Approximations for reflected fractional Brownian motion

Artagan Malsagov and Michel Mandjes
Korteweg-de Vries Institute for Mathematics, University of Amsterdam, Science Park 105, 1098 XH Amsterdam, The Netherlands

(Received 5 March 2019; revised manuscript received 25 June 2019; published 13 September 2019)

Fractional Brownian motion is a widely used stochastic process that is particularly suited to model anomalous diffusion. We focus on capturing the mean and variance of fractional Brownian motion reflected at level 0. As explicit expressions or numerical techniques are not available, we base our analysis on Monte Carlo simulation. Our main findings concern closed-form approximations of the mean and variance, with a near-perfect fit.

DOI: 10.1103/PhysRevE.100.032120

I. INTRODUCTION

Owing to its ubiquitous applicability, diffusion is one of the core concepts in a broad range of scientific disciplines. It was initially intended to describe the trajectory of a moving particle [1], with evident applications in, e.g., physics and biology. In its conventional form, it is a zero-mean Gaussian process (also known as Brownian motion) with stationary independent increments, entailing that the mean-square displacement (MSD) of the moving particle, $\langle x_t^2\rangle$, grows linearly in t.

In many relevant systems, backed by extensive experimental observations, however, $\langle x_t^2\rangle$ behaves sub- or superlinearly, which case is usually referred to as *anomalous diffusion*. In this respect we mention, without attempting to provide an exhaustive list, the overviews [2] (including a historic account) and [3] and references therein. A convenient framework that covers sub- and superdiffusions is that of *fractional Brownian motion* (FBM), a zero-mean Gaussian process with stationary (but not independent) increments. For FBM the MSD grows (in time t) as a power law: $\langle x_t^2\rangle \sim t^\alpha$, for $\alpha \in (0,2]$. For $\alpha < 1$ (the subdiffusive case) the increments are negatively correlated, whereas for $\alpha > 1$ (the superdiffusive case) they are positively correlated. The special case of $\alpha = 1$ corresponds to conventional Brownian motion. FBM is self-similar, meaning that x_t has the same distribution as $t^{\alpha/2}x_1$. In the literature $\alpha/2$ is often referred to as the Hurst parameter H.

Anomalous diffusion has been observed in a broad variety of systems, including ultra-cold atoms [4] and single-particle movements in cytoplasm [5]. The subdiffusive case in particular is relevant in the context of, e.g., the motion of tracer particles in living biological cells [6], whereas the superdiffusive case occurs, e.g., in the presence of active motion in living biological cells [7]. Apart from the fields mentioned above, FBM has found widespread use in the modeling of, e.g., traffic in communication networks [8], hydrology [9], and finance [10,11].

FBM has attracted substantial attention in the mathematical literature [12,13], but despite this research effort there are still many open questions. In particular the case that the FBM is reflected at a boundary is poorly understood, the most prominent example being the case that the FBM is enforced to take non-negative values only. As we will point out below, FBM reflected at zero can be interpreted as a queueing system with FBM input [14]. Virtually all results for queues with FBM input are in terms of tail asymptotics of the corresponding probability distribution [15–19]. Perhaps surprisingly, no explicit results are available even for the mean $m_\alpha(t)$ and variance $v_\alpha(t)$ of the position of a reflected FBM (RFBM) at time t; the only generally applicable computational approach known is Monte Carlo simulation. Recently, in the interesting contribution [20], bounds have been established for $m_\alpha(t)$: positive constants Γ_ℓ and Γ_u are identified such that, across all $\alpha \in (0,2]$,

$$\frac{\Gamma_\ell}{\sqrt{\alpha}} \leqslant \frac{m_\alpha(t)}{t^{\alpha/2}} \leqslant \frac{\Gamma_u}{\sqrt{\alpha}}. \tag{1}$$

The fact that the constants Γ_ℓ and Γ_u are a factor 81.5 apart, and the lack of results for $v_\alpha(t)$, motivate the search for accurate approximations for $m_\alpha(t)$ and $v_\alpha(t)$, covering both the sub- and the superdiffusive regime. Our main findings concern approximations $\bar{m}_\alpha(t)$ and $\bar{v}_\alpha(t)$ with a nearly perfect fit, as confirmed by extensive simulation experiments. Remarkably, $\bar{m}_\alpha(t)$ is *not* of the form $\Gamma\, t^{\alpha/2}/\sqrt{\alpha}$ (for some $\Gamma > 0$), as would be suggested by (1). Our paper can be viewed as a continuation of the recent work [21], where the focus is on a simulation-based analysis of the distribution of the position of RFBM at a given point in time; cf. also Refs. [22,23]. One of the key findings of Ref. [21] was that this distribution is highly non-Gaussian, with intrinsically different behavior for $\alpha < 1$ and $\alpha > 1$: in the latter case particles tend to stick close to the barrier 0 (so that the corresponding density blows up close to 0), whereas in the former case values close to 0 are unlikely (so that the density goes to zero in this region).

II. REFLECTED FBM AND ITS DISCRETE COUNTERPART

Let x_t denote the FBM at time t, and q_t its reflected counterpart RFBM. The reflection is at level 0, and it concerns reflection from below, which entails that q_t attains non-negative values only. A common way to construct the process q_t from the process x_t is by solving a so-called *Skorokhod problem* ([24], Sec. 9.2). This effectively means that two conditions are imposed on the way the FBM x_t is mapped onto its reflected version q_t, and then it is shown that this leads to a unique

solution ([24], Proposition 9.2.2) for the process q_t in terms of the process x_t.

