Bayesian inference of Lévy walks via hidden Markov models

OPEN ACCESS
IOP Publishing
Journal of Physics A: Mathematical and Theoretical
J. Phys. A: Math. Theor. **54** (2021) 484001 (23pp)
https://doi.org/10.1088/1751-8121/ac31a1

Bayesian inference of Lévy walks via hidden Markov models

Seongyu Park[1], Samudrajit Thapa[2,3], Yeongjin Kim[1], Michael A Lomholt[4,*] and Jae-Hyung Jeon[1,*]

[1] Department of Physics, Pohang University of Science and Technology (POSTECH), Pohang 37673, Republic of Korea
[2] Sackler Center for Computational Molecular and Materials Science, Tel Aviv University, Tel Aviv 69978, Israel
[3] School of Mechanical Engineering, Tel Aviv University, Tel Aviv 69978, Israel
[4] PhyLife, Department of Physics, Chemistry and Pharmacy, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark

E-mail: mlomholt@sdu.dk and jeonjh@postech.ac.kr

Received 7 July 2021, revised 8 October 2021
Accepted for publication 20 October 2021
Published 10 November 2021

Abstract

The Lévy walk (LW) is a non-Brownian random walk model that has been found to describe anomalous dynamic phenomena in diverse fields ranging from biology over quantum physics to ecology. Recurrently occurring problems are to examine whether observed data are successfully quantified by a model classified as LWs or not and extract the best model parameters in accordance with the data. Motivated by such needs, we propose a hidden Markov model for LWs and computationally realize and test the corresponding Bayesian inference method. We introduce a Markovian decomposition scheme to approximate a renewal process governed by a power-law waiting time distribution. Using this, we construct the likelihood function of LWs based on a hidden Markov model and the forward algorithm. With the LW trajectories simulated at various conditions, we perform the Bayesian inference for parameter estimation and model classification. We show that the power-law exponent of the flight-time distribution can be successfully extracted even at the condition that the mean-squared displacement does not display the expected scaling exponent due to the noise or insufficient trajectory length. It is also demonstrated that the Bayesian method performs remarkably inferring the LW trajectories from given unclassified trajectory data set if the noise level is moderate.

*Authors to whom any correspondence should be addressed.

Keywords: Bayesian inference, Levy walks, parameter estimation, model classification

(Some figures may appear in colour only in the online journal)

1. Introduction

Complex diffusion dynamics is often observed in diverse fields, such as transport dynamics of living macromolecules inside a biological cell [1–8], tracer particles in polymer networks [9–11], foraging dynamics of living organisms [12–14] or animals [15, 16], and stock prices in the stock market [17, 18], to name a few. Such complex dynamics usually deviate from Brownian dynamics [19–21] characterized by Gaussian displacements $P(\Delta x; \Delta t) \propto \exp\left(-\frac{\Delta x^2}{4D\Delta t}\right)$ and the linear growth of the mean-squared displacement (MSD) with time $\langle x^2(t)\rangle \propto t$. They exhibit non-Brownian characteristics in which the MSD usually increases with time as

$$\langle x^2(t)\rangle \propto t^{\alpha} \tag{1}$$

with the anomaly exponent α in the range $0 < \alpha \leqslant 2$ [22, 23] and/or the displacements have a non-Gaussian distribution that violates the central limit theorem [9, 23]. These anomalous diffusion dynamics have been explained by continuous-time random walk [24, 25], fractional Brownian motion (FBM) [26], Lévy walk (LW) [27], annealed transient time [28], and scaled Brownian motion (SBM) [29, 30].

It is often challenging to pinpoint the anomalous diffusion model underlying observed complex diffusion dynamics and determine the value of model parameters (e.g. the anomaly exponent). The former is understood as model classification and the latter as parameter estimation. To get proper information on these tasks, conventionally, one analyzes the experimental time-series data with various dynamic measures such as MSD, Van-Hove correlation functions, step-length or flight-time distribution, and ergodicity breaking parameter [31–37]. However, the interpretation of the data can be subjective and sometimes becomes highly nontrivial if the given data are not ideal. The non-ideal conditions include insufficient samples in terms of length and number, the high noise level in the data, and the spatiotemporal heterogeneity in the samples. To overcome these difficulties, recently, new approaches, e.g. Bayesian inference and machine learning, have been developed with keen interest from cross-disciplinary sciences [10, 38–46].

In this work, we develop a Bayesian inference method for analyzing LWs. Bayesian inference provides a direct comparison between the possible statistical models and finds the most likely model parameters over a given parameter space by calculating the likelihood functions of given data. In a Bayesian framework, one does not have to define point-estimators and extract the statistical features (e.g. MSD or van-Hove correlation function). From a single raw trajectory, diffusion models and model parameters can be examined to derive the best-fit model and its model parameters. Recently, some efforts have been devoted to developing the Bayesian framework for Brownian motion, FBM, and diffusing diffusivity [10, 42–45]. These studies have proven the success of the Bayesian inference approach in the trajectory analysis.

The LW model is a physical modification of Lévy flights [27, 47–51]. The corresponding trajectory of LW consists of independent ballistic flights with constant speed v and the random flight times governed by a Lévy distribution. LWs are known to model various non-equilibrium dynamic processes in biology, ecology, physics, and economics. Prominent examples include foraging dynamics of bacteria, animals, and humans [12, 15, 16, 52–59], active transport of macromolecules inside a living cell [60–62], blinking dynamics of quantum dots [63], light

in inhomogeneous media [64], a flow motion in a rotating annulus [65], and related Hamiltonian systems [66, 67]. For modeling animal's foraging dynamics in ecology [47], a Bayesian inference approach was applied to differentiate Lévy flights/walks from composite correlated random walks via calculating their likelihood functions [46]. Their method, however, was based on manually-extracted step lengths and turning angles from trajectories. Moreover, the correlation between the step length and flight time—the essential ingredient of a LW—was not taken into account in the likelihood function. Previously, we developed Bayesian inference frameworks for several anomalous diffusion models [10, 43–45]. In this work, based on these experiences, we set the Bayesian inference method for LWs. Our approach is based on a Markov decomposition method that mathematically describes a LW process without the manually-extracted distribution features. Using this method, we construct the likelihood function of a LW trajectory systematically as a function of trajectory length, the power-law exponent of the flight-time distribution, and the noise strength.

The present paper is organized in the following structure. In section 2 we give a brief overview of Bayesian inference theory and the LW model that are used throughout the paper. In section 3, we propose a theory of the hidden Markov model for LWs. This model is employed to compute the likelihood function of a one dimensional LW trajectory based on the forward algorithm (we comment on a generalization to two dimensions in the appendix B). Sections 4 and 5 present the computational results of our Bayesian inference method. Here, we demonstrate that random flight times governed by a power-law distribution are generated from the hidden Markov model. We also visualize the numerically calculated likelihood functions at various conditions. We then systematically investigate the performance of our Bayesian inference tool for extracting the power-law exponent of LW trajectories upon varying the trajectory length, the exponent α, and the signal-to-noise ratio. In section 5, we study the model classification using the Bayesian method differentiating among LW, fractional Brownian and scaled Brownian trajectories at various conditions. Finally, in section 6 we summarize our study and discuss the generalization of our Bayesian inference theory of LWs for the application to other similar anomalous dynamic processes.

2. Theoretical background

2.1. Bayesian inference method

In this subsection, we provide a short overview of Bayesian inference in connection with our Bayesian approach to LWs in the following section. Here we introduce key concepts as well as define essential probability distributions to be used later. For a complete overview of Bayesian inference, see reference [68]. Suppose that a data set $\mathcal{D}$ is given, and we have several statistical models $\mathcal{M}_i$ to explain the data. For a quantitative comparison between the models, in principle, we need to estimate the conditional probability $P(\mathcal{M}_i|\mathcal{D})$ that the model $\mathcal{M}_i$ is selected to explain $\mathcal{D}$. The best-fit model is the one that maximizes the given conditional probability. According to Bayes' theorem [69], $P(\mathcal{M}_i|\mathcal{D})$ is given by

$$P(\mathcal{M}_i|\mathcal{D}) = \frac{P(\mathcal{D}|\mathcal{M}_i)P(\mathcal{M}_i)}{P(\mathcal{D})}, \tag{2}$$

where $P(\mathcal{M}_i)$ is the prior probability that the model $\mathcal{M}_i$ is given before our observation. $P(\mathcal{D}|\mathcal{M}_i) \equiv Z_i$ is the conditional probability that the data $\mathcal{D}$ is observed when a model $\mathcal{M}_i$ is given, referred to as evidence of a model $\mathcal{M}_i$. Assuming that all the prior probabilities are the same, i.e. $P(\mathcal{M}_i) = P(\mathcal{M}_j)$ for all (i, j) pairs, then the probability that the model $\mathcal{M}_i$ is true

among N possible models is given by the ratio of the model evidence:

$$\frac{P(\mathcal{M}_i|\mathcal{D})}{\sum_{j=1}^{N} P(\mathcal{M}_j|\mathcal{D})} = \frac{P(\mathcal{D}|\mathcal{M}_i)}{\sum_{j=1}^{N} P(\mathcal{D}|\mathcal{M}_j)} = \frac{Z_i}{\sum_{j=1}^{N} Z_j}. \tag{3}$$

This is the key idea in the Bayesian approach toward model inference. The best-fit model can be found via the comparison of the model evidences.

To calculate the model evidence, we define the likelihood function $\mathcal{L}_i(\vec{\theta}_i) \equiv P(\mathcal{D}|\vec{\theta}_i, \mathcal{M}_i)$ where $\vec{\theta}_i$ is the parameter set necessary to specify the model $\mathcal{M}_i$. The model evidence (marginal likelihood) is then evaluated by marginalizing the likelihood function $\mathcal{L}_i(\vec{\theta}_i)$ over the parameter space Θ with a weight $\pi(\vec{\theta}_i)$:

$$Z_i = P(\mathcal{D}|\mathcal{M}_i) = \int_{\Theta} P(\mathcal{D}|\vec{\theta}_i, \mathcal{M}_i)\pi(\vec{\theta}_i)\mathrm{d}\vec{\theta}_i. \tag{4}$$

Here, $\pi(\vec{\theta}_i) \equiv P(\vec{\theta}_i|\mathcal{M}_i)$ is a prior probability for the parameter set $\vec{\theta}_i$ given based on additional information or intuition [68].

The Bayesian inference approach is effectively used to find the best parameter set for a given data set if a specific model is given or the best model is determined from the above model inference. The posterior probability distribution of the parameters is given by

$$P(\vec{\theta}_i|\mathcal{M}_i, \mathcal{D}) = \frac{P(\mathcal{D}|\vec{\theta}_i, \mathcal{M}_i)P(\vec{\theta}_i|\mathcal{M}_i)}{P(\mathcal{D}|\mathcal{M}_i)} = \frac{\mathcal{L}_i(\vec{\theta}_i)\pi(\vec{\theta}_i)}{Z_i}, \tag{5}$$

and we report the medians of this distribution as the estimated values of the parameters of a given model. It is computationally nontrivial to perform the marginalization of the likelihood function (equation (4)) and to sample over the posterior distribution (5). For this task, here we use the method of annealed importance sampling [70]. For further information on this method, see our MATLAB code for annealed importance sampling in the AnDi challenge GitHub [71].

