



Approximations for the Boundary Crossing Probabilities of Moving Sums of Random Variables

Jack Noonan¹ · Anatoly Zhigljavsky¹

Received: 2 October 2018 / Revised: 10 September 2019 / Accepted: 26 December 2019 /
Published online: 07 May 2020
© The Author(s) 2020

Abstract

In this paper we study approximations for the boundary crossing probabilities of moving sums of i.i.d. normal random variables. We approximate a discrete time problem with a continuous time problem allowing us to apply established theory for stationary Gaussian processes. By then subsequently correcting approximations for discrete time, we show that the developed approximations are very accurate even for a small window length. Also, they have high accuracy when the original r.v. are not exactly normal and when the weights in the moving window are not all equal. We then provide accurate and simple approximations for ARL, the average run length until crossing the boundary.

Keywords Moving sum · Boundary crossing probability · Moving sum of normal · Change-point detection

Mathematics Subject Classification (2010) Primary: 60G50, 60G35 · Secondary: 60G70, 94C12, 93E20

1 Introduction: Statement of the Problem

Let $\varepsilon_1, \varepsilon_2, \dots$ be a sequence of i.i.d. normal random variables (r.v.) with mean θ and variance $\sigma^2 > 0$. For a fixed positive integer L , the moving sums are defined by

$$S_{n,L} := \sum_{j=n+1}^{n+L} \varepsilon_j \quad (n = 0, 1, \dots). \quad (1.1)$$

The sequence of the moving sums (1.1) will be denoted by \mathbb{S} so that $\mathbb{S} = \{S_{0,L}, S_{1,L}, \dots\}$.

✉ Anatoly Zhigljavsky
ZhigljavskyAA@cardiff.ac.uk

Jack Noonan
Noonanjl@cardiff.ac.uk

¹ School of Mathematics, Cardiff University, Cardiff, CF24 4AG, UK

The main aim of this paper is development of accurate approximations for the boundary crossing probability (BCP) for the maximum of the moving sums:

$$\mathcal{P}_{\mathbb{S}}(M, H, L) := \Pr \left(\max_{n=0,1,\dots,M} S_{n,L} \geq H \right), \quad (1.2)$$

where M is a given positive integer and H is a fixed threshold. Note that the total number of r.v. ε_i used in (1.2) is $M + L$ and $\mathcal{P}_{\mathbb{S}}(M, H, L) \rightarrow 1$ as $M \rightarrow \infty$, for all H and L . We will mostly be interested in deriving accurate approximations when $M \geq L$. The case of $M \leq L$ is much simpler and is comprehensively covered in Noonan and Zhigljavsky (2019, Section 3), see Section 4.6 for discussion.

Developing accurate approximations for the BCP $\mathcal{P}_{\mathbb{S}}(M, H, L)$ for generic parameters H , M and L is very important in various areas of statistics, predominantly in applications related to change-point detection; see, for example, papers (Bauer and Hackl 1980; Chu et al. 1995; Glaz et al. 2012; Moskvina and Zhigljavsky 2003; Xia et al. 2009) and especially books (Glaz et al. 2001; Glaz et al. 2009). Engineering applications of MOSUM (moving sums charts) are extremely important and have been widely discussed in literature; see e.g. Chu et al. (1995), Eiauer and Hackl (1978), Glaz et al. (2001), Glaz et al. (2009), and Waldmann (1986). The BCP $\mathcal{P}_{\mathbb{S}}(M, H, L)$ is an $(M + 1)$ -dimensional integral and therefore direct evaluation of this BCP is hardly possible even with modern software.

To derive approximations for the BCP (1.2) one can use standard tools and approximate the sequence of moving sums with a continuous-time process and then use some continuous-time approximations, see e.g. Haiman (1999); these approximations, however, are not accurate especially for small window length L ; see discussion in Section 4.7. There is, therefore, a need for derivation of specific approximations for the BCP (1.2). Such a need was well understood in the statistical community and indeed very accurate approximations for the BCP and the Average Run Length (ARL) have been developed in a series of quality papers by J. Glaz and coauthors, see for example Glaz and Johnson (1988), Glaz et al. (2012), Wang and Glaz (2014), and Wang et al. (2014) (the methodology was also extended to the case when ε_j are integer-valued r.v., see Glaz and Naus (1991)). We will call these approximations ‘Glaz approximations’ by the name of the main author of these papers; they will be formally written down in Sections 2.2 and 7.

The accuracy of the approximations developed in the present paper is very high and similar to the Glaz approximations; this is discussed in Sections 6 and 7. The methodologies of derivation of Glaz approximations and the approximations of this paper are very different. The practical advantage of our approximations (they require approximating either a one-dimensional integral or an eigenvalue of an integral operator) is their relative simplicity as to compute the Glaz approximations one needs to numerically approximate $L + 1$ and $2L + 1$ dimensional integrals. This is not an easy task even taking into account the fact of existence of a sophisticated software; see references in Section 2.2.

The paper is structured as follows. In Section 2 we reformulate the problem, state the Glaz approximation and discuss how to approximate our discrete-time problem with a continuous-time problem. In Section 3 we provide exact formulas for the first-passage probabilities (in the continuous-time setup) due to Shepp (1971) and give their alternative representation which will be crucial for deriving some of our approximations. In Section 4 we adapt the methodology of D. Siegmund to correct Shepp’s formulas for discrete time and define a version of the Glaz approximation which we will call Glaz-Shepp-Siegmund approximation. In Section 5 we develop continuous-time approximations based on approximating eigenvalues of integral operators and subsequently correct them for discrete time.

In Sections 4.7 and 6 we present results of large-scale simulation studies evaluating the performance of the considered approximations (also, in the cases when the original r.v. ε_j are not normal and the weights in the moving window are not equal). In Section 7, we develop an approximation for ARL and compare its accuracy to the one developed in Glaz et al. (2012).

2 Boundary Crossing Probabilities: Discrete and Continuous Time

2.1 Standardisation of the Moving Sums

The first two moments of $S_{n,L}$ are $\mathbb{E} S_{n,L} = \theta L$ and $\text{var}(S_{n,L}) = \sigma^2 L$. Define

$$\xi_{n,L} := \frac{S_{n,L} - \mathbb{E} S_{n,L}}{\sqrt{\text{var}(S_{n,L})}} = \frac{S_{n,L} - \theta L}{\sigma \sqrt{L}}, \quad n = 0, 1, \dots, \tag{2.1}$$

which are the standardized versions of $S_{n,L}$. All r.v. $\xi_{n,L}$ are $N(0, 1)$; that is, they have the probability density function and c.d.f.

$$\varphi(x) := \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad \Phi(t) := \int_{-\infty}^t \varphi(x) dx. \tag{2.2}$$

Unlike the original r.v. ε_i , the r.v. $\xi_{0,L}, \xi_{1,L}, \dots$ are correlated so that for all $k = 0, 1, \dots$ we have $\text{Corr}(\xi_{0,L}, \xi_{k,L}) = \text{Corr}(\xi_{n,L}, \xi_{n+k,L})$ and

$$\text{Corr}(\xi_{n,L}, \xi_{n+k,L}) = \max\{0, 1 - k/L\} = \begin{cases} 1 - k/L & \text{for } 0 \leq k \leq L \\ 0 & \text{for } k > L. \end{cases} \tag{2.3}$$

Proof of (2.3) is straightforward, see Noonan and Zhigljavsky (2019, Lemma 1).

Set $T = M/L$ and

$$h = \frac{H - \theta L}{\sigma \sqrt{L}} \text{ so that } H = \theta L + \sigma h \sqrt{L}. \tag{2.4}$$

Define the BCP for the sequence of r.v. $\xi_{0,L}, \xi_{1,L}, \dots$:

$$P_L(T, h) := \Pr \left(\max_{n=0,1,\dots,TL} \xi_{n,L} \geq h \right). \tag{2.5}$$

From (2.1) and (2.4), the BCPs $\mathcal{P}_{\mathbb{S}}(M, H, L)$ and $P_L(T, h)$ are equal:

$$\mathcal{P}_{\mathbb{S}}(M, H, L) = P_L(T, h) \text{ for any } H, L \text{ and } T = M/L.$$

Note also that $P_L(T, h) = 1 - F_L(T, h)$, where

$$F_L(T, h) = \Pr \left(\max_{n=0,1,\dots,TL} \xi_{n,L} < h \right). \tag{2.6}$$

In accordance with the terminology of Shepp (1971) and Slepian (1961) we shall call $F_L(T, h)$ ‘first-passage probability’. In the following sections, we derive approximations for (2.5). These approximations will be based on approximating the sequence of r.v. $\{\xi_{0,L}, \xi_{1,L}, \dots, \xi_{M,L}\}$ by a continuous-time random process and subsequently correcting

the obtained approximations for discreteness. Before doing this, we formulate the approximation which is currently the state-of-the-art.