More concretely, the two conditions imposed in the Skorokhod approach are the following: (1) Given a path of x_t, its reflection q_t can be written as $x_t + \ell_t$, where ℓ_t (typically referred to as the *regulator*) is increasing in t. (2) The process ℓ_t increases only when the value of the RFBM equals 0, in the sense that, for any $T > 0$,

$$\int_0^T q_t \, d\ell_t = 0. \tag{2}$$

The regulator makes sure that q_t does not become negative. When $q_0 = 0$, Ref. ([24], Proposition 9.2.2) states that the unique solution to this Skorokhod problem (i.e., satisfying the requirements (1) and (2) that we imposed above) is

$$\ell_t = -\inf_{0\leqslant s\leqslant t} x_s, \quad q_t = \sup_{0\leqslant s\leqslant t}(x_t - x_s). \tag{3}$$

One could alternatively say that q_t is the workload level at time t of a *queue* (also sometimes called a *storage system*) with FBM input that has started empty at time 0 [14].

There is a convenient duality relation between the distribution of a reflected process (in our case the RFBM q_t) at time t on one hand, and the maximum value attained by the nonreflected process (in our case the FBM x_t) over the interval $[0, t]$ on the other hand. To this end, first observe that, due to time reversibility, $x_t - x_s$ is distributed as x_{t-s}. This implies that the supremum over $s \in [0, t]$ of $x_t - x_s$ has the same distribution as the supremum over $s \in [0, t]$ of x_s. We thus obtain the following relation between x_t and the running maximum of x_t:

$$\mathbb{P}(q_t \geqslant u) = \mathbb{P}\left(\sup_{0\leqslant s\leqslant t} x_s \geqslant u\right). \tag{4}$$

The right-hand side of the relation (4) can be simplified further, by making the dependence on t explicit. Applying the self-similarity that is inherent to FBM, we obtain that (4) equals

$$\mathbb{P}(q_t \geqslant u) = \mathbb{P}\left(\sup_{0\leqslant s\leqslant 1} x_s \geqslant \frac{u}{t^{\alpha/2}}\right) \tag{5}$$

$$= \mathbb{P}\left(q_1 \geqslant \frac{u}{t^{\alpha/2}}\right). \tag{6}$$

This means that, in order to analyze q_t for some $t \geqslant 0$, it suffices to consider q_1 only; cf. Ref. [21, Equation (4)]. In this paper our goal is to analyze $m_\alpha(t) = \langle q_t\rangle$ and $v_\alpha(t) = \langle q_t^2\rangle - \langle q_t\rangle^2$ for $t \geqslant 0$. However, noting that (6) implies that

$$m_\alpha(t) = t^{\alpha/2} m_\alpha(1), \quad v_\alpha(t) = t^\alpha v_\alpha(1), \tag{7}$$

without loss of generality we restrict ourselves to studying $m_\alpha(1) = \langle q_1\rangle$ and $v_\alpha(1) = \langle q_1^2\rangle - \langle q_1\rangle^2$. In other words, due to the self-similarity it suffices that we succeed in capturing the mean and variance of RFBM at time 1.

In our approach we first accurately estimate $m_\alpha(1)$ and $v_\alpha(1)$ (as functions of α) using Monte Carlo simulation. The second step in our procedure is to fit explicit functions of α to these simulation-based curves. To simulate q_1, we work with a discrete-time counterpart $\bar{q}_1$ [25], obtained by using a time grid consisting of n points (for n large enough to make sure that the error is sufficiently small; we get back to this issue in detail later in this paper). In this discrete-time framework, the reflection is usually not expressed as the solution to a Skorokhod problem, but rather as the solution to the *Lindley recursion* ([24], Sec. III.6). More specifically, the reflected process at time i/n can be expressed in terms of the reflected process at time $(i-1)/n$ through the recursion, for $i = 1, \ldots, n$,

$$\bar{q}_{i/n} = \max\{\bar{q}_{(i-1)/n} + \xi_i, 0\}, \tag{8}$$

with $\bar{q}_0 = 0$; note that this truncation at 0 provides a natural mechanism that enforces that the process $\bar{q}_{i/n}$ does not become negative. In the recursion the ξ_i are the increments of the FBM, which means that ξ_i is a sequence of zero-mean random variables such that

$$\langle \xi_i \, \xi_{i+j}\rangle = \frac{\sigma^2}{2n^\alpha}(|j-1|^\alpha - 2|j|^\alpha + |j+1|^\alpha). \tag{9}$$

The recursion (9) can be solved iteratively. Performing one step yields

$$\bar{q}_{i/n} = \max\{\max\{\bar{q}_{(i-2)/n} + \xi_{i-1}, 0\} + \xi_i, 0\} \tag{10}$$

$$= \max\{\bar{q}_{(i-2)/n} + \xi_{i-1} + \xi_i, \xi_i, 0\}. \tag{11}$$

Continuing along these lines, after i steps one arrives at (with $\Xi_i := \sum_{k=1}^i \xi_k$ and using $\bar{q}_0 = 0$), in self-evident notation,

$$\bar{q}_{i/n} = \max_{j\in\{1,\ldots,i\}}\left\{\sum_{k=j}^i \xi_k, 0\right\} = \max_{j\in\{0,\ldots,i\}}(\Xi_i - \Xi_j). \tag{12}$$

Observing the similarity with the solution q_t of the continuous-time Skorokhod problem as given by (3), we conclude that this discrete-time framework can be considered as the natural counterpart of the continuous-time framework that we introduced above.