2.2. The discrete-time LW

In the LW model, a flight time τ of a ballistic run with speed v and its step length Δx is given by the joint probability

$$\Psi(\Delta x, \tau) = \frac{1}{2}\psi(\tau)\delta(|\Delta x| - v\tau). \tag{6}$$

Because of the constraint between flight time and step length, the whole dynamics are solely determined by characteristics of the flight time distribution $\psi(\tau)$. In the LW we consider the class of $\psi(\tau)$ that asymptotically behaves as a power-law function at large times

$$\psi(\tau) \propto \frac{1}{\tau^{1+\gamma}} \tag{7}$$

with a power-law exponent γ (i.e. the stability index of a Lévy distribution) of $0 < \gamma \leqslant 2$. The LW has distinct diffusion dynamics depending on γ [27]: the diffusion becomes ballistic for $0 < \gamma < 1$, sub-ballistic superdiffusive for $1 < \gamma < 2$, and diffusive for $\gamma \geqslant 2$. In this work, we are interested in the sub-ballistic LW where the MSD increases with time as [27]

$$\langle x^2(t)\rangle \propto t^{3-\gamma} \text{ at } t \to \infty. \tag{8}$$

Thus, in this range of γ, the index of flight-time distribution (equation (7)) is related to the anomaly exponent via

$$\alpha = 3 - \gamma \text{ with } 1 < \alpha < 2. \tag{9}$$

For simplicity, hereafter, we denote the flight-time distribution $\psi(t)$ in terms of α (via equation (9)) throughout the text.

While the LW is defined with $\psi(\tau)$ in the continuous time-domain of $\tau > 0$, time series in the experiment, such as particle trajectories obtained from a single-particle tracking experiment, are usually given in unit of a time lag that is bound by the time resolution of a measurement. To apply our Bayesian inference approach to such data, we here introduce a discrete-time LW. Practically, the LW trajectory is often generated in this way. The anomalous diffusion simulation package (the AnDi package [72, 73]) used in this work provides the discrete-time LW.

In the discrete-time LW, the flight times are integers, sampled by the following two steps. (1) A continuous random time $\tau(> 0)$ is sampled by the transformation

$$\tau \leftarrow F^{-1}(u;\alpha) = (1-u)^{-\frac{1}{3-\alpha}} \tag{10}$$

using a uniform random number u in $(0, 1)$ and $F(t) = \int_1^t \psi(\tau)\mathrm{d}\tau = (3-\alpha)\int_1^t \tau^{\alpha-4}\mathrm{d}\tau$. The normalized probability density function for τ is

$$\psi_{\mathrm{c}}^{\mathrm{target}}(\tau;\alpha) = \frac{3-\alpha}{\tau^{4-\alpha}} \quad \text{for } \tau \geqslant 1. \tag{11}$$

(2) The integer part of τ is taken using floor function $\tau' = \lfloor \tau \rfloor$. After the rounding-off, the probability mass function for τ' is given by

$$\begin{aligned} \psi_{\mathrm{d}}^{\mathrm{target}}(\tau';\alpha) &= \int_{\tau'}^{\tau'+1} \psi_{\mathrm{c}}^{\mathrm{target}}(\tau;\alpha)\mathrm{d}\tau \quad \text{for } \tau' \in \mathbb{N} \\ &= \frac{1}{\tau'^{3-\alpha}} - \frac{1}{(\tau'+1)^{3-\alpha}} \propto \frac{1}{\tau'^{4-\alpha}}. \end{aligned} \tag{12}$$

From now on, we call equation (11) the continuous target distribution and equation (12) the discrete target distribution. With the integer random times given by $\psi_{\mathrm{d}}^{\mathrm{target}}(\tau')$, we can generate a discrete-time LW trajectory parametrized by $\vec{\theta} = \{v, \alpha, \sigma_{\mathrm{noise}}\}$ via the governing equation

$$\begin{aligned} x_t &= x_t^{\mathrm{LW}} + \xi_t \\ &= \left[\sum_{i=1}^{N} (-1)^{r_i} v\tau_i' + (-1)^{r_{N+1}} v\tau_{\mathrm{res}}'\right] + \xi_t \end{aligned} \tag{13}$$

with

$$t = \sum_{i=1}^{N} \tau_i' + \tau_{\mathrm{res}}'. \tag{14}$$

Figure 1(a) shows a schematic description of our discrete-time LW. For given observation time t, the walker (or particle) has N complete flight events and an incomplete flight event. The duration time for the latter is described by a residual time $\tau_{\mathrm{res}}' = t - \sum_{i=1}^{N} \tau_i'$. The velocity of the walker during the ith flight is $(-1)^{r_i} v$ where $r_i = 0$ or 1 with an equal probability. Additionally,

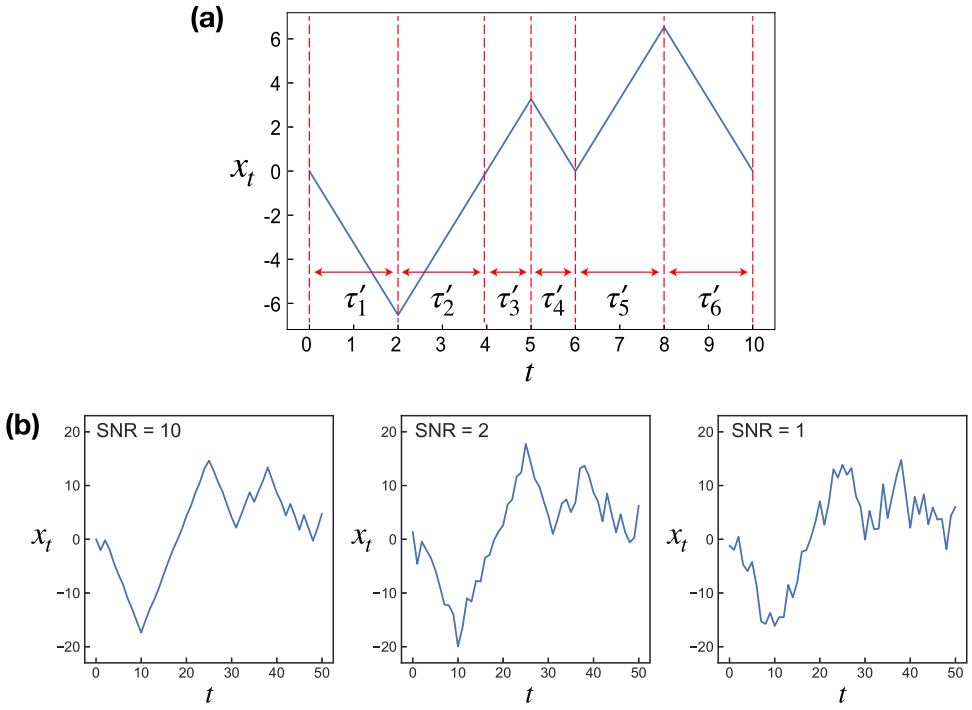

Figure 1. Sample trajectories for one-dimensional discrete-time LWs. (a) A LW comprises ballistic flights whose flight events are specified by (velocity, flight time) $= \{(v_i = \pm v, \tau_i)\}$. The flight times $\tau \in (0, \infty)$ are randomly given by a power-law PDF $\psi(\tau) \propto \tau^{\alpha-4}$ $(1 \leqslant \alpha \leqslant 2)$. In the discrete-time LW (shown in the panel), flight times are integers $\tau'(\in \mathbb{N})$ whose probability mass function is given by equation (12). (b) Noisy discrete-time LW trajectories simulated from the governing equation (13) with $\alpha = 1.36$. From left to right, the trajectories have a noise level with SNR $= 10, 2$, and 1. See the text for further information. The trajectories are generated by the AnDi package [72].

we consider a LW in a noisy environment. Physically, the noise may originate from localization errors during the tracking measurement or the thermal random agitation in the heat bath. Thus, in a complete picture, we invent the so-called *noisy* LW trajectory by superimposing a Gaussian random noise ξ_t to the noise-free trajectory in equation (13). The Gaussian noise is characterized by $\xi_t \sim \mathcal{N}(0, \sigma^2_{\text{noise}})$ where v/σ_{noise} (or $\sigma_D/\sigma_{\text{noise}}$ after standardizing the trajectory in the later section) is understood as the signal-to-noise (SNR) ratio. Figure 1(b) shows examples of the noisy LW trajectories for varying SNRs. When the noise level is significantly high (SNR $= 1$), the noisy LW no longer looks like a LW.

3. Hidden Markov model for LWs & Bayesian inference

3.1. Hidden Markov model for LWs

In this section, we develop our theoretical formalism of the Bayesian inference framework for LWs. For this, we first build up a solvable model that appropriately describes the discrete-time LW process in the scheme of the Bayesian formalism. Our approach is to establish a hidden Markov model for non-Markovian processes governed by a power-law relaxation dynamics using the Markovian decomposition method [74].

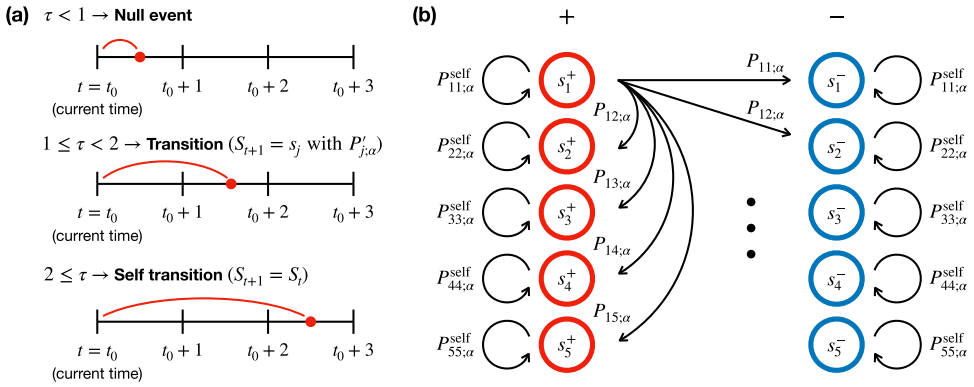

Figure 2. (a) Transition strategy of a random walker in our Markovian approximation of discrete-time LW model. A walker in a state $s_i^{\pm}$ samples transition time τ from $\phi_i(\tau)$ every time step. The probability of $\tau < 1$ event (top) is zero. If a walker samples $1 \leqslant \tau < 2$ (middle), it changes the state from $s_i^{\pm}$ to $s_j^{\pm}$ with probability $P_{j;\alpha}$ ($\frac{1}{2}P_{j;\alpha}$ for each of the two states s_j^+ and s_j^-). If a walker chooses $2 \leqslant \tau$ (bottom), it maintains its original state ($S_{t+1} = S_t$). (b) Schematic illustration of hidden Markov model for one-dimensional LW. In this model, the '+' states (red circles) have positive velocity $+v$ and '−' states (blue circles) have negative velocity $-v$. The probabilities of self-transitions and transitions for every time step are denoted by $P^{\text{self}}_{ii;\alpha}$ and $P_{ij;\alpha}$, which are given by (25) and (24), respectively.

