Multiscale Properties of Random Walk Models of Animal Movement: Lessons from Statistical Inference

REIICHIRO KAWAI AND SERGEI PETROVSKII

Abstract

The random search problem has long attracted continuous interest due to its broad interdisciplinary range of applications, including animal foraging, facilitated target location in biological system and human motion. In this paper, we address the issue of statistical inference for ordinary Gaussian, Pareto, tempered Pareto and fractional Gaussian random walk models, which are among the most studied random walk models proposed as the best strategy in the random search problem. Based on rigorous analysis of the local asymptotic normality property and the Fisher information, we discuss some issues in unbiased joint estimation of the model parameters, in particular, the maximum likelihood estimation. We present that there exist both theoretical and practical difficulties in more realistic tempered Pareto and fractional Gaussian random walk models from a statistical standpoint. We discuss our findings in the context of individual animal movement and show how our results may be used to facilitate the analysis of movement data and to improve the understanding of the underlying stochastic process.

1 Introduction

Peculiarities of individual animal movement have been attracting considerable attention over the last three decades (e.g. see Mandelbrot [25]; Kareiva and Shigesada [18]; Viswanathan et al. [54, 57]; Bartumeus et al. [3]; Reynolds et al. [43]; Codling et al. [8]) as they are thought to hold the key to the understanding of the animal dispersal, and hence to better understanding of the spatiotemporal phenomena in population dynamics such as biological invasions, pattern formation etc. [35, 51]. Moreover, since the observed patterns in the animal movement (in particular, in foraging behavior) are thought to optimize the search success [2], it may help to better understand some crucially important phenomena such as population outbreaks or species extinction as well as the evolutionary aspects of the population dynamics.

An animal path is typically continuous and that provides a considerable challenge as a consistent theoretical framework allowing for analysis of continuous paths is yet largely missing [30, 31], although some progress has recently been made [36, 37]. In contemporary practice of animal movement studies, the path is usually mapped into a broken line where the line's nodes correspond to animal position at certain observation times. Generally speaking, the time-step Δt between two subsequent observations can vary along the path; here, for the sake of simplicity, we assume it to be constant (as is indeed the case with many field and laboratory studies). The movement along the broken line can then be quantified by the probability distribution for the step size or "jump" between two subsequent positions and by the probability distribution for the turning angle [51]. Therefore, researches have to work with finite amount of data. Moreover, in order to reach a good understanding of the underlying stochastic process, it often appears necessary to work with rare events, e.g. long jumps, where the data are scarce by definition. Since scarce data are subject to a large statistical fluctuation, analysis of rare events is a challenging issue which requires application of advanced statistical tools.

Application of statistical techniques to animal movement data usually consists of two steps. The first one is the choice of the model, e.g. the type of the step size distribution. Examples of widely used models are given by the Brownian motion (also known as Gaussian random walk), fractional Brownian motion, and Lévy walk or flight (aka Pareto random walk). Note that animal movement usually consists of more than one mode [29], so that it can be a combination of different movement types or behavioral states [5]. Once the distribution is chosen using an appropriate model selection procedure [6] (e.g. the Akaike information criteria), the next step is to estimate the model parameters

Published in *Proceedings of the Royal Society A* doi:10.1098/rspa.2011.0665. Email Address: {reiichiro.kawai, sp237}@le.ac.uk. Postal Address: Department of Mathematics, University of Leicester, Leicester LE1 7RH, United Kingdom.

such as the variance, the value of cutoff at small steps (where applies) etc. This is normally done by using the maximum likelihood method.

Standard approaches, however, tell little about the possible error of the parameter estimate. Meanwhile, a large error may reduce the practical value of the estimate to zero. Consider a hypothetical example when experimental data, for instance on the jump size distribution, are nicely fitted by a theoretical curve. In case the error of the parameter estimation is small, that can be regarded as a successful interpretation of the data and hence can make it possible to identify the process behind the observed pattern. However, a large error would make this interpretation very questionable. Even more importantly, although the second step of model fitting (i.e. parameter estimation) tends to be regarded as a technical stage rather than a conceptual one, there can be a logical loop when results from parameter estimation may overturn the hypothesis made at the stage of model selection. An example of such situation is considered below.

Note that mapping a continuous curve onto a discrete set (or a broken line) is of course not a one-to-one correspondence, and hence some information is lost. A question arises as to how important this information loss can be. In particular, whether the pattern of movement may change depending on the observations frequency has been a controversial issue. Using data simulated from a mathematical model, Reynolds [44] showed that, if for a certain Δt the movement pattern along the broken line is a Lévy walk, then a "sub-sampling" with a larger Δt will also result in a Lévy walk and even with a heavier tail. Similar results were obtained by Plank and Codling [39] who also emphasized that standard statistical tools may sometimes fail to distinguish between the power law and the exponential rate of decay in the step size distribution and hence fail to distinguish between the two qualitatively different movement patterns (see also [7]). A question remains largely open as to what may happen in case of a small/decreasing Δt. Note that this is not only important but also a timely issue as the developments in relevant laboratory and field equipment makes it possible to obtain high-resolution data on animal position.

Assuming that the observation frequency is properly established, another question arises as to how large should be the total time scale of observations in order to provide robust information about the movement pattern. That revokes a recent discussion in the literature that, in some cases, the Lévy flight may be difficult to identify in the movement data as it results in the pattern similar to that of the correlated random walk [1, 45]. This issue was first addressed by Viswanathan et al. [56] who, having considered the tangling impact of the turning angles [4], showed that distinguishing between different types of movement (e.g. correlated random walk vs Lévy flight) can only only possible when the data span over a sufficiently large time-scale, i.e. when the time series of animal positions is long enough. Here we consider this problem from a different angle. Having chosen parameter estimators for the step size distribution, we reveal their convergence rate(s) for a few most commonly used random walk models. This, in principle, makes it possible to estimate the amount of data required to provide an estimate of parameters with a given accuracy.

In this paper, we address the issue of statistical analysis of animal random walk data using a mathematically rigorous approach borrowed from the general theory of stochastic processes. We revisit a few most commonly used random walk types and reveal the convergence rate for different parameters. Analysis of the power law and the truncated power law distributions lead to a conclusion that the best-fitting hypothesis may indeed depend on the time scale of observations. Moreover, we show that an increase in the observation frequency may change the whole pattern of movement unless some a posteriori biological information is used. We thus conclude that the random walk is not robust to the time scale of the observation. Finally, we discuss how our findings may help researches working on individual animal movement to reveal the movement pattern and to identify the movement type.

1.1 Outline of the Method

The local asymptotic normality property is a vital concept in asymptotically optimal statistical analysis. In short, the local asymptotic normality property for a differentiable statistical model for the parameter $\theta \in \mathbb{R}^d$ to be estimated is defined through the weak convergence of the likelihood ratio to the Gaussian shift experiment; for each $h \in \mathbb{R}^d$,

$$\frac{d\mathbb{P}_{\theta+R_n(\theta)h}}{d\mathbb{P}_\theta}\Big|_{\mathscr{F}_n} \xrightarrow{\mathscr{L}} \exp\left[\langle h, Z(\theta)\rangle - \frac{1}{2}\langle h, \mathscr{I}(\theta)h\rangle\right],$$

under $\mathbb{P}_\theta$, where $\mathbb{P}_\theta|_{\mathscr{F}_n}$ is a probability measure associated with θ restricted to the filtration $\mathscr{F}_n$, $\{R_n(\theta)\}_{n\in\mathbb{N}}$ is a sequence of diagonal matrices in $\mathbb{R}^{d\times d}$ whose diagonal entries tend to zero, $\mathscr{I}(\theta)$ is a non-negative definite deterministic matrix in $\mathbb{R}^{d\times d}$, called the Fisher information matrix, and $Z(\theta) \sim \mathscr{N}(0, \mathscr{I}(\theta))$ under $\mathbb{P}_\theta$. If the above weak convergence holds, then we say that *the local asymptotic normality (LAN) property holds at point θ with the rate*

$R_n(\theta)$ *and the Fisher information matrix* $\mathscr{I}(\theta)$. The convergence is equivalent to the following locally asymptotically quadratic structure of log likelihood function

$$\ell_n(\theta + R_n(\theta)h) - \ell_n(\theta) = \langle h, R_n(\theta)\nabla_\theta(\ell_n(\theta))\rangle - \frac{1}{2}\langle h, R_n(\theta)\mathrm{Hess}_\theta(\ell_n(\theta))R_n(\theta)h\rangle + o_{\mathbb{P}_\theta}(1), \quad (1.1)$$

where the vector $R_n(\theta)\nabla_\theta(\ell_n(\theta))$ converges in law to $\mathscr{N}(0,\mathscr{I}(\theta))$ under $\mathbb{P}_\theta$, and where the matrix $R_n(\theta)\mathrm{Hess}_\theta(\ell_n(\theta))R_n(\theta)$ converges in $\mathbb{P}_\theta$-probability to $\mathscr{I}(\theta)$. Moreover, if the LAN property holds with non-singular $\mathscr{I}(\theta)$, then a unbiased estimator $\{\widehat{\theta}_n\}_{n\in\mathbb{N}}$ of θ is said to be asymptotically efficient in a neighborhood of θ if

$$R_n(\theta)^{-1}\left(\widehat{\theta}_n - \theta\right) \stackrel{\mathscr{L}}{\to} \mathscr{N}\left(0, \mathscr{I}(\theta)^{-1}\right),$$

under $\mathbb{P}_\theta$, that is, such estimators achieve asymptotically the Cramer-Rao lower bound $\mathscr{I}(\theta)^{-1}$ for the estimation variance. (We refer the reader to [9, 23, 24, 41] for thorough details.) Let us remark that the concept of asymptotic efficiency resides within the situation where asymptotic normality of the estimator holds valid. The efficiency here is not necessarily identical to the optimality in a general sense, which we will demonstrate with concrete examples in Section 2.2 and 2.3.

