Statistical properties of the anomalous scaling exponent estimator based on time-averaged mean-square displacement

Grzegorz Sikora,[*] Marek Teuerle,[†] and Agnieszka Wyłomańska[‡]
Faculty of Pure and Applied Mathematics, Hugo Steinhaus Center, Wroclaw University of Science and Technology, Wybrzeże Wyspianskiego 27, 50-370 Wroclaw, Poland

Denis Grebenkov[§]
Laboratoire de Physique de la Matière Condensée (UMR 7643), CNRS-École Polytechnique, Université Paris-Saclay, 91128 Palaiseau, France

(Received 5 April 2017; published 14 August 2017)

The most common way of estimating the anomalous scaling exponent from single-particle trajectories consists of a linear fit of the dependence of the time-averaged mean-square displacement on the lag time at the log-log scale. We investigate the statistical properties of this estimator in the case of fractional Brownian motion (FBM). We determine the mean value, the variance, and the distribution of the estimator. Our theoretical results are confirmed by Monte Carlo simulations. In the limit of long trajectories, the estimator is shown to be asymptotically unbiased, consistent, and with vanishing variance. These properties ensure an accurate estimation of the scaling exponent even from a single (long enough) trajectory. As a consequence, we prove that the usual way to estimate the diffusion exponent of FBM is correct from the statistical point of view. Moreover, the knowledge of the estimator distribution is the first step toward new statistical tests of FBM and toward a more reliable interpretation of the experimental histograms of scaling exponents in microbiology.

DOI: 10.1103/PhysRevE.96.022132

I. INTRODUCTION

The statistical analysis of single-particle trajectories is an important and rapidly growing field of statistics and statistical physics [1–12]. Optical experiments with various tracers have revealed the anomalous character of the intracellular dynamics. In particular, the mean-square displacement (MSD) was shown to grow as a power law of time, $\mathbb{E}\{X^2(t)\} \propto t^\beta$, with the anomalous scaling exponent β distinguishing between subdiffusive ($\beta < 1$), diffusive ($\beta = 1$), and superdiffusive ($\beta > 1$) regimes [2,13–20]. The anomalous scaling exponent β is connected to target finding times [21], cellular organization [22], reaction rates [23], etc. Due to a limited (often small) number of acquired trajectories, the conventional ensemble average over multiple realizations of the stochastic process has to be replaced by the time average along a single random trajectory. As a consequence, the time-averaged mean-square displacement (TAMSD) is a *random function* of the lag time t, from which the scaling exponent β has to be estimated. Among various methods for estimating β (e.g., Bayesian techniques; see [24]), the most basic way consists of plotting the TAMSD versus the lag time at the log-log scale and thus estimating the slope of the expected straight line by the least squares method [8,11]. Although this standard technique and its eventual improvements (e.g., see [25]) remain the most commonly used, we are not aware of a detailed theoretical analysis of the statistical properties of this estimator.

In this paper, we fill this gap by investigating the β estimator for the case of fractional Brownian motion [26]. This Gaussian self-similar process with stationary increments is one of the archetypical models of anomalous dynamics, in particular, in living cells (see [16,27,28] and references therein). The Gaussian character of this process will allow us to obtain the closed formulas for the mean, the variance, and the distribution of the β estimator. We show that this estimator is asymptotically unbiased (i.e., its mean converges to the true scaling exponent), whereas its variance vanishes as the trajectory length increases. This property means that even a single (long enough) trajectory can ensure an accurate estimation. With the aid of Monte Carlo simulations, we also analyze the effect of noise on the quality of such estimations.

II. TIME-AVERAGED MEAN-SQUARE DISPLACEMENT

For a trajectory of length N, the TAMSD at the lag time τ is

$$M_N(\tau) = \frac{1}{N-\tau} \sum_{i=1}^{N-\tau} (X(i+\tau) - X(i))^2. \tag{1}$$

Here we focus on the one-dimensional case. In higher dimensions, an extension is straightforward for isotropic processes, whereas a tensor version of the TAMSD is needed for anisotropic processes.

For a centered Gaussian process, the TAMSD in Eq. (1) has a generalized chi-squared (χ^2) distribution [5,29]

$$(N-\tau) M_N(\tau) \stackrel{d}{=} \sum_{i=1}^{N-\tau} \lambda_i(\tau) U_i, \tag{2}$$

where U_i's are independent and identically distributed (i.i.d.) χ^2 with one-degree-of-freedom [$\chi(1)$] random variables, and weights $\lambda_i(\tau)$ are the eigenvalues of the $(N-\tau) \times (N-\tau)$ covariance matrix $\Sigma(\tau)$ of the vector $\{X(1+\tau) - X(1),$

[*]grzegorz.sikora@pwr.edu.pl
[†]marek.teuerle@pwr.edu.pl
[‡]agnieszka.wylomanska@pwr.edu.pl
[§]denis.grebenkov@polytechnique.edu