Now let us consider a flight-time distribution with a power-law decay up to a time T, which is referred to as a cutoff time and usually considered the maximum observation time window in the measurement. Such power-law distributions can be represented by the superposition of a finite number of independent exponential distributions (i.e. Ornstein–Uhlenbeck processes); see the method developed in reference [74]. We note that such approximation of power-laws, which can be interpreted in terms of superstatistics [75], goes back to the explanation of $1/f$ noise in semiconductors by a superposition of independent processes with different but definite relaxation times [76–79]. In our approximation scheme, the continuous target distribution can be expanded in terms of multiple exponential distributions as such:

$$\psi_{\text{c}}^{\text{target}}(\tau;\alpha) \approx \sum_{i=1}^{N_k} P_{i;\alpha} k_i e^{-k_i(\tau-1)} \quad (\tau \geqslant 1), \tag{15}$$

where the inverse of the relaxation time of the ith component and its statistical weight are, respectively, given by

$$k_i = \frac{k_{\text{fast}}}{b^i} \quad (i = 1, 2, \ldots, N_k) \tag{16}$$

$$P_{i;\alpha} = \frac{k_i^{3-\alpha} e^{-k_i}}{\sum_{j=1}^{N_k} k_j^{3-\alpha} e^{-k_j}}. \tag{17}$$

See the appendix A for the validation of equation (15). Here, b is a scale parameter, k_{fast} is the high-frequency cutoff, and N_k determines the long-time exponential cutoff $T \approx (4-\alpha)\, b^{N_k}/k_{\text{fast}}$. We set $b = 4$, $k_{\text{fast}} = 8$, and $N_k = 5$ through the work. In this scheme, we regard the power-law distributed flight times (within the observed time window) as the inter-arrival times

among N_k-component independent Markov processes. The sojourn time in the state s_i has the probability density function

$$\phi_i(\tau) = k_i e^{-k_i(\tau-1)}. \tag{18}$$

The state transition from state s_i to state s_j occurs with a probability $P_{j;\alpha}$ when the process leaves the state s_i. This process gives the inter-arrival time distribution equation (15) in the form of $\sum_{i=1}^{N_k} P_{i;\alpha}\phi_i(\tau)$.

Now we set up the proper rounding-off process to make discrete-time Markov chains of N_k components having the inter-arrival times governed by $\psi_d^{\text{target}}(\tau')$. The walker's state is renewed at every time unit such that it is changed to one of the other states (*transition*) or maintains the same state (*self-transition*). Figure 2(a) illustrates how the transition and self-transition events are determined from the continuous inter-arrival times. Let us assume that at $t = t_0$ the walker (Markov chain) is in the state $S_{t_0} = s_i$ and samples the transition time τ according to $\phi_i(\tau)$. Here S_t denotes the walker's state at time t. (i) If $1 \leqslant \tau < 2$, it is allocated to $\tau' = 1$ from the rounding-off process. Thus, the walker in the next time ($t = t_0 + 1$) should change its state to a new one, say, $S_{t_0+1} = s_j$. In this case, the state transition occurs with the transition probability

$$\left[\int_1^2 \phi_i(\tau)\mathrm{d}\tau\right] P_{j;\alpha} = \left(1 - e^{-k_i}\right) P_{j;\alpha.} \tag{19}$$

(ii) If $\tau \geqslant 2$, then $\tau' \geqslant 2$ and the walker stays in the same state at $t = t_0 + 1$ ($S_{t_0+1} = S_{t_0} = s_i$). This is the case called the self-transition, which occurs with the transition probability

$$\int_2^\infty \phi_i(\tau)\mathrm{d}\tau = e^{-k_i}. \tag{20}$$

(Beware that the event of $\tau < 1$ is the null case because such events are not allowed from $\phi(\tau)$.) The probability that a walker in the ith state has an integer sojourn time τ' is

$$\int_{\tau'}^{\tau'+1} \phi_i(\tau)\mathrm{d}\tau = e^{-k_i(\tau'-1)} - e^{-k_i\tau'}, \tag{21}$$

which can be understood in the Markov chain description as the probability that the walker maintains the same state $(\tau' - 1)$ times and then changes its state, i.e.

$$(e^{-k_i})^{\tau'-1} \cdot (1 - e^{-k_i}) = e^{-k_i(\tau'-1)} - e^{-k_i\tau'}. \tag{22}$$

Therefore, by simply repeating the above process every time step, we obtain the process having integer inter-arrival times of $\tau' \geqslant 1$ with the distribution

$$\begin{aligned}\sum_{i=1}^{N_k} P_{i;\alpha} \int_{\tau'}^{\tau'+1} \phi_i(\tau)\mathrm{d}\tau &= \int_{\tau'}^{\tau'+1} \sum_{i=1}^{N_k} P_{i;\alpha}\phi_i(\tau)\mathrm{d}\tau \quad (\tau' \in \mathbb{N}) \\ &\approx \int_{\tau'}^{\tau'+1} \psi_{\mathrm{c}}^{\text{target}}(\tau;\alpha)\mathrm{d}\tau \\ &= \psi_{\mathrm{d}}^{\text{target}}(\tau';\alpha).\end{aligned} \tag{23}$$

Figure 2(b) illustrates our Markov chain scheme based on the above formalism. Here we incorporate the fact that the LW (in 1D) has two velocity states, $\pm v$, for given flight time. This

is described in our model such that the state s_i has two substates s_i^+ (for $+v$) and s_i^- (for $-v$) with the same transition rate k_i. Thus, there are $2N_k$ states in total in the model and the transition probability from the ith state to the jth state is given by

$$P_{ij;\alpha} = \frac{(1 - e^{-k_i}) \cdot P_{j;\alpha}}{2}. \tag{24}$$

As shown in figure 2(b), the change of the velocity state from $s_i^+(s_i^-)$ to $s_i^-(s_i^+)$ is considered as a transition with the probability $P_{ii;\alpha}$. On the other hand, if the process changes its state from $s_i^+(s_i^-)$ to $s_i^+(s_i^-)$, this is effectively a self-transition. The effective probability of self-transition is then given by

$$P_{ii;\alpha}^{\text{self}} = e^{-k_i} + \frac{(1 - e^{-k_i}) \cdot P_{i;\alpha}}{2}. \tag{25}$$

We numerically confirm our hidden Markov model for discrete-time LWs in section 4.1.

3.2. The likelihood function

The above discrete-time Markov chain describes the dynamics of the latent states S_t. For given parameter set $\vec{\theta} = \{\alpha, v, \sigma_{\text{noise}}\}$ and S_t, the conditional probability of observing $\Delta x_t = x_{t+1} - x_t$, i.e. the so-called emission probability, is given by

$$P(\Delta x_t | \mathcal{M}_{\text{Lévy}}, S_t, \vec{\theta}) = \frac{1}{\sqrt{4\pi\sigma_{\text{noise}}^2}} \exp\left(-\frac{1}{2}\left(\frac{\Delta x_t \pm v}{\sqrt{2}\sigma_{\text{noise}}}\right)^2\right). \tag{26}$$

In this expression, $\Delta x_t - v$ corresponds to the state $S_t = s_i^+$ $(i = 1, \ldots, N_k)$ and $\Delta x_t + v$ to the state $S_t = s_i^-$, respectively. The factor $\sqrt{2}\sigma_{\text{noise}}$ is attributed to the fact that the displacement is $\Delta x_t = x_{t+1}^{\text{LW}} - x_t^{\text{LW}} + \xi_t'$ where ξ_t' is a Gaussian random noise of $\mathcal{N}(0, 2\sigma_{\text{noise}}^2)$. We consider this term as an approximate way to include measurement noise for the positions, i.e. we should really have had $\xi_t' = \xi_{t+1} - \xi_t$ with the ξ_t independent. This approximation ignores the correlations between steps that measurement noise induces, and we will see later that it does not work well when σ_{noise} is comparable to the Levy walk step size, i.e. when SNR is not larger than unity. For brevity, below we omit the notation of $\mathcal{M}_{\text{Lévy}}$ in the conditional probabilities.

In the discrete-time LW, a trajectory x_t can be represented by a series of the unit-time displacements $\mathcal{D} = \{\Delta x_1, \Delta x_2, \ldots, \Delta x_T\}$. Thus, when the latent state is given by $\mathcal{S} = \{S_1, S_2, \ldots, S_T\}$, the probability of finding $\mathcal{D}$ (or x_t) is written as the product of the emission probabilities

$$P(\mathcal{D}|\mathcal{S}, \vec{\theta}) = P(\Delta x_1 | S_1, \vec{\theta}) P(\Delta x_2 | S_2, \vec{\theta}) \ldots P(\Delta x_T | S_T, \vec{\theta}). \tag{27}$$

We should marginalize equation (27) over all possible sequences of the latent state $\mathcal{S} = \{S_1, S_2, \ldots, S_T\}$ to obtain the likelihood function

$$P(\mathcal{D}|\vec{\theta}) = \sum_{\mathcal{S}} P(\mathcal{D}|\mathcal{S}, \vec{\theta}) P(\mathcal{S}|\vec{\theta}). \tag{28}$$

In this work, we use the forward algorithm method to calculate equation (28) for obtaining the corresponding likelihood function [80]. The summation over $\mathcal{S}$ is carried out in an iterative way with the help of the probability function

$$\beta(S_t = s_i^{\pm}) \equiv P(\Delta x_1, \ldots, \Delta x_t, S_t = s_i^{\pm} | \vec{\theta}). \tag{29}$$

This probability function satisfies the recurrence relation

$$\begin{aligned}\beta(S_{t+1}=s_i^{\pm}) &= P(\Delta x_1,\ldots,\Delta x_t,\Delta x_{t+1},S_{t+1}=s_i^{\pm}|\vec{\theta})\\ &= P(\Delta x_{t+1}|S_{t+1}=s_i^{\pm},\vec{\theta})\\ &\quad\times\sum_{j=1}^{N_k}\left[\beta(S_t=s_j^{+})P(S_{t+1}=s_i^{\pm}|S_t=s_j^{+})\right.\\ &\quad\left.+\,\beta(S_t=s_j^{-})P(S_{t+1}=s_i^{\pm}|S_t=s_j^{-})\right].\end{aligned}\tag{30}$$

In this expression, we figure that $P(\Delta x_t|S_t=s_i^{\pm},\vec{\theta})$ is the emission probability (26) and $P(S_{t+1}=s_i^{\pm}|S_t=s_j^{\pm})$ is the transition probabilities between two distinct states (equation (24)) or between the same state (self-transition) (equation (25)). The initial condition of the recurrence relation is

$$\begin{aligned}\beta(S_1=s_i^{\pm}) &= P(\Delta x_1,S_1=s_i^{\pm}|\vec{\theta})\\ &= P(\Delta x_1|S_1=s_i^{\pm},\vec{\theta})P(S_1=s_i^{\pm}).\end{aligned}\tag{31}$$

The initial distribution of the states $P(S_1=s_i^{\pm})$ is given by equation (17) provided that the process starts at $t=0$ (or more generally, the initial time $t=0$ is the renewal time). Alternatively, if the process is observed after a long time where it has become stationary, then an extra factor of the average length of the time intervals, $C/(1-e^{-k_i})$, should be multiplied on the right-hand side of equation (17) to obtain the initial distribution. C is a normalizing constant of the distribution.

