



**HAL**  
open science

# Statistics of the number of records for random walks and Lévy flights on a 1D lattice

Philippe Mounaix, Satya Majumdar, Grégory Schehr

► **To cite this version:**

Philippe Mounaix, Satya Majumdar, Grégory Schehr. Statistics of the number of records for random walks and Lévy flights on a 1D lattice. *Journal of Physics A: Mathematical and Theoretical*, 2020, 53 (41), pp.415003. 10.1088/1751-8121/abac97 . hal-02958283

**HAL Id: hal-02958283**

**<https://hal.science/hal-02958283v1>**

Submitted on 5 Oct 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Statistics of the Number of Records for Random Walks and Lévy Flights on a $1D$ Lattice

Philippe Mounaix,<sup>1,\*</sup> Satya N. Majumdar,<sup>2,†</sup> and Grégory Schehr<sup>2,‡</sup>

<sup>1</sup>*CPHT, CNRS, Ecole Polytechnique,  
IP Paris, F-91128 Palaiseau, France.*

<sup>2</sup>*LPTMS, CNRS, Univ. Paris-Sud,  
Université Paris-Saclay, 91405 Orsay, France.*

(Dated: September 21, 2020)

## Abstract

We study the statistics of the number of records  $R_n$  for a symmetric,  $n$ -step, discrete jump process on a  $1D$  lattice. At a given step, the walker can jump by arbitrary lattice units drawn from a given symmetric probability distribution. This process includes, as a special case, the standard nearest neighbor lattice random walk. We derive explicitly the generating function of the distribution  $P(R_n)$  of the number of records, valid for arbitrary discrete jump distributions. As a byproduct, we provide a relatively simple proof of the generalized Sparre Andersen theorem for the survival probability of a random walk on a line, with discrete or continuous jump distributions. For the discrete jump process, we then derive the asymptotic large  $n$  behavior of  $P(R_n)$  as well as of the average number of records  $E(R_n)$ . We show that unlike the case of random walks with symmetric and continuous jump distributions where the record statistics is *strongly* universal (i.e., independent of the jump distribution for all  $n$ ), the record statistics for lattice walks depends on the jump distribution for any fixed  $n$ . However, in the large  $n$  limit, we show that the distribution of the scaled record number  $R_n/E(R_n)$  approaches a universal, half-Gaussian form for any discrete jump process. The dependence on the jump distribution enters only through the scale factor  $E(R_n)$ , which we also compute in the large  $n$  limit for arbitrary jump distributions. We present explicit results for a few examples and provide numerical checks of our analytical predictions.

PACS numbers: 02.50.Ga, 05.40.Fb, 05.45.Tp

---

\*Electronic address: philippe.mounaix@polytechnique.edu

†Electronic address: satya.majumdar@lptms.u-psud.fr

---

<sup>‡</sup>Electronic address: `gregory.schehr@lptms.u-psud.fr`

## I. INTRODUCTION

Consider a discrete time series  $\{x_0, x_1, x_2, \dots, x_n\}$  of  $n$  entries representing, e.g., the temperature in a city or the depth of a river or the stock price of a company, all these data being taken on, say, a daily basis. Then a record is said to happen on day  $i$  if the  $i$ th entry  $x_i$  is larger than all previous entries  $x_0, x_1, x_2, \dots, x_{i-1}$ . In general, record statistics is expected to be of interest in fields where time series are used and where the size of the entries is a relevant parameter. Such fields include meteorology [1, 2], hydrology [3, 4], insurance and risk management, trading [5–7], economics [8, 9], sports [10–14], etc. Another important application concerns current issues of climatology such as global warming where both temperature records [2, 15–19] and rainfall precipitation records [20–22] play an important role in anticipating future climatic conditions. Statistics of record events have also been found relevant in biology [23], in the theory of spin glasses [24, 25], and in models of growing networks [26].

The mathematical theory of records has been an important subject of research since the paper by Chandler in 1952 (see, e.g., [27–32]). Record statistics is now well understood in the case when the entries  $x_i$ 's are independent and identically distributed (i.i.d.) random variables. However, in most realistic situations the entries of the time series are correlated and the theory in this case is still piecemeal. For weak correlations, with a finite correlation time, one expects the record statistics for a large  $n$  series to be asymptotically similar to the uncorrelated case, but this is no longer true when there are strong correlations between the entries. One of the simplest and most natural time series with strongly correlated entries corresponds to the one of the positions of a one-dimensional random walk, for which many results are known in the mathematics literature, in particular in the context of fluctuation and renewal theory, and can be found in the classical book by Feller [33]. In the physics literature, a comprehensive theory of record statistics for random walks with continuous jumps was worked out by Majumdar and Ziff in [34] (see also [35]). Considering a time series defined by

$$x_i = x_{i-1} + \eta_i, \tag{1}$$

with  $x_0 = 0$  and where the  $\eta_i$ 's are i.i.d. random variables drawn from a symmetric and continuous jump distribution  $f(\eta)$ , they showed that the joint probability  $P(R_n, \tau_1, \tau_2, \dots, \tau_{R_n})$  of  $R_n$  records in  $n$  steps, with respective life-times  $\tau_i$  ( $1 \leq i \leq R_n$ ), is completely independent

of  $f(\eta)$  for any  $n$ , and not just asymptotically for large  $n$ . This remarkable result, which is a consequence of the so-called Sparre Andersen theorem [36], includes also Lévy flights where  $f(\eta) \sim 1/|\eta|^{\mu+1}$  for large  $\eta$ , with  $0 < \mu < 2$ . From the joint probability  $P(R_n, \tau_1, \tau_2, \dots, \tau_{R_n})$  one can derive the distribution of the number of records  $P(R_n)$  by integrating out the age degrees of freedom  $\tau_1, \tau_2, \dots, \tau_{R_n}$ , which thus is also universal for all  $n$ , i.e., independent of  $f(\eta)$ . In this paper, we will be mainly interested in  $P(R_n)$  and in particular, its first moment,  $E(R_n)$ , the average number of records up to step  $n$ . For a symmetric and continuous jump distribution, Majumdar and Ziff found that these quantities are given by the universal formulas, valid for any  $n$ ,

$$P(R_n = m) = \binom{2n - m + 1}{n} 2^{-2n+m-1}, \quad (2)$$

and

$$E(R_n) = (2n + 1) \binom{2n}{n} 2^{-2n}. \quad (3)$$

Note that the expressions in Eqs. (2) and (3) corresponds to taking  $R_0 = 1$  (the initial position is counted as the first record). In particular, for large  $n$ , Eq. (3) gives

$$E(R_n) = \sqrt{\frac{4n}{\pi}} + O\left(\frac{1}{\sqrt{n}}\right) \quad (n \rightarrow +\infty), \quad (4)$$

independently of the jump distribution  $f(\eta)$ . Similarly, in the limit  $m, n \rightarrow +\infty$  with fixed  $m/\sqrt{n}$ , the record number distribution  $P(R_n = m)$  in Eq. (2) approaches a scaling form,

$$P(R_n = m) \simeq \sqrt{\frac{\pi}{4n}} g\left(m \sqrt{\frac{\pi}{4n}}\right); \quad \text{with} \quad g(x) = \frac{2}{\pi} e^{-x^2/\pi}, \quad x \geq 0, \quad (5)$$

where the scaling function  $g(x)$  is a half-Gaussian, normalized to unity. As long as  $f(\eta)$  is symmetric and continuous, the scaling form in Eq. (5) is also universal, i.e., independent of the jump distribution  $f(\eta)$ .

Following [34], record statistics for random walks has since been studied in different, more general, settings (for a recent review on both i.i.d. and random walk cases see [37]). This includes, for instance, continuous-time symmetric random walks [38], discrete-time random walks with a constant drift [39, 40], an ensemble of  $N$  independent symmetric random walks [6], random walks in the presence of a measurement error and additive instrumental noises [41], active run-and-tumble particles [42] or random walks with correlated jumps [43]. The complete universality for all  $n$ , found in [34] for symmetric and continuous  $f(\eta)$ , does

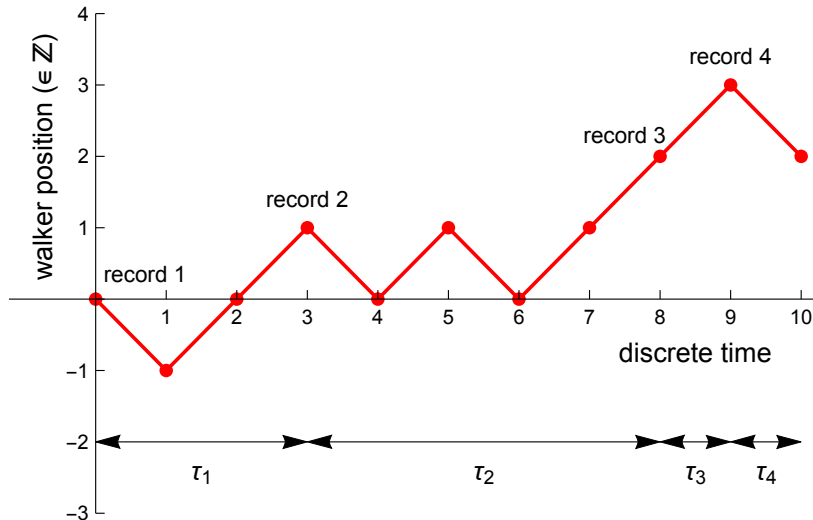


FIG. 1: Illustration of how we count records for a random walk on a  $1D$  lattice. A record is counted when the walker position is *strictly* greater than all previous positions. When the walker position is equal to the one at the last record, it is *not* counted as a new record. The initial position is counted as the first record. The time interval  $\tau_i$  represents the life-time of the  $i$ th record.

not hold in some of these examples, e.g., in the case of the discrete-time random walk in the presence of a constant drift [40]. Nevertheless, these generalizations usually concern random walks with continuous jumps and it is natural to ask what happens to record statistics for random walks with *discrete* jumps, or ‘lattice random walks’. The question was briefly brought up at the very end of [34]. There, it was found that in the particular case where the walker jumps by  $\eta = \pm 1$  at each time step, with equal probability  $1/2$ , Eq. (4) is replaced with

$$\begin{aligned}
 E(R_n) &= \frac{1}{2} \left[ \frac{(-1)^{n+1} \Gamma(n-1/2) {}_2F_1(\frac{3}{2}, -n; \frac{3}{2} - n; -1)}{2\sqrt{\pi} \Gamma(n+1)} + 1 \right] \\
 &= \sqrt{\frac{2n}{\pi}} + \frac{1}{2} + O\left(\frac{1}{\sqrt{n}}\right) \quad (n \rightarrow +\infty),
 \end{aligned} \tag{6}$$

where  ${}_2F_1$  is the hypergeometric function. This is the only known result so far about record statistics for random walks with discrete jumps. It can be seen in Eq. (6) that while the expected number of records still grows asymptotically as  $\sqrt{n}$  for large  $n$ , like in the case of a continuous jump distribution in Eq. (4), the prefactor  $\sqrt{2/\pi}$  of the leading term differs from the one in the continuous case by a factor  $1/\sqrt{2}$  and the next correction term  $1/2$  is a constant while in the continuous case it goes to zero as  $O(n^{-1/2})$ .