2.2 Glaz Approximation for $P_L(T, h)$

The approximation for the BCP $P_L(T, h)$ developed in Glaz and Johnson (1988), Glaz et al. (2012), Wang and Glaz (2014), and Wang et al. (2014) and discussed in the introduction is as follows.

Approximation 1 (Glaz approximation) For $T \geq 2$,

$$P_L(T, h) \simeq 1 - F_L(2, h) \left[\frac{F_L(2, h)}{F_L(1, h)} \right]^{T-2}, \tag{2.7}$$

where to approximate the first-passage probabilities $F_L(1, h)$ and $F_L(2, h)$, which are $L+1$ and $2L+1$ dimensional integrals respectively, it is advised to use the so-called ‘GenzBretz’ algorithm for numerical evaluation of multivariate normal probabilities; see Genz and Bretz (2009) and Genz et al. (2018).

Unless h is large (say, $h > 3$), Approximation 1 is very accurate. However, its computational cost is also high, especially for large L . Moreover, the main option in the ‘GenzBretz’ package requires the use of Monte-Carlo simulations so that for reliable estimation of high-dimensional integrals one needs to make a lot of averaging; see Sections 6.1 and 7 for more discussion on these issues.

2.3 Continuous-time (diffusion) Approximation

For the purpose of approximating the BCP $P_L(T, h)$, we replace the discrete-time process $\xi_{0,L}, \dots, \xi_{M,L}$ with a continuous process $S(t), t \in [0, T]$, where $T = M/L$ (we will then correct the corresponding first-passage probabilities for discreteness). We do this as follows.

Set $\Delta = 1/L$ and define $t_n = n\Delta \in [0, T], n = 0, 1, \dots, M$. Define a piece-wise linear continuous-time process $S_L(t), t \in [0, T]$:

$$S_L(t) = \frac{1}{\Delta} \left[(t_n - t)\xi_{n-1,L} + (t - t_{n-1})\xi_{n,L} \right] \text{ for } t \in [t_{n-1}, t_n], n = 1, \dots, M.$$

By construction, the process $S_L(t)$ is such that $S_L(t_n) = \xi_{n,L}$ for $n = 0, \dots, M$. Also we have that $S_L(t)$ is a second-order stationary process in the sense that $\mathbb{E} S_L(t), \text{ var}(S_L(t))$ and the autocorrelation function $R^{(L)}(t, t+k\Delta) = \text{Corr}(S_L(t), S_L(t+k\Delta))$ do not depend on t .

Lemma 1 Assume $L \rightarrow \infty$. The limiting process $S(t) = \lim_{L \rightarrow \infty} S_L(t)$, where $t \in [0, T]$, is a Gaussian second-order stationary process with marginal distribution $S(t) \sim N(0, 1)$ for all $t \in [0, T]$ and autocorrelation function $R(t, t+s) = R(s) = \max\{0, 1-|s|\}$.

This lemma is a simple consequence of (2.3).

2.4 Diffusion Approximations: Definition and Their Role in This Study

The above approximation of a discrete-time process $\{\xi_{0,L}, \xi_{1,L}, \dots, \xi_{M,L}\}$ with a continuous process $S(t), t \in [0, T]$, allows us to approximate the BCP $P_L(T, h)$ by a continuous-time analogue as follows.

By the definition of a diffusion approximation, the BCP $P_L(T, h)$ is approximated by

$$P(T, h) := \Pr \left\{ \max_{0 \leq t \leq T} S(t) \geq h \right\}. \tag{2.8}$$

Note that approximating the discrete process of moving sums by a continuous-time process $S(t)$ and subsequent approximation of the BCP $P_L(T, h)$ by $P(T, h)$ is by no means new. This has been done, in particular, in Haiman (1999).

We will call (2.8) and any approximation to (2.8), which does not involve the knowledge of L , ‘diffusion approximation’. These approximations can be greatly improved with the help of the methodology developed by D.Siegmund and adapted to our setup in Section 4. The importance of the discrete-time correction is illustrated by Figs. 1 and 2, where for a fixed h and T we can see a significant difference in values of the BCPs $P_L(T, h)$ for different values of L . As seen from Fig. 2, even for very large $L = 1000$, the discrete-time correction is still needed. Hence we are not recommending to use any approximation for $P(T, h)$ (including rather sophisticated ones like the one developed in Haiman (1999)) as an approximation for $P_L(T, h)$. In the next section we will discuss a diffusion approximation that, after correcting for discrete time, will be a cornerstone for all approximations developed in this paper.

In what follows, it will also be convenient to use the first-passage probability

$$F(T, h) = \Pr \left\{ \max_{0 \leq t \leq T} S(t) < h \right\} = 1 - P(T, h).$$

Since $\xi_{0,L} = S(0) \sim N(0, 1)$, we have $F(0, h) = 1 - P(0, h) = \Phi(h)$.

3 Exact Formulas for the First-passage Probabilities in the Continuous-time Case

3.1 Shepp’s Formulas

Define the conditional first-passage probability

$$F(T, h | x) := \Pr \left\{ S(t) < h \text{ for all } t \in [0, T] \mid S(0) = x \right\}. \tag{3.1}$$

Since $F(T, h | x) = 0$ for $x > h$, for the unconditional first-passage probability $F(T, h)$ we have $F(T, h) = \int_{-\infty}^h F(T, h | x) \varphi(x) dx$.

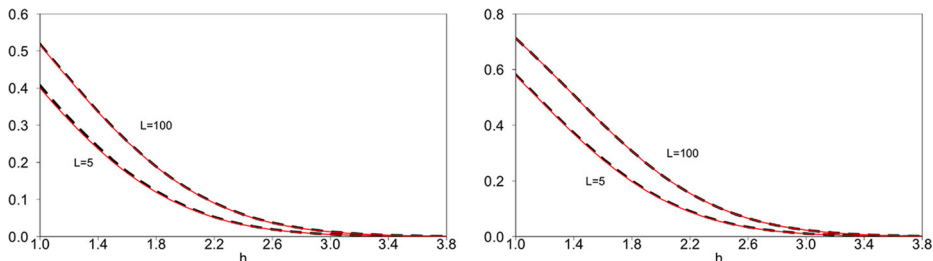


Fig. 1 Empirical probabilities of reaching the barrier h (dashed black) and corresponding versions of Approximation 2 (solid red). Left: $T = 1$ with (a) $L = M = 5$ and (b) $L = M = 100$. Right: $T = 2$ with a $L = 5, M = 10$ and b $L = 100, M = 200$

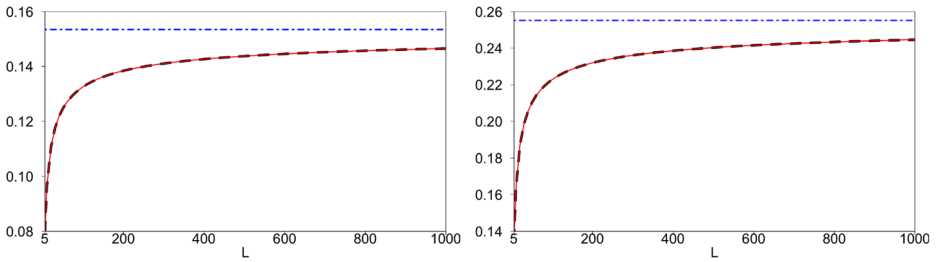


Fig. 2 Empirical probabilities of reaching the barrier $h = 2$ as a function of L (dashed black), uncorrected diffusion approximation $P(T, 2)$ (dot-dashed blue) and corresponding version of $P_L(T, h)$, which is Approximation 2 (solid red). Left: $M = L$ ($T = 1$). Right: $M = 2L$ ($T = 2$)

The result of Shepp (1971, p.949) states that if $T = n$ is a positive integer then

$$F(n, h | x) = \frac{1}{\varphi(x)} \int_{D_x} \det[\varphi(y_i - y_{j+1} + h)]_{i,j=0}^n dy_2 \dots dy_{n+1} \tag{3.2}$$

where $y_0 = 0, y_1 = h - x, D_x = \{y_2, \dots, y_{n+1} | h - x < y_2 < y_3 < \dots < y_{n+1}\}$. For non-integer $T \geq 1$, the exact formula for $F(T, h | x)$ is even more complex (the integral has the dimension $\lceil 2T \rceil$) and completely impractical for computing $P(T, h)$ with $T > 2$, see Shepp ((Shepp 1971), p.950).

For $n = 1$, we obtain

$$F(1, h) = \int_{-\infty}^h \int_{-x-h}^{\infty} \det \begin{bmatrix} \varphi(x) & \varphi(-x_2-h) \\ \varphi(h) & \varphi(-x-x_2) \end{bmatrix} dx_2 dx = \Phi(h)^2 - \varphi(h)[h\Phi(h) + \varphi(h)]. \tag{3.3}$$

For $n = 2$, (3.2) yields

$$F(2, h) = \int_{-\infty}^h \int_{-x-a}^{\infty} \int_{x_2-a}^{\infty} \det \begin{bmatrix} \varphi(x) & \varphi(-x_2-a) & \varphi(-x_3-2a) \\ \varphi(a) & \varphi(-x-x_2) & \varphi(-x-a-x_3) \\ \varphi(x_2+2a+x) & \varphi(a) & \varphi(x_2-x_3) \end{bmatrix} dx_3 dx_2 dx. \tag{3.4}$$

The three-dimensional integral in (3.4) can be reduced to a one-dimensional, see (4.11) below with $h_L = h$.