The recurrence relation equation (31) is given in terms of the transition probability, emission probability, and the initial distribution of the states that can be computable. Therefore, starting from equation (31), we can iteratively obtain $\beta(S_T=s_i^{\pm})=P(\Delta x_1,\ldots,\Delta x_T,S_T=s_i^{\pm}|\vec{\theta})$. The likelihood function is then simply obtained by marginalizing $\beta(S_T=s_i^{\pm})$ over all possible states:

$$P(\mathcal{D}|\vec{\theta})=P(\Delta x_1,\ldots,\Delta x_T|\vec{\theta})=\sum_{i=1}^{N_k}\left[\beta(S_T=s_i^{+})+\beta(S_T=s_i^{-})\right].\tag{32}$$

Note that changes in direction and observations of the Lévy walker can only happen at integer values of time. In case observations happen with a camera that is operated independently of the system it observes, this can be considered an approximation that is valid in the limit of the camera frame rate being much faster than the typical interval between changes of direction.

3.3. Prior distributions

For Bayesian inference, we need to specify the prior distributions for parameters $\vec{\theta}=\{\alpha,v,\sigma_{\text{noise}}\}$. We use a uniform prior distribution for the anomaly exponent α based on the fact that our LW describes a superdiffusive and Fickian dynamics with the anomaly exponent in the range of $1\leqslant\alpha\leqslant 2$. The prior distribution is set to be $P(\alpha)=1$ for $1\leqslant\alpha\leqslant 2$ or zero otherwise. For the walker's speed $v(>0)$, we use a folded Gaussian distribution $P(v)=\sqrt{\frac{2}{\pi}}\exp\left(-\frac{v^2}{2}\right)$ in the positive regime. We use the Gaussian distribution because the trajectories we analyze in section 4 have a standard normal distribution of speed v after standardization. The pre-processing (standardization) of the trajectories is further explained in section 4.2. In the case of the background noise strength σ_{noise}, its magnitude compared to

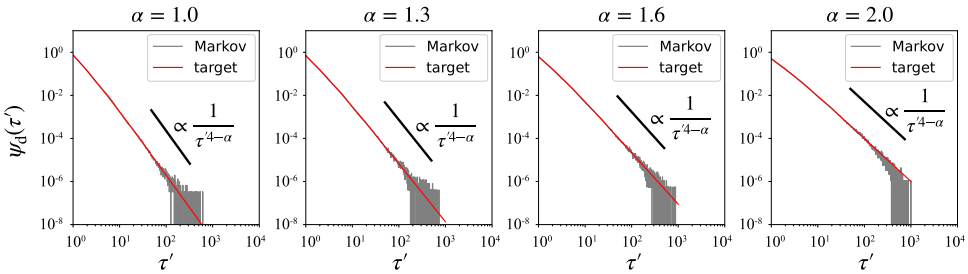

Figure 3. Flight time statistics sampled from the inter-arrival times for our Markov chain model (gray lines). In the simulation of the Markov chain model we use the parameters $N_k = 5$, $b = 4$, $k_{\text{fast}} = 8$, and $\alpha = 1.0, 1.3, 1.6$, and 2.0 (from left to right). The red lines show the discrete target distribution $\psi_d^{\text{target}}(\tau')$ shown in equation (12).

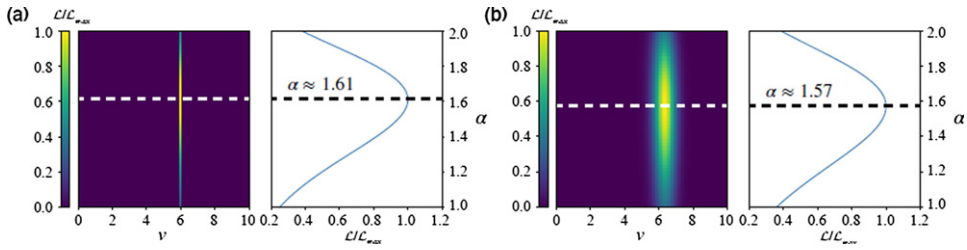

Figure 4. Heatmap of likelihood functions for a LW trajectory with $\alpha = 1.6$ and the speed $v \approx 5.92$. (a) The likelihood function for the Lévy trajectory without the noise ($\sigma_{\text{noise}} = 0$). (b) The likelihood function for the trajectory with $\sigma_{\text{noise}} = 3$. In (a) & (b), the heatmap shows the likelihood function marginalized over σ_{noise} as a function of v and α. The line plot represents the likelihood function marginalized over σ_{noise} and v, where the dashed line shows the α at which the likelihood function gives the maximum value.

the signal can be arbitrary. Here, we use a folded Cauchy distribution $P(\sigma_{\text{noise}}) = \frac{2}{\pi(1+\sigma_{\text{noise}}^2)}$ with $\sigma_{\text{noise}} > 0$ as a prior distribution. The Cauchy distribution is a heavy-tailed distribution, allowing σ_{noise} to be an arbitrarily large value.

4. Results I: numerical implementation of the Bayesian inference and parameter estimation

4.1. Numerical implementation of the hidden Markov model

We have numerically implemented our hidden Markov model for the discrete-time LW. In figure 3 we plot the flight-time distributions obtained from the simulated inter-arrival times of the Markov chain at $\alpha = 1.0, 1.3, 1.6$, and 2.0. We allocate 5000 random walkers on every Markov chain (figure 2(b)) and collect the inter-arrival times for the transition events between distinct latent states up to $t = 1000$. For all cases, we find the numerically obtained inter-arrival times (gray lines, figure 3) are in excellent agreement with the expected power-law target distribution (12) (red lines). While the longest exponential cutoff is expected to start at $T \approx O(300)$, the power-law scaling in the numerical data seems to extend longer than T. The tail of the distribution for $t \gtrsim T$ is noisy because this part is sampled from the rare events of

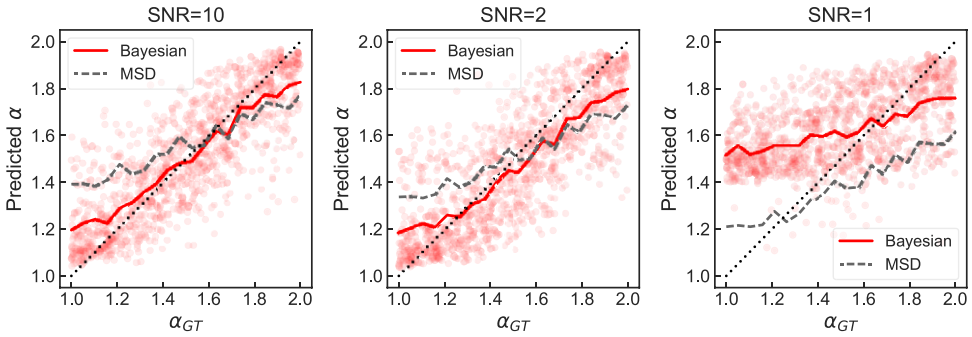

Figure 5. The predicted α values from our Bayesian method vs the ground-true value α_{GT}. From Left to Right, the plot correspond to SNR $= 10, /2$, and 1. The solid line (red) shows the average of the predicted α values for given α_{GT}. For reference, the dashed line depicts the corresponding curve for α estimated from the fitting of time-averaged MSD curves. We fit the linear part of the MSD curves in log-log plot, where the lag time Δ ranges from $2 \leqslant \Delta \leqslant 21$ for the trajectories whose lengths are over 100. For the shorter ones, we took $2 \leqslant \Delta \leqslant 8$. The dotted line is the guide line of $y = x$.

the N_kth latent state. The exponential cut-off time T gets shorter as α increases. For $\alpha = 2$, the distribution shows the expected power-law up to $T \approx 256$, and then it is slightly off the power-law due to the exponential cutoff.

We numerically obtain the likelihood function by the iterative method (equation (32)) and confirm the validity of our method. For this purpose, we generate LW trajectories of length 100 time steps with the flight-time exponent $\alpha = 1.6$ and $v \approx 5.92$ (i) without the background noise and (ii) with the noise with SNR $= 2$. We estimate the likelihood function for the trajectories over the parameter space $v \in [0, 10]$, $\alpha \in [1, 2]$, and $\sigma_{\mathrm{noise}} \in [0.01, 5]$. Figure 4 visualizes the numerically estimated likelihood functions. In the figure, the heatmaps show the likelihood functions as a function of v and α that are marginalized over the background noise σ_{noise}; the line plots are the likelihood function as a function of α marginalized over both σ_{noise} and v. All the likelihood functions are divided by their maximum value $\mathcal{L}_{\mathrm{max}}$ for normalization. (1) The noise-free LW trajectory (figure 4(a)). When such ideal LW data are given, the likelihood function expectedly produces the maximum probability at the parameter (v, α) that is very close to the true value. It is noted that the shape of the likelihood function around the maximum is dependent on the parameter. While the likelihood function appears to be a sharply peaked function for v, it is a broad distribution for α (see the line plot). This is because there are much more samples (displacement vectors) for inferring v than for inferring α that relies on fewer samples of flight times. (2) The noisy LW trajectory (figure 4(b)). We find that the noise makes the likelihood function dispersed in the parameter space v while the profile of the likelihood function for α is almost unaffected by the noise. The increased uncertainty for v is understandable. The speed of a flight can be fluctuating by the noise in the unit of time step. Despite the dispersion, the likelihood function still has the maximum at around the input speed. This suggests that the Bayesian approach can successfully estimate the true parameter value (α & v) even from noisy trajectories.

4.2. Parameter estimation via the Bayesian inference

Based on the likelihood function constructed above, we have implemented a Bayesian inference method to estimate model parameters from the trajectory data. The LW is characterized

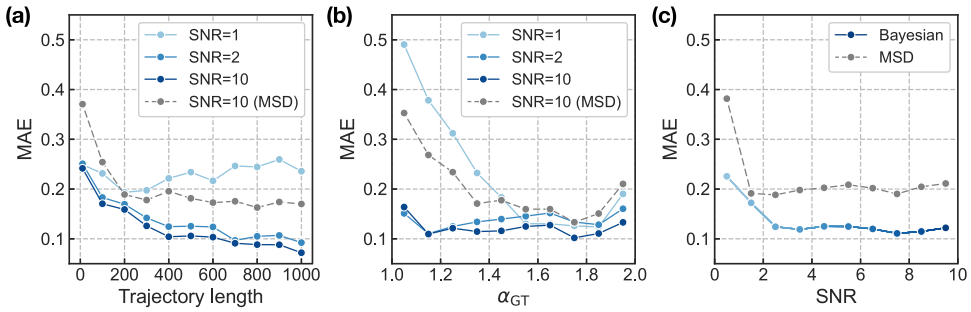

Figure 6. MAE from our Bayesian parameter estimation (line-dot) and from the fitting of MSD curves (dashed-dot). (a) MAE as a function of the trajectory length. (b) MAE as a function of α_{GT}. Each data point represents the average value for all trajectory lengths at a given α_{GT} with half bin width $|\Delta\alpha_{GT}| = 0.05$. (c) MAE as a function of SNR. Each data point represents the averaged AE for all trajectory lengths. For given N_{step}, five hundred trajectories are generated with an SNR uniformly distributed over $(0, 10]$. In x-axis, half bin width is $|\Delta\text{SNR}| = 0.5$.

by three model parameters: the power-law exponent of flight-time distribution $\gamma = 3 - \alpha$ (or, equivalently, the anomaly exponent α), the walker's flight speed v, and the background noise strength σ_{noise}. In this work, we focus on inferring the power-law exponent of the flight-time statistics α for given LW trajectories. The other parameters, such as v, can be easily inferred by available statistical approaches, e.g. by averaging absolute values of unit displacements.

