

Controlling the Time Discretization Bias for the Supremum of Brownian Motion

KRZYSZTOF BISEWSKI, CWI Amsterdam

DAAN CROMMELIN, CWI Amsterdam, KdV Institute for Mathematics, University of Amsterdam

MICHEL MANDJES, KdV Institute for Mathematics, University of Amsterdam

We consider the bias arising from time discretization when estimating the threshold crossing probability $w(b) := \mathbb{P}(\sup_{t \in [0,1]} B_t > b)$, with $(B_t)_{t \in [0,1]}$ a standard Brownian Motion. We prove that if the discretization is equidistant, then to reach a given target value of the relative bias, the number of grid points has to grow quadratically in b , as b grows. When considering non-equidistant discretizations (with threshold-dependent grid points), we can substantially improve on this: we show that for such grids the required number of grid points is independent of b , and in addition we point out how they can be used to construct a strongly efficient algorithm for the estimation of $w(b)$. Finally, we show how to apply the resulting algorithm for a broad class of stochastic processes; it is empirically shown that the threshold-dependent grid significantly outperforms its equidistant counterpart.

CCS Concepts: • **Computing methodologies** → **Rare-event simulation**; *Discrete-event simulation*;

Additional Key Words and Phrases: Rare-event simulation, hitting time, discretization error, optimal discretization, Brownian Motion, continuity correction

ACM Reference format:

Krzysztof Bisewski, Daan Crommelin, and Michel Mandjes. 2018. Controlling the Time Discretization Bias for the Supremum of Brownian Motion. *ACM Trans. Model. Comput. Simul.* 28, 3, Article 24 (July 2018), 25 pages. <https://doi.org/10.1145/3177775>

1 INTRODUCTION

Extreme values of random processes play a prominent role in a broad range of practical problems. It is often of interest to find the tail of the distribution of the supremum of a continuous-time stochastic process $(X_t)_{t \geq 0}$ over a finite time interval. In this article, the focus is on the level crossing probability

$$w(b) := \mathbb{P}\left(\sup_{t \in [0,1]} X_t > b\right).$$

This work is part of the research programme “Mathematics of Planet Earth” which is funded by the Netherlands Organisation for Scientific Research (NWO). Michel Mandjes’ research is partly funded by the NWO Gravitation Programme NETWORKS, grant number 024.002.003.

Authors’ addresses: K. Bisewski, CWI Amsterdam, Science Park 123, 1098 XG, Amsterdam, Netherlands; email: bisewski@cwi.nl; D. Crommelin, CWI Amsterdam, KdV Institute for Mathematics, University of Amsterdam, Science Park 123, 1098 XG, Amsterdam, Netherlands; email: daan.crommelin@cwi.nl; M. Mandjes, KdV Institute for Mathematics, University of Amsterdam, Science Park 105-107, 1098 XG, Amsterdam, Netherlands; email: m.r.h.mandjes@uva.nl.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2018 ACM 1049-3301/2018/07-ART24 \$15.00

<https://doi.org/10.1145/3177775>

For many classes of processes, such as the Gaussian processes (Adler 1990), typically no explicit expressions for $w(b)$ are available, with Brownian Motion and the Ornstein-Uhlenbeck process being notable exceptions. When an explicit expression for $w(b)$ is unavailable, one usually resorts to using high-dimensional numerical integration and simulation-based methods (see, e.g., Genz and Bretz (2009) for further reading).

For most of the available numerical methods, the underlying continuous-time process needs to be discretized in time. One chooses a certain *finite grid* $T \subset [0, 1]$ and then approximates $w(b)$ with $w_T(b) := \mathbb{P}(\sup_{t \in T} X_t > b)$. We note that this always leads to an underestimation, i.e., $w_T(b) \leq w(b)$. We quantify this underestimation by $\beta_T(b) := (w(b) - w_T(b))/w(b)$, the relative *discretization bias*.¹ Typically, T is chosen to be an *equidistant grid* $T = \{\frac{1}{n}, \frac{2}{n}, \dots, 1\}$ and in that case, $\beta_T(b)$ can be reduced only by changing the *grid size* n . The finer the grid, the smaller the bias, but also, the larger the computational effort to estimate $w_T(b)$. The main drawback of using equidistant grids is that typically, to reach a given target value of the discretization bias, the grid size n has to grow with the threshold b . In that case, for large b , the appropriate grid size can become so large that the computation is not feasible. Two central questions arise from these observations: How fast does n have to grow in b ? Furthermore, can we identify a more efficient family of grids?

In this article, we address these issues for standard Brownian Motion. Although in this case $w(b)$ can be computed explicitly, there are no available expressions for $\beta_T(b)$. We conduct a thorough study of the influence of the choice of the grid on the corresponding relative bias. Furthermore, we argue that exploring the case of standard Brownian Motion is a first step toward finding efficient grids for a more general class of processes. We demonstrate numerically how our analysis of efficient grids for Brownian Motion leads to a useful procedure to determine efficient grids for a broad range of other processes.

The contributions of this article are the following. (i) The first finding can be seen as a negative result: we show that to *uniformly control*² the relative bias, the size n of the equidistant grid must grow at least quadratically in b (see Theorem 1 in Section 3). (ii) The second finding is that we can do much better by using a *threshold-dependent* family of grids, meaning that grid points change their location with b (but the number of points does not increase). The discretization bias induced by this particular family of grids is uniformly controlled without having to increase the number of grid points (see Theorem 2 in Section 4). To the best of the authors' knowledge, this is the first result which shows that a careful choice of the grid can drastically increase the accuracy of the discrete estimator of $w(b)$. Using threshold-dependent grids makes it feasible to estimate $w(b)$ with moderate grid sizes even for very high thresholds b , which would be impossible to estimate using equidistant grids. In particular, in Section 5 we present a strongly efficient algorithm for the estimation of $w(b)$ that relies on threshold-dependent grids. (iii) In the third place, we point out how the ideas underlying our threshold-dependent grid can be used for a broad class of stochastic processes (including Gaussian processes, such as fractional Brownian Motion, and Lévy processes); it is empirically shown that the threshold-dependent grid significantly outperforms its equidistant counterpart.

An efficient grid (both small in size and inducing a small discretization bias) is particularly relevant for situations with large b . In this respect, the work presented here connects to the rare-event simulation literature. As b approaches infinity, $w(b)$ decays exponentially to 0 and standard simulation-based methods like Crude Monte Carlo to estimate $w(b)$ become extremely

¹As $b \rightarrow \infty$, both $w_T(b)$ and $w(b)$ tend to 0, so that the *absolute* bias is not a meaningful accuracy measure.

²In this context, *uniform control* means that for a fixed $\varepsilon > 0$, we have that $\beta_T(b) < \varepsilon$ for all $b > 0$; the grid T can change in b .

time-consuming. We emphasize that rare-event simulation methods commonly aim to control the sampling error, not the bias due to the discretization. Adler et al. (2012) develop an algorithm that is strongly efficient (with bounded relative sampling error) for estimation of $w_T(b)$ (rather than $w(b)$). We will show that combining their algorithm with the use of threshold-dependent grids provides a strongly efficient algorithm for estimation of $w(b)$.

A topic closely related to ours concerns the quantification of the difference between the supremum of the stochastic process taken over $[0, 1]$ and the supremum taken over a finite grid $T \subset [0, 1]$, i.e.,

$$\Delta(T) = \sup_{t \in [0,1]} X_t - \sup_{t \in T} X_t.$$

There are several results in the literature that study the behavior of $\Delta(T)$ for standard Brownian Motion. Asmussen et al. (1995) have shown that for the equidistant grids $T_n^{\text{eq}} = \{\frac{1}{n}, \dots, \frac{n}{n}\}$, $\sqrt{n} \Delta(T_n^{\text{eq}})$ has a tight, non-degenerate weak limit, as $n \rightarrow \infty$ and Janssen and Van Leeuwaarden (2009) derived an expansion for $\mathbb{E}\Delta(T_n^{\text{eq}})$. For *random grids* $T_n^{\text{rnd}} = \{U_1, \dots, U_n\}$, where U_1, \dots, U_n are i.i.d. uniform samples on $(0, 1)$, independent of the Brownian Motion $(X_t)_{t \in [0,1]}$, Calvin and Glynn (1997) establish the weak limit of $\sqrt{n} \Delta(T_n^{\text{rnd}})$. Finally, Calvin (1997) proposed a class of *adaptive* grids, meaning that the consecutive grid points t_{k+1} are chosen based on $((t_1, B_{t_1}), \dots, (t_k, B_{t_k}))$; given any $\delta > 0$, an adaptive grid $T_n^\delta = \{t_1^\delta, \dots, t_n^\delta\}$ is provided such that $n^{1-\delta/2} \Delta(T_n^\delta)$ has a weak limit.

In our study, we do not focus on the difference $\Delta(T)$ between the values of the maxima of the discrete and continuous-time Brownian Motion, but rather on the $\beta_T(b)$, i.e., the relative difference between the probabilities that these maxima lie above a certain fixed threshold.

There are several approaches to tackle the discretization bias available in the literature. Arguably, the most widely applicable method is *Multilevel Monte Carlo* (MLMC) (Giles 2008). It can be applied together with any numerical method that relies on discretization. The idea is to use several different *levels of discretization* and spend less computational effort (draw less samples) at the finest levels of discretization. MLMC effectively reduces the computational effort, and the time saved can be used to produce even finer levels of discretization. It could be interesting to explore the combination of MLMC method together with the idea of threshold-dependent grids, but further exploiting this procedure lies beyond the scope of this article.

One of the methods that aims to directly decrease the bias induced by equidistant grids is *continuity correction*. Since the discrete-time approximation $w_T(b)$ is always smaller than $w(b)$, one could *slightly* lower the threshold b to compensate for the underestimation. Broadie et al. (1997), using the machinery developed in Siegmund (1985), proposed a way of lowering the threshold which improves the rate of convergence of the relative bias from $O(n^{-1/2})$, cf. Proposition 1, to $O(n^{-1})$, as the number of grid points n grows large. However, in the non-Brownian case, it remains a non-trivial problem how much b should be decreased. In fact, there is no direct way of making sure whether lowering b decreases the absolute relative bias, as lowering b by *too much* leads to overcompensation and thus to an estimate that is *larger* than $w(b)$. By contrast, it is straightforward to compare the bias induced by two different grids—the larger the discrete estimator $w_T(b)$, the smaller the relative bias.

There are also several simulation-based algorithms that do not rely on pre-discretization. Li and Liu (2015) propose a strongly efficient algorithm for estimation of $w(b)$ for a large class of Gaussian processes (most prominently, processes with constant variance function). However, when the underlying process has a unique point of maximal variance (such as Brownian Motion), the algorithm requires the simulation of a random time $\tau \in [0, 1]$ from a density $f(t) \propto \mathbb{P}(X_t > b)$, which becomes a rare-event simulation problem when b is large. While for an arbitrary process,

the random discretization proposed in the algorithm requires a computational effort cubic in the number of grid points (in order to simulate a discrete Gaussian path), pre-discretization requires only quadratic effort (see the discussion in Section 5).

This article is organized as follows. Section 2 provides definitions, preliminaries, and develops a general intuition. In Section 3, we introduce useful upper and lower bounds for the discretization bias (see Lemma 1) and show that the number of points on the equidistant grid has to grow quadratically in the threshold b in order to uniformly control the discretization bias. In Section 4, as an alternative to equidistant grids, we study threshold-dependent grids, which control the relative bias with a constant grid size, independently of b . The proofs of all lemmas and a proposition are postponed to Section 8. In Section 5, we present an algorithm by Adler et al. (2012), that we use throughout the article for producing the numerical results; combining this algorithm with the use of threshold-dependent grids yields a strongly efficient algorithm for estimation of $w(b)$ (see Corollary 1). In Section 6, we apply threshold-dependent grids developed in a previous section to stochastic processes other than Brownian Motion: Brownian Motion with jumps, Ornstein-Uhlenbeck process, and fractional Brownian Motion. Lastly, in Section 7, we present concluding remarks and discuss some ideas for future research of *optimal grids*. In the appendices, we collect various technical results used throughout the article.