In this paper, we consider a generalized lattice random walk model where the walker stays on a lattice, but can jump in one step by an arbitrary number of lattice units. Our model includes the  $\pm 1$  random walk model as a special case. For this generalized model, we derive the distribution  $P(R_n)$  and its first moment  $E(R_n)$  explicitly and discuss their asymptotic behavior for large  $n$ . Our principal motivation for studying this model is to test how far the universality of record statistics for symmetric and continuous jump distribution found in Ref. [34] can be pushed, and how the discreteness of the position of the walker may modify the record statistics.

The rest of the paper is organized as follows. In section II, we define our model and summarize our main results. In section III, we determine the generating function of  $P(R_n)$  by following a renewal approach, we prove the generalized Sparre Andersen theorem (11), and we use the obtained results to derive the universal scaling form of  $P(R_n)$  in the large  $n$  limit. In section IV, we determine the generating function of the average number of records  $E(R_n)$ . In the simple case  $\eta = \pm 1$ , with equal probability  $1/2$ , we check that this generating function and the associated expression for  $E(R_n)$  coincide with the results in [34]. Then, we derive the large  $n$  asymptotic behavior of  $E(R_n)$  in the general case, and we apply the result to random walks with  $\eta = 0, \pm 1, \pm 2$ , where explicit expressions of the coefficients can be obtained. A scaling form for the large  $n$  behavior of  $E(R_n)$  is derived in the general case when the probability of staying in place ( $\eta = 0$ ) is close to 1. Section V is devoted to the verification of our analytical predictions via numerical simulations. Finally, we conclude in section VI. Some more technical points are relegated to the appendices.

## II. THE MODEL AND SUMMARY OF THE MAIN RESULTS

We consider a  $1D$  random walk starting at the origin,  $x_0 = 0$ , and evolving according to Eq. (1) with the jump distribution  $f(\eta)$  of the form

$$f(\eta) = \sum_{h \in \mathbb{Z}} P(h) \delta(\eta - h\alpha), \quad (7)$$

where  $\alpha > 0$  is a given lattice spacing and  $0 \leq P(h) \leq 1$  represents the probability to jump by  $h$  lattice units in one step, where  $h \in \mathbb{Z}$ . We further assume that  $P(h)$  is symmetric:  $P(h) = P(-h)$ . Thus, if the walker starts at the origin, it stays on the lattice with lattice spacing  $\alpha$  at any time. Indeed, the position of the walker after the  $i$ th step is  $x_i = l_i \alpha$

where  $l_i \in \mathbb{Z}$  evolves according to  $l_i = l_{i-1} + h_i$  with  $l_0 = 0$  and where the jumps  $h_i$  are i.i.d. random variables with values in  $\mathbb{Z}$  drawn from the symmetric probability  $P(h)$ . We assume that the Fourier transform of the jump distribution

$$\hat{f}(k) = \int_{-\infty}^{+\infty} e^{ik\eta} f(\eta) d\eta = \sum_{h \in \mathbb{Z}} e^{ikh\alpha} P(h) \quad (8)$$

has the small  $k$  behavior  $\hat{f}(k) \sim 1 - |ak|^\mu$  ( $k \rightarrow 0$ ), where  $a \gtrsim \alpha$  is the characteristic length scale of the jumps and  $0 < \mu \leq 2$  is the Lévy index. As the discrete Fourier transform of  $P(h)$ ,  $\hat{f}(k)$  is a periodic function of  $k$ , unlike the case of continuous jumps where  $\hat{f}(k)$  goes to zero as  $k \rightarrow +\infty$ .

### Summary of main results

Before entering the details of the calculations, it is useful to summarize our main results.

- We show that the renewal property of random walks that was used to study the record statistics for symmetric and continuous jump distributions in [34], can be extended to the present model, but with a twist. Like in the continuous case, we show that the generating functions of  $P(R_n)$  and  $E(R_n)$  for lattice random walks are respectively given by

$$\sum_{n \geq 0} P(R_n = m) s^n = [1 - (1 - s) Q_{\geq 0}(s)]^{m-1} Q_{\geq 0}(s), \quad (9)$$

and

$$\sum_{n \geq 0} E(R_n) s^n = \frac{1}{(1 - s)^2 Q_{\geq 0}(s)}, \quad (10)$$

where  $Q_{\geq 0}(s) = \sum_{n \geq 0} q_{\geq 0}(n) s^n$  is the generating function of the survival probability,  $q_{\geq 0}(n) = \text{Prob}(x_1 \geq 0, x_2 \geq 0, \dots, x_n \geq 0 | x_0 = 0)$ . The main difference is in the expression for  $Q_{\geq 0}(s)$ . While in the continuous case,  $Q_{\geq 0}(s) = 1/\sqrt{1-s}$  is universal due the Sparre Andersen theorem [36], we show that for discrete jump processes on a lattice, it is non-universal and given by

$$Q_{\geq 0}(s) \equiv \sum_{n \geq 0} q_{\geq 0}(n) s^n = \frac{Z(s)}{\sqrt{1-s}}, \quad (11)$$

where

$$Z(s) = \exp \left( -\frac{\alpha}{2\pi} \int_0^{\pi/\alpha} \ln [1 - s\hat{f}(k)] dk \right), \quad (12)$$



depends explicitly on the jump distribution  $f(\eta)$  through its Fourier transform. The result in Eq. (11) can be derived from a generalized Sparre Andersen theorem [36] valid for arbitrary jump distributions on a line, discrete or continuous. We also prove this result using an alternative method, which in fact provides a relatively simple non-combinatorial proof of this generalized Sparre Andersen theorem. In the special case of a  $\pm 1$  random walk,  $f(\eta) = \frac{1}{2} [\delta(\eta - \alpha) + \delta(\eta + \alpha)]$ , one has  $\hat{f}(k) = \cos(\alpha k)$  and  $Z(s)$  in Eq. (12) can be computed explicitly. One finds  $Z(s) = [\sqrt{1+s} - \sqrt{1-s}] / s$  independent of  $\alpha$ , thus recovering the known result [34, 37], as it should be. Our formula for  $Z(s)$  in Eq. (12) generalizes this special case of  $\pm 1$  random walks to arbitrary discrete jump processes on a lattice.

- Analyzing the general results above for large  $n$ , we find that while the expected number of records  $E(R_n)$  is non-universal as depending explicitly on  $f(\eta)$ , remarkably the asymptotic large  $n$  behavior of the distribution of the scaled record number  $R_n/E(R_n)$  turns out to be universal, i.e., independent of  $f(\eta)$ . More precisely, we show that the distribution  $P(R_n = m)$ , in the scaling limit  $m, n \rightarrow +\infty$  with  $m/E(R_n)$  fixed, takes the scaling form:

$$P(R_n = m) \sim \frac{1}{E(R_n)} g\left(\frac{m}{E(R_n)}\right); \quad \text{with } g(x) = \frac{2}{\pi} e^{-x^2/\pi}, \quad x \geq 0. \quad (13)$$

Moreover, the scaling function  $g(x)$  coincides rather surprisingly with the one of the symmetric and continuous case in Eq. (5). Thus, all the non-universal features are absorbed in  $E(R_n)$ , while the distribution of the scaled variable  $R_n/E(R_n)$  for large  $n$  is *super-universal*, i.e., independent of  $f(\eta)$  for both continuous and discrete jump processes.

- Finally, our explicit results in Eqs. (10), (11) and (12) allow us to analyze the large  $n$  asymptotic behavior of  $E(R_n)$  exactly. We find that for arbitrary discrete jump processes,  $E(R_n)$  still behaves as  $\sqrt{n}$  for large  $n$ , like in the continuous case, but now the prefactor and the next subleading term are non-universal as depending explicitly on  $f(\eta)$ . More specifically, from Eqs. (10) and (11) near  $s = 1$ , we find that the large  $n$  behavior of  $E(R_n)$  keeping the terms surviving the  $n \rightarrow +\infty$  limit only, is given by

$$E(R_n) \sim \frac{2}{Z(1)} \sqrt{\frac{n}{\pi}} + C_\mu \quad (n \rightarrow +\infty), \quad (14)$$

with

$$C_2 = \frac{\alpha N_0}{2 a Z(1)}, \text{ and } C_{0 < \mu < 2} = 0, \quad (15)$$

where  $N_0 \geq 1$  is the number of zeros of  $1 - \hat{f}(k)$  in the interval  $0 \leq k \leq \pi/\alpha$ . In Section IVD we give some simple examples where  $Z(1)$  in Eqs. (14) and (15) can be computed explicitly. For instance, in the simplest case where the walker jumps by  $h = 0, \pm 1$  at each time step, with  $P(0) = p$  and  $P(\pm 1) = (1 - p)/2$  (with  $0 \leq p < 1$ ), Eqs. (14) and (15) yield

$$E(R_n) \sim \sqrt{\frac{2(1-p)n}{\pi}} + \frac{1}{2} \quad (n \rightarrow +\infty), \quad (16)$$

which shows how the large  $n$  behavior in Eq. (6) is modified when the walker is allowed to stay in place with probability  $p$ . The result in Eq. (16) holds in the limit  $n \rightarrow +\infty$  with  $0 \leq p < 1$  fixed. It is also interesting to consider the scaling limit when  $p \rightarrow 1$ ,  $n \rightarrow +\infty$  with the product  $n(1-p)$  kept fixed. In this scaling limit, we find that

$$E(R_n) \simeq \mathcal{F}((1-p)n), \quad (17)$$

where the scaling function  $\mathcal{F}(x)$  is explicitly given in Eq. (80). We also show in Appendix B how to generalize this result to any discrete jump distributions with Lévy index  $0 < \mu \leq 2$ .

### III. DISTRIBUTION OF THE NUMBER OF RECORDS: A RENEWAL APPROACH

To compute the distribution  $P(R_n = m)$  of the number of records  $R_n$  after  $n$  steps, it turns out to be useful to exploit the renewal property of random walks [34] (whether the jumps are continuous or discrete). Namely, instead of trying to get the distribution  $P(R_n = m)$  of  $R_n$  directly (which actually seems impossible due to the nontrivial memory carried by the random walk), the strategy would be first to compute the joint probability  $P(R_n = m, \tau_1, \tau_2, \dots, \tau_m)$  of  $R_n$  and the time intervals  $\tau_1, \tau_2, \dots, \tau_m$  between successive records (see Figure 1). Note that  $\tau_m$  is the time interval between the last record and the arrival time  $n$  which does not necessarily correspond to a record. Then, by summing  $P(R_n = m, \tau_1, \tau_2, \dots, \tau_m)$  over the  $\tau_i$ 's one can obtain the distribution of records  $P(R_n = m)$ . In this section, we follow this strategy to determine the generating function of  $P(R_n = m)$  with respect to  $n \geq 0$  which encodes all the information about  $P(R_n = m)$ .