Let us close this introductory section with some notation which will be used throughout the text. We denote by $\mathbb{R}^d$ the d-dimensional Euclidean space with the norm $\|\cdot\|$ and the inner product $\langle\cdot,\cdot\rangle$. We use the incomplete gamma function

$$\Gamma(s,x) := \int_x^{+\infty} t^{s-1}e^{-t}dt,$$

for $s\in\mathbb{R}$ and $x>0$, where the case $s<0$ is well defined as long as $x>0$. Let us denote by $\mathbb{I}_d$ the identity matrix in $\mathbb{R}^{d\times d}$. We denote by $\mathscr{N}(\gamma,\Sigma)$ the Gaussian law with mean γ and covariance matrix Σ, that is, its probability density function is given by

$$\phi(x) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}}\exp\left[-\frac{1}{2}\langle x-\gamma, \Sigma^{-1}(x-\gamma)\rangle\right], \quad x\in\mathbb{R}^d,$$

and by $\mathrm{Exp}(\lambda)$ the exponential distribution with rate $\lambda>0$, that is, its probability density function is given by

$$f(x) = \lambda e^{-\lambda x}, \quad x>0.$$

2 Random Walk Models

Let us begin with a general modeling base, which applies to all the discrete time random walks models of this paper. Let $\{X_n\}_{n\in\mathbb{N}}$ be a discrete time stochastic process in $\mathbb{R}^2$. In this paper, we call it a *(discrete time) random walk* if

(i) the increments $\{X_n - X_{n-1}\}_{n\in\mathbb{N}}(=:\{Z_n\}_{n\in\mathbb{N}})$ forms a sequence of identically distributed (not necessarily independent) random vectors in $\mathbb{R}^2$,

(ii) the standardized increments $\{Z_n/\|Z_n\|\}_{n\in\mathbb{N}}$ form a sequence of independent and identically distributed (iid) uniform random vectors on the unit sphere in $\mathbb{R}^2$.

The condition (ii) means that each step has no preference in direction, which corresponds to an isotropic random walk; e.g. see [8].

2.1 Ordinary Gaussian Random Walks

We mean by *ordinary Gaussian random walk* (also known in applications as the Brownian motion) a random walk with $\{Z_n\}_{n\in\mathbb{N}}$ being iid with Gaussian distributed length and with no directional preference. To define this precisely and uniquely, we first transform, without any essential loss of information, each step Z_n in $\mathbb{R}^2$ into a random variable W_n (that is, in $\mathbb{R}$) and an angle $\eta_n\in(0,\pi]$ in the polar decomposition

$$Z_n = W_n\begin{bmatrix}\cos\eta_n\\ \sin\eta_n\end{bmatrix}. \quad (2.1)$$

By restricting the angles $\{\eta_n\}_{n\in\mathbb{N}}$ to the half circle $(0,\pi]$, we let the random variables $\{W_n\}_{n\in\mathbb{N}}$ possess a sign. In the case of the ordinary Gaussian random walk model, $\{\eta_n\}_{n\in\mathbb{N}}$ is a sequence of iid random variables in $(0,\pi]$ and $\{W_n\}_{n\in\mathbb{N}}$ is a sequence of iid centered Gaussian, that is, $\mathscr{L}(W_1)=\mathscr{N}(0,\sigma^2)$, where $\sigma>0$ is the parameter to be estimated. It is well known that the maximum likelihood estimator $\widehat{\sigma}_n^2 := n^{-1}\sum_{k=1}^n \|Z_k\|^2$ is asymptotically efficient and satisfies under $\mathbb{P}_{\sigma_0}$ that $\widehat{\sigma}_n^2 \to \sigma_0^2$, *a.s.*, and

$$\sqrt{n}\left(\widehat{\sigma}_n^2-\sigma_0^2\right) \stackrel{\mathscr{L}}{\to} \mathscr{N}\left(0,2\sigma_0^4\right).$$

We omit the proof as those are very well known results; for example, see [13, 23].

2.2 Pareto Random Walks

We call a random walk *Pareto random walk* if $\{\|Z_n\|\}_{n\in\mathbb{N}}$ forms a sequence of iid Pareto random variables whose probability density has the form

$$f(y;\alpha,0,\tau) := \frac{\alpha\tau^\alpha}{y^{\alpha+1}}, \quad y\in(\tau,+\infty), \tag{2.2}$$

where $\alpha>0$ and $\tau>0$ are parameters, $\alpha>0$ quantifying the rate of decay of the probability density at large y. In applications to animal movement, Pareto random walk as given by (2.2) is usually referred to as a "power law distribution," for instance, of the jump size y [3, 50, 54]. The corresponding pattern of individual animal movement is usually referred to as a Lévy flight. Obviously, it is only valid if the value of the 'cutoff' τ at small jumps is positive[1]. In the below, we will show that this simple observation sometimes may have crucial consequences for data interpretation.

Generation of Pareto random variables for modelling purposes is straightforward as the probability density function (2.2) has an inverse of its tail integral in closed form, that is,

$$\inf_{x\in(0,+\infty)}\left\{\int_x^{+\infty} f(y;\alpha,0,\tau)dy \ge u\right\} = \tau u^{-1/\alpha}, \quad u\in(0,1).$$

The algorithm is as simple as;

Algorithm 1

Step 1. Generate U as uniform $(0,1)$.

Step 2. Exit with $\tau U^{-1/\alpha}$.

In Figure 1, we draw typical sample paths of the rotation invariant Lévy flight in $\mathbb{R}^2$.

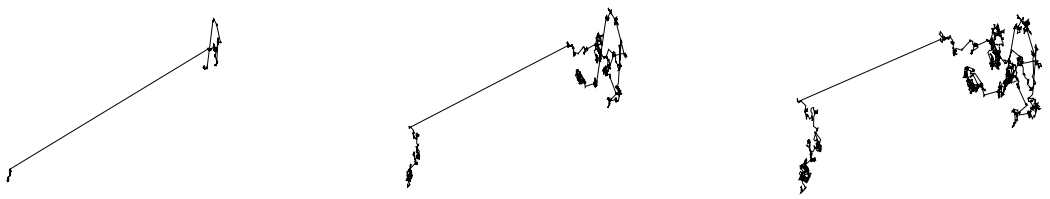

Figure 1: Typical sample paths of the rotation invariant Pareto random walk in $\mathbb{R}^2$ with $\tau=0.1$ for three different values of the probability density decay rate: $\alpha=1.0$ (left), $\alpha=1.5$ (center) and $\alpha=1.8$ (right).

The unknown parameter to be estimated in our hypothesis is $[\alpha,\tau]^\top$. To state results, let us prepare some notation. Let Θ_1 and Θ_2 be bounded convex domains satisfying

$$\Theta_1^- \subset \left\{[\alpha,\tau]^\top\in\mathbb{R}^2 |\, \alpha\in(0,+\infty),\, \tau\in(0,+\infty)\right\},$$
$$\Theta_2^- \subset \{\alpha\in\mathbb{R} |\, \alpha\in(0,+\infty)\}.$$

[1]Note that observation data on animal movement are usually distributed on the whole semiline, i.e. $f>0$ for any $y\ge 0$. It means that the events with small jump size should either be excluded for some a posteriori reasons (see Discussion) or otherwise the probability distribution function f should be defined additionally for $0\le y\le\tau$.

We denote by $\theta_0 := [\alpha_0, \tau_0]^\top \in \Theta_1$ the true value of the unknown parameter. We denote by $\mathbb{P}_\theta$ the probability measure associated with $\theta \in \Theta_1$, or with $\theta \in \Theta_2$ and τ_0 being known.

Theorem 2.1. *(i) The minimum* $\min_{k=1,\dots,n} \|Z_k\|$ *is the maximum likelihood estimator of* τ_0*. In particular, it holds under* $\mathbb{P}_{\theta_0}$*-a.s. that as* $n \uparrow +\infty$*,*

$$n\left(\min_{k=1,\dots,n} \|Z_k\| - \tau_0\right) \xrightarrow{\mathscr{L}} \mathrm{Exp}\left(\frac{\alpha_0}{\tau_0}\right). \tag{2.3}$$

(ii) Suppose that τ_0 *is known. The LAN property holds at the point* $\alpha \in \Theta_2$ *with*

$$R_n(\alpha) = \frac{1}{\sqrt{n}} =: R_n, \quad \mathscr{I}(\alpha) = \frac{1}{\alpha^2}.$$

In particular, letting

$$\widehat{\alpha}_n := \left(\frac{1}{n-1}\sum_{k=1}^{n} \ln \frac{\|Z_k\|}{\tau_0}\right)^{-1}, \quad n \in \mathbb{N},$$

it holds under $\mathbb{P}_{\theta_0}$ *that as* $n \uparrow +\infty$*,* $\widehat{\alpha}_n \to \alpha_0$*, a.s., and*

$$\sqrt{n}\,(\widehat{\alpha}_n - \alpha_0) \xrightarrow{\mathscr{L}} \mathscr{N}\left(0, \alpha_0^2\right). \tag{2.4}$$

Note that for each $n \in \mathbb{N}$, $\mathrm{Var}_\theta(\partial_\tau \sum_{k=1}^n \ln f(\|Z_k\|; \alpha, 0, \tau)) = 0$. This indicates that the Cramér-Rao bound for τ is infinite. In the meantime, due to (i) of Theorem 2.1, the simple minimum $\min_{k=1,\dots,n} \|Z_k\|$ of the sample converges to the true value at the much faster rate of n and is not asymptotically normal. It is noteworthy that this faster convergence in (2.3) is realized without knowing the true value of α. Therefore, in practice, given a certain number of sample steps, the parameter τ is estimated via (2.3), then the parameter (c, α) is estimated jointly via (2.4) with τ_0 being the estimate from the first procedure. Note that the maximum likelihood estimator of α is

$$\widehat{\alpha}_n^* = \left(\frac{1}{n}\sum_{k=1}^{n} \ln \|Z_k\| - \ln \tau_0\right)^{-1},$$

rather than $\widehat{\alpha}_n$ above. For each $n \in \mathbb{N}$, however, it is neither an unbiased estimator nor attains the Cramér-Rao bound. See Rytgaard [47] for more details. Also, once τ_0 is known, the Pareto sample is nothing but a disguised exponential sample, as $\mathscr{L}(\ln(\|Z_n\|/\tau_0)) = \mathrm{Exp}(\alpha)$. Hence, the Pareto sample with known τ_0 can be analyzed with various known techniques for exponential distributions.