As before, we can make use of time reversal, to conclude that we can represent the quantity of our interest, $\bar{q}_1$, as follows:

$$\bar{q}_1 =_{\mathrm{d}} \max_{i\in\{0,\ldots,n\}} x_{i/n}, \quad x_{i/n} := \Xi_i = \sum_{k=1}^i \xi_k, \tag{13}$$

with "$=_{\mathrm{d}}$" denoting equality in distribution; here the empty sum is defined as 0. Summarizing, the challenge that we are facing is to estimate $m_\alpha(1)$ and $v_\alpha(1)$ by simulating the partial sum process of the increments of a discrete-time version of FBM and recording its maximum value. In our approach we split the unit interval into n small intervals of length n^{-1}; we remark, however, that due to the self-similarity our procedure is conceptually equivalent to sampling a (properly rescaled version) of our FBM over a long period of time (namely, the time units 1 up to n). We refer to Fig. 1 for a simulated path of a free and reflected FBM; the top panel corresponds to the case that the increments are negatively correlated ($\alpha < 1$), whereas in the bottom panel they are positively correlated ($\alpha > 1$). The graphs show the effect of the truncation at 0.

We conclude this section with a few remarks related to the simulation of the discrete-time version of FBM. In the first place, we mention that Ref. [26] discusses various approaches to simulating the increment process ξ_i. We have used the

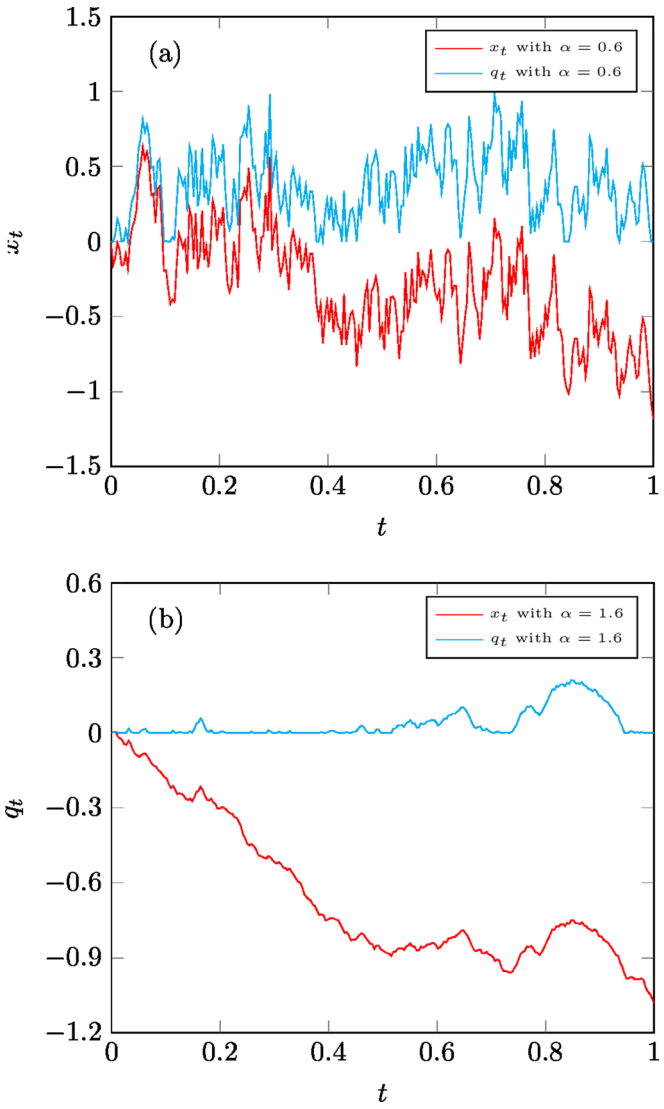

FIG. 1. Simulation of free FBM (x_t) and reflected FBM (q_t) on an equidistant grid with 2^8 points, for (a) $\alpha = 0.6$ and (b) $\alpha = 1.6$.

circular approach [27], which provides an *exact sample* of $\bar{q}_1$ (in the sense that its distribution exactly matches with that of $\bar{q}_1$). Without losing any generality, in the sequel we rescale space such that $\sigma^2 = 1$.

In the second place, it has been shown that when n is large q_1 and $\bar{q}_1$ are provably close. More concretely, the difference between q_1 and $\bar{q}_1$ is bounded by $\delta\sqrt{\log n}/n^{\alpha/2}$ for some positive constant δ [20]. This justifies the use of the above discretization procedure. Later in this paper we assess how large n should be to make sure that the gap is guaranteed to be below some predefined bound.

III. APPROXIMATIONS

In this section we first treat results for three boundary cases and then set up our approximations $\bar{m}_\alpha(1)$ and $\bar{v}_\alpha(1)$. The main results of our papers are closed-form approximations with a nearly perfect fit.

A. Boundary cases

The case $\alpha = 1$. We start with the Brownian case; recall that for $\alpha = 1$ the increments are independent. The distribution of q_1 can be explicitly computed; see, e.g., Ref. ([28], Ch. 3). It leads to

$$m_{\frac{1}{2}}(1) = \sqrt{\frac{2}{\pi}} \approx 0.79, \quad v_{\frac{1}{2}}(1) = 1 - \frac{2}{\pi} \approx 0.36. \tag{14}$$

The case $\alpha \uparrow 2$. We proceed by analyzing the case of "perfect positive correlation," i.e., $\alpha = 2$. In this case the path of x is a straight line with a random slope; put differently, an initial increment of the FBM, say, x_ϵ for some small ϵ, effectively determines x_t for all $t \geqslant 0$ and hence also q_1. As a consequence, q_1 is distributed as the maximum of a standard normal random variable and 0. Elementary calculus yields

$$m_1(1) = \sqrt{\frac{1}{2\pi}} \approx 0.40, \quad v_1(1) = \frac{1}{2} - \frac{1}{2\pi} \approx 0.34. \tag{15}$$