2 PRELIMINARY RESULTS

Let $(B_t)_{t \in [0,1]}$ be a standard Brownian Motion on the time interval $[0, 1]$ with $B_0 = 0$. We consider the probability of crossing a positive threshold b , i.e.,

$$w(b) := \mathbb{P}\left(\sup_{t \in [0,1]} B_t > b\right). \quad (1)$$

For a standard Brownian Motion, an explicit formula for the threshold-crossing probability (1) is known, namely, $w(b) = 2\mathbb{P}(B_1 > b)$, which follows directly using the *reflection principle* (see, e.g., Mörters and Peres (2010)). Given a *finite grid* T , we define a discrete-time approximation of $w(b)$:

$$w_T(b) := \mathbb{P}\left(\sup_{t \in T} B_t > b\right), \quad (2)$$

where $T = \{t_1, \dots, t_n\}$ is a finite subset of the interval $[0, 1]$, ordered such that $t_1 < \dots < t_n$. As we are mostly interested in choosing the grid T efficiently, we define the following performance measure.

Definition 1. Let T be a finite grid on $[0, 1]$, then

$$\beta_T(b) := \frac{w(b) - w_T(b)}{w(b)} = \mathbb{P}\left(\sup_{t \in T} B_t < b \mid \sup_{t \in [0,1]} B_t > b\right)$$

is called the relative bias induced by the grid T .

The second representation of relative bias in Definition 1 is especially intuitive. It means that the relative bias is *the probability that B_t stays below b on the grid T , given that its supremum over $[0, 1]$ is greater than b* . Notice that any grid which includes the endpoint $t = 1$ will induce a relative bias no greater than $\frac{1}{2}$. Indeed, if $1 \in T$, then $w_T(b) = \mathbb{P}(\sup_{t \in T} B_t > b) \geq \mathbb{P}(B_1 > b)$ and thus

$$\beta_T(b) = 1 - \frac{w_T(b)}{w(b)} \leq 1 - \frac{\mathbb{P}(B_1 > b)}{2\mathbb{P}(B_1 > b)} = \frac{1}{2}.$$

Our objective is to accurately estimate $w(b)$ using discrete approximations $w_T(b)$, in a computationally efficient manner. Brownian Motion has continuous paths and thus it is always possible

for a given b to find a fine enough grid to bound the bias up to a desired accuracy. However, the computational cost of estimating $w_T(b)$ grows in the grid size and thus it might be infeasible to numerically compute $w_T(b)$ for large grids.

At this point, we emphasize that we are not as much interested in the behavior of $\beta_T(b)$ for a fixed b or a fixed n but rather in asymptotic regimes in which b and/or n approach infinity. For every b , we allow one to use a different grid, so it seems natural to treat the grid as a function of threshold. For every b , we define a collection of grids of all possible sizes $\{T_1(b), T_2(b), \dots\}$, where $T_n(b)$ has n elements, and we denote $\beta_n(b) := \beta_{T_n(b)}(b)$. For a given family of grids, we are interested in the behavior of $\beta_n(b)$ as n or b tend to infinity. The most straightforward choice for the family of grids is the following.

Definition 2. The family $\{T_n\}_{n \in \mathbb{N}}$, where $T_n := \{t_1^n, \dots, t_n^n\}$ with $t_k^n := \frac{k}{n}$ is called the equidistant family of grids.

Notice that the location of grid points on the equidistant grid is independent of b . Since the distance between neighboring points is equal to $\frac{1}{n}$, and since Brownian paths are continuous, it follows that $\beta_n(b) \rightarrow 0$, as $n \rightarrow \infty$ for any fixed b . It has been established in Asmussen et al. (1995) that for T_n , equidistant grid, the difference between the continuous-time and discrete-time supremum $\varepsilon_n = \sup_{t \in [0,1]} B_t - \sup_{t \in T_n} B_t$ is of order $n^{-1/2}$. More precisely, the sequence $(\sqrt{n}\varepsilon_n)_{n \in \mathbb{N}}$ has a tight and non-degenerate weak limit.

PROPOSITION 1. *Let $(B_t)_{t \in [0,1]}$ denote standard Brownian Motion and $\{T_n\}_{n \in \mathbb{N}}$ be the equidistant family of grids from Definition 2 with $\beta_n(b) := \beta_{T_n(b)}$. For any threshold $b > 0$, there exist positive constants C_1, C_2 such that*

$$C_1 n^{-1/2} \leq \beta_n(b) \leq C_2 n^{-1/2}.$$

The proof of Proposition 1 is given in Appendix C, which is a part of the Supplementary Materials, available in the online version of this article. The proof we give strongly resembles the proof of Theorem 1 below in Section 3, but we remark that it is also possible to derive it using the tools developed in Broadie et al. (1997).

Proposition 1 states that $\beta_n(b)$ decays like $n^{-1/2}$, when n grows large for a fixed b , but it does not describe the behavior of the relative bias when b varies. In Theorem 1 in the following section, we derive an upper bound for $\beta_n(b)$ for n and b simultaneously.

Figure 1 shows the evolution of the relative bias for four different thresholds $b = 5, 6, 7, 8$ against the size of the equidistant grid. Even though all four graphs show the $n^{-1/2}$ decay, the graphs rise up with growing threshold. In particular, for thresholds $b = 5$ and 8 , respectively $n = 700$ and 1700 points are needed to arrive at around 10% relative bias. It indicates that, as b grows, *increasingly many grid points are needed to arrive at the target relative bias*. Using the threshold-dependent grid that we develop in Section 4, one can arrive at 10% relative bias using approximately $n = 100$ grid points, independently of the value of the threshold. This amounts to a substantial improvement of the computational efficiency.

In some cases, the equidistant family of grids is the best possible choice, in the sense that other grid families require at least equally fast asymptotic growth of n as b increases, in order to control the relative bias. Adler et al. (2012) prove that for *centered, homogeneous and twice continuously differentiable (in a mean-squared sense) Gaussian processes*, n has to grow linearly in b to uniformly control the relative bias. Moreover, if n grows sublinearly in b , then the relative bias of any family of grids (not necessarily equidistant) tends to its maximal value, as b approaches infinity. It is noted, however, that Brownian Motion does not belong to the family of Gaussian processes for which the result of Adler et al. (2012) applies.

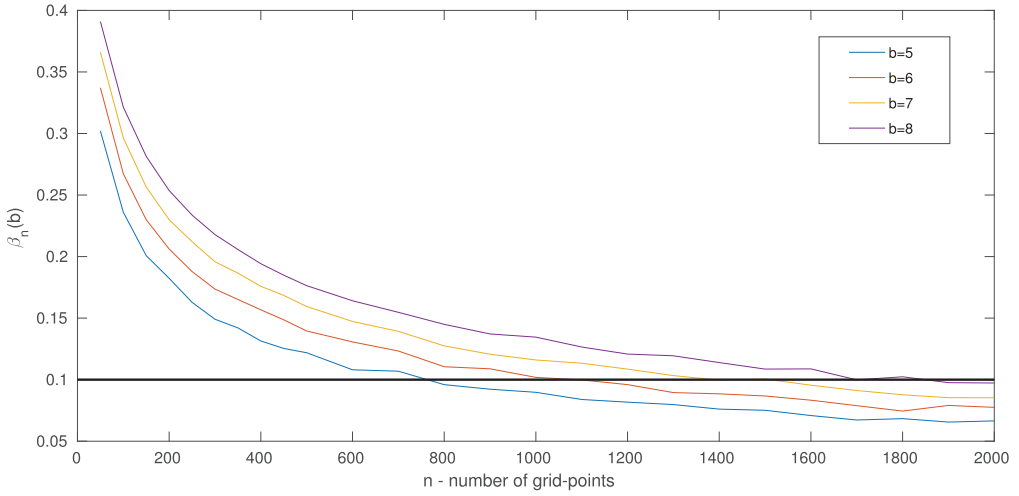


Fig. 1. Plots of the relative bias $\beta_n(b)$ against the grid size n for the equidistant family of grids for four different thresholds. The numerical results are computed using an algorithm described in Section 5.

In the following two sections, we analyze the asymptotic behavior of the relative bias $\beta_n(b)$ for two families of grids. We prove that the equidistant grid requires quadratic growth of n in b (see Theorem 1 in Section 3). As an alternative, we develop the threshold-dependent family of grids, for which we prove that the relative bias can be made arbitrarily small, uniformly in b for fixed n (see Theorem 2 in Section 4). We obtain a uniform rate of convergence in n and also provide a closed-form expression for the threshold-dependent family of grids (see Definition 9 in Section 4).

3 EQUIDISTANT FAMILY OF GRIDS FOR BROWNIAN MOTION

This section is devoted to analyzing the asymptotic behavior of the relative bias for the equidistant family of grids. The methodology developed in this section will be used later to prove Theorem 2; in particular, the crucial part of the proof concerns bounds for the relative bias induced by an arbitrary finite grid, developed in Lemma 1.

The following theorem describes the asymptotic behavior of the relative bias, under the equidistant family of grids.

THEOREM 1. *Let $(B_t)_{t \in [0,1]}$ denote standard Brownian Motion and $\{T_n\}_{n \in \mathbb{N}}$ be the equidistant family of grids from Definition 2 with $\beta_n(b) := \beta_{T_n}(b)$.*

- (a) *Let b_0 be any positive, real number. There exist positive constants C_0, C_1 , independent of b and n such that*

$$\beta_n(b) \leq C_0 \cdot bn^{-1/2},$$

for all $b \geq b_0$, and

$$\beta_n(b) \leq C_1 \cdot n^{-1/2},$$

for all $b \in (0, b_0]$.

- (b) *Let $m : (0, \infty) \rightarrow (0, \infty)$ be such that $\lim_{b \rightarrow \infty} m(b)/b^2 = 0$. Then, as $b \rightarrow \infty$,*

$$\inf_{n \leq m(b)} \beta_n(b) \rightarrow \frac{1}{2}.$$

Part (a) of Theorem 1 states that $\beta_n(b) \leq C_0 b n^{-1/2}$, so that in order to bound $\beta_n(b)$ uniformly in b it suffices to take $n = O(b^2)$. The second part of the Theorem 1 states that if $n = o(b^2)$, then $\beta_n(b) \rightarrow 1/2$, meaning that the relative bias cannot be bounded by an arbitrarily small number. Together, the two parts entail that the growth $n = O(b^2)$ is sufficient and there is no better (slower) growth which would guarantee a uniformly bounded relative bias.

The crucial part of the proof of Theorem 1 is the method of bounding the relative bias. Since no explicit expressions for $w_T(b)$ or $\beta_T(b)$ are known (even if T is an equidistant grid), we develop a general upper bound for $\beta_T(b)$ in the following lemma, in which we use the quantities

$$\begin{aligned} a_j(b) &:= \mathbb{P}\left(B_{t_j(b)-t_{j-1}(b)} < 0, \dots, B_{t_n(b)-t_{j-1}(b)} < 0\right), \quad a_{n+1}(b) := 1/2, \\ w_j(b) &:= \mathbb{P}\left(\tau_b \in (t_{j-1}(b), t_j(b)] \mid \tau_b \in (0, 1]\right), \\ \tau_b &:= \inf\{t \geq 0 : B_t > b\}. \end{aligned}$$

Notice that in this definition of $a_j(b)$ and $w_j(b)$, we allow grid points t_1, \dots, t_n to change their location with b . In the present section, which is on equidistant grids, the grid points obviously do not depend on b , but in later sections they do.