### A. Generating function of $P(R_n)$

We define the survival probability  $q_{\geq 0}(n)$  after  $n$  steps by

$$q_{\geq 0}(n) = \text{Prob}(x_1 \geq 0, x_2 \geq 0, \dots, x_n \geq 0 | x_0 = 0), \quad (18)$$

for  $n \geq 1$  and  $q_{\geq 0}(0) = 1$ . The generating function of  $q_{\geq 0}(n)$  is

$$Q_{\geq 0}(s) = \sum_{n \geq 0} q_{\geq 0}(n) s^n. \quad (19)$$

Note that, since we are considering symmetric random walks and using the translational invariance of the walk with respect to the initial position, one also has

$$\begin{aligned} q_{\geq 0}(n) &= \text{Prob}(x_1 \geq x_0, x_2 \geq x_0, \dots, x_n \geq x_0 | x_0) \\ &= \text{Prob}(x_1 \leq x_0, x_2 \leq x_0, \dots, x_n \leq x_0 | x_0), \end{aligned} \quad (20)$$

independent of the starting point  $x_0$ . We define the first-crossing time of  $x_0$  to be  $n$  if the walker strictly crosses  $x_0$  for the first time during the  $n$ -th step. The first-crossing time probability  $\varphi(n)$  is then given by

$$\varphi(n) = \text{Prob}(x_1 \leq x_0, x_2 \leq x_0, \dots, x_{n-1} \leq x_0, x_n > x_0 | x_0), \quad (21)$$

which is also independent of  $x_0$  by translation invariance of the walk. From Eqs. (20) and (21), one has the relation

$$\varphi(n) = q_{\geq 0}(n-1) - q_{\geq 0}(n), \quad (22)$$

which implies that the generating function of  $\varphi(n)$  can be written in terms of the generating function of  $q_{\geq 0}(n)$  as

$$\Phi(s) \equiv \sum_{n \geq 1} \varphi(n) s^n = 1 - (1-s)Q_{\geq 0}(s). \quad (23)$$

Now, by decomposing the walk into  $m-1$  stretches of duration  $\tau_i$  ( $1 \leq i < m$ ) between the successive records and one final stretch of duration  $\tau_m$  between the last record and the last step  $n$ , and by using the translation invariance with respect to the initial position in each stretch, one finds that the joint probability  $P(R_n = m, \tau_1, \tau_2, \dots, \tau_m)$  can be written as [34] (see also [37] for further discussion)

$$P(R_n = m, \tau_1, \tau_2, \dots, \tau_m) = \varphi(\tau_1)\varphi(\tau_2) \cdots \varphi(\tau_{m-1})q_{\geq 0}(\tau_m) \delta_{n, \sum_{i=1}^m \tau_i}, \quad (24)$$

where the Kronecker delta enforces that the sum of the record ages  $\tau_i$  equals the total walk duration  $n$ . The distribution of records  $P(R_n = m)$  is then obtained by summing this expression (24) over the  $\tau_i$ 's, yielding

$$P(R_n = m) = \sum_{\tau_1 \geq 1} \cdots \sum_{\tau_{m-1} \geq 1} \sum_{\tau_m \geq 0} \varphi(\tau_1) \varphi(\tau_2) \cdots \varphi(\tau_{m-1}) q_{\geq 0}(\tau_m) \delta_{n, \sum_{i=1}^m \tau_i}. \quad (25)$$

Taking the generating function of both sides of Eq. (25) with respect to  $n \geq 0$ , one finally obtains

$$\sum_{n \geq 0} P(R_n = m) s^n = \Phi(s)^{m-1} Q_{\geq 0}(s) = [1 - (1-s) Q_{\geq 0}(s)]^{m-1} Q_{\geq 0}(s), \quad (26)$$

which shows that the full statistics of the number of records is entirely encoded in the generating function of the survival probability,  $Q_{\geq 0}(s)$ , that we will now determine.

## B. Generalized Sparre Andersen theorem

In this subsection, we prove the results in Eqs. (11) and (12) mentioned in section II. Below, first we provide a short proof using the generalized Sparre Andersen theorem, and then we give an alternative derivation without assuming this theorem. Our derivation then constitutes a new proof of the generalized Sparre Andersen theorem.

**Derivation using the generalized Sparre Andersen theorem:** The generalized Sparre Andersen theorem [36] states that the generating function  $Q_{\geq 0}(s)$  of the survival probability is given by the formula, valid for arbitrary jump distributions (discrete or continuous),

$$Q_{\geq 0}(s) = \sum_{n \geq 0} q_{\geq 0}(n) s^n = \exp \left[ \sum_{m=1}^{\infty} \frac{s^m}{m} \text{Prob.}(x_m \leq 0) \right]. \quad (27)$$

This is a remarkable theorem as it relates the survival probability  $q_{\geq 0}(n)$  up to  $n$  steps (which depends on the full history of a trajectory and hence is nonlocal in time) to a local quantity in time, namely the probability  $\text{Prob.}(x_m \leq 0)$  that the walker *at* step  $m$  is non-positive. Using the symmetry of the walk, it is clear that  $\text{Prob.}(x_m < 0) = \text{Prob.}(x_m > 0)$ . In addition, the total probability at step  $m$  must be unity, i.e.,  $2 \text{Prob.}(x_m < 0) + \text{Prob.}(x_m = 0) = 1$ . Hence,

$$\text{Prob.}(x_m \leq 0) = \text{Prob.}(x_m < 0) + \text{Prob.}(x_m = 0) = \frac{1}{2} [1 + \text{Prob.}(x_m = 0)]. \quad (28)$$

Substituting (28) on the right-hand side of Eq. (27) and using  $\sum_{m=1}^{\infty} s^m/m = -\ln(1-s)$  gives

$$Q_{\geq 0}(s) = \frac{Z(s)}{\sqrt{1-s}}; \quad \text{where} \quad Z(s) = \exp \left[ \frac{1}{2} \sum_{m=1}^{\infty} \frac{s^m}{m} \text{Prob.}(x_m = 0) \right]. \quad (29)$$

Thus, it remains to compute  $\text{Prob.}(x_m = 0)$ , which can be done easily. For arbitrary discrete jump processes, it is clear that the position of the walker after  $m$  steps, starting from the origin, is simply  $x_m = \sum_{i=1}^m \eta_i$ , where  $\eta_i$ 's are i.i.d random variables each drawn from  $f(\eta)$ . Hence the Fourier transform of  $\text{Prob.}(x_m)$  is given by  $E(e^{ikx_m}) = [\hat{f}(k)]^m$ . Inverting this Fourier transform and putting  $x_m = 0$ , one obtains

$$\text{Prob.}(x_m = 0) = \frac{\alpha}{2\pi} \int_{-\pi/\alpha}^{\pi/\alpha} [\hat{f}(k)]^m dk, \quad (30)$$

which can be used in the expression (29) of  $Z(s)$  to get

$$Z(s) = \exp \left[ -\frac{\alpha}{2\pi} \int_0^{\pi/\alpha} \ln(1 - s \hat{f}(k)) dk \right], \quad (31)$$

where we used  $\hat{f}(k) = \hat{f}(-k)$ . This completes the proof of the results announced in Eqs. (11) and (12), in Section II.

**An alternative derivation:** Here we derive the expression of  $Q_{\geq 0}(s)$  in Eq. (11) with  $Z(s)$  given in Eq. (12) by an alternative method which does not rely on the generalized Sparre Andersen theorem in Eq. (27). In fact, the theorem in Eq. (27) was originally proved by Sparre Andersen using combinatorial arguments [36]. Our method outlined below actually provides an alternative algebraic/non-combinatorial derivation of this theorem. We proceed in two steps. First, we write

$$q_{>0}(n) = \text{Prob}(x_1 > 0, x_2 > 0, \dots, x_n > 0 | x_0 = 0), \quad (32)$$

for  $n \geq 1$  and  $q_{>0}(0) = 1$ . The generating function of  $q_{>0}(n)$  is

$$Q_{>0}(s) = \sum_{n \geq 0} q_{>0}(n) s^n. \quad (33)$$

The first step consists in finding the relation between  $Q_{\geq 0}(s)$  and  $Q_{>0}(s)$ . Let  $0 \leq n_1 \leq n$  denote the step at which the walk reaches its minimum position for the first time and  $\pi(n_1)$  the probability of  $n_1$ . Split the walk into one part from  $i = 0$  to  $i = n_1$  and a second part from  $i = n_1$  to  $i = n$  ( $i$  is the step number) (see Fig. 2 for an illustration). Taking the

minimum position as a new space origin and inverting the direction of time in the first part, one gets

$$\begin{aligned}\pi(n_1) &= q_{>0}(n_1)q_{\geq 0}(n - n_1) \\ &= \sum_{n_2 \geq 0} q_{>0}(n_1)q_{\geq 0}(n_2)\delta_{n, n_1+n_2},\end{aligned}\tag{34}$$

and  $\sum_{n_1 \geq 0} \pi(n_1) = 1$  reads

$$\sum_{n_1 \geq 0} \sum_{n_2 \geq 0} q_{>0}(n_1)q_{\geq 0}(n_2)\delta_{n, n_1+n_2} = 1.\tag{35}$$

Taking the generating function of both sides of Eq. (35) with respect to  $n \geq 0$  and using the definitions in Eqs. (19) and (33), one finds the relation

$$Q_{>0}(s) Q_{\geq 0}(s) = \frac{1}{1-s}.\tag{36}$$

Note that in the case of a continuous jump distribution, one has  $q_{>0}(n) = q_{\geq 0}(n)$ . Thus,  $Q_{>0}(s) = Q_{\geq 0}(s)$  and Eq. (36) immediately gives the Sparre Andersen result  $Q_{\geq 0}(s) = 1/\sqrt{1-s}$  used in [34], which is a very simple proof of this result.