$X(2+\tau)-X(2),\ldots,X(N)-X(N-\tau)\}$. (For noncentered Gaussian processes, see [7] and references therein.) When the increments are stationary, the covariance matrix $\Sigma(\tau)$ is the Toeplitz (or diagonal-constant) matrix

$$\Sigma(\tau)=\begin{pmatrix}\sigma_\tau(0) & \sigma_\tau(1) & \sigma_\tau(2) & \cdots\\ \sigma_\tau(1) & \sigma_\tau(0) & \sigma_\tau(1) & \cdots\\ \sigma_\tau(2) & \sigma_\tau(1) & \sigma_\tau(0) & \cdots\\ \cdots & \cdots & \cdots & \cdots\end{pmatrix}.$$

The distribution of the quadratic form $(N-\tau)M_N(\tau)$ in Eq. (2) can also be understood as a sum of independent gamma distributions with constant shape parameter $1/2$ and different scale parameters, because $\lambda_i(\tau)U_i \overset{d}{=} G(1/2,2\lambda_i(\tau))$. We remind that the probability density function (PDF) of the gamma distribution, $G(k,\theta)$, with shape parameter k and scale parameter θ reads

$$\rho(x)=\frac{x^{k-1}\exp(-x/\theta)}{\Gamma(k)\,\theta^k}\quad (x>0).$$

Therefore, the characteristic function of $(N-\tau)M_N(\tau)$ is the product of characteristic functions of gamma-distributed random variables:

$$\phi_{(N-\tau)M_N(\tau)}(k)=\prod_{j=1}^{N-\tau}\frac{1}{[1-2\lambda_j(\tau)ik]^{1/2}}.$$

Moreover, the moment generating function (MGF) of $(N-\tau)M_N(\tau)$ reads [30]

$$\mathrm{MGF}_{(N-\tau)M_N(\tau)}(s)=C(1-2\lambda_1(\tau)s)^{-(N-\tau)/2}\times\exp\left(\sum_{k=1}^{\infty}\frac{\gamma_k}{(1-2\lambda_1(\tau)s)^k}\right),$$

where $\lambda_1(\tau)$ is the smallest eigenvalue of the matrix $\Sigma(\tau)$,

$$\gamma_k=\sum_{i=1}^{N-\tau}\frac{(1-\lambda_1(\tau)/\lambda_i(\tau))^k}{2k},$$

and

$$C=\prod_{i=1}^{N-\tau}\left(\frac{\lambda_1(\tau)}{\lambda_i(\tau)}\right)^{1/2}.$$

Based on the results of [30], one can show that the PDF of $(N-\tau)M_N(\tau)$ for $x>0$ is given by

$$\rho_\tau(x)=C\sum_{k=0}^{\infty}\frac{\delta_k\,x^{\frac{N-\tau}{2}+k-1}\exp\left(-\frac{x(N-\tau)}{2\lambda_1(\tau)}\right)}{\Gamma\left(\frac{N-\tau}{2}+k\right)\left(\frac{2\lambda_1(\tau)}{N-\tau}\right)^{\frac{N-\tau}{2}+k}},\tag{3}$$

where δ_k can be calculated by the recursive formula

$$\delta_{k+1}=\frac{1}{k+1}\sum_{i=1}^{k+1}i\gamma_i\delta_{k+1-i},\quad \delta_0=1.$$

The series in Eq. (3) converges uniformly; see [30].

The mean and the variance of $M_N(\tau)$ for a general noncentered Gaussian process were analyzed in [7]. In the particular case of a centered Gaussian process with stationary increments, the mean and the variance are

$$\mathbb{E}\{M_N(\tau)\}=\sigma_\tau(0),$$
$$\mathbb{V}\mathrm{ar}\{M_N(\tau)\}=\frac{1}{(N-\tau)^2}\left[8(N-\tau)\sigma_\tau^2(0)+4\sum_{i=1}^{N-\tau-1}(N-\tau-i)\sigma_\tau^2(i)\right].\tag{4}$$

III. ESTIMATION OF FBM PARAMETERS BY USING TAMSD

A. Fractional Brownian motion

Fractional Brownian motion (FBM) with the self-similarity index (Hurst exponent) $0<H<1$ serves as a basic model that is often used for analyzing and interpreting single-particle tracking (SPT) experiments [31–37]. For our consideration, a random trajectory of FBM with N successive positions $\{X(1),X(2),\ldots,X(N)\}$, acquired at equal time steps Δt, is just a centered Gaussian vector with zero mean, $\mathbb{E}\{X(i)\}\equiv 0$, and the $N\times N$ covariance matrix

$$\mathbb{E}\{X(i)\,X(j)\}=D(i^{2H}+j^{2H}-|i-j|^{2H}),\tag{5}$$

where $D>0$ is the (dimensionless) diffusion coefficient (that includes the variance of each displacement and the time step Δt). The Hurst exponent H is just proportional to the anomalous scaling exponent: $\beta=2H$. This covariance matrix implies that the $\Sigma(\tau)$ matrix is formed by diagonals:

$$\sigma_\tau(i)=D((i+\tau)^{2H}-2i^{2H}+|i-\tau|^{2H}).$$

According to Eqs. (4) and (5), the mean and variance of the TAMSD $M_N(\tau)$ are

$$\mathbb{E}\{M_N(\tau)\}=2D\tau^{2H},\tag{6}$$

$$\mathbb{V}\mathrm{ar}\{M_N(\tau)\}=\frac{8D^2}{(N-\tau)}\tau^{4H}+\frac{4}{(N-\tau)^2}\times\sum_{i=1}^{N-\tau-1}(N-\tau-i)[\sigma_\tau(i)]^2.\tag{7}$$