LEMMA 1. *Let $T(b) = \{t_1(b), \dots, t_n(b)\} \subset [0, 1]$, where $0 < t_1(b) < \dots < t_n(b) \leq 1$, and let $t_0(b) = 0$. The following lower and upper bounds for $\beta_T(b)$ apply:*

$$\underline{\beta}_T(b) \leq \beta_T(b) \leq \bar{\beta}_T(b)$$

with

$$\underline{\beta}_T(b) := \frac{1}{2} \sum_{j=1}^n a_{j+1}(b) w_j(b), \quad \bar{\beta}_T(b) := \sum_{j=1}^n a_j(b) w_j(b).$$

A short proof of Lemma 1 is included in Section 8. The bounds consist of elements of two types: $a_j(b)$, the probability that B_t stays negative at times $t_j - t_{j-1}, \dots, t_n - t_{j-1}$, and $w_j(b)$, the probability that B_t hits b for the first time in the interval $[t_{j-1}, t_j]$ given that its supremum over $[0, 1]$ is greater than b .

For a general grid $T(b)$, the probabilities $a_j(b)$ are difficult to control. However, when $T(b)$ is equidistant (thus independent of b), then also the probabilities a_j are independent of b ; we emphasize this independence by writing a_j instead of $a_j(b)$ throughout this section. As a result, there exists a tight asymptotic bound for them (see Lemma 2 below); we were inspired to look into such quantities while reading (Mörters and Peres (2010, Section 5)). The upper and lower bounds for the probabilities $w_j(b)$ are developed using a mean value theorem (see (B.V) in Appendix B).

LEMMA 2. *There exist constants $C_1^*, C_2^* > 0$ such that*

$$C_1^* n^{-1/2} \leq \mathbb{P}(B_1 > 0, \dots, B_n > 0) \leq C_2^* n^{-1/2}$$

for all $n \in \mathbb{N}$.

In fact, the assertion in Lemma 2 is true for any symmetric random walk (see Feller (1971, Theorem 4 in Section XII.7, and Lemma 1 in Section XII.8)). Before proving Theorem 1, we present one more lemma.

LEMMA 3. *Let $T = \{t_1, \dots, t_n\}$ be such that $t_k = \frac{k}{n}$ and let $t_0 = 0$. Then the upper bound $\bar{\beta}_T(b)$ developed in Lemma 1 is an increasing function of b .*

An important implication of Lemma 3 is that for any $b_0 > 0$, we have that $\beta_T(b) \leq \bar{\beta}_T(b) \leq \bar{\beta}_T(b_0)$ uniformly for all $b \leq b_0$, which completely covers the statement on the situation that $b \leq b_0$ in part (a) of Theorem 1. The proof of Lemma 3 is included in Appendix C.

PROOF OF THEOREM 1(A). Thanks to Lemma 3, it suffices to prove the first part of Theorem 1(a), i.e., we assume that $b \geq b_0$. Without loss of generality, we put $b_0 = 1$. Exploiting the upper bound developed in Lemma 1, we decompose the sum $\sum_{j=1}^n a_j \cdot w_j(b)$ into three parts, which we treat separately:

$$\beta_n(b) \leq a_1 \cdot w_1(b) + \sum_{j=2}^{n-1} a_j \cdot w_j(b) + a_n \cdot w_n(b). \quad (3)$$

Using the definition of the equidistant grid and the scaling property of Brownian Motion, we can see that $a_j = \mathbb{P}(B_{t_j - t_{j-1}} < 0, \dots, B_{t_n - t_{j-1}} < 0) = \mathbb{P}(B_1 < 0, \dots, B_{n-j+1} < 0)$ and the bound in Lemma 2 yields $a_j \leq C_2^* (n-j+1)^{-1/2}$ for all $j \in \{1, \dots, n\}$. Since all $w_j(b) \leq 1$, we thus have a straightforward bound for the first term in inequality (3):

$$a_1 \cdot w_1(b) \leq C_2^* n^{-1/2}.$$

The second term we bound in the following fashion, relying on the upper bound stated in (B.V) in Appendix B that we have for $w_j(b)$,

$$\begin{aligned} \sum_{j=2}^{n-1} a_j \cdot w_j(b) &\leq \sum_{j=2}^{n-1} C_2^* (n-j+1)^{-1/2} \cdot \frac{b(b + \sqrt{b^2 + 4})}{4} \frac{\sqrt{n}}{(j-1)^{3/2}} e^{-\frac{b^2}{2} \cdot \left(\frac{n}{j-1}\right)} \\ &\leq C_1 \cdot bn^{-1/2} \cdot \sum_{j=2}^{n-1} \frac{1}{n} \cdot \left(\frac{b}{\sqrt{1 - \frac{j}{n}}} \cdot \left(\frac{j}{n}\right)^{-3/2} e^{-\frac{b^2}{2} \cdot \left(\frac{n}{j-1}\right)} \right) \end{aligned} \quad (4)$$

$$\begin{aligned} &\leq C_1 \cdot bn^{-1/2} \cdot \int_0^1 \frac{b}{\sqrt{1-x}} \cdot x^{-3/2} \cdot e^{-\frac{b^2}{2} (1/x-1)} dx \\ &\leq C_1 \cdot bn^{-1/2}, \end{aligned} \quad (5)$$

where C_2^* comes from Lemma 2 and C_1 is a positive constant, independent of b and n . To arrive at expression (4), we use that $2(j-1) \geq j$ for all relevant j . In the transition from (4) to (5), we use the definition of the Riemann sum for the function

$$f(b, x) := \frac{b}{\sqrt{1-x}} x^{-3/2} e^{-\frac{b^2}{2} (1/x-1)},$$

note that, since $f(b, x)$ is an increasing function of x when $b \geq 1$ (see (B.VI) in Appendix B), the Riemann sum in (4) underestimates the integral, i.e., $\sum_{j=2}^n \frac{1}{n} f(b, \frac{j-1}{n}) \leq \int_0^1 f(b, x) dx = \sqrt{2\pi}$.

Lastly, since $a_n = \mathbb{P}(B_{t_n} < 0) = \frac{1}{2}$, we have

$$a_n \cdot w_n(b) \leq \frac{1}{2} \cdot \frac{b(b + \sqrt{b^2 + 4})}{4} \frac{\sqrt{n}}{(n-1)^{3/2}} \leq C_2 \frac{b^2}{n},$$

where C_2 is a positive constant independent of n and b . Since $w_n(b) \leq 1$, this results in

$$a_n \cdot w_n(b) \leq \min \left\{ C_2 \frac{b^2}{n}, \frac{1}{2} \right\} \leq \sqrt{\min \left\{ C_2 \frac{b^2}{n}, \frac{1}{2} \right\}} \leq \sqrt{C_2} bn^{-1/2}.$$

Combining the above bounds,

$$\begin{aligned}\beta_n(b) &\leq a_1 \cdot w_1(b) + \sum_{j=2}^{n-1} a_j \cdot w_j(b) + a_n \cdot w_n(b) \leq C_2^* n^{-1/2} + C_1 b n^{-1/2} + \sqrt{C_2} b n^{-1/2} \\ &\leq C_0 b n^{-1/2},\end{aligned}$$

where C_0 is a positive constant independent of b and n . This concludes the proof. \square

PROOF OF THEOREM 1(B). Without loss of generality, we can assume $m(b) \rightarrow \infty$ as $b \rightarrow \infty$. Similar to the proof of Lemma 1 in Section 8, we obtain

$$\begin{aligned}w(b) \beta_T(b) &= \mathbb{P}\left(\sup_{t \in T} B_t < b, \sup_{t \in [0,1]} B_t > b\right) = \mathbb{P}\left(\sup_{t \in T} B_t < b, \tau_b \in [0, 1]\right) \\ &= \sum_{j=1}^n \mathbb{P}\left(\sup_{t \in (t_{j-1}, t_j]} B_t < b, \tau_b \in (t_{j-1}, t_j]\right) \geq \mathbb{P}\left(B_{t_n} < b, \tau_b \in (t_{n-1}, t_n]\right) \\ &= \int_{t_{n-1}}^{t_n} \mathbb{P}(B_{t_n} < b | B_s = b) \mathbb{P}(\tau_b \in ds) = \frac{1}{2} \mathbb{P}(\tau_b \in (t_{n-1}, t_n)).\end{aligned}$$

Dividing both sides of the inequality by $w(b)$ yields an elementary lower bound on $\beta_T(b)$:

$$\beta_T(b) \geq \frac{1}{2} \mathbb{P}\left(\tau_b \in (t_{n-1}, t_n] \mid \tau_b \in (0, 1]\right) = \frac{1}{2} \cdot \frac{\Phi(-b/\sqrt{t_n}) - \Phi(-b/\sqrt{t_{n-1}})}{\Phi(-b)}, \quad (6)$$

where $\Phi(\cdot)$ denotes the standard normal cdf, and we use the fact that $\mathbb{P}(\tau_b \leq t) = 2\mathbb{P}(B_t > b)$. In our case, $t_n = 1$ and $t_{n-1} = \frac{n-1}{n} \leq \frac{m-1}{m}$, so that due to the monotonicity of $\Phi(\cdot)$

$$\inf_{n \leq m(b)} \beta_n(b) \geq \frac{1}{2} - \frac{1}{2} \frac{\Phi(-b/\sqrt{(m-1)/m})}{\Phi(-b)}. \quad (7)$$

Taking the limit $b \rightarrow \infty$ on both sides of inequality (7) yields

$$\begin{aligned}\lim_{b \rightarrow \infty} \inf_{n \leq m(b)} \beta_n(b) &\geq \frac{1}{2} - \frac{1}{2} \lim_{b \rightarrow \infty} \frac{\Phi(-b/\sqrt{(m-1)/m})}{\Phi(-b)} \\ &= \frac{1}{2} - \frac{1}{2} \lim_{b \rightarrow \infty} \frac{\frac{\sqrt{(m-1)/m}}{b} \phi(b/\sqrt{(m-1)/m})}{\frac{1}{b} \phi(b)} \\ &= \frac{1}{2} - \frac{1}{2} \lim_{b \rightarrow \infty} e^{-b^2/(2(m-1))} = \frac{1}{2},\end{aligned} \quad (8)$$

where $\phi(\cdot)$ denotes the standard normal pdf, in Eq. (8) we apply (B.II) in Appendix B, and the last equality is a consequence of the assumption that $\lim_{b \rightarrow \infty} m(b)/b^2 = 0$. \square

In this section, we have proven that in order to uniformly control the relative bias, the size of the equidistant grid must grow at least quadratically in b , as b approaches infinity. In the next section, we present a threshold-dependent grid, which yields a uniform bound on the relative bias using a grid of given size. In other words, in order to control the relative bias with increasing b , instead of adding more and more points to the grid, it suffices to suitably shift their location.

4 THRESHOLD-DEPENDENT GRIDS FOR BROWNIAN MOTION

In this section, we prove the main result of the article. We explicitly present a threshold-dependent family of grids which uniformly (in b) bounds the relative bias.

Before we introduce the result, we give some intuition as to why it is possible to control the relative bias as b grows, without increasing n . Firstly, for any given $\varepsilon > 0$, we have that

$$\mathbb{P}\left(\sup_{t \in [0, 1-\varepsilon]} B_t > b\right) = 2\mathbb{P}(B_{1-\varepsilon} > b) = o(w(b)),$$

as $b \rightarrow \infty$. Therefore,

$$\mathbb{P}\left(\sup_{t \in [1-\varepsilon, 1]} B_t > b \mid \sup_{t \in [0, 1]} B_t > b\right) \rightarrow 1, \text{ as } b \rightarrow \infty.$$

It means that with growing b , the “hitting of the threshold” occurs closer and closer to time $t = 1$. It indicates that the grid points should be gradually shifted toward the point $t = 1$, as b is increasing. Moreover, the result in Theorem 1 indicates how fast the points should be shifted. It states that for the family of equidistant grids, the uniform bound on the bias is achieved if the number of grid points grows quadratically in b . Equivalently, the distances between neighboring points are decreasing proportionally to b^{-2} . It turns out that this is indeed the pace at which the points should be shifted toward $t = 1$.