To get the generalized Sparre Andersen theorem for discrete jumps we need to split the walk into three parts, instead of two as we did to get Eq. (36). Let  $0 \leq n_1 \leq n$  denote the step at which the walk reaches its minimum position for the first time,  $n_1 + \ell$ , with  $0 \leq \ell \leq n - n_1$ , the step at which it reaches its minimum position for the last time, and  $\pi(n_1, \ell)$  the joint probability of  $n_1$  and  $\ell$  (see Fig. 3 for an illustration). We write

$$b_{\geq 0}(n) = \text{Prob}(x_1 \geq 0, x_2 \geq 0, \dots, x_n = 0 | x_0 = 0),\tag{37}$$

for  $n \geq 1$  and  $b_{\geq 0}(0) = 1$ . The generating function of  $b_{\geq 0}(n)$  is

$$B_{\geq 0}(s) = \sum_{n \geq 0} b_{\geq 0}(n) s^n.\tag{38}$$

Split the walk into one part from  $i = 0$  to  $i = n_1$ , a second part from  $i = n_1$  to  $i = n_1 + \ell$ , and a third part from  $i = n_1 + \ell$  to  $i = n$ . Taking the minimum position as a new space origin and inverting the direction of time in the first part, one gets

$$\begin{aligned}\pi(n_1, \ell) &= q_{>0}(n_1)b_{\geq 0}(\ell)q_{>0}(n - n_1 - \ell) \\ &= \sum_{n_2 \geq 0} q_{>0}(n_1)b_{\geq 0}(\ell)q_{>0}(n_2)\delta_{n, n_1+n_2+\ell},\end{aligned}\tag{39}$$

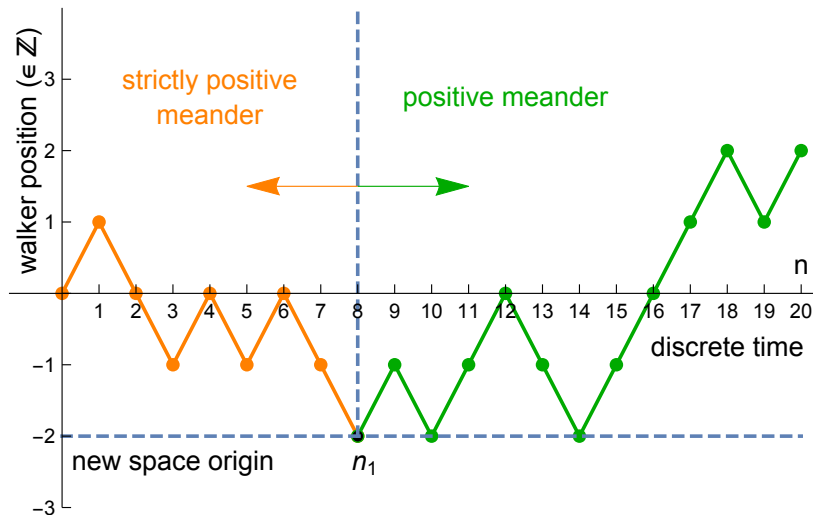


FIG. 2: A typical trajectory of a discrete jump process of  $n$  steps on a  $1D$  lattice. We denote by  $n_1$  the step at which the walk reaches for the first time its global minimum within  $n$  steps. The trajectory is split into two intervals, one on the left (shown in orange) and one on the right (shown in green) of the minimum at  $n_1$ . Seen from the minimum, the left part of the trajectory is a *strictly* positive meander, since given that the walker arrives for the *first time* at the minimum at  $n_1$ , it could not have reached this minimum before  $n_1$ . In contrast, the right part of the trajectory is just a positive meander, since the walker can reach the same minimum at some step between  $n_1$  and  $n$ , as illustrated in the figure.

and  $\sum_{n_1 \geq 0} \sum_{\ell \geq 0} \pi(n_1, \ell) = 1$  yields

$$\sum_{n_1 \geq 0} \sum_{n_2 \geq 0} \sum_{\ell \geq 0} q_{>0}(n_1) b_{\geq 0}(\ell) q_{>0}(n_2) \delta_{n, n_1+n_2+\ell} = 1. \quad (40)$$

Taking as previously the generating function of both sides of Eq. (40) with respect to  $n \geq 0$  and using the definitions in Eqs. (33) and (38), one finds

$$Q_{>0}(s)^2 B_{\geq 0}(s) = \frac{1}{1-s}. \quad (41)$$

Defining then  $Z(s) = \sqrt{B_{\geq 0}(s)}$  and using Eq. (36) on the left-hand side of Eq. (41), one obtains the generalized Sparre Andersen theorem

$$Q_{\geq 0}(s) = \frac{Z(s)}{\sqrt{1-s}}; \quad \text{with} \quad Z(s) = \sqrt{B_{\geq 0}(s)}. \quad (42)$$

Note that for continuous jumps, all the terms in Eq. (37) vanish (because the probability that the walker arrives exactly at  $x_n = 0$  is zero if the jumps are continuous), and  $B_{\geq 0}(s)$  in

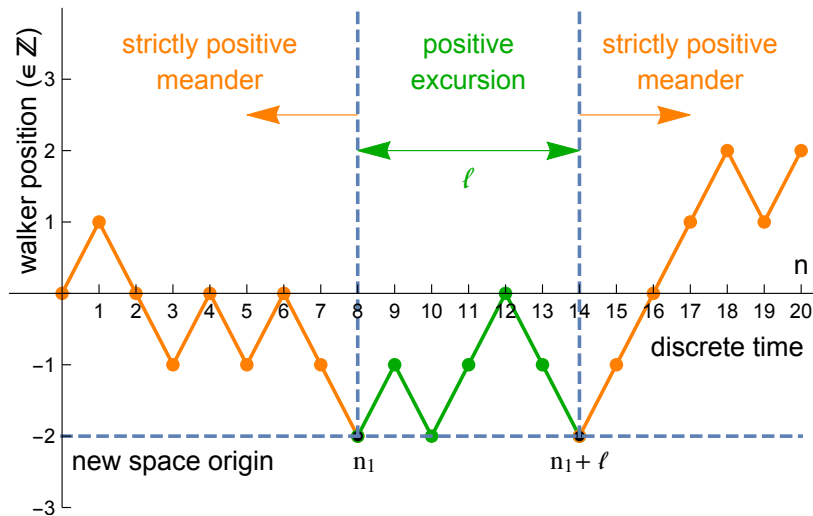


FIG. 3: A typical trajectory of a discrete jump process of  $n$  steps on a  $1D$  lattice. We denote by  $n_1$  the step at which the walk reaches for the first time its global minimum within  $n$  steps and by  $n_1 + \ell$  the step at which it reaches the global minimum for the last time. The parts of the trajectory (shown in orange), on the left of  $n_1$  and on the right of  $n_1 + \ell$ , seen from the minimum respectively at  $n_1$  and  $n_1 + \ell$ , are strictly positive meanders. Between  $n_1$  and  $n_1 + \ell$ , again seen from the minimum at  $n_1$ , we have a positive (but not necessarily *strictly* positive) excursion (shown in green).

Eq. (38) reduces to  $B_{\geq 0}(s) = b_{\geq 0}(0) = 1$  yielding  $Z(s) = 1$ , independently of the (continuous) jump distribution, as expected. For discrete jumps, the expression of  $Z(s)$  does depend on the jump distribution. A simple derivation of this expression is given in Appendix A. We indeed find that  $Z(s)$  coincides with Eq. (31), namely,

$$Z(s) = \sqrt{B_{\geq 0}(s)} = \exp\left(-\frac{\alpha}{2\pi} \int_0^{\pi/\alpha} \ln[1 - s\hat{f}(k)] dk\right). \quad (43)$$

This completes our alternative derivation of the generalized Sparre Andersen theorem. Note that Eqs. (26), (42) and (43) entirely characterize the full statistics of the number of records for both discrete and continuous jumps (with  $Z(s) = 1$  in the latter case).



### C. Universal scaling form of $P(R_n)$ in the large $n$ limit

Using the generalized Sparre Andersen theorem (42) in Eq. (26) and inverting the generating function, one obtains the following integral representation of  $P(R_n)$ ,

$$P(R_n = m) = \frac{1}{2i\pi} \oint \frac{[1 - Z(s)\sqrt{1-s}]^{m-1} Z(s)}{s^{n+1}\sqrt{1-s}} ds. \quad (44)$$

We consider the scaling regime defined by  $n, m \rightarrow +\infty$  with  $m/\sqrt{n} = O(1)$ . Making the change of variable  $s = \exp(-p/n)$  on the right-hand side of Eq. (44) and using the fact that, in the  $n \rightarrow +\infty$  limit, only the vicinity of  $s = 1$  contributes to the  $s$ -integral, one finds that at lowest order in the scaling regime, the integral representation (44) yields

$$P(R_n = m) \sim \frac{Z(1)}{2i\pi\sqrt{n}} \int_{\mathcal{L}} \frac{1}{\sqrt{p}} \exp\left(-Z(1)\frac{m}{\sqrt{n}}\sqrt{p}\right) e^p dp \quad (n, m \rightarrow +\infty, m/\sqrt{n} = O(1)), \quad (45)$$

where  $\mathcal{L}$  is a Bromwich contour. Performing the inverse Laplace transform on the right-hand side of Eq. (45) (see, e.g., Eq. 29.3.84 in [44]), one obtains

$$P(R_n = m) \sim \frac{Z(1)}{\sqrt{\pi n}} \exp\left[-\left(\frac{Z(1)m}{2\sqrt{n}}\right)^2\right] \quad (n, m \rightarrow +\infty, m/\sqrt{n} = O(1)). \quad (46)$$

From this expression it is straightforward to compute the average number of records in the scaling regime. One has

$$\begin{aligned} E(R_n) &= \sum_{m \geq 1} m P(R_n = m) \sim \sum_{m \geq 1} \frac{Z(1)m}{\sqrt{\pi n}} \exp\left[-\left(\frac{Z(1)m}{2\sqrt{n}}\right)^2\right] \\ &\sim \frac{4}{Z(1)} \sqrt{\frac{n}{\pi}} \int_0^{+\infty} x e^{-x^2} dx = \frac{2}{Z(1)} \sqrt{\frac{n}{\pi}} \quad (n \rightarrow +\infty), \end{aligned} \quad (47)$$

where we have made the change of variable  $x = Z(1)m/(2\sqrt{n})$  and replaced the Riemann sum over  $m$  with the corresponding integral over  $x$ . Rewriting then the right-hand side of Eq. (46) in terms of  $E(R_n)$  in Eq. (47), one gets the universal scaling form

$$P(R_n = m) \sim \frac{1}{E(R_n)} g\left(\frac{m}{E(R_n)}\right) \quad (n, m \rightarrow +\infty, m/\sqrt{n} = O(1)), \quad (48)$$

with  $g(x) = (2/\pi) \exp(-x^2/\pi)$ . The full statistics of the number of records in the scaling regime is thus entirely characterized (at lowest order) by a single parameter,  $E(R_n)$ , which depends on the jump distribution (through  $Z(1)$  in Eq. (47)). The scaling function itself,  $g(x)$ , is universal, i.e., independent of the jump distribution. For continuous jumps,  $Z(1)$

reduces to  $Z(1) = 1$  independently of the (continuous) jump distribution and the scaling form (48) reduces to the one mentioned in [34] (see Eq. (5)), as it should be.