2.3 Tempered Pareto Random Walks

As before, we call a random walk *tempered Pareto random walk* if $\{\|Z_n\|\}_{n\in\mathbb{N}}$ forms a sequence of iid tempered Pareto random variables, while there could exist two definitions.

2.3.1 Tempering Density Function

The first definition is given in the form of probability density function

$$f(y; \alpha, \kappa, \tau) = \frac{1}{\kappa^\alpha \Gamma(-\alpha, \kappa\tau)} \frac{e^{-\kappa y}}{y^{\alpha+1}} = \frac{e^{-\kappa y}}{\alpha(\kappa\tau)^\alpha \Gamma(-\alpha, \kappa\tau)} f(y; \alpha, 0, \tau), \quad y \in (\tau, +\infty), \tag{2.5}$$

where $\alpha \in (0, +\infty)$, $\kappa \in (0, +\infty)$ and $\tau \in (0, +\infty)$. For convenience, we interpret (2.5) as the distribution of the jump size y of animal movement along the path. Correspondingly, the distribution defined by (2.5) is usually known as a "truncated power law" or, in a more precise manner, as a power law with exponential cutoff at large distances [12, 26, 33]. In terms of the underlying stochastic process, truncated power law may indicate a transition from superdiffusive Lévy flight to a more usual, diffusive Brownian motion, e.g. due to the boundedness of the available space [12, 26].

In view of the limiting relation

$$\lim_{\kappa \downarrow 0} \kappa^\alpha \Gamma(-\alpha, \kappa\tau) = \frac{1}{\alpha\tau^\alpha}, \tag{2.6}$$

the Pareto random walk can also be thought of as a limit of this tempered Pareto random walk. This definition provides a simpler acceptance-rejection method.

Algorithm 2

Step 1. Generate U as uniform $(0,1)$.

Step 2. Generate V through **Algorithm 1**, independent of U.

Step 3. If $U \leq e^{-\kappa V}$, then exit with V. Otherwise, return to Step 1.

The acceptance rate at Step 3 is

$$\mathbb{P}\left(U \leq e^{-\kappa V}\right) = \mathbb{E}\left[e^{-\kappa V}\right] = \alpha(\kappa\tau)^{\alpha}\Gamma(-\alpha, \kappa\tau),$$

which tends up to 1 as $\kappa \downarrow 0$, while down to 0 as $\kappa \uparrow +\infty$. Hence, this acceptance-rejection method tends to terminate more quickly with smaller κ. In Figure 2, we draw typical sample paths of the rotation invariant truncated Lévy flight in $\mathbb{R}^2$. Evidently, sample paths look more like non-tempered Pareto random walk for smaller κ.

Figure 2: Typical sample paths of the rotation invariant tempered Pareto random walk ($(\alpha,\tau) = (1.5, 0.1)$) in $\mathbb{R}^2$ under two different tempering parameters; $\kappa = 0.01$ (left), $\kappa = 1.0$ (center) and $\kappa = 2.0$ (right).

The unknown parameter to be estimated in our hypothesis is $[\alpha, \kappa, \tau]^{\top}$. In a similar manner to before, let Θ_3 and Θ_4 be bounded convex domains satisfying

$$\Theta_3^- \subset \left\{ [\alpha,\kappa,\tau]^{\top} \in \mathbb{R}^3 \,|\, \alpha \in (0,+\infty),\, \kappa \in (0,+\infty),\, \tau \in (0,+\infty) \right\},$$
$$\Theta_4^- \subset \left\{ [\alpha,\kappa]^{\top} \in \mathbb{R}^2 \,|\, \alpha \in (0,+\infty),\, \kappa \in (0,+\infty) \right\}.$$

We denote by $\theta_0 := [\alpha_0, \kappa_0, \tau_0]^{\top} \in \Theta_3$ the true value of the unknown parameter and by $\mathbb{P}_{\theta}$ the probability measure associated with $\theta \in \Theta_3$, or with $\theta \in \Theta_4$ and τ_0 being known.

Theorem 2.2. *(i) The minimum* $\min_{k=1,\ldots,n} \|Z_k\|$ *is the maximum likelihood estimator of* τ_0. *In particular, it holds under* $\mathbb{P}_{\theta_0}$ *that as* $n \uparrow +\infty$,

$$n\left(\min_{k=1,\ldots,n} \|Z_k\| - \tau_0\right) \xrightarrow{\mathscr{L}} \mathrm{Exp}\left(\frac{e^{-\kappa_0\tau_0}}{\tau_0^{\alpha_0+1}\kappa_0^{\alpha_0}\Gamma(-\alpha_0, \kappa_0\tau_0)}\right). \tag{2.7}$$

(ii) Suppose that τ_0 *is known. The LAN property holds at the point* $\theta := [\alpha,\kappa]^{\top} \in \Theta_4$ *with*

$$R_n(\theta) = \frac{1}{\sqrt{n}}\mathbb{I}_2 =: R_n, \quad \mathscr{I}(\theta) = \begin{bmatrix} \mathscr{I}_{11}(\theta) & \mathscr{I}_{12}(\theta) \\ \mathscr{I}_{12}(\theta) & \mathscr{I}_{22}(\theta) \end{bmatrix},$$

where

$$\begin{aligned}
\mathscr{I}_{11}(\theta) &:= \frac{\int_{\kappa\tau_0}^{+\infty} z^{-\alpha-1}e^{-z}(\ln z)^2 dz}{\Gamma(-\alpha,\kappa\tau_0)} - \left(\frac{\int_{\kappa\tau_0}^{+\infty} z^{-\alpha-1}e^{-z}(\ln z) dz}{\Gamma(-\alpha,\kappa\tau_0)}\right)^2, \\
\mathscr{I}_{12}(\theta) &:= \frac{\int_{\kappa\tau_0}^{+\infty} z^{-\alpha-1}e^{-z}(\ln z) dz}{\kappa\Gamma(-\alpha,\kappa\tau_0)} \frac{\Gamma(1-\alpha,\kappa\tau_0)}{\Gamma(-\alpha,\kappa\tau_0)} - \frac{\int_{\kappa\tau_0} z^{-\alpha}e^{-z}(\ln z - \ln\kappa) dz}{\kappa\Gamma(-\alpha,\kappa\tau_0)}, \\
\mathscr{I}_{22}(\theta) &:= \frac{\Gamma(2-\alpha,\kappa\tau_0)}{\kappa^2\Gamma(-\alpha,\kappa\tau_0)} - \left(\frac{\Gamma(1-\alpha,\kappa\tau_0)}{\kappa\Gamma(-\alpha,\kappa\tau_0)}\right)^2.
\end{aligned} \tag{2.8}$$

In particular, the matrix $\mathscr{I}(\theta)$ is not singular.
(iii) For each $n \in \mathbb{N}$, define

$$\widehat{\theta}_n := \begin{bmatrix} \alpha_n \\ \kappa_n \end{bmatrix} := \frac{1}{(x_{1,n}-\tau_0)x_{3,n} - x_{2,n}^2 + \tau_0 x_{2,n}(x_{1,n}+\tau_0) - \tau_0^2 x_{1,n}^2} \begin{bmatrix} x_{1,n}x_{3,n} - 2x_{2,n}^2 + 2\tau_0 x_{1,n}x_{2,n} - \tau_0^2 x_{1,n}^2 \\ (x_{1,n}-2\tau_0)x_{2,n} + \tau_0^2 x_{1,n} \end{bmatrix}, \tag{2.9}$$

where $x_{p,n} := n^{-1}\sum_{k=1}^n \|Z_k\|^p$, $p = 1,2,3$, and define

$$\theta_n^* := \widehat{\theta}_n + \frac{1}{n}\mathscr{I}\left(\widehat{\theta}_n\right)^{-1} \sum_{k=1}^n \begin{bmatrix} \frac{\int_{\kappa_n\tau_0}^{+\infty} z^{-\alpha_n-1}e^{-z}(\ln z)dz}{\Gamma(-\alpha_n,\kappa_n\tau_0)} - \ln(\kappa_n\|Z_k\|) \\ \frac{\Gamma(1-\alpha_n,\kappa_n\tau_0)}{\kappa_n\Gamma(-\alpha_n,\kappa_n\tau_0)} - \|Z_k\| \end{bmatrix}. \tag{2.10}$$

Then, it holds under $\mathbb{P}_{\theta_0}$ that as $n \uparrow +\infty$, $\theta_n^* \to \theta_0$, a.s., and

$$\sqrt{n}(\theta_n^* - \theta_0) \xrightarrow{\mathscr{L}} \mathscr{N}\left(0, \mathscr{I}(\theta_0)^{-1}\right).$$

As before, estimation of the parameter τ is possible at the much faster rate n without knowing the true value of (α,κ). Efficient estimation of (α,κ) here is not a trivial problem, since the likelihood equation is highly involved and requires numerical approximation methods to solve. The condition (iii) of the above Theorem claims that the method of moments with the so-called method of scoring achieves the asymptotic efficiency. (It is not clear whether the method of moments alone achieves the asymptotic efficiency.) Let us however note two possible issues in this approach. First, we might be careful in estimation accuracy of $x_{l,n}$'s, in particular of higher orders, which may require an extremely great number of sample when the distribution has a heavy tail, that is, when κ_0 is close to zero. Also, it is a little cumbersome to compute the matrix $\mathscr{I}(\theta_n^*)$ for every n, which may be addressed by somehow avoiding to update the matrix at each iteration.

In principle, the model based on the tempered Pareto distribution is not appropriate either when κ is very large or when it is very close to zero. If κ is very large, on the one hand, the tempered Pareto distribution is nearly exponential at infinity. A simpler model based on the exponential distribution is then enough to describe sample paths. Note that the issue of higher moments in the above method of moments is likely to disappear, while estimates of α_0 tend to be very unstable in return. On the other hand, if κ is very close to 0, then the tempered Pareto distribution is nearly non-tempered. In this case, the non-tempered Pareto model should work sufficiently well. The user can therefore select the models with a posteriori knowledge.