3.2 An Alternative Representation of the Shepp’s Formula (3.2)

Set $s_i = h + y_i - y_{i+1}$ ($i = 0, 1, \dots, n$) with $s_0 = x, y_0 = 0, y_1 = h - x$. It follows from Shepp’s proof of (3.2) that s_0, s_1, \dots, s_n have the meaning of the values of the process $S(t)$ at the times $t = 0, 1, \dots, n$: $S(i) = s_i$ ($i = 0, 1, \dots, n$). The range of the variables s_i is $(-\infty, h)$. Changing the variables in (3.2), we obtain

$$F(n, h | x) = \frac{1}{\varphi(x)} \int_{-\infty}^h \dots \int_{-\infty}^h \det[\varphi(s_i + a_{i,j})]_{i,j=0}^n ds_1 \dots ds_n, \tag{3.5}$$

where

$$a_{i,j} = y_{i+1} - y_{j+1} = \begin{cases} 0 & \text{for } i = j \\ (i - j)h - s_{j+1} - \dots - s_{i+1} & \text{for } i > j \\ (i - j)h + s_{i+1} + \dots + s_j & \text{for } i < j. \end{cases}$$

3.3 Joint Density for the Values $\{S(i)\}$ and Associated Transition Densities

From (3.5), we obtain the following expression for the joint probability density function for the values $S(0), S(1), \dots, S(n)$ under the condition $S(t) < h$ for all $t \in [0, n]$:

$$p(s_0, s_1, \dots, s_n) = \frac{1}{\varphi(s_0)F(n, h | s_0)} \det[\varphi(s_i + a_{i,j})]_{i,j=0}^n. \tag{3.6}$$

From this formula, we can derive the transition density from $s_0 = x$ to s_n conditionally $S(t) < h, \forall t \in [0, n]$:

$$q_h^{(0,n)}(x \rightarrow s_n) = \frac{1}{\varphi(x)} \int_{-\infty}^h \dots \int_{-\infty}^h \det[\varphi(s_i + a_{i,j})]_{i,j=0}^n ds_1 \dots ds_{n-1}. \tag{3.7}$$

For this transition density, $\int_{-\infty}^h q_h^{(0,n)}(x \rightarrow z) dz = F(n, h | x)$. Moreover, since $S(0) \sim N(0, 1)$, the non-normalized density of $S(n)$ under the condition $S(t) < h$ for all $t \in [0, n]$ is

$$p_h^{(0,n)}(z) := \int_{-\infty}^h q_h^{(0,n)}(x \rightarrow z) \varphi(x) dx \tag{3.8}$$

with $z < h$ and $\int_{-\infty}^h p^{(0,n)}(z) dz = F(n, h)$. In the case $n = 1$, (3.7) gives

$$q_h^{(0,1)}(x \rightarrow z) = \frac{1}{\varphi(x)} \det \begin{pmatrix} \varphi(x) & \varphi(x-h+z) \\ \varphi(h) & \varphi(z) \end{pmatrix} = \varphi(z) \left[1 - e^{-(h-z)(h-x)} \right], \quad z = s_1 < h. \tag{3.9}$$

From this and (3.8) we get

$$p_h^{(0,1)}(z) = \int_{-\infty}^h q_h^{(0,1)}(x \rightarrow z) \varphi(x) dx = \Phi(h)\varphi(z) - \Phi(z)\varphi(h)$$

with $z < h$ and $\int_{-\infty}^h p^{(0,n)}(z) dz = F(1, h)$.

Rather than just recovering the transition density from $s_0 = x$ to s_n , we can also use (3.6) and (3.8) to obtain the transition density from $x = s_j$ to $z = s_n, 0 < j < n$, under the condition $S(t) < h$ for all $t \in [0, n]$:

$$q_h^{(j,n)}(x \rightarrow z) = \frac{1}{p_h^{(0,j)}(z)} \int_{-\infty}^h \dots \int_{-\infty}^h \det[\varphi(s_i + a_{i,j})]_{i,j=0}^n ds_0 ds_1 \dots ds_{j-1} ds_{j+1} \dots ds_{n-1}, \tag{3.10}$$

where $s_j = x$ and $s_n = z$. For $j = 1$ and $n = 2$ we obtain the transition density from $x = s_1$ to $z = s_2$ under the condition $S(t) < h$ for all $t \in [0, 2]$:

$$\begin{aligned} q_h^{(1,2)}(x \rightarrow z) &= \frac{1}{p_h^{(0,1)}(z)} \int_{-\infty}^h \det \begin{pmatrix} \varphi(s_0) & \varphi(s_0-h+x) & \varphi(s_0-2h+x+z) \\ \varphi(h) & \varphi(x) & \varphi(x+z-h) \\ \varphi(2h-x) & \varphi(h) & \varphi(z) \end{pmatrix} ds_0 \\ &= \frac{1}{\Phi(h)\varphi(x) - \Phi(x)\varphi(h)} \det \begin{pmatrix} \Phi(h) & \Phi(x) & \Phi(x+z-h) \\ \varphi(h) & \varphi(x) & \varphi(x+z-h) \\ \varphi(2h-x) & \varphi(h) & \varphi(z) \end{pmatrix}. \end{aligned} \tag{3.11}$$

4 Correcting Shepp's Formula (3.2) for Discrete Time

4.1 Rewriting (3.2) in Terms of the Brownian Motion

Let $W(t)$ be the standard Brownian Motion process on $[0, \infty)$ with $W(0) = 0$ and $\mathbb{E} W(t)W(s) = \min(t, s)$. Recall the conditional probability $F(T, h | x)$ defined in (3.1). Suppose $T \geq 1$ is an integer and define the event

$$\begin{aligned} \Omega &= \{W(t) < W(t + 1) + h < W(t + 2) + 2h < \dots < W(t + T) + Th, \forall 0 \leq t \leq 1\} \\ &= \{W(t) - W(t + 1) < h, \dots, W(t + T - 1) - W(t + T) < h, \forall 0 \leq t \leq 1\}. \end{aligned}$$

If $W(i) = x_i, i = 0, 1, \dots, T + 1$, we obtain from Shepp (1971, p.948)

$$\begin{aligned} F(T, h | x) &= \int \dots \int \Pr\{\Omega \mid W(i) = x_i, i = 0, 1, 2, \dots, T + 1, \\ &\quad W(0) = 0, W(0) - W(1) = x\} \\ &\quad \times \Pr\{W(i) \in dx_i, i = 0, 1, 2, \dots, T + 1, \mid W(0) = 0, W(0) - W(1) = x\}. \end{aligned} \tag{4.1}$$

It follows from the proof of (3.2) that to correct (4.1) for discrete time, one must correct the following probability for discrete time

$$\begin{aligned} &\Pr\{\Omega \mid W(i) = x_i, i = 0, 1, 2, \dots, T + 1, W(0) = 0, W(0) - W(1) = x\} \\ &= \Pr\{\sqrt{2} W_1(t) < h, \dots, \sqrt{2} W_T(t) < h, \forall 0 \leq t \leq 1 \mid W(i) = x_i, i = 0, 1, 2, \dots, T + 1, \\ &\quad W(0) = 0, W(0) - W(1) = x\} \end{aligned} \tag{4.2}$$

where $W_i(t) = \frac{\sqrt{2}}{2} [W(t+i-1) - W(t+i)], i = 1, 2, \dots, T$. Due to the conditioning on the rhs of (4.2), the processes $W_i(t)$ can be treated as independent Brownian motion processes. Therefore, the independent increments of the Brownian motion means correcting formula (3.2) for discrete time is equivalent to correcting the probability $\Pr(\sqrt{2} W(t) < h, \forall 0 \leq t \leq 1)$ for discrete time.