The case $\alpha \downarrow 0$. This case with "perfect negative correlation" has been recently considered in detail in Refs. [20,29]. There it is established that the FBM x_t converges (in terms of finite-dimensional distributions) to χ_t as $\alpha \downarrow 0$, where $\chi_t = (\zeta_t - \zeta_0)/\sqrt{2}$, with ζ_t (for $t \in [0, 1]$) independent standard normal random variables. We first look at the discrete-time approximation

$$\lim_{\alpha\downarrow 0} m_\alpha(1) \approx \lim_{\alpha\downarrow 0} \Big\langle \max_{i\in\{0,\ldots,n\}} x_{i/n} \Big\rangle, \tag{16}$$

the symbol "$\approx$" meaning that the approximation becomes increasingly accurate as $n \to \infty$. We thus obtain

$$\lim_{\alpha\downarrow 0} m_\alpha(1) \approx \frac{1}{\sqrt{2}} \langle A_n \rangle, \quad A_n := \max_{i\in\{0,\ldots,n\}} (\zeta_i - \zeta_0). \tag{17}$$

Observe that

$$\langle A_n \rangle = \langle B_n \rangle, \quad \text{where} \quad B_n := \max_{i\in\{0,\ldots,n\}} \zeta_i. \tag{18}$$

We now consider the corresponding variance. Denote the variance of a random variable r by $[r]$, i.e., $[r] := \langle r^2 \rangle - \langle r \rangle^2$. Hence,

$$\lim_{\alpha\downarrow 0} v_\alpha(1) \approx \tfrac{1}{2}[A_n] = \tfrac{1}{2}\langle A_n^2 \rangle - \tfrac{1}{2}\langle A_n \rangle^2; \tag{19}$$

again "$\approx$" indicates that the approximation becomes increasingly accurate as $n \to \infty$. It is straightforward to verify that

$$A_n^2 = B_n^2 - 2\zeta_0 B_n + \zeta_0^2. \tag{20}$$

It thus follows that the right-hand side of (19) equals

$$\tfrac{1}{2}\langle B_n^2 \rangle - \langle \zeta_0 B_n \rangle + \tfrac{1}{2} - \tfrac{1}{2}\langle B_n \rangle^2 = \tfrac{1}{2}[B_n] - \langle \zeta_0 B_n \rangle + \tfrac{1}{2}. \tag{21}$$

It directly follows from the main result of Ref. [30] that, remarkably,

$$\langle \zeta_0 B_n \rangle = \frac{1}{n+1}, \tag{22}$$

so that we end up with the approximation

$$\lim_{\alpha\downarrow 0} v_\alpha(1) \approx \frac{1}{2}[B_n] - \frac{1}{n+1} + \frac{1}{2}. \tag{23}$$

The next step is to approximate $\langle B_n \rangle$ and $[B_n]$, which we do using *extreme value theory* (EVT); we refer to, e.g., Ref. ([31], Chapter 3) for an accessible textbook treatment of this topic. EVT is a branch of probability and statistics that aims at describing the distribution of the maximum M_n of a sequence of n (independent and identically distributed) random variables, say, $Z_1, \ldots, Z_n$, in the domain that n grows large. A key result is the celebrated Fisher-Tippett-Gnedenko (FTG) theorem [31, Theorem 3.2.3] that states that there are sequences λ_n and s_n such that, as $n \to \infty$,

$$\frac{M_n - \lambda_n}{s_n} \tag{24}$$

converges in distribution to a nondegenerate random variable, say, M; this remarkable result goes back to the first half of the previous century. The nature of the limiting random variable M depends on the tail behavior of the underlying random variables $Z_1, \ldots, Z_n$. More particularly, there are three possible families of limiting distributions, namely, the Weibull, Gumbel, and Fréchet distribution.

In our setting we are interested in B_n, which corresponds to the maximum of independent random variables ζ_i that are (standard) normally distributed. In this case, the FTG theorem entails that the limiting distribution of $(B_n - \lambda_n)/s_n$ (for appropriately chosen sequences λ_n and s_n) is of Gumbel type, which concretely means the following. We let $F(x) := \exp(-e^{-x})$ denote the distribution function of the standard Gumbel random variable. Then, in line with the expressions presented in Ref. ([31], Example 3.3.29), we define

$$s_n := \frac{1}{\sqrt{2 \log n}}, \tag{25}$$

and

$$\lambda_n := \sqrt{2 \log n} - \frac{\log \log n + \log(4\pi)}{2\sqrt{2 \log n}}. \tag{26}$$

In this specific case we have that the FTG result corresponds to the following convergence in distribution:

$$\mathbb{P}\left(\frac{B_n - \lambda_n}{s_n} \leqslant u \right) \to F(u) \tag{27}$$

as $n \to \infty$ (for given u). In practical terms, (27) entails that for large n we can approximate B_n by $\lambda_n + s_n G$, with s_n given by (25), λ_n by (26), and G having a standard Gumbel distribution (i.e., G has the distribution function F defined above). As a random variable with a standard Gumbel distribution has a mean equal to the Euler-Mascheroni constant $\gamma \approx 0.5772$ and a variance equal to $\pi^2/6$, we thus find the following approximations for the mean $\langle B_n \rangle$ and the variance $[B_n]$:

$$\langle B_n \rangle \sim \lambda_n + \gamma s_n, \quad [B_n] \sim \frac{\pi^2}{6} s_n^2; \tag{28}$$

here $g_n \sim h_n$ means that $g_n/h_n \to 1$ as $n \to \infty$. Observe that, as $n \to \infty$, $\langle B_n \rangle \to \infty$ (but very slowly; roughly as $\sqrt{\log n}$) and $[B_n] \to 0$ (again rather slowly; essentially as $1/\log n$). This in particular yields, using (17), (18), and (23),

$$m_0(1) = \infty, \quad v_0(1) = \tfrac{1}{2} \tag{29}$$

[where we note that $m_0(1) = \infty$ already followed from the lower bound in (1)].