In the following result, $\Phi(\cdot)$ and $\Phi^{-1}(\cdot)$ denote the standard normal cdf and its inverse, respectively.

THEOREM 2. *Let $(B_t)_{t \in [0, 1]}$ be a standard Brownian Motion. Fix $b_0 > 0$ and let $\{T_n(b)\}_{n \in \mathbb{N}, b > 0}$ be a family of grids such that $T_n(b) = \{t_1^n(b), \dots, t_n^n(b)\}$; here $t_k^n(b) := \frac{k}{n}$ for $b \leq b_0$, and*

$$t_k^n(b) := \left(\frac{b}{\Phi^{-1}\left(\frac{k}{n}\Phi(-b)\right)}\right)^2, \quad (9)$$

for $b > b_0$. Denote $\beta_n(b) := \beta_{T_n(b)}(b)$. There exists a positive C , independent of b and n , such that

$$\beta_n(b) \leq C n^{-1/4}$$

for all $b > 0$.

We emphasize that the bound for the relative bias $\beta_n(b)$ developed above does not depend on the threshold b and thus holds *uniformly*, for all b . Figure 2 shows the comparison between the relative bias of the equidistant and the threshold-dependent grid, both of size $n = 100$. The bias induced by the threshold-dependent grid remains uniformly bounded (by circa 0.1), while the former tends to 0.5, the worst possible relative bias (cf. Theorem 1, part (b)).

Notice that for small b , $\{T_n(b)\}_{n \in \mathbb{N}, b \in (0, b_0]}$ in Theorem 2 is identical to the equidistant family of grids. In fact, this is exactly the setting of the second part of Theorem 1(a). The real contribution of Theorem 2 is the regime when $b > b_0$. The grid defined in (9) is the unique solution to the set of equations

$$\mathbb{P}\left(\tau_b \in (t_{k-1}^n(b), t_k^n(b)] \mid \tau_b \in (0, 1]\right) = \frac{1}{n} \quad (10)$$

for all $k \in \{1, \dots, n\}$ and $t_0 := 0$. To see this, we sum up the first k equations in Eq. (10) and obtain an explicit equation for $t_k^n(b)$:

$$\mathbb{P}\left(\tau_b \in (0, t_k^n(b)] \mid \tau_b \in (0, 1]\right) = \frac{k}{n}. \quad (11)$$

Since for Brownian Motion it holds that

$$\mathbb{P}(\tau_b \in (0, t_k^n(b)]) = 2\mathbb{P}(B_{t_k^n(b)} > b) = 2\Phi\left(-b/\sqrt{t_k^n(b)}\right),$$

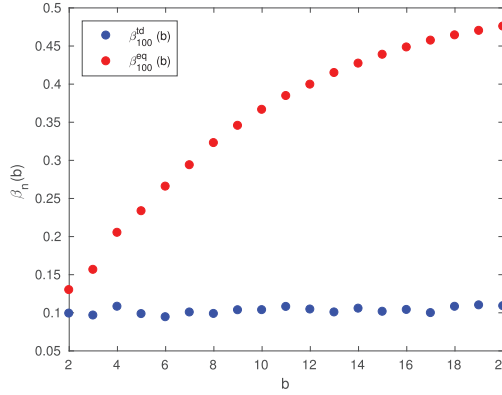


Fig. 2. A plot of the relative bias of the equidistant grid $\beta_{100}^{eq}(b)$ and the threshold-dependent grid $\beta_{100}^{td}(b)$, both with fixed grid size $n = 100$, as a function of b . Notice that $\beta_{100}^{eq}(b)$ tends to 0.5 with growing b (the largest possible bias), while $\beta_{100}^{td}(b)$ remains bounded by about 0.1. The numerical results are computed with the algorithm discussed in Section 5. The relative error due to finite sample size is negligible (smaller than 0.006).

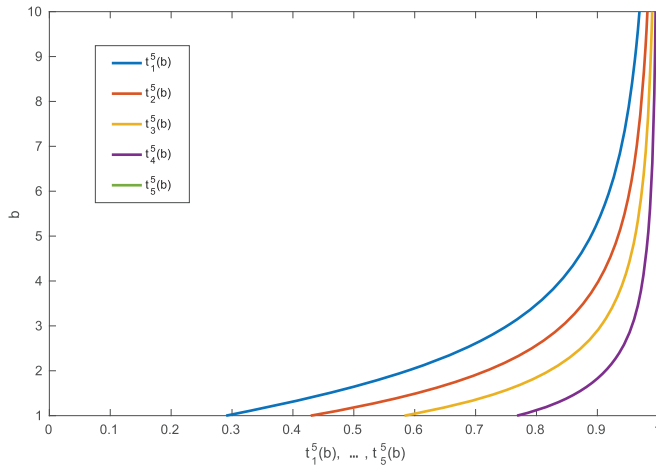


Fig. 3. Location of the grid points $t_1^5(b), \dots, t_5^5(b)$ defined in Eq. (9) with increasing threshold b . Note that with growing b , all the points are gradually shifted toward the end point $t = 1$.

and in particular $\mathbb{P}(\tau_b \in (0, 1]) = 2 \mathbb{P}(B_1 > b) = 2 \Phi(-b)$, Eq. (11) can be equivalently expressed as

$$\frac{\mathbb{P}(B_{t_k^n(b)} > b)}{\mathbb{P}(B_1 > b)} = \frac{k}{n} \tag{12}$$

or, in terms of the cdf $\Phi(\cdot)$,

$$\Phi\left(-b/\sqrt{t_k^n(b)}\right) = \frac{k}{n}\Phi(-b).$$

Finally, after taking the inverse $\Phi^{-1}(\cdot)$ from both sides of the equation above, we see that $t_k^n(b)$ satisfies Eq. (9). Figure 3 shows the placement of the grid points on the grid $T_5(b)$, as defined in (9),

for increasing b . In fact, one can prove that

$$b^2(1 - t_k^n(b)) \xrightarrow{b \rightarrow \infty} -2 \log(k/n) \quad (13)$$

and thus

$$t_k^n(b) \approx 1 - \frac{2 \log(n/k)}{b^2}$$

for large b . It means that the points of the grid (9) are clustered around $t = 1$, with distances between the points proportional to b^{-2} . Here we see an important connection with Theorem 1(a), where the distances between grid points decrease at the same pace, as already mentioned in the opening paragraph of this section.

For $b > b_0$, the points $t_1^n(b), \dots, t_n^n(b)$ of the threshold-dependent grid (9) do not coincide with the equidistant grid, entailing that we cannot directly use Lemma 2 to control the terms of type $a_j(b)$ in the upper bound developed in Lemma 1 in Section 3. The following lemma resolves this issue.

LEMMA 4. *Let $t_0 = 0 < t_1 < t_2 < \dots < t_n < \infty$. Then*

$$\mathbb{P}(B_{t_1} > 0, \dots, B_{t_n} > 0) \leq \mathbb{P}(B_1 > 0, \dots, B_N > 0),$$

for any $N \leq N_n$, where

$$N_n := \left(\frac{t_n}{\max_{k=1, \dots, n} (t_k - t_{k-1})} \right)^{1/2}.$$

A proof of this lemma is provided in Section 8. Lemma 2 applied to the upper bound in Lemma 4 yields a simple upper bound for $\mathbb{P}(B_{t_1} > 0, \dots, B_{t_n} > 0)$ for any choice of $t_0 = 0 < t_1 < t_2 < \dots < t_n < \infty$. In our case, after applying Lemma 1, we have to control probabilities of the type $\mathbb{P}(B_{t_j - t_{j-1}} < 0, \dots, B_{t_n - t_{j-1}} < 0)$, and thus we need a lower bound on

$$\frac{t_n - t_{j-1}}{\max_{k=j, \dots, n} (t_k - t_{k-1})},$$

which we give in the following lemma.

LEMMA 5. *For the grid in Eq. (9), for $k > j$, $b > 0$, and $n \in \mathbb{N}$ we have*

$$(a) \quad \frac{t_n^n(b) - t_j^n(b)}{t_k^n(b) - t_j^n(b)} \geq \frac{\log n - \log j}{\log k - \log j},$$

and when additionally $b \geq \sqrt{3}$, we have

$$(b) \quad \max_{k=j, \dots, n} (t_k^n(b) - t_{k-1}^n(b)) = t_j^n(b) - t_{j-1}^n(b).$$

Lemma 5 is proven in Section 8. We note that the lower bound in part (a) of Lemma 5 is, in fact, equal to

$$\lim_{b \rightarrow \infty} \frac{1 - t_j^n(b)}{t_k^n(b) - t_j^n(b)}.$$

With these lemmas we can prove Theorem 2.

PROOF OF THEOREM 2. Part (a) of Theorem 1 states that for any choice of b_0 there exists positive C_1 such that $\beta_n(b) \leq C_1 n^{-1/2}$ for $b \leq b_0$ and thus also $\beta_n(b) \leq C_1 n^{-1/4}$. Without the loss of generality, from now on we assume that $b > b_0 = \sqrt{3}$. Fix $n \in \mathbb{N}$ and denote $t_k := t_k^n(b)$ for notational simplicity. After combining the general upper bound from Lemma 1 with the

equivalent definition (10) of the threshold-dependent grid (9), we obtain

$$\beta_n(b) \leq \sum_{j=1}^n a_j(b) w_j(b) = \frac{1}{n} \sum_{j=1}^n a_j(b);$$

observe that in our setting $w_n(b) = \frac{1}{n}$. Moreover, Lemma 4 yields (recalling the definition of $a_n(b)$)

$$\beta_n(b) \leq \frac{1}{2n} + \frac{1}{n} \sum_{j=2}^{n-1} \mathbb{P}(B_1 > 0, \dots, B_{N_n(j)} > 0) + \frac{1}{2n},$$

where

$$N_n(j) := \left\lceil \left(\frac{t_n^n(b) - t_{j-1}^n(b)}{\max_{k \geq j} |t_k^n(b) - t_{k-1}^n(b)|} \right)^{1/2} \right\rceil.$$

Combining Lemma 2 with Lemma 5 gives

$$\beta_n(b) \leq \frac{1}{n} + C \frac{1}{n} \sum_{j=2}^{n-1} \tilde{N}_n(j)^{-1/4}, \quad \text{where } \tilde{N}_n(j) := \frac{\log n - \log(j-1)}{\log j - \log(j-1)}$$

with a constant $C > 0$ that is independent of n and b . Notice that $\tilde{N}_n(j)$ does not depend on b . For $b > b_0$, we thus obtain

$$\begin{aligned} \beta_n(b) &\leq \frac{1}{n} + C n^{-1} \sum_{j=2}^{n-1} \left(\frac{\log j - \log(j-1)}{\log n - \log(j-1)} \right)^{1/4} \\ &= \frac{1}{n} + C n^{-1} \sum_{j=2}^{n-1} \left(\frac{\log \left(1 + \frac{1}{j-1} \right)}{\log \left(\frac{n}{j-1} \right)} \right)^{1/4} \\ &\leq \frac{1}{n} + C n^{-1/4} \sum_{j=2}^{n-1} \frac{1}{n} \left(\frac{\frac{n}{j-1}}{\log \left(\frac{n}{j-1} \right)} \right)^{1/4} \end{aligned} \quad (14)$$

$$\begin{aligned} &\leq \frac{1}{n} + C n^{-1/4} \int_0^1 \left(\frac{1}{-x \log x} \right)^{1/4} dx \\ &\leq C n^{-1/4}, \end{aligned} \quad (15)$$

where C is a constant, independent from n and b , that might differ from line to line. In (14) we use the inequality $\log(1+x) \leq x$ and in (15) we use the convergence of the Riemann sum to the integral. This concludes the proof of Theorem 2. \square

Remark 1. For the purpose of showing that for any confidence level α and bias ε (see also Equation (16)) the “equiprobable” grid (as defined through Eq. (9)) requires a computational effort that is bounded in b , it suffices that the decay of the upper bound for $\beta_n(b)$ in Theorem 2 is of order $n^{-1/4}$; see Corollary 1 in Section 5. As an aside, we remark that we hypothesize that this decay is actually of order $n^{-1/2}$. This is supported by numerical experiments (see Figure 4 where plots of $\beta_n(b)$ versus n are shown for the threshold-dependent grid (9)). The step we expect to be “loose,” in obtaining the bound of Theorem 2, is the one corresponding to Lemma 4. We conjecture that Lemma 4 is valid with

$$N_n := \frac{t_n}{\max_{k=1, \dots, n} (t_k - t_{k-1})}$$

(i.e., without the square root), which suffices to yield the $n^{-1/2}$ decay of $\beta_n(b)$.