The ergodicity of the FBM ensures that the TAMSD exhibits the same scaling for long enough trajectories; i.e.,

$$M_N(\tau)\simeq\mathbb{E}\{M_N(\tau)\}=2D\tau^{2H}\quad (N\to\infty).\tag{8}$$

B. Anomalous scaling exponent estimator for FBM

Writing the scaling relation (8) as

$$\ln\left(M_N(\tau)\right)\simeq\ln(2D)+\beta\ln(\tau)\tag{9}$$

and setting $\tilde{D}=2D$, one gets an estimator for the $\tilde{D}$ parameter,

$$\hat{D}=M_N(1).$$

One can also obtain the distribution of $\hat{D}$, namely,

$$\hat{D}\sim\sum_{i=1}^{N-1}\lambda_i(1)U_i,$$

where $\lambda_i(1)$ are eigenvalues of the covariance matrix $\Sigma(1)$.

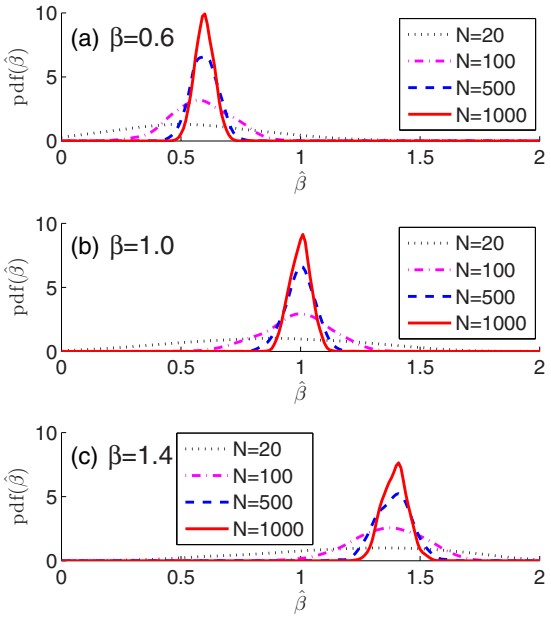

FIG. 1. Probability density function of TAMSD estimator $\hat{\beta}$ for different trajectory lengths N ($N = 20, 100, 500, 1000$), obtained via Monte Carlo simulations with 1000 independent runs, with $n = 10$ and $\tau_i = i$ ($i = 1, \ldots, 10$). The true values of β are (a) 0.6, (b) 1.0, and (c) 1.4.

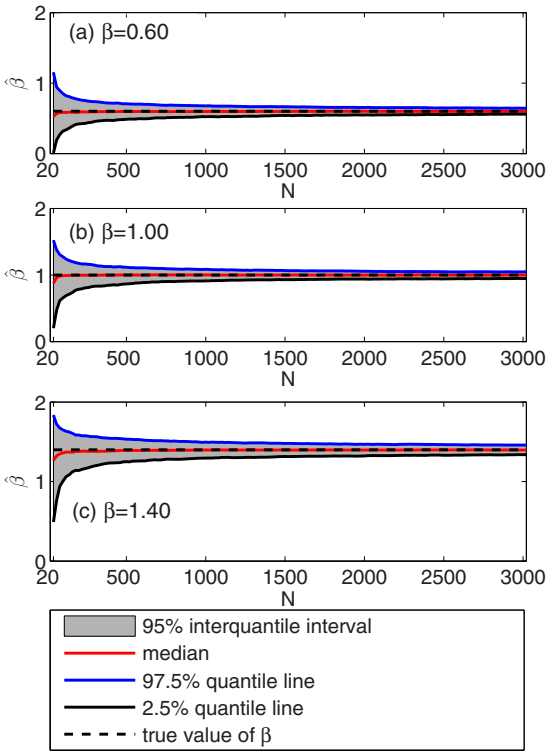

FIG. 2. Empirical quantiles of the TAMSD estimator $\hat{\beta}$ as functions of the trajectory length N, obtained via Monte Carlo simulations with 1000 independent runs, with $n = 15$ and $\tau_i = i$ ($i = 1, \ldots, 15$). The true values of β are (a) 0.6, (b) 1.0, and (c) 1.4.

One can easily check that $\mathbb{E}\{\hat{D}\} = \tilde{D}$ that means this estimator is unbiased. Moreover, $\mathbb{V}\mathrm{ar}\{\hat{D}\} = \mathbb{V}\mathrm{ar}\{M_N(1)\}$, which vanishes when $N \to \infty$, because for large N the random variable $M_N(\tau)$ tends to a constant as $N \to \infty$. Moreover, the PDF of $\hat{D}$ is given by Eq. (3) at $\tau = 1$.