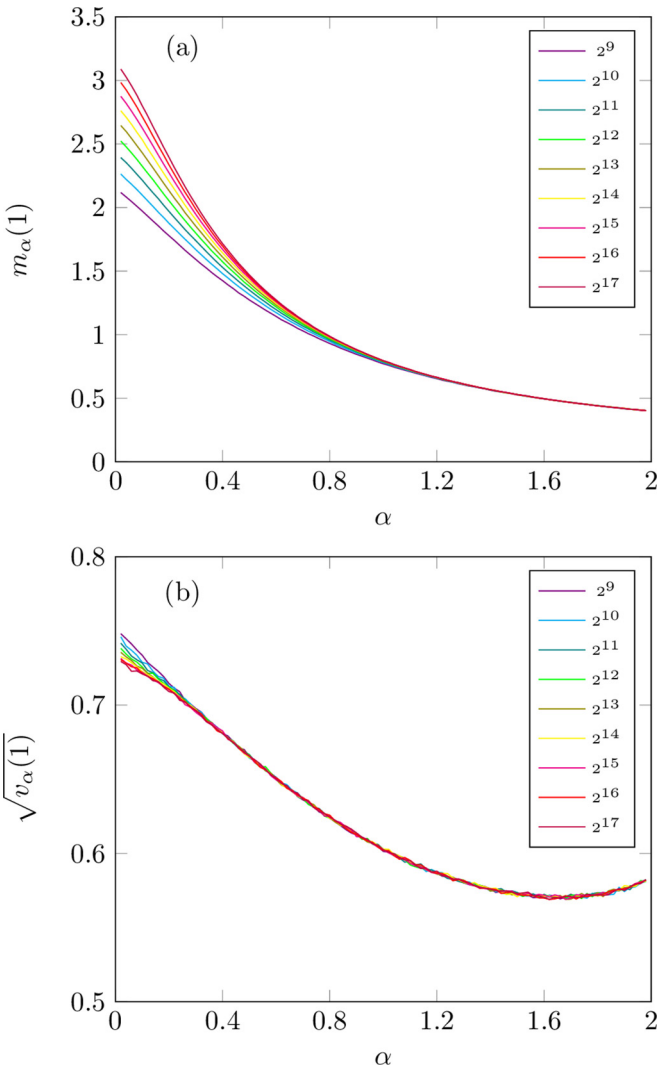

FIG. 2. Estimates of (a) $m_\alpha(1)$ and (b) $\sqrt{v_\alpha(1)}$ as functions of α, for $\alpha = 0.02, 0.04, \ldots, 1.98, 2.00$. An equidistant grid with $n = 2^k$ points has been used, with $k = 9, \ldots, 17$. Each estimate is based on 5×10^5 runs. The curves are essentially monotone in the number of grid points; in (a) the top curve corresponds to 2^{17} points and the bottom curve to 2^9 points, whereas in (b) the top curve corresponds to 2^9 points and the bottom curve to 2^{17} points.

B. Fitting procedure

Figure 2 presents estimates of $m_\alpha(1)$ and $v_\alpha(1)$ as functions of α, for different grids. We make the following observations.

Impact of number of sample points. Particularly for small α the numbers of sample points $n = 2^k$ has a substantial impact on the estimate of $m_\alpha(1)$. This aligns with the findings in the previous section for the case $\alpha \downarrow 0$. With the EVT-based results, we see that

$$\frac{\langle A_n \rangle}{\sqrt{2}} \approx \frac{\lambda_n}{\sqrt{2}} \approx \frac{\sqrt{2 \log n}}{\sqrt{2}} = \sqrt{k \log 2} \approx 0.832\sqrt{k}, \tag{30}$$

which nicely matches the simulation-based estimates (i.e., the values of the curves on the vertical axis).

Also regarding the estimates of $v_\alpha(1)$, the number of sample points has an impact, but it is less affected by n than

$m_\alpha(1)$. For large k, we observe that our estimate of $\sqrt{v_0(1)}$ approaches the $\frac{1}{2}\sqrt{2}$ that we found in (29).

Our simulation experiments show that from $k = 17$ on, the estimates of $m_\alpha(1)$ and $v_\alpha(1)$ do not significantly change anymore (for α in the interval $[0.1, 2]$).

Variance is not monotone. Perhaps surprisingly, $v_\alpha(1)$ is *not* monotone in α; this is in contrast with $m_\alpha(1)$, which is monotonically decreasing. Figure 2 shows that the variance $v_\alpha(1)$ attains a minimum around $\alpha = 1.86$. An intuitive explanation of this phenomenon is the following. For small values of α, due to the strong negative correlation, the process x_t is "zigzagging," causing it to frequently attain relatively large values. When increasing α the process becomes smoother, thus leading to a lower variance. However, when α approaches 2, it increasingly behaves as a straight line with random slope; this effectively means that the value of the maximum is determined by just a single random number, without any averaging, leading to a relatively high variance.

Approximation for $m_\alpha(1)$. Based on (1), we could first try to approximate $m_\alpha(1)$ by $\Gamma/\sqrt{\alpha}$ for some $\Gamma > 0$, but it turns out that this functional form provides a poor fit. This is due to the fact that, as was observed in Ref. [21], the shape of the distribution of q_1 when α is in the subdiffusive range (i.e., $\alpha < 1$) differs substantially from the shape when α is in the superdiffusive range (i.e., $\alpha > 1$). This motivates why we develop separate approximations for $\alpha < 1$ and $\alpha > 1$.