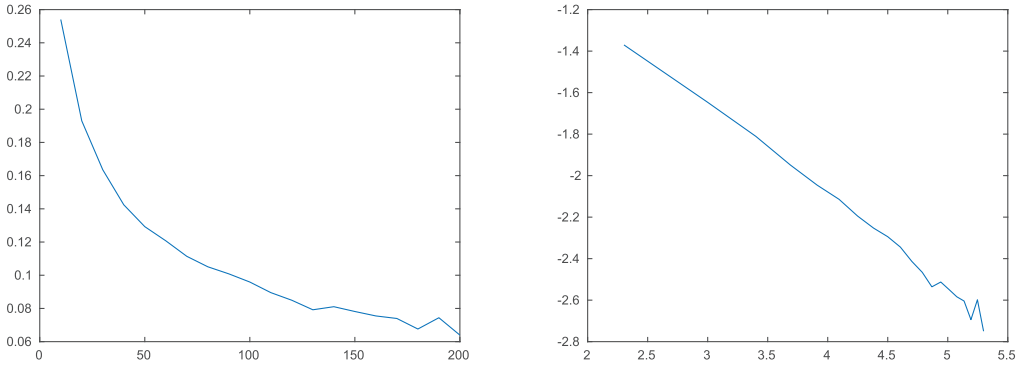


Fig. 4. Relative bias $\beta_n(b)$ versus grid size n for the threshold-dependent grid (9). The threshold is fixed at $b = 3$. The right panel shows a loglog plot, and the left panel a linear plot. The results suggest that $\beta_n(b)$ decays proportionally to $n^{-1/2}$ rather than $n^{-1/4}$ (see also Remark 1).

5 NUMERICAL ALGORITHM FOR ESTIMATION OF $w(b)$

As mentioned in the Introduction, the family of threshold-dependent grids (9) can be used to construct a strongly efficient algorithm for estimation of $w(b)$ (see Corollary 1 below). In this article, by “strongly efficient” we mean that for any given accuracy $\varepsilon > 0$ and confidence level $\alpha > 0$ the computational time of an estimator $\widehat{w}(b)$ for $w(b)$ that satisfies

$$\mathbb{P}\left(\left|\frac{\widehat{w}(b)}{w(b)} - 1\right| > \varepsilon\right) < \alpha \quad (16)$$

is bounded independently of the threshold b .

In all numerical experiments throughout this article, we used an algorithm developed by Adler et al. (2012) (see Algorithm 1 below). Although it is applicable for estimation of quantities such as $\mathbb{P}(\max_{i \in \{1, \dots, n\}} X_i > b)$, where $X \in \mathbb{R}^n$ is normally distributed with an arbitrary positive-definite covariance matrix, we present their algorithm for the specific case of Brownian Motion, as considered in this article.

ALGORITHM 1 (ADLER ET AL. 2012). *Choose a threshold b and a finite grid $T = \{t_1, \dots, t_n\} \in [0, 1]$. The estimator $\widehat{w}_T(b)$, computed according to the following algorithm, is an unbiased estimator of $w_T(b)$.*

- (1) Generate a random time τ on the grid, i.e., $\tau \in T$, according to the law

$$\mathbb{P}(\tau = t_k) = \frac{\mathbb{P}(B_{t_k} > b)}{\sum_{j=1}^n \mathbb{P}(B_{t_j} > b)}.$$

- (2) Generate B_τ under the condition $B_\tau > b$.
- (3) Generate a discrete path of the Brownian Motion $(B_{t_1}, \dots, B_{t_n})$ conditioned on the pair (τ, B_τ) generated in the previous steps.
- (4) Compute

$$\widehat{w}_T(b) := \frac{\sum_{j=1}^n \mathbb{P}(B_{t_j} > b)}{\sum_{j=1}^n \mathbb{1}(B_{t_j} > b)}.$$

Adler et al. (2012) prove that Algorithm 1 gives an unbiased estimator of $w_T(b)$ (not of $w(b)$) and that for a fixed T (independent of b), the relative variance $\text{Var}(\widehat{w}_T(b))/w_T^2(b) \rightarrow 0$, as $b \rightarrow \infty$. The authors also propose an estimator for $w(b)$, which relies on a *random discretization*. However, with

growing b , one needs increasingly many random grid points in order to control the relative bias, therefore the continuous-time algorithm *is not* strongly efficient. In order to reduce the sampling error one generates multiple replicas of the estimator and takes their average. Since every replica is based on a different grid, one must repeatedly calculate the Cholesky decomposition (whose computational time is cubic in the number of grid points) in order to sample discrete Gaussian paths in Step 3 of Algorithm 1. Choosing a predefined grid speeds up this computation, as in that case the Cholesky decomposition has to be performed *only once*, making its computational cost negligible.

Combining the threshold-dependent grids as proposed in Section 4 with Algorithm 1 yields a strongly efficient estimator for $w(b)$, which is given in the corollary below.

COROLLARY 1 (STRONGLY EFFICIENT ALGORITHM FOR THE ESTIMATION OF $w(b)$). *Fix an accuracy $\varepsilon > 0$ and a confidence level $\alpha > 0$. Choose a grid $T := T_n(b)$ from the family of grids defined in (9) such that $\beta_T(b) := \beta_n(b) < \varepsilon$ for all $b > 0$ (this is possible due to the result in Theorem 2). Let $\widehat{w}_T^{(1)}(b), \dots, \widehat{w}_T^{(N)}(b)$ be i.i.d. copies of the estimator from Algorithm 1, with*

$$N \geq \frac{n^2}{\alpha(\varepsilon - \beta_T(b))^2}.$$

Then

$$\widehat{w}(b) := \frac{1}{N} \sum_{i=1}^N \widehat{w}_T^{(i)}(b)$$

satisfies

$$\mathbb{P}\left(\left|\frac{\widehat{w}(b)}{w(b)} - 1\right| > \varepsilon\right) < \alpha, \quad (17)$$

and the computational effort to simulate $\widehat{w}(b)$ is bounded independently of b .

PROOF. First notice that since $\beta_T(b)$ is uniformly bounded in b (see Theorem 2), so that N is fixed independently of b ; it follows that $\widehat{w}(b)$ can be computed in bounded time, independently of b . It remains to prove that $\widehat{w}(b)$ satisfies the strong efficiency property (17). Note that $\widehat{w}(b)$ is an unbiased estimator of $w_T(b)$, not of $w(b)$. The relative variance of $\widehat{w}(b)$ with respect to $w_T(b)$ can be bounded independently of b for an arbitrary choice of the grid in terms of the grid size n ,

$$\frac{\text{Var}(\widehat{w}_T(b))}{(w_T(b))^2} \leq \frac{\mathbb{E}(\widehat{w}_T(b))^2}{(w_T(b))^2} \leq \left(\frac{\sum_{j=1}^n \mathbb{P}(B_{t_j} > b)}{\max_{j \in \{1, \dots, n\}} \mathbb{P}(B_{t_j} > b)}\right)^2 \leq n^2.$$

Due to Chebyshev's inequality,

$$\begin{aligned} \mathbb{P}\left(\left|\frac{\widehat{w}(b)}{w(b)} - 1\right| > \varepsilon\right) &= \mathbb{P}\left(\left|\frac{\widehat{w}(b) - w_T(b)}{w(b)} + \frac{w_T(b) - w(b)}{w(b)}\right| > \varepsilon\right) \leq \mathbb{P}\left(\left|\frac{\widehat{w}(b) - w_T(b)}{w(b)}\right| > \varepsilon - \beta_T(b)\right) \\ &\leq \frac{\text{Var}(\widehat{w}(b))}{(\varepsilon - \beta_T(b))^2 (w(b))^2} = \frac{1}{N} \cdot \frac{\text{Var}(\widehat{w}_T(b))}{(\varepsilon - \beta_T(b))^2 (w(b))^2} \\ &\leq \frac{1}{N} \cdot \frac{n^2}{(\varepsilon - \beta_T(b))^2} \leq \alpha. \end{aligned}$$

This concludes the proof. \square

We conclude this section by a remark on the simulation of the conditioned Brownian Motion in Step 3 of Algorithm 1. The naive method would be to construct the covariance matrix of the conditioned process, calculate the Cholesky decomposition of that matrix (cubic in the number of grid points), and then simulate the process in a standard manner. Notice that this step must

be repeated for every replica $\widehat{w}_T^{(i)}(b)$ and thus its computational cost scales with the number of samples. The following algorithm, which can be found, e.g., in Doucet (2010), requires only a single calculation of the Cholesky decomposition for all replicas.

ALGORITHM 2 (DOUCET 2010). *Let $X = (X_1, X_2)^T \in \mathbb{R}^n$, where $X_1 \in \mathbb{R}^{n-1}$ and $X_2 \in \mathbb{R}$, be normally distributed with mean μ and covariance matrix Σ ,*

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \text{ where } \mu_1 \in \mathbb{R}^{n-1} \text{ and } \mu_2 \in \mathbb{R},$$

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}^T & \Sigma_{22} \end{pmatrix}, \text{ where } \Sigma_{11} \in \mathbb{R}^{(n-1) \times (n-1)}, \Sigma_{12} \in \mathbb{R}^{n-1} \text{ and } \Sigma_{22} \in \mathbb{R}.$$

The following algorithm generates a sample $\bar{X} \sim (X_1 | X_2 = x_2)$:

- (1) Sample $Z = (X_1, X_2)^T \sim N(\mu, \Sigma)$.
- (2) Compute $\bar{X} = X_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - X_2)$.

Note that the computational effort to produce the conditioned Gaussian random variable \bar{X} in Step 2 of Algorithm 2 is linear in the dimension n . Thus, this algorithm significantly reduces the computation time of Step 3 of Algorithm 1 when that step is repeated for each replica.

6 EFFICIENT GRIDS FOR A BROAD CLASS OF STOCHASTIC PROCESSES

In this section, we discuss how the idea of threshold-dependent grids can be applied to stochastic processes other than Brownian Motion. We let $(X_t)_{t \in [0,1]}$ be a real-valued stochastic process and $t^*(b) := \arg \max_{t \in [0,1]} \mathbb{P}(X_t > b)$. For simplicity, we here assume that $t \mapsto \mathbb{P}(X_t > b)$ is continuous and strictly increasing so that $t^*(b) = 1$ (but situations in which $t^*(b) \in (0, 1)$ can be dealt with similarly, see also the discussion in Section 7).