2.3.2 Tempering Tail Probability

There exists another definition of the tempered Pareto distribution based on the tail probability

$$\int_x^{+\infty} f(y;\alpha,\kappa,\tau)dy = \tau^\alpha e^{\kappa\tau}\frac{e^{-\kappa x}}{x^\alpha}, \quad x \in (\tau,+\infty), \tag{2.11}$$

where $\alpha \in (0,+\infty)$, $\kappa \in [0,+\infty)$, $\tau \in (0,+\infty)$ and where the probability density function $f(y;\alpha,\kappa,\tau)$ is also available in closed form

$$f(y;\alpha,\kappa,\tau) = \tau^\alpha e^{\kappa\tau}\frac{e^{-\kappa y}}{y^{\alpha+1}}(\alpha+\kappa y) = e^{-\kappa(y-\tau)}\left(1+\frac{\kappa}{\alpha}y\right)f(y;\alpha,0,\tau), \quad y \in (\tau,+\infty), \tag{2.12}$$

with $f(y;\alpha,0,\tau)$ being the Pareto density function (2.2). Clearly, it reduces to the Pareto density as soon as $\kappa = 0$. Rather than a numerical inversion of the tail probability (2.11), the density function suggests the following acceptance-rejection method.

Algorithm 3
Step 1. Generate U as uniform $(0,1)$.
Step 2. Generate V through **Algorithm 1**, independent of U.

Step 3. If

$$U \leq \left(e^{-\max(0,1-\alpha-\kappa\tau)} \max\left(1+\frac{\kappa\tau}{\alpha}, \frac{1}{\alpha}\right)\right)^{-1} \frac{f(V;\alpha,\kappa,\tau)}{f(V;\alpha,0,\tau)}$$
$$= \frac{e^{-\kappa(V-\tau)}(1+\kappa V/\alpha)}{e^{-\max(0,1-\alpha-\kappa\tau)}\max(1+\kappa\tau/\alpha, 1/\alpha)}$$

then exit with V. Otherwise, return to Step 1.
The acceptance rate at Step 3 is

$$\mathbb{P}\left(U \leq \frac{e^{-\kappa(V-\tau)}(1+\kappa V/\alpha)}{e^{-\max(0,1-\alpha-\kappa\tau)}\max(1+\kappa\tau/\alpha, 1/\alpha)}\right) = \mathbb{E}\left[\frac{e^{-\kappa(V-\tau)}(1+\kappa V/\alpha)}{e^{-\max(0,1-\alpha-\kappa\tau)}\max(1+\kappa\tau/\alpha, 1/\alpha)}\right]$$
$$= \frac{1}{e^{-\max(0,1-\alpha-\kappa\tau)}\max(1+\kappa\tau/\alpha, 1/\alpha)},$$

which tends up to $\alpha e^{1-\alpha}\mathbb{1}(\alpha \in (0,1)) + \mathbb{1}(\alpha \in [1,+\infty))$ as $\kappa \downarrow 0$, while down to 0 as $\kappa \uparrow +\infty$.

Recently, the estimation problem of the tempered Pareto distribution of this form is addressed in [27], based on Hill-type estimators. We simply describe the LAN property and refer the reader to [27] for more details on estimation. The sets Θ_3 and Θ_4 are the same respectively as the ones in Theorem 2.2.

Theorem 2.3. *(i) The minimum* $\min_{k=1,\ldots,n}\|Z_k\|$ *is the maximum likelihood estimator of* τ_0. *In particular, it holds under* $\mathbb{P}_{\theta_0}$ *that as* $n \uparrow +\infty$,

$$n\left(\min_{k=1,\ldots,n}\|Z_k\| - \tau_0\right) \xrightarrow{\mathscr{L}} \mathrm{Exp}\left(\frac{\alpha_0}{\tau_0} + \kappa_0\right).$$

(ii) Suppose that τ_0 *is known. The LAN property holds at the point* $\theta := [\alpha,\kappa]^\top \in \Theta_4$ *with*

$$R_n(\theta) = \frac{1}{\sqrt{n}}\mathbb{I}_2, \quad \mathscr{I}(\theta) = \begin{bmatrix} \mathbb{E}\left[\frac{1}{(\alpha+\kappa\|Z_1\|)^2}\right] & \mathbb{E}\left[\frac{\|Z_1\|}{(\alpha+\kappa\|Z_1\|)^2}\right] \\ \mathbb{E}\left[\frac{\|Z_1\|}{(\alpha+\kappa\|Z_1\|)^2}\right] & \mathbb{E}\left[\frac{\|Z_1\|^2}{(\alpha+\kappa\|Z_1\|)^2}\right] \end{bmatrix}.$$

In particular, the matrix $\mathscr{I}(\theta)$ *is not singular.*

2.4 Fractional Gaussian Random Walks

We mean by *fractional Gaussian random walk* a random walk with $\{Z_n\}_{n\in\mathbb{N}}$ being identically distributed with a common Gaussian distribution, while not necessarily independently distributed. We again use the polar decomposition (2.1) to transform $\{Z_n\}_{n\in\mathbb{N}}$ in $\mathbb{R}^2$ into a sequence $\{W_n\}_{n\in\mathbb{N}}$ of random variables in $\mathbb{R}$ with $\mathscr{L}(W_n) = \mathscr{N}(0,\sigma^2)$, $n \in \mathbb{N}$, and a sequence $\{\eta_n\}_{n\in\mathbb{N}}$ of iid uniform random variables over $(0,\pi]$. In addition, we hypothesize that steps are not necessarily independent and have the autocovariance structure

$$\mathrm{Cov}\left(W_{k_1}, W_{k_2}\right) = \frac{\sigma^2}{2}\left(|k_1-k_2+1|^{2H} - 2|k_1-k_2|^{2H} + |k_1-k_2-1|^{2H}\right), \quad k_1,k_2 \in \mathbb{N},$$

where $H \in (0,1)$ is often called the *Hurst* parameter. In this sense, the fractional Gaussian random walk lies in the class of correlated random walks. However, contrary to the standard CRW [18], here the correlation is between the step sizes rather than between the turning angles. Therefore, the path does not necessarily show any directional preference. In Figure 3, we draw typical sample paths of the fractional Gaussian random walk in $\mathbb{R}^2$. Note that when $H = 1/2$, this model reduces to the ordinary Gaussian random walk discussed in Section 2.1.

We want to mention it here that, as well as the models considered above, the fractional Gaussian random walk was suggested as a possible stochastic process underlying animal movement [15, 46].

The unknown model parameters to be estimated are $\theta := (\sigma,H) \in (0,+\infty)\times(0,1)$. Let Θ_5 be bounded convex domain satisfying

$$\Theta_5^- \subset \left\{[\sigma,H]^\top \in \mathbb{R}^2 \,|\, \sigma \in (0,+\infty), H \in (0,1)\right\},$$

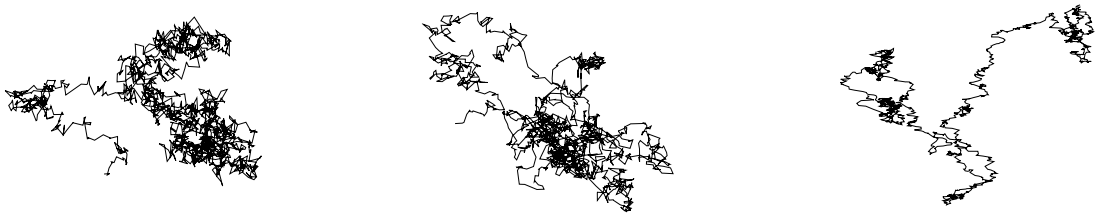

Figure 3: Typical sample paths of the fractional Gaussian random walk in $\mathbb{R}^2$ with three different Hurst parameters; $H = 0.15$ (left), $H = 0.5$ (center) and $H = 0.85$ (right).

We denote by $\theta_0 := [\sigma_0, H_0]^\top \in \Theta_5$ the true value of the unknown parameter and by $\mathbb{P}_\theta$ the probability measure associated with $\theta \in \Theta_5$. The random vector $Y_n := [W_1, \dots, W_n]$ in $\mathbb{R}^n$ has the probability density function

$$f_n(y;\theta) := \frac{1}{(2\pi)^{n/2}\sigma^n |T_n(H)|^{1/2}} \exp\left[-\frac{1}{2\sigma^2}\left\langle y, T_n(H)^{-1} y\right\rangle\right], \quad y \in \mathbb{R}^n, \quad \theta \in \Theta_5,$$

where $T_n(H)$ is the symmetric Toeplitz matrix which indicates the autocorrelation structure

$$(T_n(H))_{k_1,k_2} := \mathrm{Corr}\left(W_{k_1}, W_{k_2}\right) = \frac{1}{2}\left(|k_1 - k_2 + 1|^{2H} - 2\,|k_1 - k_2|^{2H} + |k_1 - k_2 - 1|^{2H}\right),$$

for $k_1, k_2 = 1, \dots, n$.

Theorem 2.4. *The LAN property holds at the point $\theta \in \Theta_5$ with*

$$R_n(\theta) := \frac{1}{\sqrt{n}}\mathbb{I}_2 =: R_n, \quad \mathscr{I}(\theta) := 2\begin{bmatrix} \frac{1}{\sigma^2} & \frac{\ln H}{\sigma} \\ \frac{\ln H}{\sigma} & (\ln H)^2 \end{bmatrix}.$$

In particular, the matrix $\mathscr{I}(\theta)$ is singular for every $\theta \in \Theta_5$.

Due to the singularity of the Fisher information matrix, the conventional asymptotic efficiency theory is not applicable to the full joint estimation of the two parameters. Nevertheless, as soon as either σ or H is fixed, the matrix reduces to $\mathbb{R}^{1\times 1}$ with a strictly positive entry. It is noteworthy that the singularity across σ and H occurs also under high-frequency sampling, as discussed in Kawai [19, 20]. We omit a practical estimation procedure which attains the above efficiency, as it is essentially very involved. For details on estimation based on Whittle maximum likelihood estimators and a plug-in version of maximum likelihood estimators, we refer the reader to, for example, [10, 14]. From a parameter estimation point of view, thus, this model is not necessarily the best candidate among the random walk models discussed in this paper.

3 Discussion and Conclusions

Analysis and interpretation of animal movement data is one of the most controversial issues in contemporary ecology. Importance of rare events such as long-distance jumps and, on the other hand, availability of large amount of high-resolution data due to recent developments in the methodology and equipment for field and laboratory experiments [26, 31] pose a considerable challenge for data analysis. Commonly used statistical approaches do not always allow distinguishing unambiguously between qualitatively different processes such as, for instance, correlated random walk and the Lévy flight, and that has caused a heated debate [5, 12, 39, 45, 52]. Moreover, it remains a largely open question to what extent the inherently continuous animal movement can be adequately described by discrete random walk models as it is intuitively clear (cf. Turchin [51]) that the results may depend on the time-step of the observation [7, 39, 44].