For our study, we generate a total of 1100 LW trajectories with various trajectory lengths $N_{step} \in \{10, 100, 200, \ldots, 900, 1000\}$. The trajectories are produced by the AnDi package [72]. The anomaly exponent α and the speed v are randomly given such that $\alpha \sim \text{unif}(1, 2)$ and $v \sim \text{unif}(0, 10)$, respectively. We standardize the trajectories using the protocol described in reference [81], in which the trajectories are regularized so that the standard deviation of the displacement distribution has $\sigma_D \sim \mathcal{N}(0, 1)$. This pre-processing is required to avoid the unexpected effect of step sizes on both the parameter estimation and model classification. We then add the three different levels of Gaussian noises to the 1100 trajectory with an $\text{SNR} = \sigma_D/\sigma_{noise} \in \{1, 2, 10\}$, thus in total generating 3300 noisy Lévy trajectories.

We perform our Bayesian inference analysis to these noisy LW trajectories and estimate the anomaly exponent α, more precisely, the power-law exponent of the flight-time distribution. In figure 5, we compare the predicted α to the ground-true value α_{GT} used for generating the data. In each scatter plot, the solid line is the average of the predicted α for given α_{GT}. When the background noise is not significant ($\text{SNR} = 10, 2$), the predicted α (solid line) is in good agreement with α_{GT}. We confirm that the estimation of α via the Bayesian inference is more accurate than that from the fitting of the MSD curve (the dashed line). When the noise becomes comparable with the signal ($\text{SNR} = 1$), both methods show poor performance.

For a more quantitative examination of the performance of our Bayesian inference method, we measure the absolute error (AE), $|\alpha_i - \alpha_{GT,i}|$, as a function of N_{step}, α_{GT}, and SNR. In AE, α_i and $\alpha_{GT,i}$ signify the predicted α and α_{GT} for the ith trajectory, respectively. In figure 6(a), we show the mean absolute error (MAE) as a function of N_{step}. For $\text{SNR} = 10$ and 2, the increase of trajectory length leads to more accurate estimation. For all trajectory lengths, the Bayesian method performs much better than the fitting from MSDs. For the case of $\text{SNR} = 1$, on the contrary, the inference goodness is not improved even if the length is increased. In this case, intense noises impede the accurate inference of flight times.

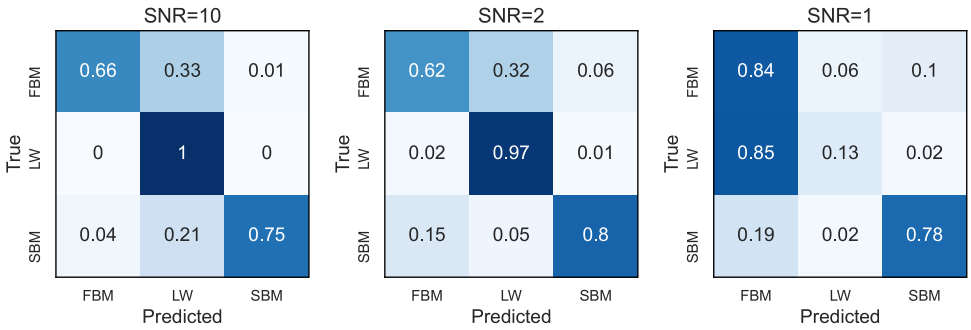

Figure 7. Confusion matrices obtained from the Bayesian model classification. From left to right, the matrices show the results at SNR $= 10$, 2, and 1, respectively. The (i, j)-component in a matrix indicates the portion of trajectories classified into the model $\mathcal{M}_j$ upon the true model $\mathcal{M}_i$.

Figure 6(b) shows the MAE from all trajectory lengths as a function of α_{GT}. For SNR $= 10$ and 2, the inference goodness is almost insensitive to α_{GT} although the error seems to be relatively larger at $\alpha_{\mathrm{GT}} = 1$ and 2. Note that even at SNR $= 10$, the MAE from the fitting method (MSD) is considerably high for these cases. When the noise level is too high (SNR $= 1$), the inference performance tends to be poorer as α_{GT} is close to unity. This tendency occurs because the zigzag flight patterns frequently occurring in LW trajectories with $\alpha_{\mathrm{GT}} \sim 1$ are easily confused with the noisy dynamics (when an SNR is small), so the short flight times cannot be correctly extracted. Consequently, as seen in figure 5 (SNR $= 1$), the LW trajectory with $\alpha_{\mathrm{GT}} \approx 1$ is inferred to the LW with a predicted $\alpha > 1$.

In figure 6(c) we examine the effect of SNR on MAE. The result shows that MAE abruptly decreases with increasing an SNR from 1 to ≈2, and after then it saturates for SNR $\gtrsim 2$. This suggests that under a noisy environment the Bayesian inference method extracts the information of α (i.e. the power-law exponent in the flight-time distribution) with the accuracy at the noise-free condition. Surprisingly, such a robust performance is achievable against the noise level as significant as SNR ≈ 2. We also notice that even in the strong noise condition (SNR $\lesssim 1$), the Bayesian inference is better than the fitting of MSD method in the parameter estimation.

5. Results II: model classification

In this section, we apply our Bayesian inference method for classifying models and comparing their likelihoods. Our task is to compare several diffusion models for a given trajectory in the scheme of Bayesian inference and find the most probable model in accordance with the data. For this task, we additionally consider two diffusion models—apart from the LW —describing anomalous diffusion with $0 < \alpha \leqslant 2$. The two anomalous models are FBM [26] and SBM [29, 30]. We refer to a recent review paper [1] for the mathematical introduction of these models and their applications to biology and various physical systems. Although the three models (LW, FBM, and SBM) can describe the superdiffusive process quantified by the MSD (1), they are based on distinct theoretical context and physical mechanisms. Accordingly, their statistical characteristics are distinguished, enabling us to use the Bayesian approach for model inference.

For a numerical test, we generate trajectory data for the three diffusion models. The data preparation is the same as in section 4.2 for the noisy LW trajectory. For each model, there are a total of 1100 trajectories with $N_{\mathrm{step}} \in \{10, 100, \ldots, 900, 1000\}$ (and 100 trajectories for

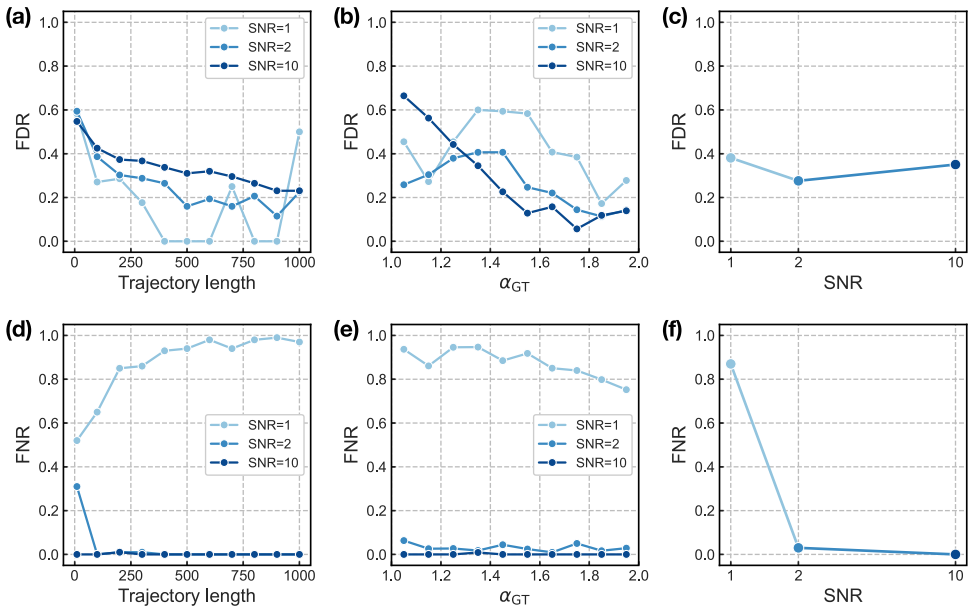

Figure 8. Error rates obtained from confusion matrices as a function of the trajectory length, α_{GT}, and SNR. (a) & (d) FDR and FNR of LWs as a function of the trajectory length. (b) & (e) FDR and FNR of LWs as a function of α_{GT} with half bin width $|\Delta\alpha_{GT}| = 0.05$. (c) & (f) FDR and FNR of LWs as a function of SNR. Each data point represents the averaged value over trajectory length and α_{GT}.

given N_{step}). The α is randomly sampled from [1, 2] for the three diffusion models. All of these trajectories are standardized and superimposed with background noise in the same way described in section 4.2. The likelihood functions for FBM [44] and SBM [82] are available from our previous work. We refer to reference [44] for our Bayesian inference of FBM. Based on that, we have implemented the MATLAB code for the Bayesian model inference for LW, FBM, and SBM [71].

Figure 7 shows the confusion matrices summarizing the performance of the Bayesian model classification upon the above-prepared data set. When SNR $\geqslant 2$, our Bayesian inference gives high true positive rates for all three processes. Especially, the LW trajectory is remarkably well classified by our Bayesian method. On the contrary, if the noise level is significantly high (SNR $= 1$), the LW is the most poorly inferred model among the three models. In this case, the Gaussian noise tends to make the LW falsely detected as FBM (a Gaussian and stationary-incremental process). The high true positive rates for FBM and SBM at SNR $= 1$, comparable to those at SNR $= 2$ and 10, are attributed to the same effect.

For more quantitative analysis of the model classification, we study error rates of the LW model as a function of N_{step}, α, and SNR. We estimate the false discovery rate (FDR) and false negative rate (FNR) from the confusion matrices. The former is the portion of FBM and SBM trajectories in the set of trajectories classified into LW, and the latter is the portion of misidentified LW trajectories (as FBM or SBM) in the set of LW trajectories.

Figures 8(a) and (d), show the FDR and FNR as a function of trajectory length. The FDR tends to decrease with trajectory length irrespective of an SNR. Interestingly, FDR at SNR $= 1$ seems to be smaller than that for SNR $\geqslant 2$ for most trajectory lengths. It appears that non-LW trajectories have an increased chance of being classified into our LW model when an SNR is

sufficiently large. This tendency can be seen in figures 7 and 8(c). The FNR is shown to be very small (except for SNR $= 1$) and tends to decrease with the trajectory length. We note that FNR behaves against the noise in the opposite way with FDR, such that a higher noise level leads to an increased FNR. As explained, at the highest noise level (SNR $= 1$), the LW trajectories are apt to be classified into FBM. Here, increasing trajectory length elevates the FNR because longer trajectories contain more evidence on the noise contribution and, subsequently, increase the chance of being falsely detected as FBM.