As argued in the previous sections, it is efficient to let the position of the grid points depend on b . We constructed for Brownian Motion a grid by finding $T(b) = \{t_1(b), \dots, t_n(b)\}$ such that

$$\mathbb{P}(\tau_b \in (0, t_k(b)] | \tau_b \in (0, 1]) = \frac{k}{n} \quad (18)$$

(cf. (11)). An inherent problem is that the class of processes for which the distribution of τ_b is known is very limited, so that the approach does not seem to be useful for relevant stochastic processes other than Brownian Motion. We saw, however, that for Brownian Motion the $t_k(b)$ satisfying Eq. (18) also solve

$$\frac{\mathbb{P}(X_{t_k(b)} > b)}{\mathbb{P}(X_1 > b)} = \frac{k}{n} \quad (19)$$

(cf. (12)). The idea now is to use the level-dependent (or, “equiprobable”) grid (19) for general real-valued processes. The major advantage of the grid (19) is that to calculate the position of the grid points t_k the sole prerequisite is that the process’ *marginals* are known (rather than the distribution of τ_b). In addition, even if the marginal distributions of X_t are not available, but the *asymptotics* of $\mathbb{P}(X_t > b)$ (as $b \rightarrow \infty$) are, then a good approximation of this grid can be found. (In the sequel we write, for brevity, $T = \{t_1, \dots, t_n\}$ instead of $T(b) = \{t_1(b), \dots, t_n(b)\}$) and t^* instead of $t^*(b)$.)

We now provide the rationale behind the grid (19). Let T be a grid such that $t^* \in T$. Evidently, by the union bound,

$$\mathbb{P}(X_{t^*} > b) \leq w_T(b) \leq \sum_{t \in T} \mathbb{P}(X_t > b).$$

Now notice that if the grid T is such that for $t \in T \setminus \{t^*\}$

$$\mathbb{P}(X_t > b) = o(\mathbb{P}(X_{t^*} > b)), \quad \text{as } b \rightarrow \infty, \quad (20)$$

then it does not make sense to include the point t for large b . Property (20) clearly compromises the performance of equidistant grids as $b \rightarrow \infty$. Considering, however, the grid points t_k of the threshold-dependent grid, as defined by Eq. (19), these will by design not experience (20).

To assess the performance of the above threshold-dependent grid (19), we introduce a measure of performance closely related to the relative bias. Note that when no formulas for $w(b)$ are available, nor is it known how to reliably approximate $w(b)$, we cannot determine the exact value of the relative bias. We now make the following two observations. (1) As $w_T(b) < w(b)$ for any choice of T , the larger $w_T(b)$ is, the better; if $w_{T_1}(b) > w_{T_2}(b)$ for grids T_1, T_2 , then also $\beta_{T_1}(b) < \beta_{T_2}(b)$. (2) The crude lower bound $w(b) \geq \mathbb{P}(X_{t^*} > b)$ provides us with a useful benchmark. Combining these two thoughts motivates the following performance measure of a grid T :

$$\gamma_T(b) := \frac{w_T(b)}{\mathbb{P}(X_{t^*} > b)}.$$

Notice that for any T such that $t^* \in T$, we have

$$\gamma_T(b) \in \left[1, \frac{w(b)}{\mathbb{P}(X_{t^*} > b)} \right].$$

What is more, for any two grids T_1, T_2 we have $\gamma_{T_1}(b) \geq \gamma_{T_2}(b)$ if and only if $\beta_{T_1}(b) \leq \beta_{T_2}(b)$; this means that the bigger the $\gamma_T(b)$ is, the better. As our main aim is to efficiently approximate $w(b)$ using discrete-time approximations $w_T(b)$, we see that if $\gamma_T(b) \approx 1$, then there is little gain from using $w_T(b)$ over a deterministic estimator $\mathbb{P}(X_{t^*} > b)$.

In a series of examples we compare $\gamma_T(b)$ induced by (i) the threshold-dependent (equiprobable) grid and (ii) the equidistant grid of the same size; we consistently use $n = 100$ grid points. In all cases $t \mapsto \mathbb{P}(X_t > b)$ is a continuous, strictly increasing function (so that $t^* = 1$). The most important conclusion is that the experiments below uniformly indicate that the equiprobable grid outperforms the equidistant one, not only in the asymptotic regime, as threshold b grows large, but already for moderate values of b . This shows how the ideas we developed earlier in this article, that have provable optimality properties for Brownian Motion, lead to an efficient estimation procedure for a much broader class of stochastic processes. In all examples, we observe that $\gamma_T(b)$ induced by the equidistant grid converges to 1, thus the corresponding $w_T(b)$ is asymptotically equivalent to $\mathbb{P}(X_{t^*} > b)$, as $b \rightarrow \infty$.

Example 1 (Brownian Motion with Jumps). Let $(X_t)_{t \in [0,1]}$ be a Brownian Motion with jumps, i.e.,

$$X_t := B_t + N_t, \quad (21)$$

where B_t is a standard Brownian Motion and N_t is a standard Poisson process with intensity $\lambda = 1$.

Even though there are no closed-form expressions for $w(b)$, it is still possible to generate exact samples from $\sup_{t \in [0,1]} X_t$ (see Dębicki and Mandjes (2015, Section 10.1)). We can use this to construct an unbiased estimator of $w(b)$ and thus can estimate the relative bias of the tested grids. The results in Figure 5 show the substantial gain achieved by the level-dependent grid. The graphs look similar to those of Brownian Motion, which is indicative of the threshold-dependent grid having a uniformly bounded relative bias.

Example 2 (Ornstein-Uhlenbeck Process). Let $(X_t)_{t \in [0,1]}$ be an Ornstein-Uhlenbeck process, i.e., a strong solution to the following SDE: with $X_0 = 0$,

$$dX_t = -X_t dt + dW_t.$$

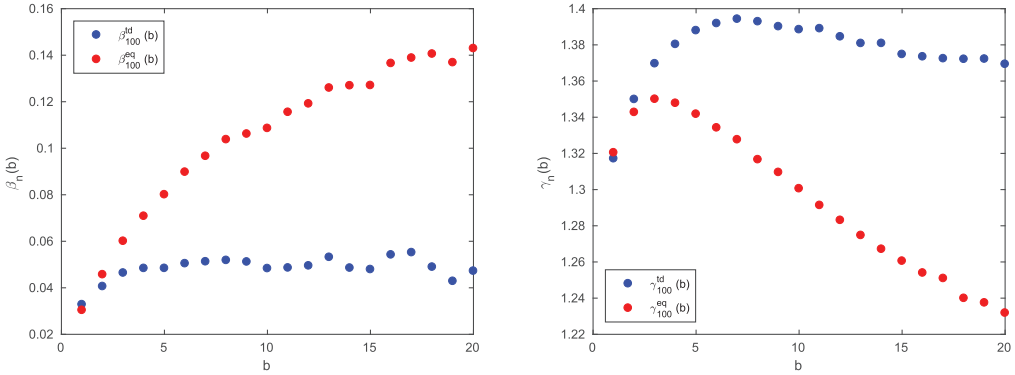


Fig. 5. Brownian Motion with jumps. Plots of $\beta_n(b)$ (left) and $\gamma_n(b)$ (right) as a function of the threshold b for threshold-dependent and equidistant grids of size $n = 100$.

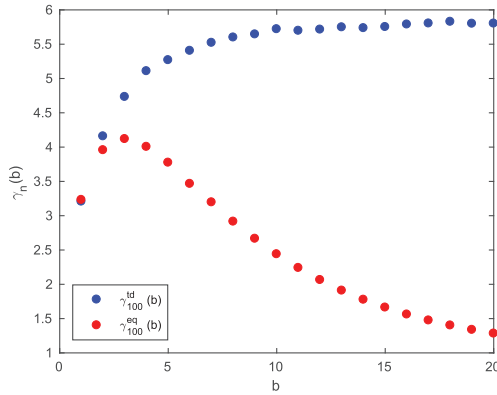


Fig. 6. Ornstein-Uhlenbeck process. Plot of $\gamma_n(b)$ as a function of the threshold b for threshold-dependent and equidistant grids of size $n = 100$. Notice that with growing b the equidistant estimator converges to $\mathbb{P}(X_1 > b)$.

Then $(X_t)_{t \in [0, 1]}$ is a zero-mean Markovian Gaussian process with covariance function

$$c(s, t) := \text{Cov}(X_s, X_t) = \frac{1}{2} \left(e^{-|t-s|} - e^{-(t+s)} \right).$$

The exact value of $w(b)$ is known only in terms of special functions (see Alili et al. (2005)) and it is not straightforwardly evaluated. However, the exact asymptotics of $w(b)$, as b grows large, are known:

$$w(b) = C \mathbb{P}(X_1 > b)(1 + o(1)), \text{ as } b \rightarrow \infty,$$

where C is a positive constant independent of b (see, e.g., Dębicki and Mandjes (2003, Theorem 5.1) or the original theorem by Piterbarg and Prisyazhnyuk (1978)); this explains why for the level-dependent grid $\gamma_n(b)$ goes to a constant in Figure 6. Again the equidistant grid is significantly outperformed by the threshold-dependent grid.

Example 3 (Fractional Brownian Motion). Let $(X_t)_{t \in [0, 1]}$ be a fractional Brownian Motion (fBM) with a Hurst parameter $H \in (0, 1)$, that is a zero-mean Gaussian process with the covariance

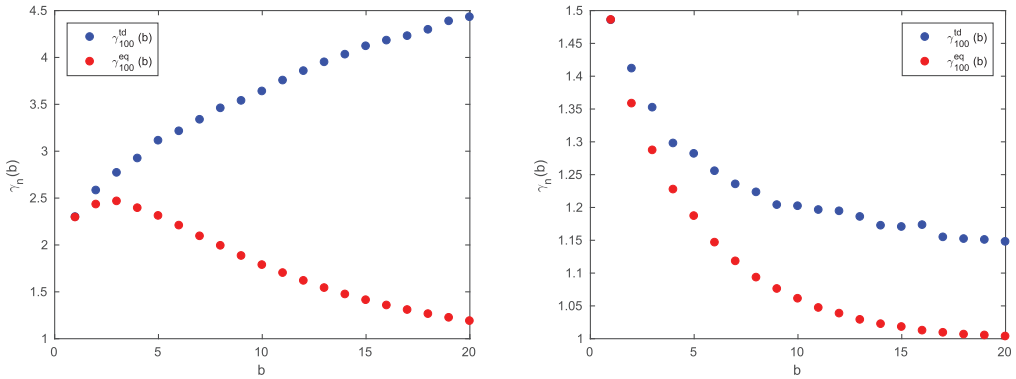


Fig. 7. fBm with Hurst parameter $H = 0.4$ (left) and $H = 0.6$ (right). Plot of $\gamma_n(b)$ as a function of the threshold b for threshold-dependent and equidistant grids of size $n = 100$.

function

$$C_H(s, t) := \text{Cov}(X_s, X_t) = \frac{1}{2} \left(s^{2H} + t^{2H} - |t - s|^{2H} \right).$$

Observe that fBm with Hurst parameter $H = 1/2$ is a standard Brownian Motion. For any H , we have $C_H(t, t) = t^{2H}$ (strictly increasing variance in time) and thus $t^* = 1$.