On the basis of Eq. (9), the anomalous diffusion exponent β can also be estimated. More precisely, one calculates the TAMSD for $\tau_1, \ldots, \tau_n$ and then fits $\ln(M_N(\tau_i))$ by the linear function $\ln(\tilde{D}) + \beta \ln(\tau_i)$, $i = 1,2, \ldots, n$. For a linear regression model without the intercept term,

$$y_i = b x_i + \epsilon_i, \quad i = 1,2, \ldots, n,$$

the ordinary least squares estimator of the b parameter is

$$\hat{b} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}. \tag{10}$$

In the considered case, we have

$$\ln\left(\frac{M_N(\tau)}{M_N(1)}\right) \sim \beta \ln(\tau);$$

thus by using Eq. (10) we obtain the estimator

$$\hat{\beta} = \frac{\sum_{i=1}^{n} \ln(\tau_i) \ln\left(\frac{M_N(\tau_i)}{M_N(1)}\right)}{\sum_{i=1}^{n} \ln^2(\tau_i)}. \tag{11}$$

Using the results presented in the previous section we deduce the distribution of $\hat{\beta}$,

$$\hat{\beta} \overset{d}{=} \frac{\sum_{i=1}^{n} \ln(\tau_i) \ln\left(\frac{(N-1)\sum_{j=1}^{N-\tau_i} \lambda_j(\tau_i) U_j}{(N-\tau_i)\sum_{k=1}^{N-1} \lambda_k(1) U_k}\right)}{\sum_{i=1}^{n} \ln^2(\tau_i)},$$

where U_j's are i.i.d. random variables from χ^2 distribution with one degree of freedom and $\lambda_j(\tau_i)$ are eigenvalues of the matrix $\Sigma(\tau_i)$.

To illustrate the properties of the $\hat{\beta}$ distribution we perform Monte Carlo simulations (see Sec. IV for more details). Figure 1 presents the probability density function of $\hat{\beta}$ for several values of sample length N ($N = 20, 100, 500, 1000$). The graphs were obtained via numerical approximation of the density functions based on Monte Carlo samples. We observe that the distribution of $\hat{\beta}$ becomes narrower when the sample length N increases. Moreover, the shapes of the three PDFs for the same N are similar. These theoretical plots can also be compared to the experimental histograms of scaling exponents from [38,39].

Figure 2 shows the quantile lines of order 2.5%, 50%, and 97.5% of the $\hat{\beta}$ distribution for various values of $\beta \in \{0.6, 1.0, 1.4\}$. The grey region between 2.5% and 97.5% quantile lines represents the 95% interquantile interval around the median. We observe that for all β parameters the longer the FBM trajectory length is, the narrower the considered interquantile interval is, in agreement with the reduction of $\hat{\beta}$ variance for increasing length of FBM trajectory.

Taking advantage of Eq. (11) one obtains

$$\mathbb{E}\{\hat{\beta}\} = \frac{\sum_{i=1}^{n} \ln(\tau_i) \mathbb{E}\{\ln(M_N(\tau_i))\}}{\sum_{i=1}^{n} \ln^2(\tau_i)} - \frac{\sum_{i=1}^{n} \ln(\tau_i) \mathbb{E}\{\ln(M_N(1))\}}{\sum_{i=1}^{n} \ln^2(\tau_i)}.$$

Using the Taylor expansions for the moments of the natural logarithm of a random variable X around its mean,

$$\mathbb{E}\{\ln(X)\} \sim \ln(\mathbb{E}\{X\}) - \frac{1}{2\mathbb{E}\{X\}}\mathbb{V}\mathrm{ar}\{X\},$$

and setting $X = M_N(\tau)$, we get, with the aid of Eqs. (6) and (7),

$$\mathbb{E}\{\ln(M_N(\tau))\} \sim \ln(\tilde{D}\tau^{\beta}) - \frac{1}{2\tilde{D}\tau^{\beta}}\mathbb{V}\mathrm{ar}\{M_N(\tau)\} \overset{N\to\infty}{\to} \ln(\tilde{D}) + \beta\ln(\tau).$$

This implies that the estimator $\hat{\beta}$ is asymptotically unbiased:

$$\mathbb{E}\{\hat{\beta}\} \overset{N\to\infty}{\to} \ln(\tilde{D})\frac{\sum_{i=1}^{n}\ln(\tau_i)}{\sum_{i=1}^{n}\ln^2(\tau_i)} + \beta\frac{\sum_{i=1}^{n}\ln^2(\tau_i)}{\sum_{i=1}^{n}\ln^2(\tau_i)} - \ln(\tilde{D})\frac{\sum_{i=1}^{n}\ln(\tau_i)}{\sum_{i=1}^{n}\ln^2(\tau_i)} = \beta. \tag{12}$$