In figures 8(b) and (d), we examine the variation of FDR and FNR as a function of $\alpha_{\rm GT}$. There is a clear tendency that the FDR increases as the ground true α approaches unity (figure 8(b)). This is because when $\alpha = 1$ the three models converge to the same process, i.e. Brownian motion. Hence, when a given (anomalous) diffusion process has $\alpha \approx 1$, it is very difficult to pinpoint the correct underlying diffusion model. Ideally, the FDR should be $2/3$ at $\alpha_{\rm GT} = 1$ if there are three models. However, FNR (at SNR $= 2$ and 10) is rarely affected by the ground truth anomaly exponent. We find from further analysis of the data that ambiguous Brownian-like trajectories around $\alpha_{\rm GT} \sim 1$ tends to be classified into LW than FBM or SBM. The rise of FNR at SNR $= 1$ is the falsely detected FBM from the LW trajectories.

6. Discussion & summary

In this work, we have proposed a hidden Markov model for LWs and, using this model, developed Bayesian inference tools for the analysis of LW-like trajectory data. The essential part of our theory was to approximate a renewal process governed by a power-law sojourn time distribution with a Markovian decomposition method [74]. We have shown that a power-law flight-time distribution can be generated by this method, enabling us to construct the likelihood functions of an LW. Using the Bayesian inference tool, we have demonstrated to extract the value of the power-law exponent of flight-time distribution ($\gamma = 3 - \alpha$) or, equivalently, the anomaly exponent α from given LW-like trajectories (containing background noises). It has been confirmed that the Bayesian inference method outperforms the traditional fitting method by extracting α from an MSD. When the noise level is negligible (SNR $= 10$), the accuracy of parameter estimation is as high as that by the top-performing machine learning methods recently developed [81, 83]. For the case of 2D LW trajectories, remarkably, our Bayesian inference method is shown to perform better than the top-performing machine learning tool (which will be separately reported elsewhere). Apart from the parameter estimation, we have employed our Bayesian method to classify anomalous diffusion models. For three distinct diffusion processes (LW, FBM, and SBM) superimposed with noises, our Bayesian method successfully infers the correct model with a high true positive rate when the noise level is low (SNR $= 10, 2$).

Our Bayesian inference method based on the hidden Markov model can be extended to analyzing multi-dimensional data or potentially applied to other similar problems. Below we provide some remarks on these issues.

Firstly, it is straightforward to generalize the current model for the analysis of two-dimensional (2D) LWs. The 2D LWs are applicable to the modeling of cell motility on the surface [84, 85], the projected 2D motion of actively transported cargoes in the cell [62, 86], swarming dynamics of bacteria [12, 87], and various foraging dynamics in ecology [15, 52–54]. We can define a LW in 2D and construct a hidden Markov model for 2D LWs. The method is briefly sketched in the appendix B. A full investigation of the Bayesian inference method to 2D LWs will be reported elsewhere. One remarkable result is that the estimation of α is more accurate in 2D than in 1D. This is attributed to the fact that successive flight events

are more easily identified in 2D, which leads to the increased performance in the parameter estimation of 2D LWs at a higher SNR.

Secondly, our LW Bayesian model can be easily expanded to include composite correlated random walk (CCRW) processes via an analogous Markovian decomposite method introduced in this work. The CCRW is a multi-state random walk in which each state has a short-term (Markovian) directional memory leading to an exponential step-length distribution [46, 88]. The foraging dynamics of animals is an important example of the CCRW [16, 89, 90]. The stochastic properties of a CCRW are governed by the superposition rule of the multiple Markovian states. In the framework of CCRWs, the LW can be regarded as a special model where the superposition is specified by equations (16) and (17), leading to a power-law flight-time (or step-length) distribution. The CCRW may have a different step-length distribution depending on how multiple Markovian states are superposed. By properly modeling this part, a specific CCRW can be constructed and inferred by the corresponding Bayesian method. A potential application is to identify a correct diffusion model for animal's foraging dynamics. There is a stimulating effort in ecology for modeling the foraging motion [16, 89–93]. Often the analysis of the data was based on the step-length distribution (or the inverse cumulative distribution of the step lengths), which led to a debate on whether the data follows a LW having a power-law step length distribution or a different CCRW [16, 46, 89, 90]. Using the Bayesian method, one can analyze the data in a different manner and efficiently differentiate between several candidate models.

Thirdly, our hidden Markov model is applicable to construct Bayesian inference methods for variant LW models. For example, one can expand the current hidden Markov model for a LW with fluctuating velocities or a LW with a rest [27]. The latter model was shown to explain the fluid dynamics in a rotating plate and the transport of mRNA-protein complex particles in a neuron cell [60]. The Bayesian method will be helpful for extracting the information of parameters in the flight-time and rest-time distributions. Beyond the LW, once the noise correlation in the successive displacement is adequately incorporated, the current hidden Markov model can be applied to continuous-time random walks with a power-law waiting time distribution.

Finally we comment on some potential technical issues of the current Bayesian approach. The high computational cost is often required for Bayesian inference, which needs to sample parameters from high dimensional posterior distributions and calculate likelihood functions for every sampled parameter. It costs hundreds to thousands of seconds to quantify a trajectory of hundreds of steps. Notwithstanding, it is computationally feasible to handle thousands of trajectories of which step length is of <1000. In our study, we analyzed such trajectory data within a day using a 40-multicore workstation. Additionally, a relevant task is to improve the Bayesian inference method under a strong noise signal. In this work, we have found that the Bayesian inference analysis is vulnerable to noise when SNR ~ 1. We anticipate that if the noise effect term in the likelihood function (equation (26)) was incorporated without approximation, the Bayesian inference would have a better performance to noisy trajectories under the condition of strong noise signals. We leave this task as future work.

Acknowledgments

This work was supported by the National Research Foundation (NRF) of Korea (Grant No. 2020R1A2C4002490). S Thapa acknowledges support in the form of a Sackler postdoctoral fellowship and funding from the Pikovsky-Valazzi matching scholarship, Tel Aviv University.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://github.com/mlomholt/andi.

Appendix A. Continuous-time Markov chain approximation of continuous target distribution

Here we briefly show the mathematical derivation of equation (15). The continuous target distribution is given by the inter-arrival time distributions and the weights as

$$\begin{aligned}\psi_{\mathrm{c}}^{\mathrm{target}}(\tau;\alpha) &\approx \sum_{i=1}^{N_k} P_{i;\alpha}\phi_i(\tau) = \sum_{i=1}^{N_k}\left(\frac{k_i^{3-\alpha}e^{-k_i}}{\sum_{j=1}^{N_k}k_j^{3-\alpha}e^{-k_j}}\right)k_i e^{-k_i(\tau-1)} \\ &= \frac{1}{\sum_{j=1}^{N_k}k_j^{3-\alpha}e^{-k_j}}\sum_{i=1}^{N_k}k_i^{4-\alpha}e^{-k_i\tau}.\end{aligned} \tag{A.1}$$

In this expression, the denominator $\sum_{j=1}^{N_k}k_j^{3-\alpha}e^{-k_j}$ is estimated under the assumption of a sufficiently large N_k to

$$\begin{aligned}\sum_{j=1}^{N_k}\left(\frac{k_{\mathrm{fast}}}{b^j}\right)^{3-\alpha}e^{-\frac{k_{\mathrm{fast}}}{b^j}} &\approx \int_{\frac{1}{2}}^{\infty}\mathrm{d}x\left(\frac{k_{\mathrm{fast}}}{b^x}\right)^{3-\alpha}e^{-\frac{k_{\mathrm{fast}}}{b^x}} \\ &= \frac{1}{\ln b}\gamma\left(3-\alpha, k_{\mathrm{fast}}/\sqrt{b}\right)\end{aligned}, \tag{A.2}$$

where $\gamma(s,x) = \int_0^x t^{s-1}e^{-t}\,\mathrm{d}t$ is the lower incomplete gamma function. The numerator can be calculated in the same way. This approximation yields

$$\sum_{i=1}^{N_k}P_{i;\alpha}\phi_i(\tau) \approx \frac{\gamma\left(4-\alpha, k_{\mathrm{fast}}\tau/\sqrt{b}\right)}{\gamma\left(3-\alpha, k_{\mathrm{fast}}/\sqrt{b}\right)}\frac{1}{\tau^{4-\alpha}}. \tag{A.3}$$

When k_{fast} becomes considerably large, the prefactor approaches to

$$\frac{\gamma\left(4-\alpha, k_{\mathrm{fast}}\tau/\sqrt{b}\right)}{\gamma\left(3-\alpha, k_{\mathrm{fast}}/\sqrt{b}\right)} \approx \frac{\Gamma(4-\alpha)}{\Gamma(3-\alpha)} = 3-\alpha. \tag{A.4}$$

Thus, $\sum_{i=1}^{N_k}P_{i;\alpha}\phi_i(\tau)$ becomes the normalized continuous target distribution given by equation (11).

Appendix B. Modeling two-dimensional Lévy walks

In the Markov chain approximation of 1D LW in the main text, we need two independent latent state sets (s_i^+ and s_i^- with $i = 1, 2, \ldots, N_k$) for the flight of positive and negative directions, respectively. In the 2D LW [94], we set there are a total of N_θ velocity orientations (see figure B1(a)). A random walker can choose a random direction from the velocity vectors $\{\vec{v_1}, \vec{v_2}, \ldots, \vec{v_{N_\theta}}\}$ with a probability of $1/N_\theta$. We allow a velocity vector $\vec{v_n}$ to have uncertainty

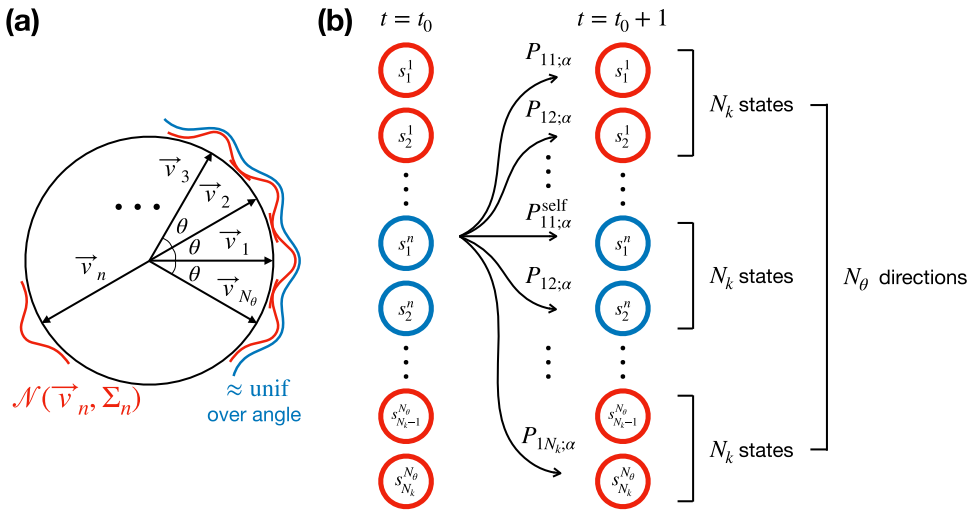

Figure B1. (a) The velocity states in our 2D LW. There are N_θ discrete velocity states specified by $\vec{v}_n$ $(n = 1, 2, \ldots, N_\theta)$. We allow the orientation of $\vec{v}_n$ to fluctuate around the unit vector $(\cos((n-1)\theta), \sin((n-1)\theta))$, which is modeled by a Gaussian fluctuation $\mathcal{N}(\vec{v}_n, \Sigma_n)$. The superposition of the Gaussian fluctuations (the blue solid line) approximates the uniformly distributed velocity states over angle. (b) The Markov chain model for the 2D discrete-time LW. Each velocity states in (a) have the latent state s_i^n with $i = 1, 2, \ldots, N_k$. The superscript n denotes the velocity state. The transition probability from the state s_i^m to s_j^n is given by $P_{ij;\alpha}$ (equation (B.2)), and the self-transition probability is given by $P_{ii;\alpha}^{\text{self}}$ (equation (B.1)).

in its direction, which is modeled by a Gaussian fluctuation (see the red lines in figure B1(a)). We obtain a nearly uniform distribution over the velocity direction (blue line in figure B1(a)) once all the Gaussian fluctuations $\mathcal{N}(\vec{v}_n, \Sigma_n)$ are superposed.