The next section is devoted to a detailed study of  $E(R_n)$ . We show in particular that in the large  $n$  limit and for  $\mu = 2$ , there is a constant correction to the leading behavior in Eq. (47). For  $0 < \mu < 2$  there is no such a correction and the leading term on the right-hand side of Eq. (47) is the only term surviving the  $n \rightarrow +\infty$  limit.

#### IV. AVERAGE NUMBER OF RECORDS

As we have just seen, the average number of records,  $E(R_n)$ , is an important quantity since it is the only non-universal parameter that enters in the full distribution of the number of records in the large  $n$  limit. In this section, we study  $E(R_n)$  in more details, going further than the simple leading asymptotic behavior (47). As we did for  $P(R_n)$ , our starting point is the generating function of  $E(R_n)$  with respect to  $n \geq 0$  which encodes all the information about  $E(R_n)$ .

##### A. Generating function of $E(R_n)$

Multiplying both sides of Eq. (26) by  $m$ , summing over all  $m$  and using the relation  $\sum_{m \geq 0} m x^{m-1} = 1/(1-x)^2$ , one gets

$$\sum_{n \geq 0} E(R_n) s^n = \frac{1}{(1-s)^2 Q_{\geq 0}(s)}. \quad (49)$$

Replacing then  $Q_{\geq 0}(s)$  by the generalized Sparre Andersen result (11), one obtains the generating function of  $E(R_n)$  as

$$\sum_{n \geq 0} E(R_n) s^n = \frac{1}{(1-s)^{3/2} Z(s)}, \quad (50)$$

with  $Z(s)$  given in Eq. (12).

In the case of a continuous jump distribution,  $Z(s) = 1$  and expanding  $1/(1-s)^{3/2}$  in power series of  $s$ , one gets the universal expression of  $E(R_n)$  in Eq. (4) (see Eq. (8) in [34]). Now, we can go further and use the equations (50) and (43) to see how this universal result is affected when the random walk takes place on a lattice instead of on the line.

## B. Average number of records for a Bernoulli random walk ( $h_i = \pm 1$ )

It turns out that both  $\sum_{n \geq 0} E(R_n) s^n$  and  $E(R_n)$  can be obtained explicitly when the random walk is a symmetric Bernoulli random walk where the walker jumps by  $h = \pm 1$  at each step, with  $P(\pm 1) = 1/2$ . This simple example of lattice random walk was dealt with at the end of [34]. Here, we check that the equations (50) and (43) do give the same results as the ones in [34], as it should be.

As mentioned in Section II, in this case  $\hat{f}(k)$  reduces to  $\hat{f}(k) = \cos(\alpha k)$  and the integral on the right-hand side of Eq. (43) yields the explicit expression

$$Z(s) = \left( \frac{2}{1 + \sqrt{1-s^2}} \right)^{1/2} = \frac{2}{\sqrt{1+s} + \sqrt{1-s}}, \quad (51)$$

independent of  $\alpha$ . Using Eq. (51) in Eq. (50), one gets

$$\sum_{n \geq 0} E(R_n) s^n = \frac{\sqrt{1+s} + \sqrt{1-s}}{2(1-s)^{3/2}}, \quad (52)$$

which coincides with the equation (12) in [34]. Expanding the right-hand side of (52) in power series of  $s$ , one gets

$$\sum_{n \geq 0} E(R_n) s^n = \sum_{n \geq 1} \frac{1}{2} \left[ \frac{1}{\Gamma(n+1)} \sum_{j=0}^n \binom{n}{j} \frac{\Gamma(3/2+j)}{\Gamma(3/2+j-n)} + 1 \right] s^n, \quad (53)$$

and using the relation (see, e.g., Eq. 15.4.1 in [44])

$${}_2F_1 \left( \frac{3}{2}, -n; \frac{3}{2} - n; -1 \right) = (-1)^{n+1} \frac{2\sqrt{\pi}}{\Gamma(n-1/2)} \sum_{j=0}^n \binom{n}{j} \frac{\Gamma(3/2+j)}{\Gamma(3/2+j-n)},$$

one obtains the expected expression in the first line of Eq. (6) (see Eq. (13) in [34]),

$$E(R_n) = \frac{1}{2} \left[ \frac{(-1)^{n+1} \Gamma(n-1/2) {}_2F_1 \left( \frac{3}{2}, -n; \frac{3}{2} - n; -1 \right)}{2\sqrt{\pi} \Gamma(n+1)} + 1 \right]. \quad (54)$$

The large  $n$  behavior of Eq. (54) is more easily obtained from the behavior of Eq. (52) near its dominant singularity at  $s = 1$ . One has

$$\sum_{n \geq 0} E(R_n) s^n = \frac{1}{\sqrt{2}(1-s)^{3/2}} + \frac{1}{2(1-s)} + O \left( \frac{1}{\sqrt{1-s}} \right) \quad (s \rightarrow 1), \quad (55)$$

which translates into the large  $n$  behavior in the second line of Eq. (6),

$$E(R_n) = \sqrt{\frac{2n}{\pi}} + \frac{1}{2} + O \left( \frac{1}{\sqrt{n}} \right) \quad (n \rightarrow +\infty). \quad (56)$$

### C. Large $n$ behavior of $E(R_n)$ in the general case

An equation like Eq. (54) is a remarkable result and, in most cases, it is not possible to derive such an explicit expression of  $E(R_n)$  for all  $n$ . Nevertheless, as we will see in this section, one can always get the large  $n$  behavior of  $E(R_n)$ , for any lattice random walks and Lévy flights with index  $0 < \mu \leq 2$ .

Again, the large  $n$  behavior of  $E(R_n)$  is more easily obtained from the behavior of its generating function in Eq. (50) near its dominant singularity at  $s = 1$ . We thus need to determine the behavior of  $Z(s)$  near  $s = 1$ . To this end, we rewrite Eq. (43) as

$$\begin{aligned} Z(s) &= Z(1) \exp \left( -\frac{\alpha}{2\pi} \int_0^{\pi/\alpha} \ln [1 + (1-s)\hat{F}(k)] dk \right) \\ &= Z(1) \exp \left( -\frac{\alpha N_0}{2\pi} \int_0^{K/2} \ln [1 + (1-s)\hat{F}(k)] dk \right), \end{aligned} \quad (57)$$

with  $\hat{F}(k) = \hat{f}(k)/[1 - \hat{f}(k)]$  and  $N_0 = 2\pi/(\alpha K)$ , where  $K$  is the period of  $\hat{f}(k)$ . For  $s \rightarrow 1$ , the correction to  $Z(s) = Z(1)$  in Eq. (57) depends on the behavior of  $\hat{F}(k)$  near the zeros of  $1 - \hat{f}(k)$ , where  $\hat{F}(k)$  is large. It can be proved that in the interval  $0 \leq k \leq K/2$ , there is no other zero than the one at  $k = 0$ . Thus, in the limit  $s \rightarrow 1$ , the correction to  $Z(s) = Z(1)$  depends on the behavior of  $\hat{F}(k)$  near  $k = 0$ , only. For  $\mu = 2$ , we make the change of variable  $k = \sqrt{1-s} q/a$  and letting  $s \rightarrow 1$ , one gets

$$Z(s) \sim Z(1) \left( 1 - \frac{\alpha N_0}{2a} \sqrt{1-s} \right) \quad (s \rightarrow 1), \quad (58)$$

where we have used the small  $k$  behavior  $\hat{F}(k) \sim |ak|^{-2}$  and  $\int_0^{+\infty} \ln(1 + 1/q^2) dq = \pi$ . Similarly, for  $1 < \mu < 2$ , we make the change of variable  $k = (1-s)^{1/\mu} q/a$  and using the small  $k$  behavior  $\hat{F}(k) \sim |ak|^{-\mu}$ , one finds

$$Z(s) \sim Z(1) [1 - O(1-s)^{1/\mu}] \quad (s \rightarrow 1). \quad (59)$$

For  $0 < \mu < 1$ ,  $\hat{F}(k)$  is integrable at  $k = 0$  and writing  $\ln [1 + (1-s)\hat{F}(k)] \simeq (1-s)\hat{F}(k)$  on the right-hand side of Eq. (57), one gets

$$Z(s) \sim Z(1) [1 - O(1-s)] \quad (s \rightarrow 1). \quad (60)$$

The case  $\mu = 1$  needs a slightly more careful treatment. Skipping the details, one finds that the algebraic singularity on the right-hand side of Eq. (59) changes to a logarithmic

singularity, and one has

$$Z(s) \sim Z(1) \left[ 1 - O(1-s) \ln \left( \frac{1}{1-s} \right) \right] \quad (s \rightarrow 1). \quad (61)$$

Putting the asymptotic behavior (58) on the right-hand side of Eq. (50), one gets

$$\sum_{n \geq 0} E(R_n) s^n \sim \frac{1}{(1-s)^{3/2} Z(1)} + \frac{\alpha N_0}{(1-s) 2aZ(1)} \quad (s \rightarrow 1), \quad (62)$$

which translates into the following large  $n$  behavior of  $E(R_n)$  for  $\mu = 2$ ,

$$E(R_n) \sim \frac{2}{Z(1)} \sqrt{\frac{n}{\pi}} + \frac{\alpha N_0}{2aZ(1)} \quad (n \rightarrow +\infty). \quad (63)$$

By doing the same with the asymptotic behaviors (59) to (61), one obtains the large  $n$  behavior of  $E(R_n)$  for  $0 < \mu < 2$ . Keeping the terms surviving the  $n \rightarrow +\infty$  limit only, one finds

$$E(R_n) \sim \frac{2}{Z(1)} \sqrt{\frac{n}{\pi}} \quad (n \rightarrow +\infty). \quad (64)$$

Note that the subleading terms of  $E(R_n)$  corresponding to the small corrections on the right-hand side of Eqs. (59) to (61) go to zero when  $n \rightarrow +\infty$ . Finally, putting the equations (63) and (64) together, one obtains the large  $n$  behavior of  $E(R_n)$  as

$$E(R_n) \sim \frac{2}{Z(1)} \sqrt{\frac{n}{\pi}} + C_\mu \quad (n \rightarrow +\infty), \quad (65)$$

with

$$C_2 = \frac{\alpha N_0}{2aZ(1)}, \text{ and } C_{0 < \mu < 2} = 0, \quad (66)$$

valid for any lattice random walks and Lévy flights with index  $0 < \mu \leq 2$ . Since the equation  $\hat{f}(k) = 1$  has no other solution than  $k = 0$  in the interval  $0 \leq k \leq K/2$ , and since  $\hat{f}(k)$  is an even periodic function with period  $K$ ,  $N_0 = 2\pi/(\alpha K)$  is equal to the number of zeros of  $1 - \hat{f}(k)$  in the interval  $0 \leq k \leq \pi/\alpha$ . The presence of  $N_0$  in Eq. (66) is crucial to ensure the consistency of the results when there are several different representations of the same physical situations. For instance,  $h = \pm 1$  with  $\alpha = 2$  and  $h = \pm 2$  with  $\alpha = 1$  correspond to the same random walk. Both have the same  $a$  and  $Z(1)$ , but not the same  $\alpha$ . Without  $N_0$  one would get two different values of  $C_2$ , which would make no sense. It can be checked that  $\alpha N_0 = 2$  in both cases, yielding the same value of  $C_2$ . Similarly,  $h = \pm 1$  with  $\alpha = 1$  and  $h = \pm 2$  with  $\alpha = 1$  correspond to a mere rescaling of the same walk, which should not affect the statistics of records. Both have the same  $\alpha$  and  $Z(1)$ , but not the same  $a$ . Again, it

can be checked that  $N_0$  compensates for the difference (one has  $N_0/a = \sqrt{2}$  in both cases), yielding the same value of  $C_2$ .