Moreover, using the formula

$$\mathbb{V}\mathrm{ar}\{\ln(X)\} \sim \frac{1}{\mathbb{E}^2(X)}\mathbb{V}\mathrm{ar}\{X\},$$

and the fact that $\mathbb{V}\mathrm{ar}\{M_N(\tau)\} \to 0$ as $N \to \infty$, we obtain

$$\mathbb{V}\mathrm{ar}\{\ln(M_N(\tau))\} \overset{N\to\infty}{\to} 0. \tag{13}$$

From Eq. (13) and Cauchy-Schwarz inequality we have the following result:

$$\begin{aligned}
\mathbb{V}\mathrm{ar}\{\hat{\beta}\} &= \frac{1}{\left(\sum_{i=1}^{n}\ln^2(\tau_i)\right)^2}\sum_{i,j=1}^{n}\ln(\tau_i)\ln(\tau_j)\mathbb{C}\mathrm{ov}\left\{\ln\left(\frac{M_N(\tau_i)}{M_N(\tau_1)}\right),\ln\left(\frac{M_N(\tau_j)}{M_N(\tau_1)}\right)\right\}\\
&= \frac{1}{\left(\sum_{i=1}^{n}\ln^2(\tau_i)\right)^2}\sum_{i,j=1}^{n}\ln(\tau_i)\ln(\tau_j)[\mathbb{C}\mathrm{ov}\{\ln(M_N(\tau_i)),\ln(M_N(\tau_j))\} - 2\mathbb{C}\mathrm{ov}\{\ln(M_N(\tau_i)),\ln(M_N(\tau_1))\}\\
&\quad + \mathbb{V}\mathrm{ar}\{\ln(M_N(\tau_1))\}]\\
&\leqslant \frac{1}{\left(\sum_{i=1}^{n}\ln^2(\tau_i)\right)^2}\sum_{i,j=1}^{n}\ln(\tau_i)\ln(\tau_j)[\sqrt{\mathbb{V}\mathrm{ar}\{\ln(M_N(\tau_i))\}}\sqrt{\mathbb{V}\mathrm{ar}\{\ln(M_N(\tau_j))\}}\\
&\quad - 2\sqrt{\mathbb{V}\mathrm{ar}\{\ln(M_N(\tau_i))\}}\sqrt{\mathbb{V}\mathrm{ar}\{\ln(M_N(\tau_1))\}} + \mathbb{V}\mathrm{ar}\{\ln(M_N(\tau_1))\}] \overset{N\to\infty}{\longrightarrow} 0.
\end{aligned} \tag{14}$$

Figure 3 shows that the bias $\mathbb{E}(\hat{\beta}) - \beta$ and the variance $\mathbb{V}\mathrm{ar}(\hat{\beta})$ of the estimator $\hat{\beta}$, obtained via Monte Carlo simulations, both tend to zero as N increases. These numerical results illustrate our theoretical predictions (12) and (14).

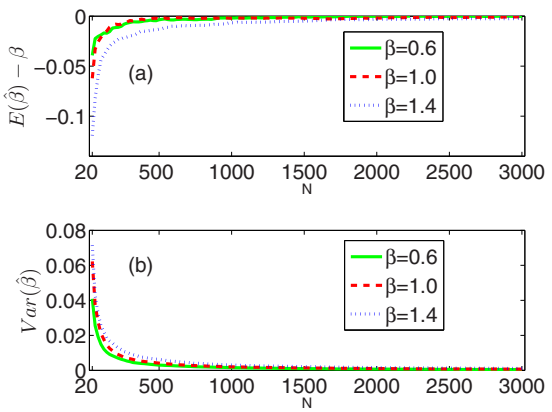

FIG. 3. The (a) bias and (b) variance of the estimator $\hat{\beta}$ as functions of the trajectory length N, obtained via Monte Carlo simulations with 1000 independent runs, with $n = 10$ and $\tau_i = i$ ($i = 1,\ldots,10$). The results are presented for three values of β: 0.6, 1.0, and 1.4.

We also show that the estimator $\hat{\beta}$ is consistent; i.e., for any $\epsilon > 0$,

$$\mathbb{P}\{|\hat{\beta} - \beta| > \epsilon\} \leqslant \frac{\mathbb{E}\{(\hat{\beta} - \beta)^2\}}{\epsilon^2} \overset{N\to\infty}{\longrightarrow} 0,$$

as follows from the Chebyshev inequality. The consistency of the estimator $\hat{\beta}$ is an important statistical property: when the trajectory length N increases, the probability of an incorrect estimation vanishes. The consistency is also confirmed by Monte Carlo simulations (see Fig. 4), as detailed in the next section.

IV. SIMULATIONS

In order to illustrate the statistical properties of the TAMSD estimator, we generate FBM trajectories and investigate the efficiency of the estimator $\hat{\beta}$. In this section we explain the details of our numerical experiment.