In this paper, we addressed this issue by considering the problem of parameter estimation for random walk data. Having applied methods of the stochastic processes theory (the Fisher information for unbiased joint estimation of

the model parameters) we provided ready-to-use formulas to estimate parameter values for a few widely used random walk models such as ordinary and fractional Gaussian random walk, Pareto random walk (also known as power law) and tempered Pareto random walk (aka truncated power law). We have shown that there exist both theoretical and computational difficulties in the "more realistic" models. In particular, in the tempered Pareto model, a root of its likelihood equation is unavailable in closed form. In the fractional Gaussian random walk model, the conventional asymptotic optimality theory is not applicable to the full joint estimation of its two model parameters as the Fisher information matrix is singular. This fact suggests that much care should be taken in the choice of the underlying model. Moreover, a posteriori knowledge may be important; see the comments at the end of section 2.3.1.

Note that here we focused on Gaussian, Pareto and tempered Pareto random walk models as they are widely regarded as paradigms of a few qualitatively different types of individual animal movement. However, our approach is not reduced to these cases only, and its extension onto some other cases may be possible. For instance, a relatively straightforward yet meaningful extension of the proposed framework can be done by introducing some preference in directions for Gaussian and fractional Gaussian random walks, instead of them being equiprobable. It is readily seen that, if the distribution of the turning angles is symmetric with respect to zero (e.g. like in the standard correlated random walk), then our results remain essentially the same.

3.1 Theoretical implications

One of the main problems in individual animal movement studies is to identify the pattern of movement. Now we are going to discuss how our findings may help to solve this problem. However, before we move to a further discussion, a general remark seems to be necessary in order to clarify the issue. In the literature on individual animal movement, the power law distribution $f(y) \sim y^{-(\alpha+1)}$ (cf. Eq. (2.2)) is often called "scale-free" for the reason that, if $\alpha \in (0,1]$, then the average step size $\langle y \rangle$ does not exist as the corresponding integral is diverging. However, we believe that this terminology is confusing and misleading. We want to emphasize it here that, whatever is the type of the individual movement (e.g. Brownian motion or a Lévy flight or walk), the movement is never really scale-free[2]. There *always* exists a characteristic step size, although it may be defined somewhat differently. Below we mention just a few possibilities:

Median of the distribution. Consider y_* such as

$$\int_0^{y_*} f(y)dy \;=\; \int_{y_*}^{\infty} f(y)dy \;=\; \frac{1}{2}\,. \tag{3.1}$$

Obviously, y_* gives a characteristic scale of the random walk with a very clear meaning.

Dimensions analysis 1. Another argument resulting in a characteristic scale is based on the observation that, if we consider the distribution $f(y)$ of the step size, then y is not an abstract number but has the dimension of length, which we denote as $[y] = L$. Correspondingly, the probability density distribution f is measured not in abstract units but has the dimension of inverse length, i.e. $[f] = L^{-1}$. It means that, apart from a senseless case of $\alpha = 0$, the step size probability distribution function must always include a parameter with the dimension of length, which then can be regarded as a characteristic scale. In particular, we immediately observe that Eq. (2.2) contains the cutoff parameter $[\tau] = L$.

Dimensions analysis 2. Note that the step size probability distribution with an asymptotic power law behaviour at large y can be written in a different form (known as a Cauchy-type distribution):

$$f(y) \;=\; \frac{C_{p,\alpha}}{p^{\alpha+1}+y^{\alpha+1}} \;\sim\; \frac{C_{p,\alpha}}{y^{\alpha+1}}, \quad y \uparrow +\infty, \tag{3.2}$$

where $p > 0$ is a parameter and $C_{p,\alpha}$ is the normalizing constant. Equation (3.2) does not contain a cutoff but it has got another parameter p instead. It is readily seen that $[p] = L$, so it can be regarded as a characteristic scale of the movement.

Similar arguments apply to truncated power law as well. Note that here we talk not about an external spatial scale which can arise due to the boundedness of the space available for the random walk (consider, for instance, dispersal of terrestrial birds in a small island). Although its importance for the movement pattern has been widely recognized

[2]The situation is different in time-continuous stochastic processes, e.g. see [48].

[55] so that there is even an opinion that it may turn the power law into a truncated power law [12], the existence of an external spatial scale is in many cases obvious. Instead, here we talk about something much less obvious, i.e. about an inherent spatial scale that originates in the power law itself and hence would have been present even in an idealistic case of movement in unbounded space.

The actual issue with the power law distribution is not the absence of a characteristic scale – which, as we have shown it above, always exists – but whether this scale is representative of the process(es) resulting in a heavy-tailed distribution. It is this issue that our findings seem to make a significant contribution to the understanding of. For instance, in case of Pareto distribution (2.2) with α close to zero, the frequency of large jumps can be high but the scale parameter τ presumably remains small being estimated as the minimum value observed in the data; see (2.3). We have therefore shown that, in fact, parameter τ controls the type of the distribution. Indeed, the power law and truncated power law distributions are only formally valid when the minimum jump size is positive, whatever good may be the fitting of the data at large and intermediate jumps. This leads to a counter-intuitive conclusion that, for an individual movement described by Pareto distribution, we cannot separate the apparently different scales of small and large jumps and the result of the analysis of long-distance dispersal essentially depend on animal movements on small scale. We refer to this situation as a *genuine* Lévy flight. It is not scale-free, as it was suggested by earlier studies, but its specifics is that the small and large scales appear to be equally important and cannot be separated.

Applying this result to movement data, we note that animal movement is never uninterrupted, usually alternating periods of fast movement with periods of slow movement or rest [21, 22, 26, 32]. Therefore, having the time step chosen sufficiently small, which seems to be prescribed by a natural desire to have the data as detailed as possible, the data set inevitably includes periods of rest when the increment in the animal position is zero (up to the measurement error). In this case, a power law-type distribution is not valid since the estimation procedure gives $\tau = 0$, see the equations (2.3) and (2.7). However, by means of coarsening the time-grid, the zero-valued increments can be eliminated, which restore validity of the power law (or truncated power law) distribution.

A question then arises as to what should be the course of action if the data fitting at large and intermediate jumps do suggest a power law but the estimate (2.3) results in $\tau = 0$. We hypothesize that, in this situation, the movement pattern is not actually a Lévy flight and fitting the data with a power law is descriptive but not insightful. Such an 'artificial' power law may appear as a result of heterogeneity of the movement data, e.g. when movement tracks of different animals are pooled together [38] or when a few different movement modes are mixed [11], even when each of the tracks/modes may be a Brownian motion [17]. We refer to this situation as a *superficial* Lévy flight. We want to mention it here that, in this situation, a power law may appear to be the best fit simply because the range of hypothesis from which the choice is made was not sufficiently broad [11, 17]. It will then be outperformed by the true model (e.g. a composite Brownian motion) once this true probability distribution is added into the range [17]. Our conclusions are in a good agreement with an earlier study by Benhamou [5] who showed that one should distinguish between the Lévy flight with the Lévy pattern.

For similar reasons, we believe that the Cauchy-type distribution is not an appropriate model for a genuine Lévy flight. Unlike (2.2), distribution (3.2) only predicts a power law at very large jump size, but the deviation from the power law becomes significant already at the intermediate scale $y \sim p$. If (3.2) appears to be the best fit to explain movement data, that may indicate that the data is a mixture of different movement modes.

We therefore demonstrated that the properties of a given random walk can be essentially different on a different time scale, i.e. for a different frequency of observations. Note that the zero-valued increments or jumps can be removed from the data set by distinguishing between different movement modes, e.g. by separating the periods of actual movement (also known as bouts) from the periods of slow movement or rest [26]. However, this procedure is purely heuristic and usually based on a conventional value of a "threshold" jump value separating the movement and the rest. Although it is often used in practice, such a separation is lacking a solid theoretical basis, and the question about the possible impact of the threshold value on the type of the distribution as well as on the corresponding parameter values is largely open. Our findings suggest that not only parameter estimation but, in some cases, also the model itself can be quite sensitive to the choice of this threshold.

3.2 Implications for data analysis

On a more technical side, we have shown that, even within the same model, the convergence rate for different parameters can be different. Especially the latter seems to bring a practically important message that, given required accuracy of parameter estimation, the constraint on the minimum number of the data-points originates in the parameter with the slowest convergence rate. Taking into account recent growth in availability of high-resolution data (cf. Nathan et

al. [31]), this seems to be a useful and timely result.

In conclusion, we show how our results on the convergence rate and the residual error distribution may help to evaluate the amount of data that is required to estimate parameter values with a give accuracy or tolerance. Consider Pareto distribution (2.2) as an important example. Let $\tau_{(n)}$ is an estimate of τ obtain from a dataset containing n datapoints, i.e.

$$\tau_{(n)} \;=\; \min_{k=1,\dots,n} \|Z_k\|,$$

see (2.3). Suppose we wish to ensure with the $(100\sigma)\%$-confidence that the difference between the estimate $\tau_{(n)}$ and the actual value τ is less than the prescribed tolerance $\varepsilon > 0$. Then, we need at least the following number of observations:

$$n(\varepsilon,\sigma;\alpha,\tau) := \mathrm{ceil}\left(\frac{\ln(1-\sigma)}{\alpha(\ln(\tau)-\ln(\tau+\varepsilon))}\right), \tag{3.3}$$

where the function $\mathrm{ceil}(A)$ rounds A to the nearest integer greater than or equal to A. Indeed,

$$\begin{aligned}
\sigma = \mathbb{P}\left(\min_{k=1,\dots,n}\|Z_k\|-\tau\le\varepsilon\right) &= 1-\mathbb{P}\left(\min_{k=1,\dots,n}\|Z_k\|-\tau>\varepsilon\right)\\
&= 1-(\mathbb{P}(\|Z_1\|>\tau+\varepsilon))^n\\
&= 1-\left(\int_{\tau+\varepsilon}^{+\infty} f(y;\alpha,0,\tau)dy\right)^n\\
&= 1-\left(\frac{\tau}{\tau+\varepsilon}\right)^{\alpha n},
\end{aligned}$$

from which we immediately obtain (3.3).