The exact value of the probability $w(b)$ for $H \neq 1/2$ remains unknown. However, like in Example 2, the exact asymptotics of $w(b)$ are known:

$$w(b) = \begin{cases} C_H b^{1/H-2} \mathbb{P}(X_1 > b)(1 + o(1)), & \text{for } H \in (0, \frac{1}{2}) \\ \mathbb{P}(X_1 > b)(1 + o(1)), & \text{for } H \in (\frac{1}{2}, 1), \end{cases}$$

where C_H is a constant only depending on H ; we again refer to Dębicki and Mandjes (2003, Theorem 5.1) or the original theorem by Piterbarg and Prisyazhnyuk (1978). We apply threshold-dependent grids in these two different asymptotic regimes for $H = 0.4$ and $H = 0.6$ see the results in Figure 7. Again the threshold-dependent grid performs considerably better. In case $H = 0.4$, the above asymptotic result explains why for the level-dependent grid $\gamma_n(b)$ keeps increasing ($w(b)/\mathbb{P}(X_1 > b)$ behaves as the increasing function $b^{1/H-2}$). In case $H = 0.6$, again using the asymptotic result, $\gamma_n(b) \rightarrow 1$ as b grows large, both for the equidistant grid and for the threshold-dependent grid (equivalently, the relative bias vanishes for both as $b \rightarrow \infty$). Note, however, that with the threshold-dependent grid, $\gamma_n(b)$ tends to 1 slower than with the equidistant grid, as can be seen in Figure 7 (right panel), showing the more favorable performance of the threshold-dependent grid.

7 CONCLUDING REMARKS AND DISCUSSION

In this article, we have demonstrated that the errors due to time discretization when estimating threshold-crossing probabilities $w(b)$ can be significantly reduced by using other grids than the commonly used equidistant grid. We have analyzed this in considerable detail for the case of standard Brownian Motion. In particular, we have shown that in order to control the error as b grows large, it suffices to properly *shift* the grid points instead of refining the grid with more and more points. At the same time, controlling the error using equidistant grids requires *quadratic* growth of the number of grid points, as b grows large.

Numerical estimation is evidently not needed for Brownian Motion due to the availability of analytical results. Our article, however, indicates that the underlying ideas can be used to construct

efficient grids for a broad class of stochastic processes (notably, Lévy processes and Gaussian processes, such as fractional Brownian Motion). The results presented in this article are intended to develop valuable insight and useful heuristics for tackling the estimation of tail probabilities of these more general classes of processes. We have demonstrated such heuristics for several processes in Section 6. There, we presented a procedure that is empirically shown to work well for stochastic process $(X_t)_{t \in [0,1]}$ of which the marginal distributions are known:

(i) Identify

$$t^*(b) := \arg \max_{t \in [0,1]} \mathbb{P}(X_t > b);$$

in case $(X_t)_{t \in [0,1]}$ is a zero-mean Gaussian process, t^* is a point of maximal variance, i.e., $\arg \max_{t \in [0,1]} \text{Var } X_t$. As argued, for many key models we have that $t^* = 1$.

(ii) Construct a grid $T = \{t_1, \dots, t_n\}$ clustered around $t^*(b)$, such that t_k solves Eq. (19), for $k \in \{1, \dots, n\}$.

As we pointed out, even if the marginal distribution of X_t is not available but only the corresponding asymptotics, as $b \rightarrow \infty$, this procedure can be applied. It is also noted that it is straightforward to compare two different grids: the larger the value of $w_T(b)$, the closer it is to the target quantity $w(b)$.

A natural question that arises in relation to Theorem 2 is whether we can find a grid that is even better than the one defined in Eq. (9). Constructing an *optimal n -grid* $T_n^*(b)$, i.e., a grid of size n that minimizes the relative bias for a given b , remains elusive. However, we have been able to find an explicit formula for an optimal 2-grid, namely, $T_2^*(b) = \{t_1^*(b), t_2^*(b)\}$, with

$$t_1^*(b) = \frac{\pi b^2}{4} \left(\sqrt{1 + \frac{8}{\pi b^2}} - 1 \right), \quad \text{and} \quad t_2^*(b) = 1,$$

where $\lim_{b \rightarrow \infty} \beta_{T_2^*(b)}(b) = 1 - \frac{1}{2} \Phi(\sqrt{2/\pi}) - \frac{1}{4} e^{-1/\pi} \approx 0.4244$. For comparison, the threshold-dependent grid defined in Eq. (9) yields $\lim_{b \rightarrow \infty} \beta_2(b) = \frac{3}{8} + \frac{1}{2} \Phi(-\sqrt{2 \log 2}) \approx 0.4348$, hence the grid (9) is not minimizing the bias (although the difference with the optimal 2-grid is small). Additionally, we were able to prove that for an optimal n -grid, $T_n^*(b) = \{t_1^*(b), \dots, t_n^*(b)\}$, the limits $\lim_{b \rightarrow \infty} b^2(1 - t_k^*(b))$ must exist, and are all finite and pairwise distinct. As a result, we were able to numerically calculate the limit $\lim_{b \rightarrow \infty} \beta_{T_3^*(b)} \approx 0.3796$. Finding optimal grids for larger n remains an open problem. We note, however, that with the threshold-dependent grid we can bound the relative bias uniformly in b (see Theorem 2) and in this sense the grid (9) is already (asymptotically) optimal.

8 PROOFS OF LEMMAS 1, 4, AND 5

PROOF OF LEMMA 1. Notice that the events $\{\sup_{t \in [0,1]} B_t > b\}$ and $\{\tau_b \in (0, 1]\}$ are equivalent. We thus find

$$\begin{aligned} w(b) \beta_T(b) &= \mathbb{P} \left(\sup_{t \in T} B_t < b, \sup_{t \in [0,1]} B_t > b \right) = \mathbb{P} \left(\sup_{t \in T} B_t < b, \tau_b \in [0, 1] \right) \\ &= \sum_{j=1}^n \mathbb{P} \left(\sup_{t \in \{t_j, \dots, t_n\}} B_t < b, \tau_b \in (t_{j-1}, t_j] \right) = \sum_{j=1}^n \int_{t_{j-1}}^{t_j} \mathbb{P} \left(\sup_{t \in \{t_j, \dots, t_n\}} B_t < b \mid B_s = b \right) \mathbb{P}(\tau_b \in ds) \\ &= \sum_{j=1}^n \int_{t_{j-1}}^{t_j} \mathbb{P}(B_{t_j-s} < 0, \dots, B_{t_n-s} < 0) \mathbb{P}(\tau_b \in ds). \end{aligned}$$

To prove the upper bound, we use the fact that $\mathbb{P}(B_{t_j-s} < 0, \dots, B_{t_n-s} < 0)$ is a non-increasing function of $s \in [t_{j-1}, t_j]$ (see Appendix A, Transformation (T2)), so that

$$\begin{aligned} w(b) \beta_T(b) &\leq \sum_{j=1}^n \int_{t_{j-1}}^{t_j} \mathbb{P}\left(B_{t_j-t_{j-1}} < 0, \dots, B_{t_n-t_{j-1}} < 0\right) \mathbb{P}(\tau_b \in ds) \\ &= \sum_{j=1}^n \mathbb{P}\left(B_{t_j-t_{j-1}} < 0, \dots, B_{t_n-t_{j-1}} < 0\right) \cdot \mathbb{P}(\tau_b \in (t_{j-1}, t_j]). \end{aligned}$$

Dividing both sides of the inequality by $w(b) = \mathbb{P}(\tau_b \in (0, 1])$ gives $\beta_T(b) \leq \bar{\beta}_T(b)$. To prove the lower bound, we use (B.IV) in Appendix B, so as to obtain

$$\begin{aligned} w(b) \beta_T(b) &= \sum_{j=1}^n \int_{t_{j-1}}^{t_j} \mathbb{P}\left(B_{t_j-s} < 0, \dots, B_{t_n-s} < 0\right) \mathbb{P}(\tau_b \in ds) \\ &\geq \sum_{j=1}^{n-1} \int_{t_{j-1}}^{t_j} \frac{1}{2} \mathbb{P}\left(B_{t_{j+1}-t_j} < 0, \dots, B_{t_n-t_j} < 0\right) \mathbb{P}(\tau_b \in ds) + \frac{1}{2} \mathbb{P}(\tau_b \in (t_{n-1}, t_n]) \quad (22) \\ &\geq \sum_{j=1}^{n-1} \frac{1}{2} \mathbb{P}\left(B_{t_{j+1}-t_j} < 0, \dots, B_{t_n-t_j} < 0\right) \cdot \mathbb{P}(\tau_b \in (t_{j-1}, t_j]) + \frac{1}{2} \mathbb{P}(\tau_b \in (t_{n-1}, t_n]). \end{aligned}$$

Dividing both sides of the inequality by $w(b)$ leads to $\beta_T(b) \geq \underline{\beta}_T(b)$ and concludes the proof. \square

PROOF OF LEMMA 4. Let $h := \max_{k=1, \dots, n} (t_k - t_{k-1})$. We transform the grid $T = \{t_1, \dots, t_n\}$ with Transformations (T1)–(T3) (see Appendix A) in such a way that after all transformations we end up with $\{h, \dots, Nh\}$.

- (1) Using Transformation (T2), translate the grid to the right by $h - t_1$, i.e., put

$$t_j := t_j + h - t_1 \quad \text{for all } j \in \{1, \dots, n\}.$$

- (2) Put $\sigma_1 := 1$, $c_1 = 1$, and $k := 2$. While $k \leq N$ do the following:

- Put $\sigma_k := \inf\{j : t_j \geq kh\}$.
- Using Transformation (T3), contract the grid after time $t_{\sigma_{k-1}}$ by a factor c_k , where c_k is defined by $h/(t_{\sigma_k} - t_{\sigma_{k-1}})$. Formally, we put

$$t_j := \begin{cases} t_j, & j \in \{1, \dots, \sigma_{k-1}\} \\ t_{\sigma_{k-1}} + c_k(t_j - t_{\sigma_{k-1}}), & j \in \{\sigma_{k-1} + 1, \dots, n\}. \end{cases}$$

Notice that after this operation, $t_{\sigma_k} = kh$.

- Put $k := k + 1$.

- (3) Using Transformation (T1), delete all the points t_k such that $t_k \notin \{h, \dots, hN\}$.

Now we prove that the algorithm is well-defined, more precisely, we confirm that all σ_k 's exist. First, see that σ_1 is well-defined. By induction, assume that σ_k is well-defined and prove that σ_{k+1} is well-defined as well. Notice that after the k th loop in Step 2 of the algorithm, the distances between the points shrunk at most by a factor $p_k = \prod_{j=1}^k c_j$ compared with the initial maximal distance h . Moreover, we observe that

$$c_k = \frac{h}{t_{\sigma_k} - t_{\sigma_{k-1}}} \geq \frac{h}{h + (t_{\sigma_k} - t_{\sigma_{k-1}})} \geq \frac{h}{h + \max_{j>\sigma_{k-1}} |t_{j+1} - t_j|} \geq \frac{1}{1 + \prod_{j=1}^{k-1} c_j}. \quad (23)$$

We prove by induction that $p_k = \prod_{j=1}^k c_j \geq \frac{1}{k}$ for all $k \in \{1, \dots, N\}$. Obviously, $p_2 = c_2 \geq \frac{1}{2}$. Assume that $p_{k-1} \geq \frac{1}{k-1}$. After multiplying inequality (23) by p_{k-1} , we obtain

$$p_k \geq \frac{p_{k-1}}{1 + p_{k-1}} = 1 - \frac{1}{1 + p_{k-1}} \geq 1 - \frac{1}{1 + \frac{1}{k-1}} = \frac{1}{k},$$

which ends the inductive proof. Next, in order to show that σ_{k+1} is well-defined for $k \in \{1, \dots, N-1\}$ it suffices to prove that the endpoint t_n , after the k th loop of Step 2, is greater than $h(k+1)$. We prove a stronger statement, namely, that the endpoint t_n after being shrunk by a factor p_k is still greater than $h(k+1)$, i.e., $h(k+1) \leq t_n p_k$. By the definition of N , h satisfies the inequality $h \leq t_n / N^2$, thus

$$h(k+1) \leq \frac{t_n(k+1)}{N^2} = \frac{t_n(k+1)}{N^2 p_k} p_k = \frac{k(k+1)}{N^2} t_n p_k \leq t_n p_k,$$

which concludes the proof that σ_{k+1} is well-defined. As all transformations used in steps 1–3 have property (27), we have

$$\mathbb{P}(B_{t_1} > 0, \dots, B_{t_n} > 0) \leq \mathbb{P}(B_h > 0, \dots, B_{N_h} > 0).$$