4.2 Discrete-time Correction for the BCP of Cumulative Sums

Let X_1, X_2, \dots be i.i.d. $N(0, 1)$ r.v.'s and set $Y_n = X_1 + X_2 + \dots + X_n$. Consider the sequence of cumulative sums $\{Y_n\}$ and define the stopping time $\tau_{Y,a,b} = \inf\{n \geq 1 : Y_n \geq a + bn\}$ for $a > 0$ and $b \in \mathbb{R}$. Consider the problem of evaluating

$$\Pr(\tau_{Y,a,b} \leq N) = \Pr(Y_n \geq a + bn \text{ for at least one } n \in \{1, 2, \dots, N\}). \tag{4.3}$$

Exact evaluation of (4.3) is difficult even if N is not very large but it was accurately approximated by D.Siegmund see e.g. Siegmund (1986, p.19). Let $W(t)$ be the standard Brownian Motion process on $[0, \infty)$. For $a > 0$ and $b \in \mathbb{R}$, define $\tau_{W,a,b} = \inf\{t : W(t) \geq a + bt\}$ so that

$$\Pr(\tau_{W,a,b} \leq N) = P_W(N, a + bt) := \Pr\{W(t) > a + bt \text{ for at least one } t \in [0, N]\}. \tag{4.4}$$

In Siegmund (1986), (4.4) was used to approximate (4.3) after translating the barrier $a + bt$ by a suitable scalar $\rho \geq 0$. Specifically, the following approximation has been constructed:

$$P(\tau_{Y,a,b} \leq N) \cong P_W(N, (a + \rho) + bt),$$

where the constant ρ approximates the expected excess of the process $\{Y_n\}$ over the barrier $a + bt$. From Siegmund (1985, p. 225)

$$\rho = -\pi^{-1} \int_0^\infty \lambda^{-2} \log\{2(1 - \exp(-\lambda^2/2))/\lambda^2\} d\lambda \simeq 0.582597. \tag{4.5}$$

4.3 Discretised Brownian Motion

Define $\epsilon = 1/L$ and let $t'_n = n\epsilon \in [0, 1], n = 0, 1, \dots, L$. Let X_1, X_2, \dots be i.i.d. $N(0, 1)$ r.v.'s and set $W(t'_n) = \sqrt{\epsilon} \sum_{i=1}^n X_i$. For $a > 0$ define the stopping time

$$\tau_{W,a,b} = \inf\{t'_n : \sqrt{2}W(t'_n) \geq a\} \tag{4.6}$$

and consider the problem of approximating

$$\Pr(\tau_{W,a,b} > 1) = \Pr\left(\sqrt{2}W(t'_n) < a \text{ for all } t'_n \in \{0, \epsilon, \dots, L\epsilon = 1\}\right). \tag{4.7}$$

As $L \rightarrow \infty$, the piecewise linear continuous-time process $W^\epsilon(t), t \in [0, 1]$, defined by:

$$W^\epsilon(t) := \frac{1}{\epsilon} [(t'_n - t)W(t'_{n-1}) + (t - t'_{n-1})W(t'_n)] \text{ for } t \in [t'_{n-1}, t'_n], n = 1, \dots, L,$$

converges to $W(t)$ on $[0, 1]$ as so we can refer to $W(t'_n)$ as discretised Brownian motion. We make the following connection between $\sqrt{2}W(t'_n)$ and the random walk Y_n :

$$\sqrt{2}W(t'_n) = \sqrt{2\epsilon} Y_n = \frac{\sqrt{2}}{\sqrt{L}} Y_n, n = 1, 2, \dots M.$$

Then by using (4.5), we approximate the expected excess over the boundary for the process $\sqrt{2}W(t'_n)$ by

$$\omega_L := \frac{0.82}{\sqrt{L}} \simeq \frac{\sqrt{2}\rho}{\sqrt{L}}.$$

We have deliberately rounded the value $\sqrt{2}\rho \simeq 0.8239\dots$ to 0.82 as for small h and small L it provides marginally better approximation (4.9).

4.4 Corrected Version of (3.2)

Set $h_L = h + \omega_L$. To correct (3.2) for discrete time we substitute the barrier h with h_L . From this and the relation $F(T, h) = \int_{-\infty}^h F(T, h | x)\varphi(x)dx$, the discrete-time corrected form of $F(T, h)$ is

$$\begin{aligned} F(T, h, h_L) &:= \int_{-\infty}^h F(T, h_L | x)\varphi(x)dx \\ &= \int_{-\infty}^h \int_{D_x} \det[\varphi(y_i - y_{j+1} + h_L)]_{i,j=0}^T dy_2 \dots dy_{T+1} dx, \end{aligned} \tag{4.8}$$

where $y_0 = 0, y_1 = h_L - x$, and $D_x = \{y_2, \dots, y_{T+1} | h_L - x < y_2 < y_3 < \dots < y_{T+1}\}$.

4.5 A Generic Approximation Involving Corrected Shepp's Formula

Approximation 2 For integral $T \geq 1$, the discrete-time correction for the BCP (2.5) is

$$P_L(T, h) \cong P(T, h, h_L) := 1 - F_L(T, h, h_L), \tag{4.9}$$

where $F_L(T, h, h_L)$ is given in (4.8).

Whilst Approximation 2 is very accurate (see the next subsection), computation of $P(T, h, h_L)$ requires numerical evaluation of a $T + 1$ dimensional integral which is impractical for large T . To overcome this, in Section 5.2 we develop approximations that can be easily used for any $T > 0$ (which is not necessarily integer).

4.6 Particular Cases: $T = 1$ and $T = 2$

For $T = 1$, evaluation of (4.8) yields

$$F(1, h, h_L) = \Phi(h)\Phi(h_L) - \varphi(h_L)[h\Phi(h) + \varphi(h)]. \tag{4.10}$$

In our previous work (Noonan and Zhigljavsky 2019) we have derived approximations $\hat{P}_L(T, h)$ for the BCP $P_L(T, h)$ with $T \leq 1$. The approximations $\hat{P}_L(T, h)$ developed in Noonan and Zhigljavsky (2019) are also discrete-time corrections of the continuous-time probabilities $P(T, h)$ but they are based almost exclusively on the fact that the process $S(t)$ is conditionally Markov on the interval $t \in [0, 1]$; hence the technique of Noonan and Zhigljavsky (2019) cannot be extended for intervals $t \in [0, T]$ with $T > 1$. The approximation $\hat{P}_L(1, h)$ of Noonan and Zhigljavsky (2019) is different from $P(1, h, h_L) = 1 - F(1, h, h_L)$ of (4.10). It appears that $\hat{P}_L(1, h)$ is more complicated and less accurate approximation than $P(1, h, h_L)$.

For $T = 2$, (4.8) can be expressed (after some manipulations) as follows:

$$\begin{aligned} F(2, h, h_L) = & \frac{\varphi^2(h_L)}{2} \left[(h^2 - 1 + \sqrt{\pi}h)\Phi(h) + (h + \sqrt{\pi})\varphi(h) \right] \\ & - \varphi(h_L) \Phi(h_L) [(h + h_L)\Phi(h) + \varphi(h)] \\ & + \Phi(h) \Phi^2(h_L) + \int_0^\infty \Phi(h - y) \\ & \times \left[\varphi(h_L + y)\Phi(h_L - y) - \sqrt{\pi}\varphi^2(h_L)\Phi(\sqrt{2}y) \right] dy. \end{aligned} \tag{4.11}$$

Only a one-dimensional integral has to be numerically evaluated for computing $F(2, h, h_L)$.

4.7 Simulation Study

In this section, we assess the quality of the approximations (4.10) and (4.11) as well as the sensitivity of the BCP $P_L(T, h)$ to the value of L . In Figs. 1 and 2, the black dashed line corresponds to the empirical values of the BCP $P_L(T, h)$ (for $T = M/L = 1, 2$) computed from 100 000 simulations with different values of L and M (for given L and M , we simulate $L + M$ normal random variables 100 000 times). The solid red line corresponds to Approximation 2. The axis are: the x -axis shows the value of the barrier h in Fig. 1 and value of L in Fig. 2; the y -axis denotes the probabilities of reaching the barrier. The graphs, therefore, show the empirical probabilities of reaching the barrier h (for the dashed line) and values of considered approximations for these probabilities. From these graphs we can conclude that Approximation 2 is very accurate, at least for $T = 1, 2$. We can also conclude that the BCP $P_L(T, h)$ is very sensitive to the value of L . From Fig. 2 we can observe a counter-intuitive fact that even for very high value $L = 1000$, the BCP $P_L(T, h)$ is not even close to $P_\infty(T, h) = P(T, h)$ from (2.8). This may be explained by the fact that for any fixed T and h , the inaccuracy $|P_L(T, h) - P(T, h)|$ decreases with the rate const/\sqrt{L} as $L \rightarrow \infty$.

4.8 The Glaz-Shepp-Siegmund Approximation

Combining (2.7) and the approximation (4.9) for Shepp’s formula (3.2), we arrive at the following approximation to which we suggest the name ‘Glaz-Shepp-Siegmund approximation’.

Approximation 3 For all $T > 0$,

$$P_L(T, h) \simeq 1 - F(2, h, h_L) \cdot \mu_L(h)^{T-2} \text{ with } \mu_L(h) = \frac{F(2, h, h_L)}{F(1, h, h_L)}, \tag{4.12}$$

where $F(1, h, h_L)$ and $F(2, h, h_L)$ are defined in (4.10) and (4.11) respectively.

Approximations 1 and 3 look similar but computing Approximation 1 is very hard and Approximation 3 is very easy (only a one-dimensional integral should be numerically computed).