Note also that in the case of a continuous jump distribution, one has  $Z(1) = 1$  and  $\alpha = 0$ , and Eq. (65) reduces to the large  $n$  behavior in Eq. (4), valid for all (continuous) jump distribution with  $0 < \mu \leq 2$ , as it should be.

#### D. Application to random walks with $h = 0, \pm 1, \pm 2$

As an application, we now consider the random walk on  $\mathbb{Z}$ ,  $\alpha = 1$ , defined by  $P(0) = p_0$ ,  $P(\pm 1) = p_1$ ,  $P(\pm 2) = p_2$ , and  $P(h) = 0$  otherwise ( $h \neq 0, \pm 1, \pm 2$ ), with  $p_0 + 2p_1 + 2p_2 = 1$ . Writing  $p = p_0$  and  $u = p_2/p_1$ , one has  $p_1 = (1 - p)/[2(1 + u)]$  and  $p_2 = u(1 - p)/[2(1 + u)]$ . For this random walk, the Fourier transform of the jump distribution,  $\hat{f}(k)$ , reads

$$\begin{aligned} \hat{f}(k) &= p_0 + 2p_1 \cos k + 2p_2 \cos 2k \\ &= 1 - 2p_1(1 - \cos k) - 2p_2(1 - \cos 2k) \\ &= 1 - \left(\frac{1-p}{1+u}\right) [(1 - \cos k) + u(1 - \cos 2k)] \\ &\sim 1 - \frac{(1+4u)(1-p)}{2(1+u)} k^2 \quad (k \rightarrow 0), \end{aligned} \tag{67}$$

yielding

$$a = \sqrt{\frac{(1+4u)(1-p)}{2(1+u)}}. \tag{68}$$

From the first line of Eq. (67) it is clear that the period of  $\hat{f}(k)$  is  $K = 2\pi$  if  $p_1 \neq 0$  and  $K = \pi$  if  $p_1 = 0$ . Since  $\alpha = 1$ , we then have  $N_0 = 2\pi/(\alpha K) = 1 + \delta_{p_1,0}$ . The large  $n$  behavior of  $E(R_n)$  is then given by the equation (65) for  $\mu = 2$  with  $\alpha = 1$ ,  $N_0 = 1 + \delta_{p_1,0}$ ,  $a$  given in Eq. (68), and

$$Z(1) = \sqrt{\frac{1+u}{1-p}} z(u), \tag{69}$$

where

$$z(u) = \exp\left(-\frac{1}{2\pi} \int_0^\pi \ln[1 - \cos k + u(1 - \cos 2k)] dk\right). \tag{70}$$

##### 1. Large $n$ behaviors of $E(R_n)$ for fixed $p$ and arbitrary $p_1$ and $p_2$

It turns out that the function  $z(u)$  in Eq. (70) can be computed explicitly. An easy way to do it is to compute its derivative  $z'(u)$  as given by the derivative of the right-hand side

of Eq. (70) with respect to  $u$ , and then integrate the result from 0 to  $u$  with  $z(0) = \sqrt{2}$ . For arbitrary  $u \geq 0$ , one obtains

$$z(u) = \frac{2\sqrt{2}}{1 + \sqrt{1 + 4u}}. \quad (71)$$

Therefore, substituting (69), (71), and  $N_0 = 1 + \delta_{p_1,0}$  in Eqs. (65) and (66) one finds

$$E(R_n) \sim \sqrt{n(1-p)} A(u) + B(u)(1 + \delta_{p_1,0}) \quad (n \rightarrow +\infty), \quad (72)$$

with

$$A(u) = \frac{1 + \sqrt{1 + 4u}}{\sqrt{2\pi(1 + u)}}, \quad B(u) = \frac{1}{4} \left( 1 + \frac{1}{\sqrt{1 + 4u}} \right), \quad (73)$$

where we recall that  $u = p_2/p_1$ .

It is clear that for fixed  $p_0 = p$ , the statistics of the records for walks with steps  $h = 0, \pm 1$  (i.e.  $p_2 = 0$ ) and the one with steps  $h = 0, \pm 2$  (i.e.  $p_1 = 0$ ), should be exactly the same, for any value of  $n$ . This implies that taking  $u = 0$  or  $u = +\infty$  in Eq. (72) should give exactly the same result. Indeed, one can check from Eq. (73) that  $A(0) = A(+\infty) = \sqrt{2/\pi}$  and  $B(u)(1 + \delta_{p_1,0})|_{u=0, +\infty} = 1/2$ , as it should be. Note that since  $B(+\infty) = 1/4$ , the term  $\propto \delta_{p_1,0}$  is crucial, as mentioned in the discussion below Eq. (66). It follows in particular that taking  $u < +\infty$  arbitrarily large, one gets  $\lim_{u \rightarrow +\infty} B(u)(1 + \delta_{p_1,0}) \neq B(u)(1 + \delta_{p_1,0})|_{u=+\infty}$ , which suggests that the limit  $p_1 \rightarrow 0$  is singular. This might indicate that the limits  $n \rightarrow \infty$  and  $p_1 \rightarrow 0$  do not commute, since one can check that for any finite  $n$ ,  $E(R_n)$  is a smooth function of  $p_1$ . Another interesting consequence of the simple fact that  $A(0) = A(+\infty)$  is that, since  $A(u)$  is a continuous function of  $u$ , it admits (at least) one extremum. It turns out that  $A(u)$  admits a single maximum at  $u = 2$  for which  $A(2) = (2/3)\sqrt{6/\pi}$ . Note that it is not a priori obvious that (i) the function  $A(u)$  is a non-monotonic function of  $u$  and (ii) that it has a single maximum.

## 2. Scaling form of $E(R_n)$ for large $n$ and $n(1-p) = O(1)$ : $p_2 = 0$ and general case

We note that the asymptotic expression in Eq. (72) requires  $(1-p)n \gg 1$  to be valid (not just  $n \gg 1$ ). In the opposite limit  $(1-p)n \ll 1$  corresponding to letting  $p \equiv P(h=0) \rightarrow 1$  first, then  $n \rightarrow +\infty$ , the walker gets stuck at the initial position  $x_0 = 0$  and there is only one record (the first one at the initial position), yielding  $E(R_n) = 1$ . To conclude this section, we derive a uniform scaling form of  $E(R_n)$  in the large  $n$  limit which describes the crossover

between Eq. (65) for  $(1-p)n \gg 1$  and  $E(R_n) \sim 1$  for  $(1-p)n \ll 1$ . For simplicity, here we give the details of the calculations for the simple case  $h = 0, \pm 1$  only. The derivation of the scaling form in the general case is explained in detail in Appendix B.

First, we determine the large  $n$  behavior of  $E(R_n)$  as given by Eq. (65) for  $n \rightarrow +\infty$  with fixed  $p < 1$ . The case  $h = 0, \pm 1$  corresponds to taking  $p_2 = 0$  (i.e.  $u = 0$ ) in Eq. (72). It reads, using  $A(0) = \sqrt{2/\pi}$  and  $B(0) = 1/2$ ,

$$E(R_n) \sim \sqrt{\frac{2(1-p)n}{\pi}} + \frac{1}{2} \quad (n \rightarrow +\infty), \quad (74)$$

which shows how the large  $n$  behavior in Eq. (6) is modified when the walker is allowed to stay in place with probability  $p$ .

The scaling form of  $E(R_n)$  is obtained by considering the limits  $n \rightarrow +\infty$  and  $p \rightarrow 1$ , keeping the scaling variable  $(1-p)n$  fixed. Injecting  $\hat{f}(k) = p + (1-p)\cos k$ , which corresponds to Eq. (67) with  $u = 0$ , into Eq (43), one finds

$$Z(s) = \frac{2}{\sqrt{(1-s) + 2s(1-p)} + \sqrt{1-s}}. \quad (75)$$

Using Eq. (75) on the right-hand side of Eq. (50), one obtains

$$\sum_{n \geq 0} E(R_n) s^n = \frac{\sqrt{(1-s) + 2s(1-p)} + \sqrt{1-s}}{2(1-s)^{3/2}}, \quad (76)$$

from which one gets the following integral representation for  $E(R_n)$ ,

$$E(R_n) = \frac{1}{2i\pi} \oint \frac{\sqrt{(1-s) + 2s(1-p)} + \sqrt{1-s}}{2s^{n+1}(1-s)^{3/2}} ds. \quad (77)$$

Making the change of variable  $s = \exp(-\lambda/n)$  in Eq. (77) and using the fact that, in the  $n \rightarrow +\infty$  limit, only the vicinity of  $s = 1$  contributes to the  $s$ -integral, one finds that at lowest order in the scaling regime, the integral representation of  $E(R_n)$  reads

$$E(R_n) \sim \frac{1}{2i\pi} \int_{\mathcal{L}} \frac{\sqrt{\lambda + 2n(1-p)} + \sqrt{\lambda}}{2\lambda^{3/2}} e^{\lambda} d\lambda \quad (n \rightarrow +\infty, p \rightarrow 1), \quad (78)$$

where  $\mathcal{L}$  is a Bromwich contour. Performing then the inverse Laplace transform on the right-hand side of Eq. (78), one gets the scaling form

$$E(R_n) \simeq \mathcal{F}[(1-p)n], \quad (79)$$

valid for  $n \rightarrow +\infty$ ,  $p \rightarrow 1$ , and fixed  $(1-p)n = O(1)$ , with the scaling function

$$\mathcal{F}(x) = \left[ \left( \frac{1}{2} + x \right) I_0(x) + x I_1(x) \right] e^{-x} + \frac{1}{2}, \quad (80)$$



where  $I_\nu(x)$  is the modified Bessel function of order  $\nu$ . From the large and small argument behaviors of  $I_\nu(x)$ , one readily obtains

$$\mathcal{F}(x) \sim \begin{cases} \sqrt{2x/\pi} + 1/2 & (x \rightarrow +\infty), \\ 1 + x/2 & (x \rightarrow 0), \end{cases} \quad (81)$$

from which it can be checked that the scaling form in Eq. (79) describes the crossover between Eq. (74) for  $(1-p)n \gg 1$ , and  $E(R_n) \sim 1$  in the opposite limit  $(1-p)n \ll 1$ . Let us note that a similar scaling limit for  $E(R_n)$  was studied recently in Ref. [22] for time series with i.i.d entries  $\{x_i\}$  in the context of the Bernoulli model of rainfall precipitation records, where a given day is dry with probability  $p$  and wet with probability  $(1-p)$ , and a record is counted only for a wet day. Due to the presence of strong correlations between the  $x_i$ 's in the case of random walks with discrete jumps, our scaling function  $\mathcal{F}(x)$  in Eq. (81) is however quite different from that in the Bernoulli model of i.i.d variables.