In our numerical investigation we estimate β via Monte Carlo technique with 1000 independent runs. We simulate FBM trajectories with $\beta \in \{0.6, 1.0, 1.4\}$ by using a Davis-Harte method [40], which was generalized by Wood and Chan [41]. It uses the Cholesky decomposition of the covariance matrix but it is much faster than the classical one. The three

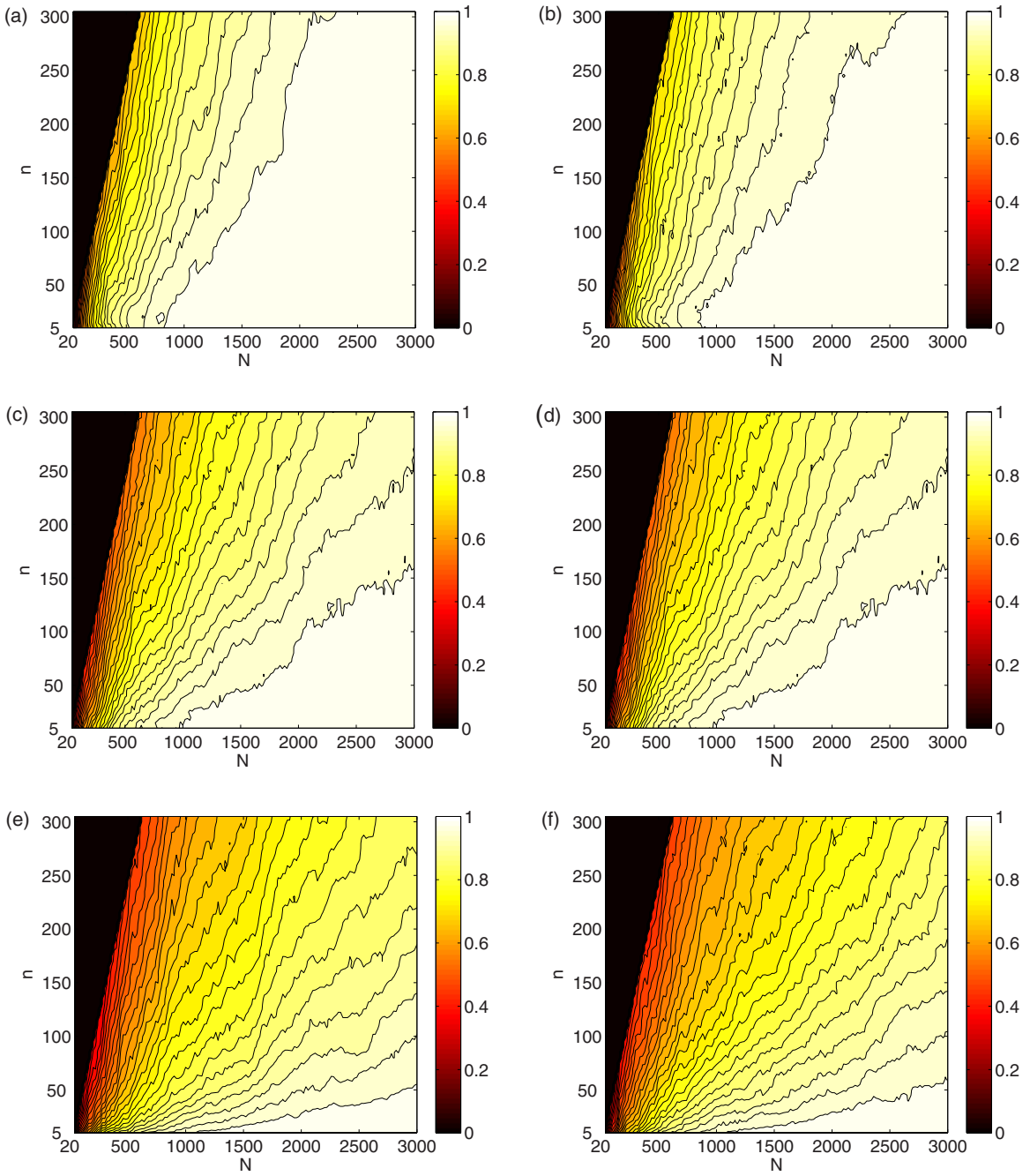

FIG. 4. The probability for the estimator $\hat{\beta}$ to be within the interval $[\beta-0.1,\beta+0.1]$ as a function of the trajectory length N and of the parameter n, obtained via Monte Carlo simulations with 1000 independent runs. (a), (c), (e) FBM trajectories and (b), (d), (f) FBM perturbed by white noise $\mathcal{N}(0,0.01)$. Top, middle, and bottom rows present the cases $\beta=0.6$, $\beta=1.0$, and $\beta=1.4$, respectively.

values of β represent distinct regimes: subdiffusion, normal diffusion, and superdiffusion, respectively. We use equidistant lag times and set $\tau_i=i$ $(i=1,\ldots,n)$ in the estimation procedure. The efficiency of the TAMSD estimator $\hat{\beta}$ is checked for trajectories of length N ($N=20,40,60,\ldots,3000$), and for the values of parameter n ($n=5,10,15,\ldots,300$), but keeping n below $N/2$. To assess the accuracy of the TAMSD estimator we compute the probability that the value of $\hat{\beta}$ is within the interval $(\beta-0.1,\beta+0.1)$, the choice of that interval length being reasonable in the biophysical context [11]. Furthermore, we investigate how the results are changed when FBM trajectories are perturbed by Gaussian white noise $\mathcal{N}(0,0.01)$. (Technically, to every point of the FBM trajectory a random realization of white noise with mean zero and variance 0.01 is added.)