5 Approximations for the BCP $P_L(T, h)$ Through Eigenvalues of Integral Operators

5.1 Continuous Time: Approximations for $F(T, h)$

Let m be a positive integer, and $q(x \rightarrow z)$ be the transition density $q_h^{(m-1,m)}(x \rightarrow z)$ defined by (3.9) for $m = 1$ (3.11) for $m = 2$ and (3.10) for $m > 2$.

Let us approximate the distributions of the values $s_i = S(i)$ for integral $i > m$ in the following way. Let $p_i(x)$ be the density of $S(i)$ under the condition that $S(t)$ does not reach h for $t \in [0, i]$. By ignoring the past values of $S(t)$ in $[0, i]$, the non-normalized density of $S(i + 1)$ under the conditions that $S(i) \sim p_i(x)$ and $S(t)$ does not reach h for $t \in [i, i + 1]$ is

$$\tilde{p}_{i+1}(x) = \int_{-\infty}^h q_h(x \rightarrow z) p_i(y) dy, \text{ for } x < h. \tag{5.1}$$

We can then define $p_{i+1}(x) = \tilde{p}_i(x)/c_i$, $x < h$, where $c_i = \int_{-\infty}^h \tilde{p}_i(x) dx$. We then replace formula (5.1) with

$$\tilde{p}_i(x) = \int_{-\infty}^h q_h(x \rightarrow z) p(y) dy, \text{ for } x < h, \tag{5.2}$$

where $p(x)$ is an eigenfunction of the integral operator with kernel (3.9) corresponding to the maximum eigenvalue $\lambda_m(h)$:

$$\lambda_m(h) p(x) = \int_{-\infty}^h p(y) q_h^{(m-1,m)}(x \rightarrow z) dy, \text{ } x < h. \tag{5.3}$$

This eigenfunction $p(x)$ is a probability density on $(-\infty, h]$ with $p(x) > 0$ for all $x \in (-\infty, h)$ and $\int_{-\infty}^h p(x) dx = 1$. Moreover, the maximum eigenvalue $\lambda_m(h)$ of the operator with kernel $K(x, y) = q_h^{(m-1,m)}(x \rightarrow z)$ is simple and positive. The fact that such maximum eigenvalue $\lambda_m(h)$ is simple and real (and hence positive) and the eigenfunction $p(x)$ can be chosen as a probability density follows from the Ruelle-Krasnoselskii-Perron-Frobenius theory of bounded linear positive operators, see e.g. Theorem XIII.43 in Reed and Simon (1979).

Using (5.2) and (5.3), we derive recursively: $F(i + 1, h) \simeq F(i, h)\lambda_m(h)$ ($i = m, m + 1, \dots$). By induction, for any integer $T \geq m$ we then have

$$F(T, h) \simeq F(m, h) \cdot [\lambda_m(h)]^{T-m} . \tag{5.4}$$

The approximation (5.4) can be used for any $T > 0$ which is not necessarily an integer. The most important particular cases of (5.4) are with $m = 1$ and $m = 2$. In these two cases, the kernel $q_h^{(m-1,m)}(x \rightarrow z)$ and hence the approximation (5.4) will be corrected for discrete time in the next section.

5.2 Correcting Approximation (5.4) for Discrete Time

To correct the approximation (5.4) for discrete time we need to correct: (a) the first-passage probability $F(m, h)$ and (b) the kernel $q_h^{(m-1,m)}(x \rightarrow z)$. The discrete-time correction of $F(m, h)$ can be done using $F_L(m, h, h_L)$ from (4.8) so that what is left is to correct the kernel $q_h^{(m-1,m)}(x \rightarrow z)$ and hence $\lambda_m(h)$.

5.2.1 Correcting the Transition Kernels for Discrete Time

As explained in Section 4, to make a discrete-time correction in the Shepp’s formula (3.2) we need to replace the barrier h with $h_L = h + \omega_L$ in all places except for the upper bound for the initial value $S(0)$. Therefore, using the notation of Section 3.2, the joint probability density function for the values $S(0), S(1), \dots, S(m)$ under the condition $S(t) < h$ for all $t \in [0, m]$ corrected for discrete time is:

$$\hat{p}(s_0, s_1, \dots, s_m) = \frac{1}{\varphi(s_0)F(m, h | s_0)} \det[\varphi(s_i + \hat{a}_{i,j})]_{i,j=0}^m \tag{5.5}$$

with $-\infty < s_0 < h, -\infty < s_j < h_L$ ($j = 1, \dots, m$),

$$\hat{a}_{i,j} = y_{i+1} - y_{j+1} = \begin{cases} 0 & \text{for } i = j \\ (i - j)h_L - s_{j+1} - \dots - s_{i+1} & \text{for } i > j \\ (i - j)h_L + s_{i+1} + \dots + s_j & \text{for } i < j . \end{cases}$$

This gives us the discrete-time corrected transition density from $s_0 = x$ to s_m conditionally $S(t) < h, \forall t \in [0, m]$:

$$q_{h_L}^{(0,m)}(x \rightarrow s_m) = \frac{1}{\varphi(x)} \int_{-\infty}^{h_L} \dots \int_{-\infty}^{h_L} \det[\varphi(s_i + \hat{a}_{i,j})]_{i,j=0}^m ds_1 \dots ds_{m-1} ; \tag{5.6}$$

which is exactly (3.7) with h_L is substituted for h . In a particular case $m = 1$, the corrected transition density is

$$q_{h_L}^{(0,1)}(x \rightarrow s_1) = \frac{1}{\varphi(x)} \det \begin{pmatrix} \varphi(x) & \varphi(x - h_L + s_1) \\ \varphi(h_L) & \varphi(s_1) \end{pmatrix} = \varphi(s_1) \left[1 - e^{-(h_L - s_1)(h_L - x)} \right] \tag{5.7}$$

with $s_1 < h_L$.

Let us now make the discrete-time correction of the transition density $q_h^{(1,2)}(x \rightarrow z)$. Denote by $p_{h,L}^{(0,1)}(z), z < h$, the non-normalized density of $S(1)$ under the condition $S(t) < h$ for all $t \in [0, 1]$ corrected for discrete time; it satisfies $\int_{-\infty}^h p_{h,L}^{(0,1)}(z) dz = F(1, h, h_L)$. Using (5.7), we obtain

$$p_{h,L}^{(0,1)}(z) = \int_{-\infty}^h q_{h_L}^{(0,1)}(x \rightarrow z)\varphi(x)dx = \varphi(z)\Phi(h) - \varphi(h_L)\Phi(h - h_L + z).$$

From (5.5) and (5.7), the transition density from $x = s_1$ to $z = s_2$ under the condition $S(t) < h$ for all $t \in [0, 2]$ corrected for discrete time (the corrected form of (3.11)) is given by

$$\begin{aligned}
 q_{h,L}^{(1,2)}(x \rightarrow z) &= \frac{1}{p_{h,L}^{(0,1)}(x)} \int_{-\infty}^h \det \begin{pmatrix} \varphi(s_0) & \varphi(s_0 - h_L + x) & \varphi(s_0 - 2h_L + x + z) \\ \varphi(h_L) & \varphi(x) & \varphi(x + z - h_L) \\ \varphi(2h_L - x) & \varphi(h_L) & \varphi(z) \end{pmatrix} ds_0. \\
 &= \frac{1}{p_{h,L}^{(0,1)}(x)} \det \begin{pmatrix} \Phi(h) & \Phi(h - h_L + x) & \Phi(h - 2h_L + x + z) \\ \varphi(h_L) & \varphi(x) & \varphi(x + z - h_L) \\ \varphi(2h_L - x) & \varphi(h_L) & \varphi(z) \end{pmatrix} \quad (5.8)
 \end{aligned}$$

Unlike the transition density (5.6) (and (5.7) in the particular case $m = 1$), which only depends on h_L and not on h , the transition density $q_{h,L}^{(1,2)}(x \rightarrow z)$ depends on both h and h_L and hence the notation. The dependence on h has appeared from integration over the $s_0 \in (-\infty, h)$.

5.2.2 Approximations for the BCP $P_L(T, h)$

With discrete-time corrected transition densities $q_h^{(0,1)}(x \rightarrow z)$ and $q_h^{(1,2)}(x \rightarrow z)$, we obtain the corrected versions of the approximations (5.4).

Approximation 4 $P_L(T, h) \simeq 1 - F(1, h, h_L) \cdot [\lambda_{L,1}(h)]^{T-1}$, where $F(1, h, h_L)$ is given in (4.10) and $\lambda_{L,1}(h)$ is the maximal eigenvalue of the integral operator with kernel $K(x, z) = q_{h_L}^{(0,1)}(x \rightarrow z)$ defined in (5.7).

Approximation 5 $P_L(T, h) \simeq 1 - F(2, h, h_L) \cdot [\lambda_{L,2}(h)]^{T-2}$, where $F(2, h, h_L)$ is given in (4.11) and $\lambda_{L,2}(h)$ is the maximal eigenvalue of the integral operator with kernel $K(x, z) = q_{h,L}^{(1,2)}(x \rightarrow z)$ defined in (5.8).