There are a total of N_θ number of velocity states in the 2D LW, where there are N_k latent states with independent transition rates $k_i (i = 1, 2, \ldots, N_k)$ for each velocity state. The probability of self-transition is

$$P_{ii;\alpha}^{\text{self}} = e^{-k_i} + \frac{(1 - e^{-k_i}) \cdot P_{i;\alpha}}{N_\theta}, \tag{B.1}$$

and the transition probability to other states reads

$$P_{ij;\alpha} = \frac{(1 - e^{-k_i}) \cdot P_{j;\alpha}}{N_\theta}. \tag{B.2}$$

The emission probability of an observation $\Delta\vec{x}_t = \vec{x}_{t+1} - \vec{x}_t$ from the velocity state $\vec{v}_n$ is given by

$$P(\Delta\vec{x}_t | \mathcal{M}_{\text{Lévy}}, \vec{v}_n, \Sigma_n) = \frac{\exp\left(-\frac{1}{2}(\Delta\vec{x}_t - \vec{v}_n)^{\mathrm{T}} \Sigma_n^{-1} (\Delta\vec{x}_t - \vec{v}_n)\right)}{\sqrt{(2\pi)^d |\Sigma_n|}}, \tag{B.3}$$

where $\Sigma_n = R(\theta_n)\Sigma_1 R(\theta_n)^{\mathrm{T}}$ is a covariance matrix for Gaussian fluctuations of $\vec{v}_n$, θ_n is an angle from x-axis of a vector $\vec{v}_n$, and $R(\theta_n)$ is a rotation matrix. From the Markov chain illustrated in figure B1(b) with transition probabilities (B.1) and (B.2) and the emission probability

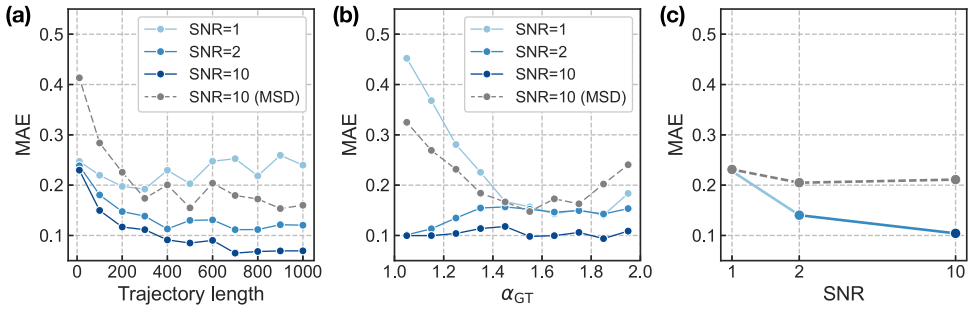

Figure B2. Parameter estimation for 2D LW trajectories. (a) MAE as a function of the trajectory length. (b) MAE as a function of the ground-truth anomaly exponent α_{GT}. (c) MAE as a function of SNR. The dotted line is the result from the Bayesian inference and the dot-dashed line from the fitting of MSD curves.

(B.3), the likelihood function for 2D LWs can be calculated using forward algorithm as in equations (27)–(32). The likelihood function for a 3D LW model can also be obtained in a similar way.

Figure B2 shows the performance of the parameter estimation. While most of the results are similar to the one-dimensional case, we notice some interesting features: the accuracy of parameter estimation has been improved at SNR $= 10$ compared to the corresponding case in 1D. Figure B2 shows that MAE decays faster than 1D as trajectory length increases. Moreover, the MAE does not change upon the variation of anomaly exponents (figure B2(b)). At SNR $= 10$, the overall MAE decreases by $\sim$0.02 compared to the 1D case (figure B2(c)).

ORCID iDs

Seongyu Park https://orcid.org/0000-0002-6478-660X
Samudrajit Thapa https://orcid.org/0000-0002-4340-5900
Michael A Lomholt https://orcid.org/0000-0002-3698-0958
Jae-Hyung Jeon https://orcid.org/0000-0003-0658-9672

References

[1] Metzler R, Jeon J-H, Cherstvy A G and Barkai E 2014 Anomalous diffusion models and their properties: non-stationarity, non-ergodicity, and ageing at the centenary of single particle tracking *Phys. Chem. Chem. Phys.* **16** 24128–64
[2] Saxton M J and Jacobson K 1997 Single-particle tracking:applications to membrane dynamics *Annu. Rev. Biophys. Biomol. Struct.* **26** 373–99
[3] Golding I and Cox C E 2006 Physical nature of bacterial cytoplasm *Phys. Rev. Lett.* **96** 098102
[4] Jeon J-H, Tejedor V, Burov S, Barkai E, Selhuber-Unkel C, Berg-Sørensen K, Oddershede L and Metzler R 2011 *In vivo* anomalous diffusion and weak ergodicity breaking of lipid granules *Phys. Rev. Lett.* **106** 048103
[5] Di Rienzo C, Piazza V, Gratton E, Beltram F and Cardarelli F 2014 Probing short-range protein brownian motion in the cytoplasm of living cells *Nat. Commun.* **5** 1–8
[6] Krapf D and Metzler R 2019 Strange interfacial molecular dynamics *Phys. Today* **72** 48–54
[7] Lee D S W, Wingreen N S and Brangwynne C P 2021 Chromatin mechanics dictates subdiffusion and coarsening dynamics of embedded condensates *Nat. Phys.* **17** 531–8

[8] Park S, Lee O-c., Durang X and Jeon J-H 2021 A mini-review of the diffusion dynamics of dna-binding proteins: experiments and models *J. Korean Phys. Soc.* **78** 408–26
[9] Wang B, Kuo J, Bae S C and Granick S 2012 When Brownian diffusion is not Gaussian *Nat. Mater.* **11** 481–5
[10] Cherstvy A G, Thapa S, Wagner C E and Metzler R 2019 Non-Gaussian, non-ergodic, and non-Fickian diffusion of tracers in mucin hydrogels *Soft Matter* **15** 2526–51
[11] Kim W K, Chudoba R, Milster S, Roa R, Kanduč M and Dzubiella J 2020 Tuning the selective permeability of polydisperse polymer networks *Soft Matter* **16** 8144–54
[12] Gil A, Rabani A, Benisty S, Partridge J D, Harshey R M and Be'Er A 2015 Swarming bacteria migrate by Lévy walk *Nat. Commun.* **6** 8396
[13] Perez Ipiña E, Otte S, Pontier-Bres R, Czerucka D and Peruani F 2019 Bacteria display optimal transport near surfaces *Nat. Phys.* **15** 610–5
[14] Patteson A E, Gopinath A, Goulian M and Arratia P E 2015 Running and tumbling with e. coli in polymeric solutions *Sci. Rep.* **5** 1–11
[15] Humphries N E *et al* 2010 Environmental context explains Lévy and Brownian movement patterns of marine predators *Nature* **465** 1066–9
[16] de Jager M, Weissing F J, Herman P M J, Nolet B A and van de Koppel J 2011 Lévy walks evolve through interaction between movement and environmental complexity *Science* **332** 1551–3
[17] Wergen G, Bogner M and Krug J 2011 Record statistics for biased random walks, with an application to financial data *Phys. Rev.* E **83** 051109
[18] Plerou V, Gopikrishnan P, Nunes Amaral L A, Gabaix X and Eugene Stanley H 2000 Economic fluctuations and anomalous diffusion *Phys. Rev.* E **62** R3023–6
[19] Brown R 1828 XXVII. A brief account of microscopical observations made in the months of June, July and August 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies *Phil. Mag.* **4** 161–73
[20] Einstein A 1956 *Investigations on the Theory of the Brownian Movement* ed R Fürth (New York: Dover)
[21] von Smoluchowski M 1906 Zur kinetischen theorie der brownschen molekularbewegung und der suspensionen *Ann. Phys.* **326** 756–80
[22] Reverey J F, Jeon J-H, Bao H, Leippe M, Metzler R and Selhuber-Unkel C 2015 Superdiffusion dominates intracellular particle motion in the supercrowded cytoplasm of pathogenic acanthamoeba castellanii *Sci. Rep.* **5** 11690
[23] Lampo T J, Stylianidou S, Backlund M P, Wiggins P A and Spakowitz A J 2017 Cytoplasmic RNA-protein particles exhibit non-Gaussian subdiffusive behavior *Biophys. J.* **112** 532–42
[24] Scher H and Montroll E W 1975 Anomalous transit-time dispersion in amorphous solids *Phys. Rev.* B **12** 2455–77
[25] Montroll E W 1969 Random walks on lattices: III. Calculation of first-passage times with application to exciton trapping on photosynthetic units *J. Math. Phys.* **10** 753–65
[26] Mandelbrot B B and Van Ness J W 1968 Fractional Brownian motions, fractional noises and applications *SIAM Rev.* **10** 422–37
[27] Zaburdaev V, Denisov S and Klafter J 2015 Lévy walks *Rev. Mod. Phys.* **87** 483
[28] Massignan P, Manzo C, Torreno-Pina J A, García-Parajo M F, Lewenstein M and Lapeyre G J 2014 Nonergodic subdiffusion from brownian motion in an inhomogeneous medium *Phys. Rev. Lett.* **112** 150603
[29] Lim S C and Muniandy S V 2002 Self-similar Gaussian processes for modeling anomalous diffusion *Phys. Rev.* E **66** 021114
[30] Jeon J-H, Chechkin A V and Metzler R 2014 Scaled brownian motion: a paradoxical process with a time dependent diffusivity for the description of anomalous diffusion *Phys. Chem. Chem. Phys.* **16** 15811–7
[31] Metzler R, Tejedor V, Jeon J-H, He Y, Deng W H, Burov S and Barkai E 2009 Analysis of single particle trajectories: from normal to anomalous diffusion *Acta Phys. Pol.* B **40** 1315–31
[32] Jeon J-H, Leijnse N, Oddershede L B and Metzler R 2013 Anomalous diffusion and power-law relaxation of the time averaged mean squared displacement in worm-like micellar solutions *New J. Phys.* **15** 045011
[33] Safdari H, Cherstvy A G, Chechkin A V, Thiel F, Sokolov I M and Metzler R 2015 Quantifying the non-ergodicity of scaled brownian motion *J. Phys. A: Math. Theor.* **48** 375002
[34] Cherstvy A G and Metzler R 2015 Ergodicity breaking and particle spreading in noisy heterogeneous diffusion processes *J. Chem. Phys.* **142** 144105

