. A new approach for open-end sequential change point monitoring. *arXiv preprint arXiv:1906.03225*, 2019.

[41] A. Polunchenko, A. Tartakovsky, and N. Mukhopadhyay. Nearly optimal change-point detection with an application to cybersecurity. *Sequential Analysis*, 31(3):409–435, 2012.

[42] D. Siegmund. *Sequential Analysis: Tests and Confidence Intervals*. Springer, 1985.

[43] G. Moustakides, A. Polunchenko, and A. Tartakovsky. Numerical comparison of cusum and shiryaev–roberts procedures for detecting changes in distributions. *Communications in Statistics—Theory and Methods*, 38(16-17):3225–3239, 2009.

[44] A. Tartakovsky. Asymptotic performance of a multichart CUSUM test under false alarm probability constraint. In *Proceedings of the 44th IEEE Conference on Decision and Control*, pages 320–325. IEEE, 2005.

[45] M. Pollak. Average run lengths of an optimal method of detecting a change in distribution. *The Annals of Statistics*, pages 749–779, 1987.

[46] M. Pollak and A. Tartakovsky. Optimality properties of the Shiryaev-Roberts procedure. *Statistica Sinica*, pages 1729–1739, 2009.

[47] T. Lai. Information bounds and quick detection of parameter changes in stochastic systems. *IEEE Transactions on Information Theory*, 44(7):2917–2929, 1998.

[48] A. Tartakovsky. Discussion on "is average run length to false alarm always an informative criterion?" by Yajun Mei. *Sequential Analysis*, 27(4):396–405, 2008.

[49] M. Hogan and D. Siegmund. Large deviations for the maxima of some random fields. *Advances in Applied Mathematics*, 7(1):2–22, 1986.

[50] B. Levin and J. Kline. The CUSUM test of homogeneity with an application in spontaneous abortion epidemiology. *Statistics in Medicine*, 4(4):469–488, 1985.

[51] D. Siegmund. Boundary crossing probabilities and statistical applications. *The Annals of Statistics*, pages 361–404, 1986.

[52] D. Siegmund. Approximate tail probabilities for the maxima of some random fields. *The Annals of Probability*, pages 487–501, 1988.

[53] Q. Yao. Large deviations for boundary crossing probabilities of some random fields. In *Journal of Mathematical Research and Exposition*, volume 9, pages 181–192, 1989.

[54] D. Slepian. First passage time for a particular Gaussian process. *The Annals of Mathematical Statistics*, 32(2):610–612, 1961.

[55] L. Shepp. First passage time for a particular Gaussian process. *Ann. Math. Stat.*, 42(3):946–951, 1971.

[56] J. Noonan and A. Zhigljavsky. Approximating Shepp's constants for the Slepian process. *Statistics & Probability Letters*, 2019.

[57] J. Naus. Approximations for distributions of scan statistics. *Journal of the American Statistical Association*, 77(377):177–183, 1982.

[58] J. Glaz and J. Naus. Tight bounds and approximations for scan statistic probabilities for discrete data. *The Annals of Applied Probability*, 1(2):306–318, 1991.

[59] J. Glaz, J. Naus, and S. Wallenstein. *Scan statistics*. Springer, 2001.

[60] M. Joner Jr, W. Woodall, and M. Reynolds Jr. Detecting a rate increase using a Bernoulli scan statistic. *Statistics in medicine*, 27(14):2555–2575, 2008.

[61] Q. Yao. Boundary-crossing probabilities of some random fields related to likelihood ratio tests for epidemic alternatives. *Journal of Applied Probability*, pages 52–65, 1993.

[62] A. Račkauskas and C. Suquet. Hölder norm test statistics for epidemic change. *Journal of statistical planning and inference*, 126(2):495–520, 2004.

[63] B. Eichinger and C. Kirch. A MOSUM procedure for the estimation of multiple random change points. *Bernoulli*, 24(1):526–564, 2018.

[64] C. Han, P. Willett, and D. Abraham. Some methods to evaluate the performance of Page's test as used to detect transient signals. *IEEE transactions on signal processing*, 47(8):2112–2127, 1999.

[65] B. Bakhache and I. Nikiforov. Reliable detection of faults in measurement systems. *International Journal of adaptive control and signal processing*, 14(7):683–700, 2000.

[66] S. McAleese. *Operational aspects of oil and gas well testing*. Elsevier, 2000.

[67] N. Golyandina, V. Nekrutkin, and A. Zhigljavsky. *Analysis of Time Series Structure: SSA and related techniques*. Chapman & Hall/CRC, 2001.

[68] N. Golyandina and A. Zhigljavsky. *Singular Spectrum Analysis for time series*. Springer, 2020.

Jack Noonan
School of Mathematics, Cardiff University, Cardiff, CF24 4AG, UK
E-mail address: Noonanj1@cardiff.ac.uk