The main results are presented in Fig. 4, which shows the probability of the TAMSD estimator $\hat{\beta}$ to be within the interval $(\beta-0.1,\beta+0.1)$ as a function of the trajectory length N and of the parameter n used for the least squares fitting. The results for FBM trajectories [Figs. 4(a), 4(c), and 4(e)] are compared

to those for FBM perturbed by white noise $\mathcal{N}(0,0.01)$ [Figs. 4(b), 4(d), and 4(f)]. The top, middle, and bottom rows correspond to $\beta = 0.6$, $\beta = 1.0$, and $\beta = 1.4$, respectively.

For FBM without noise, one can see a convex-shaped pattern of the estimation accuracy depending on parameters N and n. For instance, for $\beta = 0.6$, a convex curve between pairs $(N = 750, n = 5)$ and $(N = 2000, n = 300)$ distinguishing the probability of 90% accuracy is clearly visible.

For all cases the estimation accuracy changes slightly when FBM trajectories are perturbed by white noise. Namely, we can observe a shift of contour lines for respective accuracy levels towards the right and bottom-right directions, meaning that for a fixed n one needs a slightly larger sample length N to obtain the same level of accuracy in estimation.

The results presented in Fig. 4 can be used to optimize the estimation: for a trajectory of a given length N, one can choose the parameter n to achieve an estimation accuracy at the desired level.

V. DISCUSSION AND CONCLUSION

In this paper an exact distribution of the anomalous scaling exponent estimator based on the TAMSD is derived for FBM. We have expressed the distribution of the β estimator by an explicit function of independent gamma-distributed random variables. It is worth mentioning that the obtained results can be extended to all centered Gaussian processes. We have shown that the estimator is asymptotically unbiased and consistent, and its variance tends to zero as the trajectory length increases. These properties ensure an accurate estimation of the scaling exponent even from a single (long enough) trajectory. As a consequence, we prove that the usual way to estimate the diffusion exponent is correct from the statistical point of view, at least in the case of FBM. The theoretical results are supported by the Monte Carlo simulations that indicate that the variance and the interquantile interval of the β estimator tend to zero as the trajectory length grows. One of the main practical results of the paper is presented in Fig. 4, which shows how to recognize the appropriate number of time points used for β estimation (n) for a given trajectory length (N), in order to estimate the anomalous scaling exponent with 10% accuracy. The effect of white Gaussian noise on the β estimator has also been investigated.

The presented theoretical results are crucial for a proper anomalous diffusion model testing [42]. The knowledge of the scaling exponent estimator distribution is a starting point for building new statistical tests to answer whether the FBM is an appropriate model for an acquired experimental trajectory. Moreover, the exact form of the distribution presents the first step towards the solution of another important problem. The recent progress in imaging techniques opened an experimental way to access the distribution of scaling exponents from single-particle trajectories. Several authors have already published such distributions (e.g., see [38,39]) which provide new insights into intracellular transport. However, the dispersion of the scaling exponents obtained from fitting the TAMSD has two sources of randomness: (i) the biologically relevant variability between measurements (e.g., trajectories acquired in different cells, in different cell locations, under different cell cycles or conditions, etc.) and (ii) the intrinsic uncertainty related to the finite length of the trajectory, i.e., to the distribution of the scaling exponent estimator. Obviously, the second one is an "artifact" of the estimation procedure, or the price one pays for the advantage of dealing with single trajectories. The two sources of randomness are superimposed, which prohibits a reliable interpretation of such distributions. To overcome this fundamental problem, one has to disentangle these two sources. The knowledge of the distribution of the scaling exponent estimator (i.e., the full characterization of the second source of randomness) is the necessary step towards the solution of this complicated statistical problem in microbiology.

ACKNOWLEDGMENTS

The work of A.W. is supported by NCN OPUS Grant No. 2016/21/B/ST1/00929. The work of G.K. and M.T. is supported by NCN Maestro Grant No. 2012/06/A/ST1/00258. D.G. acknowledges the support under Grant No. ANR-13-JSV5-0006-01 of the French National Research Agency.