If, by way of example, we now set $\varepsilon = \tau/10$ and $\sigma = 0.99$ (i.e. to estimate τ with 10% tolerance and 99% confidence), then we obtain:

$$n(\tau/10,0.99;\alpha,\tau) = \mathrm{ceil}\left(-\frac{\ln(0.01)}{\alpha\ln(1+0.1)}\right) \approx \mathrm{ceil}\left(\frac{48.317716}{\alpha}\right). \tag{3.4}$$

Note that the above estimate is obtained without any initial guess or a priori information about the true value of τ. For a few hypothetical values such as $\alpha = 0.5$, $\alpha = 1.0$ and $\alpha = 1.75$, we obtain, respectively:

$$n(\tau/10,0.99;0.5,\tau) = 97, \quad n(\tau/10,0.99;1.0,\tau) = 49, \quad n(\tau/10,0.99;1.75,\tau) = 33. \tag{3.5}$$

Interestingly, the smaller is α (i.e. the heavier is the tail of the distribution) the larger is the required amount of data.

Now we recall that the rate of convergence for τ is $n^{-1/2}$ while the rate of convergence for parameter α is n^{-1}. Omitting calculations details, roughly speaking it means that, in order to evaluate α, the number of datapoints required to obtain an estimate for τ (as given by Eqs. (3.4–3.5)) should be squared. Therefore, for Pareto distribution (2.2) with a relatively fast rate of decay ($\alpha = 1.75$), a dataset of about 10^3 datapoints should be sufficient to estimate both parameters with a good accuracy, while a heavier tail ($\alpha = 0.5$) may require a much larger data set of about 10^4.

Acknowledgements SP gratefully acknowledges funding from The Leverhulme Trust through grant F/00 568/X.

Appendix: Proofs

This section is devoted to proofs of the results. The complete proof entails rather lengthy arguments of somewhat routine nature. To avoid overloading the paper, we omit non-essential details in some instances.

Proof of Theorem 2.1 and 2.2. We prove the results under the setting of Theorem 2.2, as those of Theorem 2.1 can be easily recovered with the help of the asymptotics (2.6).
(i) The result follows from the straightforward algebra

$$\begin{aligned}\mathbb{P}_{\theta_0}\left(n\left(\min_{k=1,\dots,n}\|Z_k\|-\tau_0\right)\le x\right) &= 1-\left(\int_{\tau_0+x/n}^{+\infty} f(y;\alpha_0,\kappa_0,\tau_0)dy\right)^n \\ &= 1-\left(1-\frac{\int_{\kappa_0\tau_0}^{\kappa_0\tau_0+\kappa_0 x/n} z^{-\alpha_0-1}e^{-z}dz}{\Gamma(-\alpha_0,\kappa_0\tau_0)}\right)^n \\ &\to 1-\exp\left[-\frac{e^{-\kappa_0\tau_0}}{\tau_0^{\alpha_0+1}\kappa_0^{\alpha_0}\Gamma(-\alpha_0,\kappa_0\tau_0)}x\right],\end{aligned}$$

where the first equality holds since $\{\|Z_n\|\}_{n\in\mathbb{N}}$ is a sequence of iid random variables. The limiting function (in x) indicates the exponential distribution with the desired rate.
(ii) It holds that for each $n\in\mathbb{N}$,

$$\ln\frac{d\mathbb{P}_{\theta_0+R_nh}}{d\mathbb{P}_{\theta_0}}\Big|_{\mathscr{F}_n} = \ell_n(\theta_0+R_nh)-\ell_n(\theta_0),\quad \mathbb{P}_{\theta_0}\text{-}a.s.,$$

where for $\theta\in\Theta_4$,

$$\ell_n(\theta)=\ln\left(\prod_{k=1}^n f(\|Z_k\|;\alpha,\kappa,\tau_0)\right)=n\left(-\alpha\ln\kappa-\ln\Gamma(-\alpha,\kappa\tau_0)\right)-\sum_{k=1}^n\left(\kappa\|Z_k\|+(\alpha+1)\ln\|Z_k\|\right).$$

We can show with the pathwise Taylor expansion and the mean value theorem that

$$\begin{aligned}&\ell_n(\theta_0+R_nh)-\ell_n(\theta_0)= \\ &\qquad\left\langle h,R_n\nabla_\theta\left(\ell_n(\theta)\right)\big|_{\theta=\theta_0}\right\rangle+\left\langle R_nh,\left(\int_0^1\int_0^\delta \mathrm{Hess}_\theta\left(\ell_n(\theta)\right)\big|_{\theta=\theta_0+\varepsilon R_nh}d\varepsilon\,d\delta\right)R_nh\right\rangle\end{aligned}$$

where

$$\nabla_\theta(\ell_n(\theta))=\sum_{k=1}^n\begin{bmatrix}\frac{\int_{\kappa\tau_0}^{+\infty}z^{-\alpha-1}e^{-z}(\ln z)dz}{\Gamma(-\alpha,\kappa\tau_0)}-\ln(\kappa\|Z_k\|)\\ \frac{\Gamma(1-\alpha,\kappa\tau_0)}{\kappa\Gamma(-\alpha,\kappa\tau_0)}-\|Z_k\|\end{bmatrix},$$

and $-R_n\mathrm{Hess}_\theta(\ell_n(\theta))R_n$ is independent of n and is no longer a random matrix. First, it holds by the central limit theorem and by the continuity of the summands in (α,κ) that for each $\varepsilon\in(0,1)$, as $n\uparrow+\infty$,

$$-R_n\mathrm{Hess}_\theta\left(\ell_n(\theta)\right)\big|_{\theta=\theta_0+\varepsilon R_nh}R_n\to\mathscr{I}(\theta_0).$$

Note that

$$\sup_{\theta\in\Theta_4}\left\|\mathrm{Hess}_\theta\left(\ell_n(\theta)\right)\right\|_o<+\infty,$$

due to the compactness of Θ_4 and the continuity of the Hessian matrix in θ, where $\|A\|_o:=\sup_{\|x\|\le1}\|Ax\|$ denotes the operator norm of a linear transform, Hence, it holds that

$$R_n\left(\int_0^1\int_0^\delta \mathrm{Hess}_\theta\left(\ell_n(\theta)\right)\big|_{\theta=\theta_0+\varepsilon R_nh}d\varepsilon\,d\delta\right)R_n\to-\frac{1}{2}\mathscr{I}(\theta).$$

Next, we can deduce that for each $n\in\mathbb{N}$,

$$\mathbb{E}_\theta\left[\left(R_n\nabla_\theta\left(\ell_n(\theta)\right)\right)^{\otimes2}\right]=\mathbb{E}_\theta\left[\left(\nabla_\theta\left(\ell_{n,1}(\theta)\right)\right)^{\otimes2}\right]=\mathscr{I}(\theta),$$

where we denote by $\nabla_\theta(\ell_{n,1}(\theta))$ the independent summand in $\nabla_\theta(\ell_n(\theta))$, that is, $\nabla_\theta(\ell_n(\theta))=\sum_{k=1}^n\nabla_\theta(\ell_{n,k}(\theta))$. (See Lemma 2.5 of [27].) This shows the desired local asymptotic normality property. Note that $\nabla_\theta(\ell_{n,1}(\theta))$, which

we have defined just above, is independent of n. Hence, we can show that $\mathscr{I}(\theta)$ is not singular, with the help of the Cauchy-Schwarz inequality.

The rest of Theorem 2.1 is obvious as $\mathscr{I}(\theta)$ is not singular. For the part (iii) of Theorem 2.2, observe that

$$x_p := \mathbb{E}\left[\|Z_1\|^p\right] = \int_{\tau_0}^{+\infty} y^p f(y;\alpha,\kappa,\tau)\,dy = \frac{\Gamma(p-\alpha,\kappa\tau_0)}{\kappa^p\Gamma(-\alpha,\kappa\tau_0)}, \quad p=1,2,3.$$

Using the recurrence relation $\Gamma(s,z) = (s-1)\Gamma(s-1,z) + z^{s-1}e^{-z}$, we get

$$\begin{aligned}
x_1 &= \frac{-\alpha}{\kappa} + \frac{\kappa^{-\alpha-1}\tau_0^{-\alpha}e^{-\kappa\tau_0}}{\Gamma(-\alpha,\kappa\tau_0)},\\
x_2 &= \frac{1-\alpha}{\kappa}\frac{\Gamma(1-\alpha,\kappa\tau_0)}{\kappa\Gamma(-\alpha,\kappa\tau_0)} + \tau_0\frac{\kappa^{-\alpha-1}\tau_0^{-\alpha}e^{-\kappa\tau_0}}{\Gamma(-\alpha,\kappa\tau_0)} = \left(\frac{1-\alpha}{\kappa}+\tau_0\right)x_1 + \frac{\alpha\tau_0}{\kappa},\\
x_3 &= \frac{2-\alpha}{\kappa}\frac{\Gamma(2-\alpha,\kappa\tau_0)}{\kappa^2\Gamma(-\alpha,\kappa\tau_0)} + \tau_0^2\frac{\kappa^{-\alpha-1}\tau_0^{-\alpha}e^{-\kappa\tau_0}}{\Gamma(-\alpha,\kappa\tau_0)} = \frac{2-\alpha}{\kappa}x_2 + \tau_0^2\left(x_1+\frac{\alpha}{\kappa}\right).
\end{aligned}$$

By solving the above for α and κ, we get the equations (2.9). The desired results follows from the method of scoring. The proof is complete. □

Proof of Theorem 2.4. We have

$$\frac{d\mathbb{P}_{\theta_0+R_nh}}{d\mathbb{P}_{\theta_0}}\Big|_{\mathscr{F}_n} = \ell_n(\theta_0+R_nh) - \ell_n(\theta_0), \quad \mathbb{P}_{\theta_0}\text{-}a.s.,$$

where

$$\ell_n(\theta) = \ln\left(f_n(Y_n;\theta)\right) = -\frac{n}{2}\ln(2\pi) - n\ln\sigma - \frac{1}{2}\ln|T_n(H)| - \frac{1}{2\sigma^2}\left\langle Y_n, T_n(H)^{-1}, Y_n\right\rangle.$$