By experimenting with the data depicted in Fig. 2, e.g., by using log plots and log-log plots, we concluded that in the subdiffusive case $\log \bar{m}_\alpha(1)$ is approximately linear in α, whereas in the superdiffusive case $\log \bar{m}_\alpha(1)$ is approximately linear in $\log \alpha$. Based on this, for $\alpha < 1$ we propose the functional form

$$\log \bar{m}_\alpha(1) = x^- \alpha + y^- \tag{31}$$

(where the scalars x^- and y^- are to be estimated); for $\alpha > 1$ we propose the functional form

$$\log \bar{m}_\alpha(1) = x^+ \log \alpha + y^+ \tag{32}$$

(for x^+ and y^+ to be estimated). We estimated the coefficients by ordinary least squares, each based on 50 data points (recalling that we have estimates of $m_\alpha(1)$ for $\alpha = 0.02, 0.04, \ldots, 1.98, 2.00$ to our disposal).

For $\alpha < 1$ we thus obtained, based on $n = 2^{17}$ points, the estimates $x^- = -1.4204$ and $y^- = 1.1282$, with the R^2 being as high as 0.994. This leads to the highly accurate approximation

$$\bar{m}_\alpha(1) = 3.0900\, e^{-1.4204\alpha}. \tag{33}$$

Experiments with other values of the grid size n reveal that increasing n hardly affects the approximation. We also tried "richer" functional forms, such as

$$\log \bar{m}_\alpha(1) = x^- \alpha^{z^-} + y^- \tag{34}$$

(with x^-, y^-, and z^- to be estimated), but this hardly improved the R^2; we therefore propose the use of the more parsimonious form (31).

Our approximation of $m_\alpha(1)$ for the range $\alpha < 1$, being of the form $\exp(x^-\alpha + y^-)$, does not contradict the fact that $m_\alpha(1) \to \infty$ as $\alpha \downarrow 0$, as we set up our approximation to be valid in the range $[\alpha_0, 1]$ with $\alpha_0 = 0.02$. There is a natural trade-off between the value of α_0 and the goodness of the fit, in that increasing α_0 improves the fit. The value of α_0 we chose offers a good compromise: being close to 0, but still providing a near optimal fit.

For $\alpha > 1$, again based on $n = 2^{17}$ points, we find $x^+ = -1.0054$ and $y^+ = -0.2275$, with R^2 even equal to 1.000, so that we have the near-perfect approximation

$$\bar{m}_\alpha(1) = \frac{0.7965}{\alpha^{1.0054}}; \tag{35}$$

even the elementary approximation $\bar{m}_\alpha(1) = 0.8/\alpha$ is remarkably accurate.

Approximation for $v_\alpha(1)$. For the variance $v_\alpha(1)$ we follow the same procedure. Through visual inspection of the log plots and log-log plots, we again arrived at the functional form

$$\log \bar{v}_\alpha(1) = x^- \alpha + y^- \tag{36}$$

for $\alpha < 1$ (with x^- and y^- to be estimated). For $\alpha = 2H > 1$, however, log plots and log-log plots do not yield an appropriate functional form, mainly as a consequence of the minimum around $\alpha = 1.86$. After trying various candidates, we concluded that in this case a good functional form is

$$\bar{v}_\alpha(1) = H^{x_1^+} + H^{x_2^+} + y^+ \tag{37}$$

(with x_1^+, x_2^+, and y^+ to be estimated).

Using the above functional forms we obtain, by performing a least-squares estimation procedure, the approximations

$$\bar{v}_\alpha(1) = 0.543\, e^{-0.410\alpha} \tag{38}$$

for $\alpha < 1$ (with $R^2 = 0.997$), and

$$\bar{v}_\alpha(1) = \left(\frac{\alpha}{2}\right)^{-0.380} + \left(\frac{\alpha}{2}\right)^{0.451} - 1.670 \tag{39}$$

for $\alpha > 1$ (with $R^2 = 0.970$). In particular, the latter curve correctly predicts the minimum around $\alpha = 1.86$. Evidently, the fits can be improved upon at the expense of adding a parameter in the functional form, but this hardly leads to an increase of R^2.

IV. DISCUSSION

Our major contribution is a set of highly accurate and easily applicable approximations for the mean and variance of the position of RFBM at a given point in time. We have followed a Monte Carlo-based approach, as there is little hope for obtaining exact expressions (or tight explicit bounds) for these objects.

There are various topics that could be explored further. As was also mentioned in Ref. [21], one of them concerns the *biased* case. In this setting, for some constant c, the value of the RFBM at time t is given by

$$q_{t,\alpha}(c) = \sup_{0 \leqslant s \leqslant t} (x_t - x_s - c(t-s)). \tag{40}$$

For $c < 0$ and large t, the value of q_t will be in the same range as that of x_t, as the effect of the reflection will not be very pronounced. In the case that $c > 0$, however, $q_{t,\alpha}(c)$ converges to a limiting random variable $q_\alpha^\infty(c)$ as $t \to \infty$; in queueing lingo one says that the queue is *stable*. Interestingly, for any

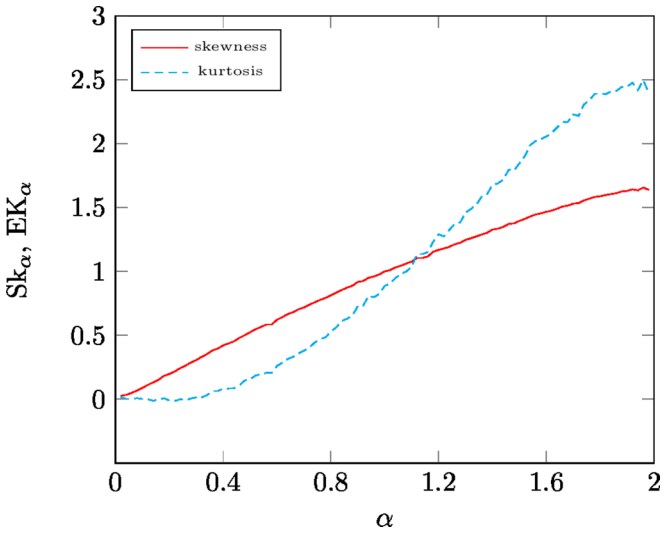

FIG. 3. Estimates of skewness Sk_α and (excess) kurtosis EK_α as a function of α. An equidistant grid with $n = 2^{17}$ points has been used, with each estimate being based on 5×10^5 runs.