[35] Burov S, Jeon J-H, Metzler R and Barkai E 2011 Single particle tracking in systems showing anomalous diffusion: the role of weak ergodicity breaking *Phys. Chem. Chem. Phys.* **13** 1800–12
[36] Ernst D, Köhler J and Weiss M 2014 Probing the type of anomalous diffusion with single-particle tracking *Phys. Chem. Chem. Phys.* **16** 7686–91
[37] Kepten E, Weron A, Sikora G, Burnecki K and Garini Y 2015 Guidelines for the fitting of anomalous diffusion mean square displacement graphs from single particle tracking experiments *PLoS One* **10** e0117722
[38] Jamali V, Hargus C, Ben-Moshe A, Aghazadeh A, Dong Ha H, Mandadapu K K and Paul Alivisatos A 2021 Anomalous nanoparticle surface diffusion in lctem is revealed by deep learning-assisted analysis *Proc. Natl Acad. Sci.* **118** e2017616118
[39] Muñoz-Gil G, Garcia-March M A, Manzo C, Martín-Guerrero J D and Lewenstein M 2020 Single trajectory characterization via machine learning *New J. Phys.* **22** 013010
[40] Cichos F, Gustavsson K, Mehlig B and Volpe G 2020 Machine learning for active matter *Nat Mach Intell* **2** 94–103
[41] Granik N, Weiss L E, Nehme E, Levin M, Chein M, Perlson E, Roichman Y and Shechtman Y 2019 Single-particle diffusion characterization by deep learning *Biophys. J.* **117** 185–92
[42] Lysy M, Pillai N S, Hill D B, Forest M G, Mellnik J W R, Vasquez P A and McKinley S A 2016 Model comparison and assessment for single particle tracking in biological fluids *J. Am. Stat. Assoc.* **111** 1413–26
[43] Thapa S, Lomholt M A, Krog J, Cherstvy A G and Metzler R 2018 Bayesian analysis of single-particle tracking data using the nested-sampling algorithm: maximum-likelihood model selection applied to stochastic-diffusivity data *Phys. Chem. Chem. Phys.* **20** 29018–37
[44] Krog J, Jacobsen L H, Lund F W, Wüstner D and Lomholt M A 2018 Bayesian model selection with fractional brownian motion *J. Stat. Mech.* 093501
[45] Krog J and Lomholt M A 2017 Bayesian inference with information content model check for Langevin equations *Phys. Rev.* E **96** 062106
[46] Auger-Méthé M, Derocher A E, Plank M J, Codling E A and Lewis M A 2015 Differentiating the Lévy walk from a composite correlated random walk *Methods Ecol. Evol.* **6** 1179–89
[47] Palyulin V V, Chechkin A V and Metzler R 2014 Levy flights do not always optimize random blind search for sparse targets *Proc. Natl Acad. Sci.* **111** 2931–6
[48] Lomholt M A, Ambjörnsson T and Metzler R 2005 Optimal target search on a fast-folding polymer chain with volume exchange *Phys. Rev. Lett.* **95** 260603
[49] Koren T, Lomholt M A, Chechkin A V, Klafter J and Metzler R 2007 Leapover lengths and first passage time statistics for Lévy flights *Phys. Rev. Lett.* **99** 160602
[50] Palyulin V V, Blackburn G, Lomholt M A, Watkins N W, Metzler R, Klages R and Chechkin A V 2019 First passage and first hitting times of Lévy flights and Lévy walks *New J. Phys.* **21** 103028
[51] Shlesinger M F and Klafter J 1986 Lévy walks versus Lévy flights *On Growth and Form* (Berlin: Springer) pp 279–83
[52] Raichlen D A, Wood B M, Gordon A D, Mabulla A Z P, Marlowe F W and Pontzer H 2014 Evidence of Levy walk foraging patterns in human hunter-gatherers *Proc. Natl Acad. Sci.* **111** 728–33
[53] Reynolds A M and Rhodes C J 2009 The Lévy flight paradigm: random search patterns and mechanisms *Ecology* **90** 877–87
[54] Garg K and Kello C T 2021 Efficient Lévy walks in virtual human foraging *Sci. Rep.* **11** 5242
[55] Wosniack M E, Santos M C, Raposo E P, Viswanathan G M and da Luz M G E 2017 The evolutionary origins of lévy walk foraging *PLOS Comput. Biol.* **13** 10–31
[56] Rhee I, Shin M, Hong S, Lee K, Kim S J and Chong S 2011 On the Levy-walk nature of human mobility *IEEE/ACM Trans. Netw.* **19** 630–43
[57] Focardi S, Montanaro P and Pecchioli E 2009 Adaptive Lévy walks in foraging fallow deer *PLOS ONE* **4** 1–6
[58] Huo H, He R, Zhang R, Yuan J and Alexandre G 2021 Swimming Escherichia coli cells explore the environment by Lévy walk *Appl. Environ. Microbiol.* **87** e02429
[59] Sims D W *et al* 2008 Scaling laws of marine predator search behaviour *Nature* **451** 1098–102
[60] Song M S, Moon H C, Jeon J-H and Park H Y 2018 Neuronal messenger ribonucleoprotein transport follows an aging Lévy walk *Nat. Commun.* **9** 344
[61] Gal N and Weihs D 2010 Experimental evidence of strong anomalous diffusion in living cells *Phys. Rev.* E **81** 020903
[62] Chen K, Wang B and Granick S 2015 Memoryless self-reinforcing directionality in endosomal active transport within living cells *Nat. Mater.* **14** 589–93

[63] Stefani F D, Hoogenboom J P and Barkai. E 2009 Beyond quantum jumps: blinking nanoscale light emitters *Phys. Today* **62** 34–9
[64] Barthelemy P, Bertolotti J and Wiersma D S 2008 A Lévy flight for light *Nature* **453** 495–8
[65] Solomon T H, Weeks E R and Swinney H L 1993 Observation of anomalous diffusion and Lévy flights in a two-dimensional rotating flow *Phys. Rev. Lett.* **71** 3975–8
[66] Klafter J and Zumofen G 1994 Lévy statistics in a Hamiltonian system *Phys. Rev.* E **49** 4873–7
[67] Geisel T, Nierwetberg J and Zacherl A 1985 Accelerated diffusion in Josephson junctions and related chaotic systems *Phys. Rev. Lett.* **54** 616–9
[68] Gelman A, Carlin J B, Stern H S, Dunson D B, Vehtari A and Rubin D B 2013 *Bayesian Data Analysis* (Boca Raton, FL: CRC Press)
[69] Bayes T 1763 An essay towards solving a problem in the doctrine of chances *Philos. Trans. R. Soc. London* **53** 370–418
[70] Neal R M 2001 Annealed importance sampling *Stat. Comput.* **11** 125–39
[71] Park S, Thapa S, Kim Y, Lomholt M A and Jeon J-H 2021 Bayesian inference for Andi challenge https://github.com/mlomholt/andi (accessed 3 June 2021)
[72] Muñoz-Gil G, Requena B, Volpe G, Garcia-March M A and Manzo C 2021 The anomalous diffusion challenge data set https://github.com/AnDiChallenge/ANDI_datasets
[73] Muñoz-Gil G, Volpe G, García-March M A, Metzler R, Lewenstein M and Manzo C 2020 The anomalous diffusion challenge: single trajectory characterisation as a competition *Emerging Topics in Artificial Intelligence 2020*
[74] Goychuk I 2009 Viscoelastic subdiffusion: from anomalous to normal *Phys. Rev.* E **80** 046125
[75] Beck C and Cohen E G D 2003 Superstatistics *Physica* A **322** 267–75
[76] Kogan S 2008 *Electronic Noise and Fluctuations in solids* (Cambridge: Cambridge University Press)
[77] Surdin M 1939 Fluctuations de courant thermionique et le «Flicker effect» *J. Phys. Radium* **10** 188–9
[78] Du Pré F K 1950 A suggestion regarding the spectral density of flicker noise *Phys. Rev.* **78** 615
[79] Van Der Ziel A 1950 On the noise spectra of semi-conductor noise and of flicker effect *Physica* **16** 359–72
[80] Rabiner L R 1989 A tutorial on hidden Markov models and selected applications in speech recognition *Proc. IEEE* **77** 257–86
[81] Muñoz-Gil G *et al* 2021 Objective comparison of methods to decode anomalous diffusion (arXiv:2105.06766)
[82] Thapa S 2021 Bayesian inference with scaled Brownian motion *J. Phys.* A (unpublished)
[83] The ANDI challenge 2021 Andi interactive tool https://andi-challenge.org/interactive-tool/ (accessed 3 June 2021)
[84] Yang T D, Park J-S, Choi Y, Choi W, Ko T-W and Lee K J 2011 Zigzag turning preference of freely crawling cells *PLoS One* **6** e20255
[85] Li H, Qi S, Jin H, Qi Z, Zhang Z, Fu L and Luo Q 2015 Zigzag generalized Lévy walk: the in vivo search strategy of immunocytes *Theranostics* **5** 1275
[86] Fedotov S, Korabel N, Waigh T A, Han D and Allan V J 2018 Memory effects and Lévy walk dynamics in intracellular transport of cargoes *Phys. Rev.* E **98** 042136
[87] Be'er A and Gil A 2019 A statistical physics view of swarming bacteria *Mov. Ecol.* **7** 1–17
[88] Codling E A, Plank M J and Benhamou S 2008 Random walk models in biology *J. R. Soc. Interface.* **5** 813–34
[89] Jansen V A A, Mashanova A and Petrovskii S 2012 Comment on 'Lévy walks evolve through interaction between movement and environmental complexity' *Science* **335** 918
[90] Reynolds A M 2014 Mussels realize Weierstrassian Lévy walks as composite correlated random walks *Sci. Rep.* **4** 4409
[91] Reynolds A M 2014 Distinguishing between Lévy walks and strong alternative models: reply *Ecology* **95** 1109–12
[92] Benhamou S 2006 Detecting an orientation component in animal paths when the preferred direction is individual-dependent *Ecology* **87** 518–28
[93] Benhamou S 2007 How many animals really do the Lévy walk? *Ecology* **88** 1962–9
[94] Zaburdaev V, Fouxon I, Denisov S and Barkai E 2016 Superdiffusive dispersals impart the geometry of underlying random walks *Phys. Rev. Lett.* **117** 270601