In what follows, we will use the notation

$$\nabla_\theta\left(\ell_n(\theta)\right) := \begin{bmatrix} \partial_\sigma\ell_n(\theta) \\ \partial_H\ell_n(\theta)\end{bmatrix}, \quad \mathrm{Hess}_\theta\left(\ell_n(\theta)\right) := \begin{bmatrix} \partial_\sigma^2\ell_n(\theta) & \partial_\sigma\partial_H\ell_n(\theta) \\ \partial_\sigma\partial_H\ell_n(\theta) & \partial_H^2\ell_n(\theta)\end{bmatrix},$$

where

$$\begin{aligned}
\partial_\sigma\ell_n(\theta) &:= -\frac{n}{\sigma} + \frac{1}{\sigma^3}\left\langle Y_n, T_n(H)^{-1}Y_n\right\rangle,\\
\partial_H\ell_n(\theta) &:= -\frac{1}{2}\partial_H\ln|T_n(H)| - \frac{1}{2\sigma^2}\left\langle Y_n, \partial_H(T_n(H)^{-1})Y_n\right\rangle,\\
\partial_\sigma^2\ell_n(\theta) &:= \frac{n}{\sigma^2} - \frac{3}{\sigma^4}\left\langle Y_n, T_n(H)^{-1}Y_n\right\rangle,\\
\partial_H^2\ell_n(\theta) &:= -\frac{1}{2}\partial_H^2\ln|T_n(H)| - \frac{1}{2\sigma^2}\left\langle Y_n, \partial_H^2(T_n(H)^{-1})Y_n\right\rangle,\\
\partial_\sigma\partial_H\ell_n(\theta) &:= \frac{1}{\sigma^3}\left\langle Y_n, \partial_H(T_n(H)^{-1})Y_n\right\rangle.
\end{aligned}$$

We denote by $T_n(H)^{-1/2}$ the matrix in $\mathbb{R}^{n\times n}$ satisfying $(T_n(H)^{-1/2})^\top T_n(H)^{-1/2} = T_n(H)^{-1}$ and write $T_n(H)^{1/2}$ for the matrix in $\mathbb{R}^{n\times n}$ satisfying $T_n(H)^{1/2}T_n(H)^{-1/2} = \mathbb{I}_n$. We denote by Z_n be a standard normal random vector in $\mathbb{R}^n$ under $\mathbb{P}_\theta$. With those notations, it holds under $\mathbb{P}_\theta$ that for each $n\in\mathbb{N}$, $(\sigma^2T_n(H))^{-1/2}Y_n \overset{\mathscr{L}}{=} Z_n$, that is,

$$\begin{aligned}
\partial_\sigma\ell_n(\theta) &\overset{\mathscr{L}}{=} \frac{1}{\sigma}\left(\|Z_n\|^2 - n\right),\\
\partial_H\ell_n(\theta) &\overset{\mathscr{L}}{=} -\frac{1}{2}\partial_H\ln|T_n(H)| - \frac{1}{2}\left\langle Z_n, (T_n(H)^{1/2})^\top\partial_H(T_n(H)^{-1})T_n(H)^{1/2}Z_n\right\rangle,\\
\partial_\sigma^2\ell_n(\theta) &\overset{\mathscr{L}}{=} \frac{1}{\sigma^2}\left(n - 3\|Z_n\|^2\right),\\
\partial_H^2\ell_n(\theta) &\overset{\mathscr{L}}{=} -\frac{1}{2}\partial_H^2\ln|T_n(H)| - \frac{1}{2}\left\langle Z_n, (T_n(H)^{1/2})^\top\partial_H^2(T_n(H)^{-1})T_n(H)^{1/2}Z_n\right\rangle,\\
\partial_\sigma\partial_H\ell_n(\theta) &\overset{\mathscr{L}}{=} \frac{1}{\sigma}\left\langle Z_n, (T_n(H)^{1/2})^\top\partial_H(T_n(H)^{-1})T_n(H)^{1/2}Z_n\right\rangle.
\end{aligned}$$

Based on the above expressions, we can show that $\mathbb{E}_\theta[\nabla_\theta(\ell_n(\theta))] = 0$, where we have used the fact that $\mathscr{L}(\|Z_n\|^2)$ is chi-square with n degrees of freedom, the identity

$$\partial_H \ln|T_n(H)| + \mathrm{tr}\left[\partial_H(T_n(H)^{-1})T_n(H)\right] = 0, \tag{3.6}$$

and the well known property $\mathrm{tr}[ABC] = \mathrm{tr}[CAB]$. Next, again since $\mathscr{L}(\|Z_n\|^2)$ is chi-square with n degrees of freedom, we get

$$\frac{1}{n}\mathbb{E}_\theta\left[(\partial_\sigma \ell_n(\theta))^2\right] = \frac{1}{n\sigma^2}\mathbb{E}_\theta\left[\left(\|Z_n\|^2 - n\right)^2\right] = \frac{2}{\sigma^2} = \mathscr{I}_{11}(\theta),$$

where we let $\mathscr{I}_{k_1k_2}(\theta)$ indicate the (k_1,k_2)-th entry of the matrix $\mathscr{I}(\theta)$, with a slight notational abuse for simplicity. Next, by letting $B_n := (T_n(H)^{1/2})^\top \partial_H(T_n(H)^{-1})T_n(H)^{1/2}$ and by using $\mathrm{Var}_\theta(\langle Z_n, B_nZ_n\rangle) = \mathrm{tr}(B_n(B_n+B_n^\top))$ and $B_n^\top = B_n$, we can show that

$$\begin{aligned}\frac{1}{n}\mathbb{E}_\theta\left[(\partial_H \ell_n(\theta))^2\right] &= \frac{1}{2n}\mathrm{tr}\left[T(H)^{-1}\partial_H(T_n(H))T(H)^{-1}\partial_H(T_n(H))\right] \\ &\to \frac{1}{4\pi}\int_{-\pi}^{+\pi}(2\ln H)^2 dx = 2(\ln H)^2 = \mathscr{I}_{22}(\theta),\end{aligned}$$

where we have used Theorem 5.1 of Dahlhaus [10] and

$$\partial_H(T_n(H)^{-1})T_n(H) + T_n(H)^{-1}\partial_H\left(T_n(H)\right) = 0 \ (\in \mathbb{R}^{n\times n}). \tag{3.7}$$

By using the well known identity $\mathbb{E}_\theta[\langle Z_n, AZ_n\rangle\langle Z_n, BZ_n\rangle] = \mathrm{tr}[A(B+B^\top)] + \mathrm{tr}[A]\mathrm{tr}[B]$, we can derive that

$$\frac{1}{n}\mathbb{E}_\theta\left[\left(\partial_\sigma \ell_n(\theta)\right)\left(\partial_H \ell_n(\theta)\right)\right] = -\frac{1}{\sigma n}\mathrm{tr}\left[\partial_H(T_n(H)^{-1})T_n(H)\right] \to \mathscr{I}_{12}(\theta),$$

where we have again used Theorem 5.1 of Dahlhaus [10] and the identities (3.6) and (3.7).

Next, we show the convergence of the Hessian matrix. It is straightforward that for each $n \in \mathbb{N}$,

$$\frac{1}{n}\partial_\sigma^2 \ell_n(\theta) \overset{\mathscr{L}}{=} \frac{1}{\sigma^2}\left(1 - \frac{3}{n}\|Z_n\|^2\right) \overset{\mathbb{P}_\theta}{\to} -\frac{2}{\sigma^2} = -\mathscr{I}_{11}(\theta),$$

where we have used the law of large numbers of the chi-square distribution. As before, we can show that

$$\begin{aligned}\frac{1}{n}\partial_H^2 \ell_n(\theta) &\sim \frac{1}{2n}\mathrm{tr}\left[\partial_H(T_n(H)^{-1})\partial_H(T_n(H))\right] \to -\frac{1}{4\pi}\int_{-\pi}^{+\pi}(2\ln H)^2 dx = -\mathscr{I}_{22}(\theta), \\ \frac{1}{n}\partial_\sigma\partial_H \ell_n(\theta) &\sim -\frac{1}{n\sigma}\mathrm{tr}\left[T_n(H)^{-1}\partial_H(T_n(H))\right] \to -\frac{1}{2\sigma\pi}\int_{-\pi}^{+\pi} 2(\ln H)dx = -\mathscr{I}_{12}(\theta),\end{aligned}$$

where the asymptotics hold in $\mathbb{P}_\theta$-probability as $n \uparrow +\infty$ and where we have again used Theorem 5.1 of Dahlhaus [10], the identities (3.6) and (3.7), and

$$\partial_H^2 \ln|T_n(H)| + \mathrm{tr}\left[\partial_H^2\left(T_n(H)^{-1}\right)T_n(H)\right] + \mathrm{tr}\left[\partial_H\left(T_n(H)^{-1}\right)\partial_H\left(T_n(H)\right)\right] = 0.$$

We can also show that

$$R_n\left(\int_0^1\int_0^\delta \mathrm{Hess}_\theta\left(\ell_n(\theta)\right)\big|_{\theta=\theta_0+\varepsilon R_n h} d\varepsilon\, d\delta\right)R_n \overset{\mathbb{P}_\theta}{\to} -\frac{1}{2}\mathscr{I}(\theta),$$

and that $R_n\nabla_\theta(\ell_n(\theta)) \overset{\mathscr{L}}{\to} \mathscr{N}(0, \mathscr{I}(\theta))$, as $n \uparrow +\infty$, in a similar manner to the derivation of the matrix $\mathscr{I}(\theta)$ above. We get the desired result by applying this to the pathwise Taylor expansion with the mean value theorem

$$\begin{aligned}&\ell_n(\theta_0 + R_nh) - \ell_n(\theta_0) = \\ &\qquad\left\langle h, R_n\nabla_\theta\left(\ell_n(\theta)\right)\big|_{\theta=\theta_0}\right\rangle + \left\langle R_nh, \left(\int_0^1\int_0^\delta \mathrm{Hess}_\theta\left(\ell_n(\theta)\right)\big|_{\theta=\theta_0+\varepsilon R_nh} d\varepsilon\, d\delta\right)R_nh\right\rangle.\end{aligned}$$

The singularity of the matrix $\mathscr{I}(\theta)$ is evident. The proof is complete. □

References

[1] Austin, D., Bowen, W.D., McMillan, J.I. 2004 Intraspecific variation in movement patterns: modeling individual behaviour in a large marine predator. *Oikos*, **105**, 15-30.