Similarly to $\lambda_m(h)$ from (5.3), the maximum eigenvalues $\lambda_{L,1}(h)$ and $\lambda_{L,2}(h)$ of the operators with kernels $K(x, z) = q_{h_L}^{(0,1)}(x \rightarrow z)$ and $K(x, z) = q_{h,L}^{(1,2)}(x \rightarrow z)$ are simple and positive; the corresponding eigenfunctions $p(x)$ can be chosen as probability densities. Both approximations can be used for any $T > 0$.

In numerical examples below we approximate the eigenvalues $\lambda_{L,k}(h)$ ($k = 1, 2$) using the methodology described in Mohamed and Delves (1985), p.154. This methodology is based on the Gauss-Legendre discretization of the interval $[-c, h]$, with some large $c > 0$, into an N -point set x_1, \dots, x_N (the x_i 's are the roots of the N -th Legendre polynomial on $[-c, h]$), and the use of the Gauss-Legendre weights w_i associated with points x_i ; $\lambda_{L,k}(h)$ and $p(x)$ are then approximated by the largest eigenvalue and associated eigenvector of the matrix $D^{1/2}AD^{1/2}$, where $D = \text{diag}(w_i)$ and $A_{i,j} = K(x_i, x_j)$ with the respective kernel $K(x, z)$. If N is large enough then the resulting approximation to $\lambda_{L,k}(h)$ is arbitrarily accurate. With modern software, computing Approximations 4 and 5 (as well as Approximation 3) with high accuracy takes only milliseconds on a regular laptop.

As discussed in the next section, Approximation 5 is more accurate than Approximation 4, especially for small h ; the accuracies of Approximations 3 and 5 are very similar. Note also that a version of Approximation 4 has been developed in our previous work (Noonan

Table 1 Values of $\lambda_{L,1}(h)$, $\lambda_{L,2}(h)$ and $\mu_L(h)$ with $L = 20$ for different h

	$h=0$	$h=0.5$	$h=1$	$h=1.5$	$h=2$	$h=2.5$	$h=3$	$h=3.5$	$h=4$
$\lambda_{L,1}(h)$	0.28494	0.46443	0.65331	0.81186	0.91687	0.97090	0.99209	0.99835	0.99974
$\lambda_{L,2}(h)$	0.25744	0.43811	0.63472	0.80239	0.91348	0.97005	0.99195	0.99833	0.99974
$\mu_L(h)$	0.25527	0.43677	0.63432	0.80241	0.91353	0.97007	0.99195	0.99833	0.99974

and Zhigljavsky 2019); this version was based on a different discrete-time approximation (discussed in Section 4.6) of the continuous-time BCP probability $P(T, h)$.

6 Simulation Study

6.1 Accuracy of Approximations for the BCP $P_L(T, h)$

In this section we study the quality of Approximations 4 and 5 for the BCP $P_L(T, h)$ defined in (2.5). Approximation 3 is visually indistinguishable from Approximation 5 and is therefore not plotted (see Table 1). Without loss of generality, ε_j in (1.1) are normal r.v.'s with mean 0 and variance 1. The style of Fig. 3 is exactly the same as of Fig. 1 and is described in the beginning of Section 4.7. In Fig. 3, the dashed green line corresponds to Approximation 4 and the solid red line corresponds to Approximation 5.

From Fig. 3 we see that the performance of Approximations 4 and 5 is very strong even for small L . For small h , Approximation 5 is more precise than Approximation 4 in view of its better accommodation to the non-Markovian nature of the process $S(t)$.

In Table 1, we display the values of $\lambda_{L,1}(h)$, $\lambda_{L,2}(h)$ and $\mu_L(h)$ with $L = 20$ for a number of different h . From this table, we see only a small difference between $\lambda_{L,2}(h)$ and $\mu_L(h)$; this difference is too small to visually differentiate between Approximations 3 and 5 in Fig. 3.

In Tables 2, 3 and 4 we numerically compare the performance of Approximations 1 and 3 for approximating $P_L(T, h)$ across different values of L and h . Since Approximation 1 relies on Monte-Carlo methods, we present the average over 100 evaluations and denote this by \bar{x} . We have also provided values for the standard deviation and maximum and minimum of the 100 runs to illustrate the randomised nature of this approximation. These are denoted by s , $Max(x_i)$ and $Min(x_i)$ respectively. The values of $P_L(T, h)$ presented in the

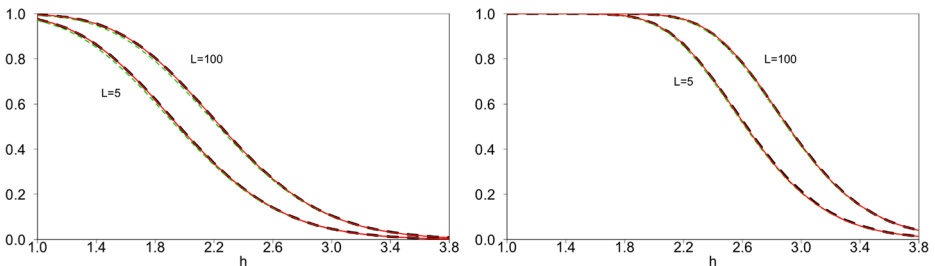


Fig. 3 Empirical probabilities of reaching the barrier h (dashed black), Approximation 4 (dashed green) and Approximation 5 (solid red). Left: $T = 10$ with **a** $L = 5$ and **b** $L = 100$. Right: $T = 50$ with **(a)** $L = 5$ and **(b)** $L = 100$

Table 2 Average values from 100 evaluations of Approximation 1 for different h along with maximum and minimum with $L = 5$ and $T = 100$

	$h=2.5$	$h=2.75$	$h=3$	$h=3.25$	$h=3.5$	$h=3.75$	$h=4$
\bar{x}	0.855957	0.627299	0.376337	0.191122	0.086253	0.033769	0.013156
s	0.004127	0.008588	0.013805	0.015181	0.012826	0.008510	0.005131
$Max(x_i) - \bar{x}$	0.010665	0.023748	0.029819	0.027066	0.025629	0.016208	0.011609
$\bar{x} - Min(x_i)$	0.012176	0.021268	0.033211	0.041322	0.041350	0.022650	0.018146
Approximation 3	0.854844	0.625113	0.373863	0.188933	0.083981	0.033833	0.012551
$P_L(T, h)$	0.855429	0.627463	0.376681	0.191625	0.085697	0.034675	0.013116

tables below are the empirical probabilities of reaching the barrier h obtained by 10^6 simulations. We have not included Approximation 5 in these tables as results are identical to Approximation 3 up to four decimal places.

From Tables 2, 3 and 4 we see that with this choice of $T = 100$, the errors of approximating $F_L(2, h)$ and $F_L(1, h)$ via the ‘GenzBretz’ algorithm can accumulate and lead to a fairly significant variation of Approximation 1. This demonstrates the need to average the outcomes of Approximation 1 over a significant number of runs, should one desire an accurate approximation. This may require rather high computational cost and run time, especially if L is large. On the other hand, evaluation of Approximation 3 is practically instantaneous for all L . Even for a very small choice of $L = 5$, Table 2 shows that Approximation 3 still remains very accurate. As L increases from 5 to 20, Table 3 shows that the accuracy of Approximation 3 increases. The averaged Approximation 1 is also very accurate but a larger L appears to produce a larger range for $Max(x_i)$ and $Min(x_i)$ when h is large; this is seen in Table 4. Note we have not included empirical values of $P_L(T, h)$ in Table 4 due to the large computational cost.

6.2 Approximation for the BCP in the Case of Non-normal Moving Sums

Approximations 3, 4 and 5 remain very accurate when then the original ε_i in (1.1) are not exactly normal. We consider two cases: (a) ε_i are uniform r.v’s on $[0,1]$ and (b) ε_i are Laplace r.v’s with mean zero and scale parameter 1. Simulation results are shown in Fig. 4; this figure has the same style as figures in Sections 4.7 and 6.1.