It is actually possible to generalize the scaling form in Eq. (79) to any discrete jump distribution with Lévy index  $0 < \mu \leq 2$ . The interested reader will find the details of the calculations in Appendix B. Writing  $Z_0(1) = Z(1)\sqrt{1-p}$  independent of  $p$  (for fixed ratios  $P(h=m)/P(h=1)$ ,  $m \geq 1$ ), one finds

$$E(R_n) \sim \mathcal{G}[(1-p)n], \quad (82)$$

valid for  $n \rightarrow +\infty$  and  $p \rightarrow 1$  with fixed  $(1-p)n = O(1)$ , where the scaling function  $\mathcal{G}(x)$  is given by

$$\mathcal{G}(x) = \frac{\sqrt{x}}{2i\pi Z_0(1)} \int_{\mathcal{L}} \exp\left(\frac{\alpha}{2\pi} \int_0^{\pi/\alpha} \ln\left[1 + \frac{\lambda/x}{1 - \hat{f}_0(k)}\right] dk\right) \frac{e^\lambda}{\lambda^{3/2}} d\lambda. \quad (83)$$

In Eq. (83),  $\hat{f}_0(k)$  denotes the Fourier transform of the jump distribution corresponding to  $p=0$  with fixed ratios  $P(h=m)/P(h=1)$ ,  $m \geq 1$ . (Note that  $Z_0(1)$  is nothing but  $Z(1)$  in Eq. (43) with  $\hat{f}_0(k)$  instead of  $\hat{f}(k)$ ). It can then be checked from the large and small argument behaviors of  $\mathcal{G}(x)$  given in Eqs. (B7) and (B8) that the scaling form in Eq. (82) matches smoothly between Eq. (65) for  $(1-p)n \gg 1$  and  $E(R_n) \sim 1$  for  $(1-p)n \ll 1$ . This result gives the scaling form of the large  $n$  behavior of  $E(R_n)$  when  $P(h=0) \rightarrow 1$  in the general case of any discrete jump distribution with Lévy index  $0 < \mu \leq 2$ .

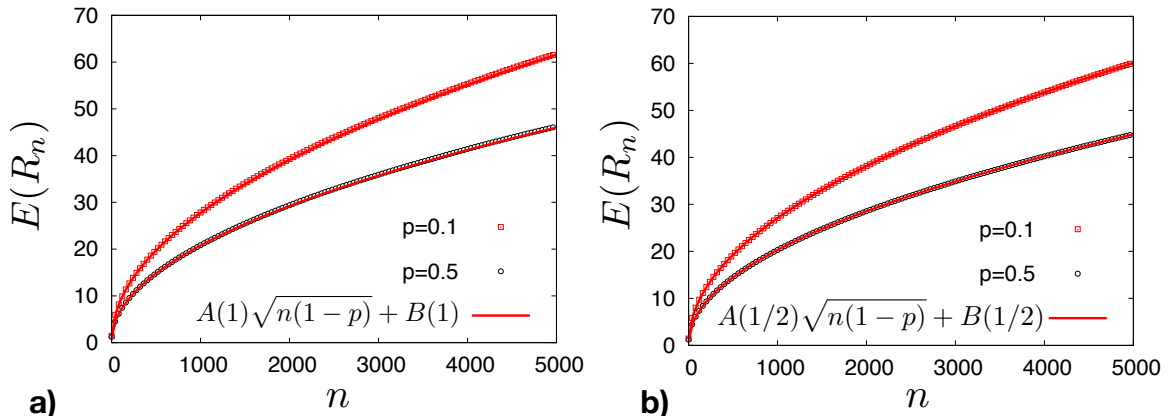


FIG. 4: Plot of the average number of records  $E(R_n)$  as a function of  $n$ , for random walks with discrete jumps  $h = 0, \pm 1, \pm 2$  corresponding to two different values of  $u = p_2/p_1$ : **a)**  $u = 1$  and **b)**  $u = 1/2$ . In each panel, the symbols (red squares and black circles) correspond respectively to  $p = 0.1$  and  $p = 0.5$  while the solid red line corresponds to the analytical (asymptotic) results given in Eqs. (72) and (73) with  $A(1) = (1 + \sqrt{5})/(2\sqrt{\pi})$  and  $B(1) = (5 + \sqrt{5})/20$  for  $u = 1$  (left panel), and  $A(1/2) = (1 + \sqrt{3})/\sqrt{3\pi}$  and  $B(1/2) = (3 + \sqrt{3})/12$  for  $u = 1/2$  (right panel). On this scale, we can see that the asymptotic estimate in Eqs. (72) and (73) is very accurate on the whole range of values of  $n$ .

## V. NUMERICAL SIMULATIONS

In this section, we present numerical simulations of discrete random walks in the case considered in Sec. IVD, i.e.  $\alpha = 1$  and  $h = 0, \pm 1, \pm 2$ , where exact analytical results can be obtained for the average number of records  $E(R_n)$  in the limit  $n \gg 1$  (see Eqs. (72) and (73)). Writing, like in Sec. IVD,  $p = P(0)$ ,  $p_1 = P(\pm 1)$ , and  $p_2 = P(\pm 2)$ , we study three different cases corresponding to three different values of the ratio  $u = p_2/p_1$ . Namely,  $u = 1$ ,  $u = 1/2$ , and  $u = 0$ . In each case, we have simulated random walks up to  $n = 5000$  steps for two different values of  $p$  (the probability of staying in place):  $p = 0.1$  and  $p = 0.5$ . The data presented here have been obtained by averaging over  $10^5$  independent realizations of the random walk.

In Fig 4 a) and b) we show our data for  $u = 1$  and  $u = 1/2$ , respectively. As we can see, the agreement between the numerics and our analytical (asymptotic) results, given in Eqs. (72) and (73), is very good. In Fig. 5 we show our data for the case  $u = 0$ , i.e.  $p_2 = 0$ .

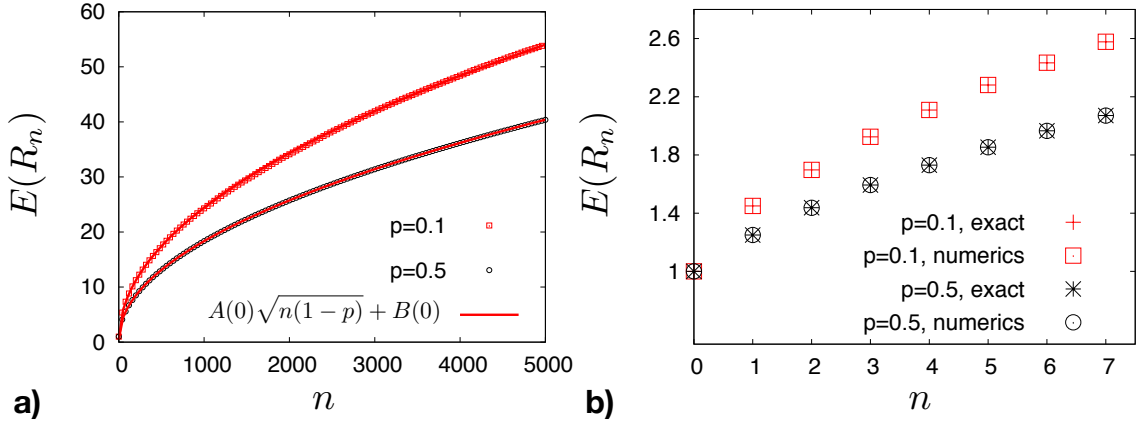


FIG. 5: Plot of the average number of records  $E(R_n)$  as a function of  $n$  for random walks with discrete jumps  $h = 0, \pm 1$ , corresponding to  $u = 0$ . **a)**: The symbols (red squares and black circles) correspond respectively to  $p = 0.1$  and  $p = 0.5$  while the solid red line corresponds to the analytical (asymptotic) results given in Eqs. (72) and (73) with  $A(0) = \sqrt{2/\pi}$  and  $B(0) = 1/2$ . **b)**: Comparison between our numerical estimates (open symbols as in the left panel) and our exact analytical results (cross symbols) given in Eq. (85) for  $p = 0.1$  and  $p = 0.5$ .

Figure 5 a) shows a comparison between our data for the average number of records  $E(R_n)$  and our asymptotic prediction in Eqs. (72) and (73) as a function of  $n$ , with  $u = 0$ , for  $p = 0.1$  and  $p = 0.5$ . Here again the agreement is excellent.

Figure 5 b) shows a comparison between our numerical computation of  $E(R_n)$  and the analytical values obtained from the generating function for  $u = 0$  and  $n = O(1)$ . More specifically, using the expression (43) of  $Z(s)$  with  $\alpha = 1$  and  $\hat{f}(k) = p + (1-p)\cos k$  (which corresponds to  $u = 0$ ) on the right-hand side of Eq. (50), one gets

$$\sum_{n=0}^{\infty} E(R_n) s^n = \frac{1}{(1-s)^{3/2}} \exp \left[ \frac{1}{2\pi} \int_0^{\pi} \ln(1 - sp - s(1-p)\cos k) \right], \quad (84)$$

and by expanding the right-hand side in powers of  $s$  (which can be done very easily with,