[2] Bartumeus, F., Catalan, J. 2009 Optimal search behavior and classic foraging theory. *Journal of Physics A: Mathematical and Theoretical*, **42**, 434002.

[3] Bartumeus, F., Peters, F., Pueyo, S., Marrase, C., Catalan, J. 2003 Helical Lévy walks: Adjusting searching statistics to resource availability in microzooplankton. *Proc. Natl. Acad. Sci. USA*, **100**, 12771-12775.

[4] Bartumeus, F., Catalan, J., Viswanathan, G.M., Raposo, E.P., da Luz, M.G.E. 2008 The influence of turning angles on the success of non-oriented animal searches. *J. Theor. Biol.*, **252**, 43-55.

[5] Benhamou, S. 2007 How many animals really do the Lévy walk? *Ecology*, **88**, 1962-1969.

[6] Burnham, K.P., Anderson, D.R. 2002 *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach* (2nd edition), Springer-Verlag: Berlin.

[7] Codling, E.A., Plank, M.J. 2011 Turn designation, sampling rate and the misidentification of power laws in movement path data using maximum likelihood estimates. *Theor. Ecol.*, **4**, 397-406.

[8] Codling, E.A., Plank, M.J., Benhamou, S. 2008 Random walk models in biology, *J. R. Soc. Interface*, **5**, 813-834.

[9] Cramér, H. 1946 *Mathematical Methods of Statistics*. Princeton University Press: Princeton.

[10] Dahlhaus, R. 1989 Efficient parameter estimation for self-similar processes, *Annals of Statistics*, **17**(4) 1749-1766.

[11] De Jager, M. et al. (2011) Lévy walks evolve through interaction between movement and environmental complexity. *Science*, **332**, 1551-1553.

[12] Edwards, A.M., et al. 2007 Revisiting Lévy flight search patterns of wandering albatrosses, bumblebees and deer, *Nature*, **499**, 1044-8.

[13] Ferguson, T.S. 1996 *A Course in Large Sample Theory*, Chapman & Hall/ CRC.

[14] Fox, R., Taqqu, M.S. 1986 Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series, *Annals of Statistics*, **14**, 517-532.

[15] Gardner, R.H., O'Neill, R.V., Turner, M.G., Dale, V.H. 1989 Quantifying scale-dependent effects of animal movement with simple percolation models, *Landscape Ecology*, **3**, 217-227.

[16] Humphries, N.E., et al. 2010 Environmental context explains Lévy and Briwnian movement patterns of marine predators, *Nature*, **465**, 1066-1069.

[17] Jansen, V.A.A., Mashanova, A., Petrovskii, S.V. 2011 Model selection and animal movement: why mussels do not do a Lévy walk. *Science*, in press.

[18] Kareiva, P.M., Shigesada, N. 1983 Analyzing insect movement as a correlated random walk, *Oecologia*, **56**, 234-238.

[19] Kawai, R., Fisher information for fractional Brownian motion under high-frequency discrete sampling, *Communications in Statistics - Theory and Methods*, doi:10.1080/03610926.2011.594540.

[20] Kawai, R., On singularity of Fisher information matrix for stochastic processes under high frequency sampling, *Numerical Mathematics and Advanced Applications 2011*, A. Cangiani et al. (Eds.) Springer-Verlag, forthcoming.

[21] Knell, A.S., Codling E.A. Classifying area-restricted search (ARS) using a partial sum approach. *Theor. Ecol.*, in press.

[22] Kramer D.L., McLaughlin, R.L. 2001 The behavioural ecology of intermittent behaviour, *American Zoologist*, **41**, 137-153.

[23] Le Cam, L. 1960 Locally asymptotically normal families of distributions. Certain approximations to families of distributions and their use in the theory of estimation and testing hypotheses. *Univ. California Publ. Statist.* **3**, 37–98.

[24] Le Cam, L., Yang, G. L. (1990) *Asymptotics in Statistics. Some Basic Concepts. Second Edition*, Springer Series in Statistics. Springer-Verlag: New York.

[25] Mandelbrot, B. 1977 *The Fractal Geometry of Nature*, Freeman: New York.

[26] Mashanova, A., Oliver, T.H., Jansen, V.A.A. 2010 Evidence for intermittency and a truncated power law from highly resolved aphid movement data, *Journal of the Royal Society Interface*, **7**(42) 199-208.

[27] Meerscheart, M.M., Roy, P., Shao, Q. 2010 Parameter estimation for tempered power law distributions, *Communications in Statistics - Theory and Methods*, forthcoming.

[28] Metzler, R., Koren, T., van den Broek, B., Wuite, G.J.L., Lomholt, M.A. 2009 And did he search for you, and could not find you? *Journal of Physics A: Mathematical and Theoretical*, **42**, 434005.

[29] Morales, J.M., Haydon, D.T., Frair, J., Holsinger, K.E., Fryxell, J.M. 2004 Extracting more out of relocation data: building movement models as mixtures of random walks. *Ecology*, **85**, 2436-2445.

[30] Nathan, R. 2001 The challenges of studying dispersal, *TREE*, **16**, 481-483.

[31] Nathan, R. et al. 2008 *Proc. Natl. Acad. Sci. USA*, **105**, 19052.

[32] Newlands, N.K., Lutcavage, M.E., Pitcher, T.J. 2004 Analysis of foraging movements of Atlantic bluefin tuna (Thunnus thynnus): individuals move with two modes of searching, *Popul. Ecol.*, **46**, 39-53.

[33] Newman, M.E.J. 2005 Power laws, Pareto distributions and Zipf's law, *Contemporary Physics*, **46**, 323-351.

[34] Nolan, J.P. 2001 Maximum likelihood estimation and diagnostics for stable distributions, In *Lévy Processes - Theory and Applications* (ed. Barndorff-Nielsen, O.-E., Mikosch, T., Resnick, S.I.), pp. 379-400, Birkhäuser: Boston.

[35] Okubo, A., Levin, S. 2001 *Diffusion and Ecological Problems: Modern Perspectives.* Berlin: Springer.

[36] Patterson, T.A., Thomas, L., Wilcox, C., Ovaskainen, O., Matthiopoulos, J. 2008 State 穆 pace models of individual animal movement. *TREE*, **23**, 87-94.

[37] Pedersen, M.W., Patterson, T.A., Thygesen, U.H., Madsen, H. 2011 Estimating animal behavior and residency from movement data. *Oikos*, **120**, 1281-1290.

[38] Petrovskii, S.V., Mashanova, A., Jansen, V.A.A. 2011 Variation in individual walking behavior creates the impression of a Lévy flight. *PNAS*, **108**, 8704-8707.

[39] Plank, M.J., Codling, E.A. 2009 Sampling rate and misidentification of Lévy and non-Lévy movement paths. *Ecology*, **90**, 3546-3553.

[40] Plank, M.J., James, A. 2008 Optimal foraging: Lévy pattern or process? *J. R. Soc. Interface*, **5**, 1077-1086.

[41] Rao, C.R. 1973 *Linear Statistical Inference and Its Applications*, 2nd ed. Wiley: New York.

[42] Raposo, E.P., Buldyrev, S.V., da Luz, M.G.E., Viswanathan, G.M., Stanley, H.E. 2009 Lévy flights and random searches, *Journal of Physics A: Mathematical and Theoretical*, *42*, 434003.

[43] Reynolds, A.M., et al. 2007 Displaced honey bees perform optimal scale-free search flights, *Ecology*, **88**(8), 1955-1961.

[44] Reynolds, A.M. 2008 How many animals really do the Lévy walk? Comment. *Ecology*, **89**(8), 23472351

[45] Reynolds, A.M. 2010 Bridging the gulf between correlated random walks and Lévy walks: autocorrelation as a source of Lévy walk movement patterns, *J. R. Soc. Interface*, **7**, 1753-1758.

[46] Reynolds, A.M. 2009 Scale-free animal movement patterns: Lévy walks outperform fractional Brownian motions and fractional Lévy motions in random search scenarios, *Journal of Physics A: Mathematical and Theoretical*, **42**, 434006.

[47] Rytgaard, M. 1990 Estimation in the Pareto distribution, ASTIN bulletin, **20**(2), 201-216.

[48] Samorodnitsky, G., Taqqu, M.S. (1994) *Stable non-Gaussian random processes*, Chapman & Hall, New York.

[49] Shlesinger, M.F., Zaslavsky, G.M., Frisch, U. (eds) 1995 *Lévy Flights and Related Topics in Physics*, Springer: Berlin.

[50] Sims, D. et al. 2008 Scaling laws of marine predator search behavior, *Nature*, **451**, 1098-102.

[51] Turchin, P. 1988 *Quantitative Analysis of Movement: Measuring and Modeling Population Redistribution in Animals and Plants*, Sunderland, MA: Sinauer.

[52] Turchin P 1996 Fractal analyses of animal movement: A critique. *Ecology*, **77**, 2086-2090.

[53] Viswanathan, G.M. 2010 Fish in Lévy-flight foraging, *Nature*, **465**, 1018-1019.

[54] Viswanathan G.M., Afanasyev V., Buldyrev S.V., Murphy E.J., Prince P.A., Stanley H.E. 1996 Lévy flight search patterns of wandering albatrosses, *Nature*, **381**, 413-415.

[55] Viswanathan, G.M., da Luz, M.G.E., Raposo, E.P., Stanley, H.E. 2011 *The Physics of Foraging: An Introduction to Random Searches and Biological Encounters.* Cambridge University Press, Cambridge UK.

[56] Viswanathan, G.M., Raposo, E.P., Bartumeus, F., Catalan, J., da Luz, M.G.E. 2005 Necessary criterion for distinguishing true superdifusion from correlated random walk processes. *Phys. Rev. E* **72**, 011111(6).

[57] Viswanathan, G.M., Raposo, E.P., da Luz, M.G.E. 2008 Lévy flights and superdiffusion in the context of biological encounters and random searches, *Phys. Life Rev.*, **5**, 133-150.