$c > 0$ we can express $q_\alpha^\infty(c)$ in terms of $q_\alpha^\infty(1)$, as follows. Renormalizing time yields, with $\gamma = 1/(\alpha - 2)$,

$$q_\alpha^\infty(c) = \sup_{t \geqslant 0}(x_t - ct) = \sup_{t \geqslant 0}(x_{c^{2\gamma}t} - c^{\alpha\gamma}t), \tag{41}$$

where the first equality is in the distributional sense, and the second due to a time renormalization. As a consequence of the self-similarity the latter expression is, again in the distributional sense, equal to

$$\sup_{t \geqslant 0}(c^{\alpha\gamma}x_t - c^{\alpha\gamma}t) = c^{\alpha\gamma}\sup_{t \geqslant 0}(x_t - t) = c^{\alpha\gamma}q_\alpha^\infty(1). \tag{42}$$

In other words, in order to estimate $q_\alpha^\infty(c)$ for any $c > 0$, we are left with analyzing just $q_\alpha^\infty(1)$. The next step is then to define

$$n_\alpha(1) := \langle q_\alpha^\infty(1)\rangle, \quad w_\alpha(1) := \langle (q_\alpha^\infty(1))^2\rangle - \langle q_\alpha^\infty(1)\rangle^2 \tag{43}$$

and use empirical techniques (similar to those relied on in the present paper) to assess these functions. A complication when estimating $n_\alpha(1)$ and $w_\alpha(1)$ by simulation, however, is that $q_\alpha^\infty(1)$ involves the supremum over an *infinite* time interval. As it concerns a stochastic process with negative drift, in the simulations this could be dealt with by truncating the time interval at some sufficiently large value T. It is, however, not *a priori* clear how this horizon T should be chosen.

A second topic concerns the assessment of the epoch at which, in the unbiased model, the maximum (in the definition of $\bar{q}_1$) is attained. Initial experiments reveal that its distribution is symmetric around $i = n/2$. For small α, $i = 0$ and $i = n$ are the least likely values; due to the negative correlation, very likely a value in between leads to a higher value. For high α, $i = 0$ and $i = n$ are the most likely values; the process essentially behaves like a straight line, attaining a maximum at $i = 0$ if the drift is negative, and at $i = n$ if the drift is positive.

An intriguing feature appearing in our work concerns the nonmonotone behavior of the variance $v_\alpha(1)$ as a function of $\alpha \in (0, 2]$. There is a potential connection with the critical value $\alpha = \frac{3}{2}$ of the so-called ergodicity breaking parameter;

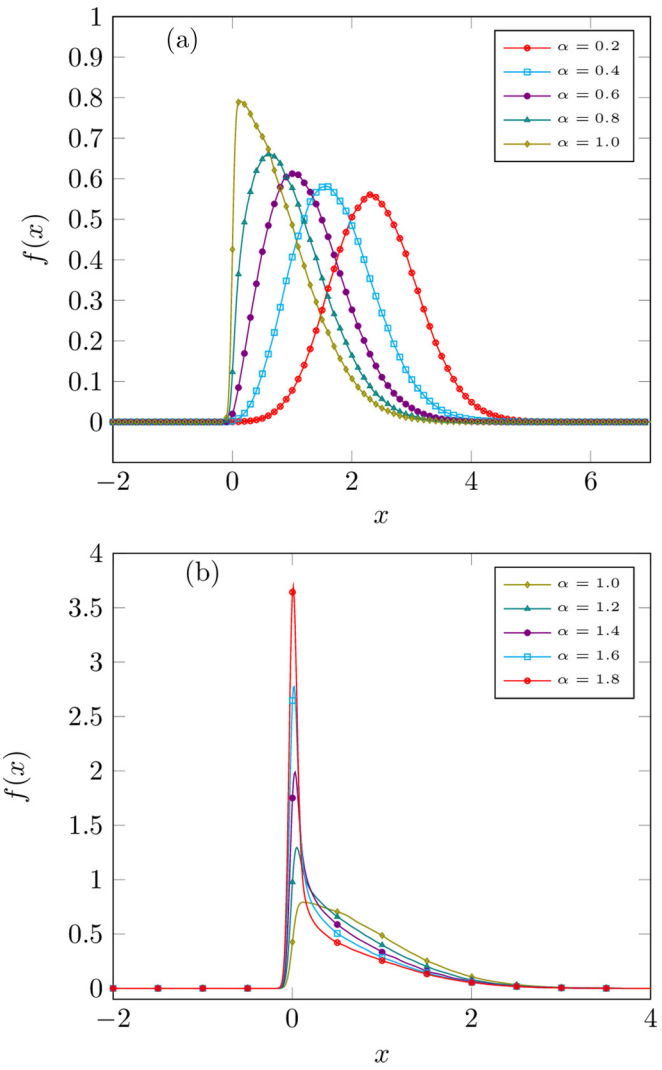

FIG. 4. Estimates of the density $f(x)$ of $\bar{q}_1$ for $x \geqslant 0$ and (a) $0.2 \leqslant \alpha \leqslant 1.0$ and (b) $1.0 \leqslant \alpha \leqslant 1.8$. An equidistant grid with $n = 2^{17}$ points has been used, with each estimate being based on 5×10^5 runs.