We finish the proof by observing that $\mathbb{P}(B_h > 0, \dots, B_{N_h} > 0) = \mathbb{P}(B_1 > 0, \dots, B_N > 0)$, due to the scaling property of Brownian Motion. \square

PROOF OF LEMMA 5. Notice that the grid points $t_k^n(b)$ defined in Eq. (9) depend only on the threshold b and the ratio $\frac{k}{n} \in [0, 1]$. We are able to extend the definition of $t_k^n(b)$ to $t : (0, 1] \times (0, \infty) \rightarrow [0, 1]$,

$$t(s, b) := \left(\frac{b}{\Phi^{-1}(s \Phi(-b))} \right)^2$$

such that $t_k^n(b) = t(\frac{k}{n}, b)$. Equivalently, $t(s, b)$ can be defined as the unique solution to

$$\Phi\left(-\frac{b}{\sqrt{t(s, b)}}\right) = s \Phi(-b). \quad (24)$$

This extension makes it possible to inspect the derivative of $t_k^n(b)$ with respect to the ratio $\frac{k}{n}$. Using the extension function of $t_k^n(b)$, we aim to prove the more general statement that for $0 < s_1 < s_2 < 1$,

$$\frac{1 - t(s_1, b)}{t(s_2, b) - t(s_1, b)} \geq \frac{-\log s_1}{\log s_2 - \log s_1} \iff \frac{1 - t(s_1, b)}{-\log s_1} \leq \frac{1 - t(s_2, b)}{-\log s_2}. \quad (25)$$

Moreover, using the definition (24) we may substitute

$$s = \Phi\left(-\frac{b}{\sqrt{t(s, b)}}\right) / \Phi(-b)$$

and arrive at another equivalent form of inequality (25):

$$\frac{1 - t(s_1, b)}{\log(\Phi(-b)) - \log(\Phi(-b/\sqrt{t(s_1, b)})} \leq \frac{1 - t(s_2, b)}{\log(\Phi(-b)) - \log(\Phi(-b/\sqrt{t(s_2, b)})}, \quad (26)$$

which follows from (B.IX) in Appendix B. For part (b), see that the density of the first hitting time,

$$\mathbb{P}(\tau_b \in ds) = \frac{b}{\sqrt{2\pi}} s^{-3/2} e^{-b^2/(2s)} ds, \quad \text{for } s > 0,$$

is an increasing function on the interval $s \in [0, \frac{b^2}{3}]$ and thus part (b) follows from the second definition of the grid points $t_k^n(b)$ in Eq. (10). \square

APPENDICES

A GRID TRANSFORMATIONS

Let $T = \{t_1, \dots, t_n\}$ with $0 < t_1 < \dots < t_n < \infty$. We introduce three *grid transformations*, i.e., operations $T \mapsto \tilde{T}$ satisfying

$$\mathbb{P}(B_t > 0 \text{ for all } t \in T) \leq \mathbb{P}(B_t > 0 \text{ for all } t \in \tilde{T}). \quad (27)$$

(T1) **Deleting.** For any $k \in \{1, \dots, n\}$

$$\mathbb{P}(B_{t_1} > 0, \dots, B_{t_n} > 0) \leq \mathbb{P}(B_{t_1} > 0, \dots, B_{t_{k-1}} > 0, B_{t_{k+1}} > 0, \dots, B_{t_n} > 0).$$

(T2) **Translation to the right of the whole sequence.** For any $s > 0$

$$\mathbb{P}(B_{t_1} > 0, \dots, B_{t_n} > 0) \leq \mathbb{P}(B_{t_1+s} > 0, \dots, B_{t_n+s} > 0).$$

(T3) **Contraction of time after some point.** For any $k \in \{1, \dots, n-1\}$ and $c \in (0, 1)$

$$\mathbb{P}(B_{t_1} > 0, \dots, B_{t_n} > 0) \leq \mathbb{P}(B_{t_1} > 0, \dots, B_{t_k} > 0, B_{t_k+c(t_{k+1}-t_k)} > 0, \dots, B_{t_k+c(t_n-t_k)} > 0).$$

The proof that transformations (T1)–(T3) satisfy (27) is included in Appendix C.

B MISCELLANEOUS RESULTS

Let $\Phi(\cdot)$ denote the standard normal cumulative distribution function and $\phi(\cdot)$ the standard normal density function. Below we list various results that we use throughout this article. Results (B.I)–(B.III) are standard; the proofs of the remaining results are included in Appendix C.

(B.I) For $x > 0$:

$$\frac{x}{1+x^2} \leq \frac{\Phi(-x)}{\phi(x)} \leq \frac{1}{x}. \quad (28)$$

(B.II) As $x \rightarrow \infty$,

$$\lim_{x \rightarrow \infty} \frac{\Phi(-x)}{\frac{1}{x}\phi(x)} \rightarrow 1.$$

(B.III) (Szarek and Werner 1999). For $x > -1$:

$$\frac{2}{x + (x^2 + 4)^{1/2}} \leq \frac{\Phi(-x)}{\phi(x)} \leq \frac{4}{3x + (x^2 + 8)^{1/2}}. \quad (29)$$

(B.IV) Let $0 < t_1 < \dots < t_n < \infty$, then

$$\mathbb{P}(B_{t_1} > 0, \dots, B_{t_n} > 0) \geq \frac{1}{2} \mathbb{P}(B_{t_2-t_1} > 0, \dots, B_{t_n-t_1} > 0).$$

(B.V) Let $T = \{t_1, \dots, t_n\}$, where $t_j := \frac{j}{n}$, $\tau_b := \inf\{t \geq 0 : B_t \geq b\}$ and $b > 0$, then

$$\mathbb{P}(\tau_b \in (t_{j-1}, t_j] \mid \tau_b \in (0, 1]) \leq \frac{b(b + \sqrt{b^2 + 4})}{4} \cdot \frac{\sqrt{n}}{(j-1)^{3/2}} e^{-\frac{b^2}{2} \cdot (\frac{n}{j-1})}$$

and

$$\mathbb{P}(\tau_b \in (t_{j-1}, t_j] \mid \tau_b \in (0, 1]) \geq \frac{b(3b + \sqrt{b^2 + 8})}{8} \cdot \frac{\sqrt{n}}{j^{3/2}} e^{-\frac{b^2}{2} \cdot (\frac{n}{j-1})}$$

for $j \in \{2, \dots, n\}$.

(B.VI) Let $f : (0, \infty) \times (0, 1) \rightarrow (0, \infty)$ such that

$$f(b, x) := \frac{b}{\sqrt{1-x}} x^{-3/2} e^{-\frac{b^2}{2}(1/x-1)}.$$

Then $f(b, x)$ is an increasing function of x , when $b \geq 1$.

(B.VII) Let $f : (0, \infty) \rightarrow (0, \infty)$ such that

$$f(x) := \frac{\Phi(-x)}{\phi(x)}.$$

Then f is a strictly decreasing function.

(B.VIII) Let $f : (0, \infty) \rightarrow (0, \infty)$ such that

$$f(x) := \frac{\Phi(-x)}{\frac{1}{x}\phi(x)}.$$

Then f is a strictly increasing function.

(B.IX) Let $f : [0, 1] \rightarrow [0, \infty)$ be such that

$$f(t) := \begin{cases} 0, & t = 0; \\ \frac{1-t}{\log(\Phi(-b)) - \log(\Phi(-b/\sqrt{t}))}, & t \in (0, 1); \\ \frac{2\Phi(-b)}{b\phi(b)}, & t = 1. \end{cases}$$

Then f is continuous and increasing.

C SUPPLEMENTARY MATERIALS

Supplementary materials consist of (i) proof of Proposition 1, (ii) proof of Lemma 3, (iii) proof that Transformations (T1)–(T3) have property (27), and (iv) proofs of Results (B.IV)–(B.IX) and are available in the online version of this manuscript.

ACKNOWLEDGMENTS

The authors would like to thank Ankush Agarwal and Johan van Leeuwen for useful discussions and suggestions. In addition, the reviewers' reports helped improve the quality of our work considerably.

REFERENCES

- R. J. Adler. 1990. *An Introduction to Continuity, Extrema, and Related Topics for General Gaussian Processes*. Institute of Mathematical Statistics.
- R. J. Adler, J. H. Blanchet, and J. Liu. 2012. Efficient Monte Carlo for high excursions of Gaussian random fields. *The Annals of Applied Probability* 22, 3 (2012), 1167–1214.
- L. Alili, P. Patie, and J. L. Pedersen. 2005. Representations of the first hitting time density of an Ornstein-Uhlenbeck process. *Stochastic Models* 21, 4 (2005), 967–980.
- S. Asmussen, P. Glynn, and J. Pitman. 1995. Discretization error in simulation of one-dimensional reflecting Brownian motion. *The Annals of Applied Probability* 5 (1995), 875–896.
- M. Broadie, P. Glasserman, and S. Kou. 1997. A continuity correction for discrete barrier options. *Mathematical Finance* 7, 4 (1997), 325–349.
- J. M. Calvin. 1997. Average performance of a class of adaptive algorithms for global optimization. *The Annals of Applied Probability* 7 (1997), 711–730.
- J. M. Calvin and P. W. Glynn. 1997. Average case behavior of random search for the maximum. *Journal of Applied Probability* 34, 3 (1997), 632–642.
- K. Dębicki and M. Mandjes. 2003. Exact overflow asymptotics for queues with many Gaussian inputs. *Journal of Applied Probability* 40, 3 (2003), 704–720.
- K. Dębicki and M. Mandjes. 2015. *Queues and Lévy Fluctuation Theory*. Springer.

- A. Doucet. 2010. A note on efficient conditional simulation of Gaussian distributions. *Departments of Computer Science and Statistics, University of British Columbia*.
- W. Feller. 1971. *An Introduction to Probability Theory and Its Applications*. Vol. II (2nd ed.). John Wiley & Sons.
- A. Genz and F. Bretz. 2009. *Computation of Multivariate Normal and t Probabilities*. Vol. 195. Springer Science & Business Media.
- M. B. Giles. 2008. Multilevel Monte Carlo path simulation. *Operations Research* 56, 3 (2008), 607–617.
- A. J. E. M. Janssen and J. S. H. Van Leeuwaarden. 2009. Equidistant sampling for the maximum of a Brownian motion with drift on a finite horizon. *Electronic Communications in Probability* 14 (2009), 143–150.
- X. Li and J. Liu. 2015. Rare-event simulation and efficient discretization for the supremum of Gaussian random fields. *Advances in Applied Probability* 47, 3 (2015), 787–816.
- P. Mörters and Y. Peres. 2010. *Brownian Motion*. Vol. 30. Cambridge University Press.
- V. I. Piterbarg and V. Prisyazhnyuk. 1978. Asymptotic analysis of the probability of large excursions for a nonstationary Gaussian process. *Theory of Probability and Mathematical Statistics* 18 (1978), 121–134.
- D. Siegmund. 1985. *Sequential Analysis: Tests and Confidence Intervals*. Springer-Verlag.
- S. J. Szarek and E. Werner. 1999. A nonsymmetric correlation inequality for Gaussian measure. *Journal of Multivariate Analysis* 68, 2 (1999), 193–211.

Received March 2017; revised September 2017; accepted January 2018