Some selected values used for plots in Fig. 4 are:
 $h = 2, L = 20$: Emp: 0.6045 ± 0.0030 (0.6123 ± 0.0030) [0.5894 ± 0.003]; Ap. 4(5): $0.5921(0.6054)$;

Table 3 Average values from 100 evaluations of Approximation 1 for different h along with maximum and minimum with $L = 20$ and $T = 100$

	$h=2.5$	$h=2.75$	$h=3$	$h=3.25$	$h=3.5$	$h=3.75$	$h=4$
\bar{x}	0.952007	0.802073	0.554613	0.315085	0.155331	0.066113	0.025608
s	0.001479	0.004856	0.012540	0.015050	0.015160	0.011647	0.008129
$Max(x_i) - \bar{x}$	0.004746	0.013360	0.027078	0.030940	0.033991	0.024111	0.030014
$\bar{x} - Min(x_i)$	0.003662	0.010894	0.031463	0.037715	0.041021	0.043283	0.016997
Approximation 3	0.952475	0.802100	0.555109	0.316076	0.153803	0.066438	0.026143
$P_L(T, h)$	0.952818	0.803078	0.555530	0.315784	0.153446	0.066642	0.026244

Table 4 Average values from 100 evaluations of Approximation 1 for different h along with maximum and minimum with $L = 100$ and $T = 100$

	$h=2.5$	$h=2.75$	$h=3$	$h=3.25$	$h=3.5$	$h=3.75$	$h=4$
\bar{x}	0.979027	0.878031	0.661247	0.402887	0.211894	0.093329	0.039110
s	0.000884	0.005502	0.014418	0.021283	0.018493	0.020459	0.015536
$Max(x_i) - \bar{x}$	0.001995	0.009243	0.039695	0.040615	0.063578	0.064306	0.037958
$\bar{x} - Min(x_i)$	0.002414	0.020613	0.025530	0.093876	0.038484	0.05694	0.033748
Approximation 3	0.979119	0.878481	0.660662	0.405674	0.209313	0.094517	0.038529

$h = 2, L = 100$: Emp: 0.6771 ± 0.0029 (0.6801 ± 0.0029) [0.6722 ± 0.003]; Ap. 4(5): $0.6633(0.6775)$;
 $h = 3, L = 20$: Emp: 0.0788 ± 0.0017 (0.0710 ± 0.0016) [0.0915 ± 0.002]; Ap. 4(5): $0.0777(0.0789)$;
 $h = 3, L = 100$: Emp: 0.1039 ± 0.0019 (0.1033 ± 0.0019) [0.1048 ± 0.002]; Ap. 4(5): $0.1022(0.1034)$.

Here we provided means and 95% confidence intervals for the empirical (Emp) values of the BCP $P_L(T, h)$ (with $T = M/L = 10$) computed from 100 000 Monte-Carlo runs of the sequences of the moving sums (1.1) with normal (no brackets), uniform (regular brackets) and Laplace (square brackets) distributions for ε_i in (1.1). Values of Approximations (Ap.) 4 and 5 are also given.

From Fig. 4 and associated numbers we can make the following conclusions: (a) the BCP $P_L(T, h)$ for the case where ε_i in (1.1) are uniform is closer to the case where ε_i are normal, than for the case where ε_i have Laplace distribution; (b) as L increases, the probabilities $P_L(T, h)$ in the cases of uniform and Laplace distributions of ε_i become closer to the BCP for the case of normal ε_i and hence the approximations to the BCP become more precise; (c) accuracy of Approximation 5 is excellent for the case of normal ε_i and remains very good in the case of uniform ε_i ; it is also rather good in the case when ε_i have Laplace distribution; (d) Approximation 4 is slightly less accurate than Approximation 5 (and Approximation 3) for the case of normal and uniform ε_i (this is in a full agreement with discussions in Sections 5.2.2 and 6.1); however, Approximation 4 is very simple and can still be considered as rather accurate.

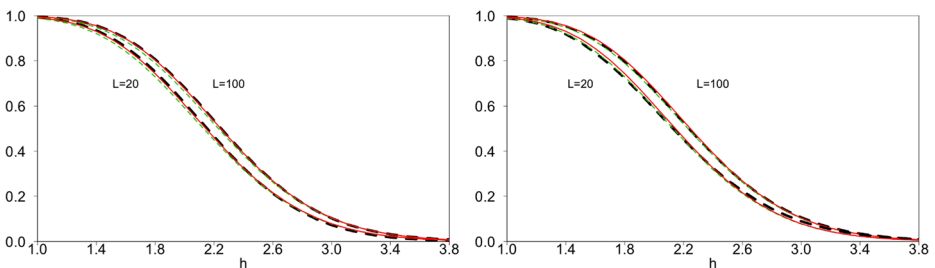


Fig. 4 Empirical probabilities of reaching the barrier h (dashed black), Approximation 4 (dashed green) and Approximation 5 (solid red). Left: $\varepsilon_i \sim \text{Uniform}[0, 1]$ and $T = 10$ with (a) $L = 20$ and (b) $L = 100$. Right: $\varepsilon_i \sim \text{Laplace}[0, 1]$ and $T = 10$ with (a) $L = 20$ and (b) $L = 100$

6.3 Approximation for the BCP in the Case of Moving Weighted Sums

We have also investigated the performance of Approximation 5 (and 3) after introducing particular weights into (1.1). We explored the following two ways of incorporating weights:

- (i) L random weights w_1, w_2, \dots, w_L , with w_i i.i.d. uniform on $[0, 2]$, are associated with a position in the moving window; this results in the moving weighted sum

$$S_{n,w,L} := \sum_{j=n+1}^{n+L} w_{j-n} \varepsilon_j \quad (n = 0, 1, \dots, M);$$

- (ii) $M + L$ random weights w_1, \dots, w_{M+L} are associated with r.v. $\varepsilon_1, \dots, \varepsilon_{M+L}$; here w_j are i.i.d. uniform r.v.'s on $[0,2]$; this gives the moving weighted sum

$$S_{n,w,L} := \sum_{j=n+1}^{n+L} w_j \varepsilon_j \quad (n = 0, 1, \dots, M).$$

Simulations results are shown in Fig. 5. In both cases, we have repeated simulations 1,000 times and plotted all the curves representing the BCP as functions of h in grey colour and Approximation 5 for the BCP for the non-weighted case (when all weights $w_j = 1$) as red dashed line. We can see that for both scenarios the Approximation 5 for the BCP in the non-weighted case gives fairly accurate approximation for the weighted BCP. Similar results have been observed for other values of L and T .

7 Approximating Average Run Length (ARL)

In this section, we provide approximations to the probability distribution of the moment of time $\tau_H(\mathbb{S}) := \min\{n \geq 0 : S_{n,L} \geq H\}$ when the sequence $\mathbb{S} = \{S_{0,L}, S_{1,L}, \dots\}$ reaches the threshold H for the first time. Note that $\tau_H(\mathbb{S}) = \tau_h(\mathbb{X})$, where $\tau_h(\mathbb{X}) := \min\{n \geq 0 : \xi_{n,L} \geq h\}$ and $\mathbb{X} = \{\xi_{0,L}, \xi_{1,L}, \dots\}$. The BCP $\mathcal{P}_{\mathbb{S}}(M, H, L)$, considered as a function of M , is the c.d.f. of this probability distribution: $\mathcal{P}_{\mathbb{S}}(M, H, L) = \Pr(\tau_H(\mathbb{S}) \leq M)$. The average run length (ARL) until \mathbb{S} reaches H for the first time is

$$ARL_H(\mathbb{S}) := \sum_{n=0}^{\infty} n \Pr\{\tau_H = n\} = \int_0^{\infty} M d\mathcal{P}_{\mathbb{S}}(M, H, L). \tag{7.1}$$

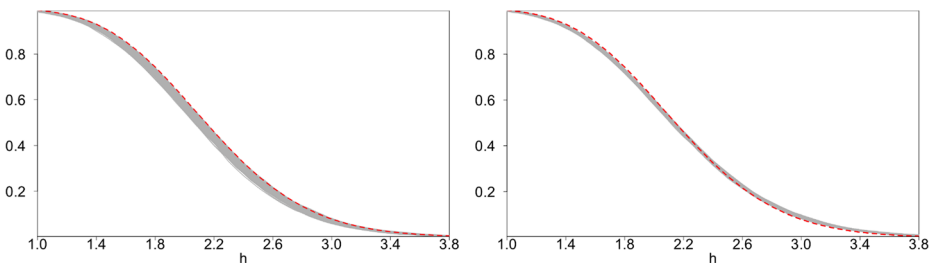


Fig. 5 BCP for the weighted sums (grey) against Approximation 5 for the BCP for non-weighted moving sums (red dotted line). Left: case (i) with $L = 20, M = 200, T = 10$. Right: case (ii) with $L = 20, M = 200, T = 10$

Note that $ARL_H(\mathbb{S}) = ARL_h(\mathbb{X})$. The diffusion approximation to the time moment $\tau_h(\mathbb{X})$ is $\tau_h(S(t)) := \min\{t \geq 0 : S(t) \geq h\}$, which is the time moment when the process $S(t)$ reaches h . The distribution of $\tau_h(S(t))$ has the form:

$$(1 - \Phi(h))\delta_0(ds) + q(s, h, S(t))ds, s \geq 0,$$

where $\delta_0(ds)$ is the delta-measure concentrated at 0 and

$$q(s, h, S(t)) = \frac{d}{ds}P(s, h), \quad 0 < s < \infty. \tag{7.2}$$

The function $q(s, h, S(t))/\Phi(h)$, considered as a function of s , is a probability density function on $(0, \infty)$ since

$$\int_0^\infty q(s, h, S(t))ds = 1 - P(0, h) = \Phi(h).$$

From this, the diffusion approximation for $ARL_H(\mathbb{X})/L$ is

$$ARL_h(S(t)) = \mathbb{E}(\tau_h(S(t))) = \int_0^\infty s q(s, h, S(t))ds. \tag{7.3}$$

The diffusion approximation (7.3) should be corrected for discrete time; otherwise it is poor, especially for small L . As shown in Section 6, Approximations 3 and 5 are very accurate approximations for $P_L(T, h)$ and can be used for all $T > 0$. We shall use Approximation 3 to formulate our approximations but note that the use of Approximation 5 would give very similar results.