[1] D. Arcizet, B. Meier, E. Sackmann, J. O. Rädler, and D. Heinrich, Phys. Rev. Lett. **101**, 248103 (2008).
[2] R. Metzler, V. Tejedor, J.-H. Jeon, Y. He, W. H. Deng, S. Burov, and E. Barkai, Acta Phys. Pol. B **40**, 1315 (2009).
[3] V. Tejedor, O. Bénichou, R. Voituriez, R. Jungmann, F. Simmel, C. Selhuber-unkel, L. B. Oddershede, and R. Metzler, Biophys. J. **98**, 1364 (2010).
[4] M. Magdziarz and A. Weron, Phys. Rev. E **84**, 051138 (2011).
[5] D. S. Grebenkov, Phys. Rev. E **84**, 031124 (2011).
[6] A. Andreanov and D. S. Grebenkov, J. Stat. Mech. (2012) P07001.
[7] D. S. Grebenkov, Phys. Rev. E **88**, 032140 (2013).
[8] N. Gal, D. Lechtman-Goldstein, and D. Weihs, Rheol. Acta **52**, 425 (2013).
[9] Y. Meroz, I. M. Sokolov, and J. Klafter, Phys. Rev. Lett. **110**, 090601 (2013).
[10] S. Türkcan and J.-B. Masson, PLoS ONE **8**, 0082799 (2013).
[11] E. Kepten, A. Weron, G. Sikora, K. Burnecki, and Y. Garini, PLoS ONE **10**, e0117722 (2015).
[12] Y. Lanoiselée and D. S. Grebenkov, Phys. Rev. E **93**, 052146 (2016).
[13] I. M. Tolić-Nørrelykke, E.-L. Munteanu, G. Thon, L. Oddershede, and K. Berg-Sørensen, Phys. Rev. Lett. **93**, 078102 (2004).
[14] I. Golding and E. C. Cox, Phys. Rev. Lett. **96**, 098102 (2006).
[15] C. Wilhelm, Phys. Rev. Lett. **101**, 028101 (2008).
[16] J. Szymanski and M. Weiss, Phys. Rev. Lett. **103**, 038102 (2009).
[17] E. Sackmann, F. Keber, and D. Heinrich, Annu. Rev. Condens. Matter Phys. **1**, 257 (2010).
[18] J.-H. Jeon, V. Tejedor, S. Burov, E. Barkai, C. Selhuber-Unkel, K. Berg-Sørensen, L. Oddershede, and R. Metzler, Phys. Rev. Lett. **106**, 048103 (2011).

[19] E. Bertseva, D. S. Grebenkov, P. Schmidhauser, S. Gribkova, S. Jeney, and L. Forró, Eur. Phys. J. E **35**, 63 (2012).
[20] P. C. Bressloff and J. M. Newby, Rev. Mod. Phys. **85**, 135 (2013).
[21] R. Metzler, T. Koren, B. Van den Broek, G. J. Wuite, and M. A. Lomholt, J. Phys. A **42**, 434005 (2009).
[22] E. Barkai, Y. Garini, and R. Metzler, Phys. Today **65**, 29 (2012).
[23] M. J. Saxton, Biophys. J. **103**, 2411 (2012).
[24] N. Makarava, S. Benmehdi, and M. Holschneider, Phys. Rev. E **84**, 021109 (2011).
[25] E. Kepten, I. Bronshtein, and Y. Garini, Phys. Rev. E **87**, 052713 (2013).
[26] B. B. Mandelbrot and J. W. Van Ness, SIAM Rev. **10**, 422 (1968).
[27] M. Weiss, Phys. Rev. E **88**, 010101 (2013).
[28] R. Metzler, J.-H. Jeon, A. Cherstvy, and E. Barkai, Phys. Chem. Chem. Phys. **16**, 24128 (2014).
[29] R. B. Davies, Appl. Stat. **29**, 323 (1980).
[30] P. G. Moschopoulos, Ann. Inst. Stat. Math. **37**, 541 (1985).
[31] F. Höfling and T. Franosch, Rep. Prog. Phys. **76**, 046602 (2013).
[32] Y. Meroz, I. M. Sokolov, and J. Klafter, Phys. Rev. Lett. **107**, 260601 (2011).
[33] I. M. Sokolov, Soft Matter **8**, 9043 (2012).
[34] K. Burnecki, E. Kepten, J. Janczura, I. Bronshtein, Y. Garini, and A. Weron, Biophys. J. **103**, 1839 (2012).
[35] M. Hellmann, J. Klafter, D. W. Heermann, and M. Weiss, J. Phys.: Condens. Matter **23**, 234113 (2011).
[36] M. Magdziarz, A. Weron, K. Burnecki, and J. Klafter, Phys. Rev. Lett. **103**, 180602 (2009).
[37] J.-H. Jeon, E. Barkai, and R. Metzler, J. Chem. Phys. **139**, 121916 (2013).
[38] M. H. G. Duits, Y. Li, S. A. Vanapalli, and F. Mugele, Phys. Rev. E **79**, 051910 (2009).
[39] M. Otten, A. Nandi, D. Arcizet, M. Gorelashvili, B. Lindner, and D. Heinrich, Biophys. J. **102**, 758 (2012).
[40] R. B. Davies and D. S. Harte, Biometrika **74**, 95 (1987).
[41] A. T. A. Wood and G. Chan, J. Comput. Graph. Stat. **3**, 409 (1994).
[42] G. Sikora, K. Burnecki, and A. Wyłomańska, Phys. Rev. E **95**, 032110 (2017).