see Ref. [32]. In addition, one could explore other, more detailed, features of the distribution of $\bar{q}_1$ (again for the unbiased model that was the object of study in this paper). Figure 3 provides estimates of metrics that are related to the third and fourth moment of $\bar{q}_1$, namely, the skewness and (excess) kurtosis. The skewness, defined as

$$\mathrm{Sk}_\alpha := \left\langle \left[\frac{\bar{q}_1 - \bar{m}_\alpha(1)}{\sqrt{\bar{v}_\alpha(1)}} \right]^3 \right\rangle, \tag{44}$$

is a measure of asymmetry; the graph shows that the skewness grows in α. The (excess) kurtosis, defined as

$$\mathrm{EK}_\alpha := \left\langle \left[\frac{\bar{q}_1 - \bar{m}_\alpha(1)}{\sqrt{\bar{v}_\alpha(1)}} \right]^4 \right\rangle - 3, \tag{45}$$

measures the distribution's "tailedness," which also increases in α. Figure 4 provides insight into the shape of the density of $\bar{q}_1$ for a range of values of α, using kernel density estimates with Gaussian kernels. From the graphs we observe

that as α increases, the estimated density becomes more skewed towards zero, in line with the findings reported in, e.g., Ref. [21]. At the same time the density shifts its mass more and more to zero as α increases, in line with the increase of the (excess) kurtosis. Note that the estimates for the skewness and (excess) kurtosis lose accuracy as $\alpha \downarrow 0$; this is a consequence of the fact that these metrics correspond to higher-order moments. To improve the accuracy, one could obviously sample more points than the current 2^{17}, and/or one could simulate more runs than the current 5×10^5. As another topic for further exploration, it could be studied to what extent in the regime $\alpha \downarrow 0$ the estimated skewness and (excess) kurtosis match with those of the approximating Gumbel random variable.

ACKNOWLEDGMENT

The authors would like to thank Krzysztof Dębicki and Ilkka Norros for useful comments.

[1] A. Einstein, *Investigations on the Theory of the Brownian Movement* (Dover, New York, 1956).
[2] R. Metzler, J.-H. Jeon, A. G. Cherstvy, and E. Barkai, Phys. Chem. Chem. Phys. **16**, 24128 (2014).
[3] Y. Meroz and I. M. Sokolov, Phys. Rep. **573**, 1 (2015).
[4] Y. Sagi, M. Brook, I. Almog, and N. Davidson, Phys. Rev. Lett. **108**, 093002 (2012).
[5] B. Regnerand, D. Vucinić, C. Domnisoru, T. Bartol, M. Hetzer, D. Tartakovsky, and T. Sejnowski, Biophys. J. **104**, 1652 (2013).
[6] M. Weiss, M. Elsner, F. Kartberg, and T. Nilsson, Biophys. J. **87**, 3518 (2004).
[7] A. Caspi, R. Granek, and M. Elbaum, Phys. Rev. Lett. **85**, 5655 (2000).
[8] M. Taqqu, W. Willinger, and R. Sherman, Comput. Comm. Rev. **27**, 5 (1997).
[9] A. Montanari, in *Theory and Applications of Long-Range Dependence*, edited by P. Doukhan, G. Oppenheim, and M. Taqqu (Birkhäuser, Boston, 2003), pp. 461–472.
[10] R. Baillie, J. Econometrics **73**, 5 (1996).
[11] R. Cont, in *Fractals in Engineering: New Trends in Theory and Applications*, edited by J. Lévy-Véhel and E. Lutton (Springer, New York, 2005), pp. 159–179.
[12] J. Beran, *Statistics for Long-Memory Processes* (Chapman & Hall, New York, 1994).
[13] F. Biagini, Y. Hu, B. Øksendal, and T. Zhang, *Stochastic Calculus for Fractional Brownian Motion* (Springer, Berlin, 2008).
[14] M. Mandjes, *Large Deviations for Gaussian Queues* (Wiley, New York, 2007).
[15] N. Duffield and N. O'Connell, Math. Proc. Cambridge Philos. Soc. **118**, 363 (1996).
[16] V. I. Piterbarg, Extremes **4**, 147 (2001).
[17] J. Hüsler and V. Piterbarg, Stoch. Proc. Appl. **114**, 231 (2004).
[18] O. Narayan, Adv. Perf. Anal. **1**, 39 (1998).
[19] K. Dębicki and M. Mandjes, J. Appl. Prob. **40**, 704 (2004).
[20] K. Borovkov, Y. Mishura, A. Novikov, and M. Zhitlukhin, Stochastics **89**, 21 (2017).
[21] A. H. O. Wada and T. Vojta, Phys. Rev. E **97**, 020102(R) (2018).
[22] T. Guggenberger, G. Pagnini, T. Vojta, and R. Metzler, New J. Phys. **21**, 022002 (2019).
[23] W. Holmes, Biophys. J. **116**, 1538 (2019).
[24] S. Asmussen, *Applied Probability and Queues* (Springer, New York, 2003).
[25] H. Qian, Lect. Notes Phys. **621**, 22 (2003).
[26] J.-F. Coeurjolly, J. Stat. Softw. **5**, 1 (2000).
[27] S. Asmussen and P. Glynn, *Stochastic Simulation* (Springer, New York, 2007).
[28] S. Shreve, *Stochastic Calculus for Finance II: Continuous-Time Models* (Springer, New York, 2004).
[29] K. Borovkov and M. Zhitlukhin, Electron. Commun. Probab. **23**, 65 (2018).
[30] Y. Rinott and E. Samuel-Cahn, Stat. Prob. Lett. **21**, 153 (1994).
[31] P. Embrechts, C. Klüppelberg, and T. Mikosch, *Modelling Extremal Events: For Insurance and Finance* (Springer, New York, 2013).
[32] M. Schwarzl, A. Godec, and R. Metzler, Sci. Rep. **7**, 3878 (2017).