We define the approximation $\hat{q}(s, h)$ for the probability density function of $\tau_h(\mathbb{X})/L$ by

$$\hat{q}(s, h) = \frac{d}{ds} \left\{ 1 - F(2, h, h_L) \cdot \mu_L(h)^{s-2} \right\} = -F(2, h, h_L) \log(\mu_L(h)) \cdot \mu_L(h)^{s-2}, \quad s > 0.$$

The corresponding approximation for $ARL_h(\mathbb{X})$ is

$$ARL_h(\mathbb{X}) = \mathbb{E} \tau_h(\mathbb{X}) \cong L \int_0^\infty s \hat{q}(s, h) ds = -\frac{L \cdot F(2, h, h_L)}{\mu_L(h)^2 \log(\mu_L(h))}. \tag{7.4}$$

The standard deviation of $\tau_h(\mathbb{X})$, denoted $SD(\tau_h(\mathbb{X}))$, is approximated by:

$$SD(\tau_h(\mathbb{X})) \cong L \left[\int_0^\infty s^2 \hat{q}(s, h) ds - \left(\int_0^\infty s \hat{q}(s, h) ds \right)^2 \right]^{1/2}. \tag{7.5}$$

In this paper, we define ARL in terms of the number of random variables $\xi_{n,L}$ rather than number of random variables ε_j . This means we have to modify the approximation for ARL of Glaz et al. (2012) by subtracting L . The standard deviation approximation in Glaz et al. (2012) is not altered.

Table 5 Approximations for $ARL_h(\mathbb{X})$ and $SD(\tau_h(\mathbb{X}))$ with $L = 10$

h	2	2.25	2.5	2.75	3	3.25	3.5
(7.4)	126	217	395	759	1551	3375	7837
(7.6)	126 ± 1	218 ± 2	394 ± 5	756 ± 17	1545 ± 65	3388 ± 300	7791 ± 1100
$ARL_h(\mathbb{X})$	127	218	396	757	1550	3344	7721
(7.5)	129	220	397	761	1553	3377	7839
(7.7)	129 ± 1	220 ± 2	397 ± 5	758 ± 17	1549 ± 65	3389 ± 300	7793 ± 1100
$SD(\tau_h(\mathbb{X}))$	129	221	395	758	1550	3341	7716

Table 6 Approximations for $ARL_h(\mathbb{X})$ and $SD(\tau_h(\mathbb{X}))$ with $L = 50$

h	2	2.25	2.5	2.75	3	3.25	3.5
(7.4)	471	791	1392	2587	5099	10695	23918
(7.6)	471 ± 3	791 ± 7	1393 ± 25	2597 ± 75	5101 ± 270	10708 ± 1250	24639 ± 5800
$ARL_h(\mathbb{X})$	472	792	1397	2588	5085	10749	24131
(7.5)	485	804	1404	2598	5109	10704	23924
(7.7)	481 ± 3	802 ± 7	1404 ± 25	2608 ± 75	5147 ± 270	10716 ± 1250	24649 ± 5800
$SD(\tau_h(\mathbb{X}))$	485	804	1407	2600	5093	10762	24105

The Glaz approximations for $ARL_h(\mathbb{X})$ and $SD(\tau_h(\mathbb{X}))$ are as follows:

$$\mathbb{E}_G(\tau_h(\mathbb{X})) = \sum_{j=L}^{2L} F_L((j/L - 1), h) + \frac{F_L(1, h)}{F_L(1, h) - F_L(2, h)} \sum_{j=1}^L (F_L(1 + j/L, h)), \quad (7.6)$$

$$SD_G(\tau_h(\mathbb{X})) = \left[L(L - 1) + 2 \sum_{j=L}^{3L} j(F_L(j/L - 1, h)) + \frac{2Lx(3 - 2x)}{(1 - x)^2} \sum_{j=1}^L F_L(1 + j/L, h) + \frac{2x}{1 - x} \sum_{j=1}^L j(F_L(1 + j/L, h)) + \mathbb{E}_G(\tau_h(\mathbb{X})) - \mathbb{E}_G(\tau_h(\mathbb{X}))^2 \right]^{1/2}, \quad (7.7)$$

where $x = F_L(2, h)/F_L(1, h)$.

In Tables 5 and 6 we assess the accuracy of the approximations (7.4) and (7.5) and also Glaz approximations (7.6) and (7.7). In these tables, the values of $ARL_h(\mathbb{X})$ and $SD(\tau_h(\mathbb{X}))$ have been calculated using 100,000 simulations. Since the Glaz approximations rely on Monte Carlo methods, in the tables we have reported value $2s$ -confidence intervals computed from 150 evaluations.

Tables 5 and 6 show that the approximations developed in this paper perform strongly and are similar, for small or moderate h , to the Glaz approximations. For $h \geq 3$, the Glaz approximation produces rather large uncertainty intervals and the uncertainty quickly deteriorates with the increase of h . This is due to the fairly large uncertainty intervals formed by Approximation 1 when approximating $P_L(T, h)$ with large h and hence small $P_L(T, h)$, as discussed in Section 6.1. The approximations developed in this paper are deterministic and are much simpler in comparison to the Glaz approximations. Moreover, they do not deteriorate for large h .

Acknowledgment The authors are grateful to the referees for careful reading of the manuscript and useful comments.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Bauer P, Hackl P (1980) An extension of the MOSUM technique for quality control. *Technometrics* 22(1): 1–7
- Chu CSJ, Hornik K, Kaun CM (1995) MOSUM Tests for parameter constancy. *Biometrika* 82(3):603–617
- Eiauer P, Hackl P (1978) The use of MOSUMS for quality control. *Technometrics* 20(4):431–436
- Genz A, Bretz F (2009) Computation of Multivariate Normal and t Probabilities. *Lecture Notes in Statistics*. Springer, Heidelberg
- Genz A, Bretz F, Miwa T, Mi X, Leisch F, Scheipl F, Hothorn T (2018) mvtnorm: Multivariate Normal and t Distributions. <https://CRAN.R-project.org/package=mvtnorm>. R package version 1.0-8
- Glaz J, Johnson B (1988) Boundary crossing for moving sums. *J Appl Probab* 25(1):81–88
- Glaz J, Naus J, Wang X (2012) Approximations and inequalities for moving sums. *Methodol Comput Appl Probab* 14(3):597–616
- Glaz J, Naus JI (1991) Tight bounds and approximations for scan statistic probabilities for discrete data. *Ann Appl Probab* 1(2):306–318
- Glaz J, Naus JI, Wallenstein S, Wallenstein S, Naus JI (2001) *Scan Statistics*. Springer
- Glaz J, Pozdnyakov V, Wallenstein S (2009) *Scan Statistics: Methods and Applications*. Birkhäuser, Boston
- Haiman G (1999) First passage time for some stationary processes. *Stoch Process Appl* 80(2):231–248
- Mohamed J, Delves L (1985) *Computational methods for integral equations*. Cambridge University Press
- Moskvina V, Zhigljavsky A (2003) An algorithm based on Singular Spectrum Analysis for change-point detection. *Commun Stat Simul Comput* 32(2):319–352
- Noonan J, Zhigljavsky A (2019) Approximations for the boundary crossing probabilities of moving sums of normal random variables. *Communications in Statistics-Simulation and Computation*, 1–22
- Reed M, Simon B (1979) *Methods of Modern Mathematical Physics: Scattering theory*. Vol. 3 Academic Press
- Shepp L (1971) First passage time for a particular Gaussian process. *Ann Math Stat* 42(3):946–951
- Siegmund D (1985) *Sequential analysis: Tests and confidence intervals*. Springer Science & Business Media
- Siegmund D (1986) Boundary crossing probabilities and statistical applications. *Ann Stat* 14(2):361–404
- Slepian D (1961) First passage time for a particular Gaussian process. *Ann Math Stat* 32(2):610–612
- Waldmann KH (1986) Bounds to the distribution of the run length in general quality-control schemes. *Statistische Hefte* 27(1):37
- Wang X, Glaz J (2014) Variable window scan statistics for normal data. *Commun Stat Theory Meth* 43(10–12):2489–2504
- Wang X, Zhao B, Glaz J (2014) A multiple window scan statistic for time series models. *Stat Probab Lett* 94:196–203
- Xia Z, Guo P, Zhao W (2009) Monitoring structural changes in generalized linear models. *Commun Stat Theory Meth* 38(11):1927–1947

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.