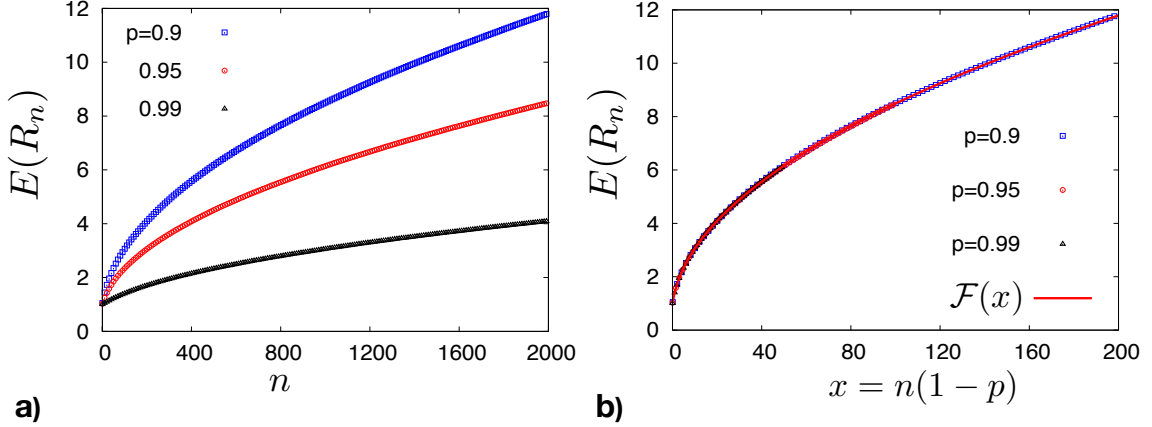


FIG. 6: **a)**: Plots of  $E(R_n)$  vs  $n$  for three different values of  $p$  close to one, namely  $p = 0.9, 0.95$  and  $0.99$  (squares, circles and triangles respectively as indicated in the keys of the main panel). **b)**: Same data as in the left panel but plotted as a function of  $x = (1-p)n$ , demonstrating a very good collapse of the three curves on a single master curve, confirming the scaling form predicted in Eq. (79). The solid red curve is a plot of  $\mathcal{F}(x)$  as given in Eq. (80).

e.g., Mathematica) one obtains, for the first values of  $n$ ,

$$\begin{aligned}
E(R_0) &= 1, E(R_1) = \frac{3}{2} - \frac{p}{2}, E(R_2) = \frac{1}{4}(7 - 2p - p^2) \\
E(R_3) &= 2 - \frac{3p}{4} - \frac{p^3}{4}, E(R_4) = \frac{1}{16}(35 - 12p - 6p^2 + 4p^3 - 5p^4) \\
E(R_5) &= \frac{1}{16}(38 - 15p - 10p^3 + 10p^4 - 7p^5) \\
E(R_6) &= \frac{1}{32}(81 - 30p - 15p^2 + 20p^3 - 45p^4 + 42p^5 - 21p^6) \\
E(R_7) &= \frac{1}{32}(86 - 35p - 35p^3 + 70p^4 - 105p^5 + 84p^6 - 33p^7).
\end{aligned} \tag{85}$$

In Fig. 5 b) we compare these exact analytical values with our numerical simulations for  $p = 0.1$  and  $p = 0.5$ . The comparison shows a perfect agreement.

Finally, in Fig. 6, we show plots of  $E(R_n)$  for three different values of  $p$  close to one, namely  $p = 0.9, p = 0.95$ , and  $p = 0.99$ , which confirm the scaling form predicted in Eq. (79) together with the expression of the scaling function  $\mathcal{F}(x)$  given in Eq. (80) – see Fig. 6 b).

## VI. CONCLUSION

In this paper we have studied a general question concerning the universality of the statistics of the number of records  $R_n$  for a discrete-time series whose entries represent the successive positions of an  $n$ -step random walk in one dimension. Starting at the origin, at each discrete-time step the walker jumps by a certain distance drawn from a symmetric distribution  $f(\eta)$ . If the jump distribution  $f(\eta)$  is continuous, it was known [34] that the statistics of records is *strongly* universal, i.e., independent of  $f(\eta)$  for all  $n$ . What happens to this universality if the walk takes place on a discrete lattice, i.e., if  $f(\eta)$  is nonzero for integer values of  $\eta$  only, was still largely unknown. For the special case of nearest neighbour lattice walk when  $\eta = \pm 1$ , previous exact results [34] (see also [37]) show that while the average number of records  $E(R_n)$  still grows asymptotically as  $\sqrt{n}$  for large  $n$ , like in the continuous case, the prefactor of this square root growth as well as the next subleading term differ from the ones in the continuous case. This raises the natural question of whether the results from  $\pm 1$  walk can be generalized to more general jump processes on a  $1D$  lattice, where the walker may jump an arbitrary number of lattice units in one step, without being necessarily restricted to nearest neighbor walk. This question about the record statistics for general discrete jump processes may have relevance, for instance, when the entries of the time series represent the price of a stock on different days. The day-to-day jump in the stock price is often rounded to the nearest integer and stock price effectively moves on a lattice. The purpose of this paper was to derive the record statistics for such a general discrete jump process on a  $1D$  lattice and investigate whether there exist universal results at all, or perhaps the universality holds only asymptotically for large  $n$ .

Our main conclusion is that the *strong* universality, valid for *arbitrary*  $n$ , in the case of continuous jump distributions, no longer holds for discrete jump processes. However, a vestige of universality still remains asymptotically for large  $n$ . We found that for arbitrary symmetric discrete jump processes, the distribution of the scaled record number  $R_n/E(R_n)$  converges, for large  $n$ , to a universal half-Gaussian distribution, like in the case of continuous jumps. The dependence on the jump distribution  $f(\eta)$  is entirely absorbed in the expected number of records  $E(R_n)$ , leaving the scaled distribution itself independent of  $f(\eta)$ . By computing the large  $n$  behavior of  $E(R_n)$  for general discrete jump processes, we found that while it still grows as  $\sqrt{n}$  for large  $n$  irrespective of  $f(\eta)$ , the prefactor of  $\sqrt{n}$  as well as the

next subleading term depend explicitly on the jump distribution.

A technical bonus of our work is an alternative derivation of the celebrated generalized Sparre Andersen theorem that was originally derived using combinatorial arguments [36]. Our derivation relies more on an algebraic and non-combinatorial approach. Finally, the techniques developed in this paper will hopefully be useful to study other observables for discrete jump processes on a lattice, such as the statistics of the maximum displacement, the time at which the maximum/minimum occurs etc.

In this paper, we have restricted ourselves to the study of record statistics for symmetric lattice random walks. A natural question is to ask how this record statistics will be affected by the presence of a nonzero drift that would make the jump distribution on the lattice asymmetric. Indeed, for continuous jump distribution (not necessarily symmetric), the record statistics has been studied using the generalized Sparre Andersen theorem. In this case, even though the full universality of the record statistics at all step  $n$  no longer holds for asymmetric jumps, there remains a vestige of universality in the limit of large  $n$  [40]. It would be interesting to see how the drift affects the record statistics in the case of lattice random walks studied in this paper.

### Appendix A: Derivation of the expression of $Z(s)$

In this appendix, we derive the expression of the function  $Z(s)$  given in Eq. (43). We write

$$b_{>0}(n) = \text{Prob}(x_1 > 0, x_2 > 0, \dots, x_n = 0 | x_0 = 0), \quad (\text{A1})$$

for  $n \geq 1$  and  $b_{>0}(0) = 1$ . The generating function of  $b_{>0}(n)$  is

$$B_{>0}(s) = \sum_{n \geq 0} b_{>0}(n) s^n. \quad (\text{A2})$$

Similar expressions for  $b_{\geq 0}(n)$  and  $B_{\geq 0}(s)$  are given in Eqs. (37) and (38), respectively. For  $n \geq 1$ , one has the relation

$$b_{\geq 0}(n) = \sum_{m=0}^{n-1} b_{\geq 0}(m) b_{>0}(n-m), \quad (\text{A3})$$

which is easily seen by taking for  $m$  the last time at which the walker touches  $x = 0$  before arriving at  $x_n = 0$ . Writing

$$\sum_{m=0}^{n-1} b_{\geq 0}(m)b_{>0}(n-m) = \sum_{m_1 \geq 0} \sum_{m_2 \geq 1} b_{\geq 0}(m_1)b_{>0}(m_2) \delta_{n, m_1+m_2}$$

and taking the generating function of both sides of Eq. (A3) with respect to  $n \geq 1$  one gets  $B_{\geq 0}(s) - 1 = B_{\geq 0}(s)(B_{>0}(s) - 1)$  which yields

$$B_{\geq 0}(s) = \frac{1}{2 - B_{>0}(s)}. \quad (\text{A4})$$

Now, we consider a bridge of length  $n \geq 1$ . Let  $n_1$  and  $n_2$  respectively denote the first and last time at which it reaches its minimum before arriving at  $x_n = 0$ . Clearly  $0 \leq n_1 \leq n_2 \leq n-1$ . For  $n_1 = 0$ , the bridge is a positive excursion of probability  $b_{\geq 0}(n)$  which we write like in Eq. (A3) with  $m = n_2$ . For  $1 \leq n_1 \leq n-1$ , we do the Vervaat construction [45] by cutting at  $n_1$ . One obtains

$$\begin{aligned} \text{Prob}(x_n = 0) &= \sum_{n_1=0}^{n-1} \sum_{n_2=n_1}^{n-1} b_{\geq 0}(n_2 - n_1)b_{>0}(n - n_2 + n_1) \\ &= \sum_{n_1=0}^{n-1} \sum_{\ell=0}^{n-n_1-1} b_{\geq 0}(\ell)b_{>0}(n - \ell), \end{aligned} \quad (\text{A5})$$

where  $\ell = n_2 - n_1$ . Now, write  $\sum_{\ell=0}^{n-n_1-1} = \sum_{\ell=0}^n - \sum_{\ell=n-n_1}^n$  on the right-hand side of Eq. (A5), use the relation

$$\sum_{\ell=0}^n b_{\geq 0}(\ell)b_{>0}(n - \ell) = 2b_{\geq 0}(n),$$

(which follows trivially from Eq. (A3)) in the sum  $\sum_{\ell=0}^n$ , and make the change of variable  $\ell \rightarrow n - \ell$  in the sum  $\sum_{\ell=n-n_1}^n$ . One gets

$$\begin{aligned} \text{Prob}(x_n = 0) &= \sum_{n_1=0}^{n-1} \left( 2b_{\geq 0}(n) - \sum_{\ell=0}^{n_1} b_{\geq 0}(n - \ell)b_{>0}(\ell) \right) \\ &= 2n b_{\geq 0}(n) - \sum_{\ell=0}^{n-1} (n - \ell) b_{\geq 0}(n - \ell)b_{>0}(\ell) \\ &= 2n b_{\geq 0}(n) - \sum_{\ell_1 \geq 1} \sum_{\ell_2 \geq 0} \ell_1 b_{\geq 0}(\ell_1)b_{>0}(\ell_2) \delta_{n, \ell_1+\ell_2}. \end{aligned} \quad (\text{A6})$$

The factor  $(n - \ell)$  in the sum over  $\ell$  in the second line comes from the fact that the double sum over  $\ell$  and  $n_1$  is over the domain  $0 \leq \ell \leq n_1 \leq n-1$ : for each value of  $\ell$ ,  $n_1$  takes  $(n - \ell)$

values. Taking the generating function of both sides of Eq. (A6) with respect to  $n \geq 1$ , one finds

$$\sum_{n \geq 1} \text{Prob}(x_n = 0) t^n = t B'_{\geq 0}(t) (2 - B_{>0}(t)) = \frac{t B'_{\geq 0}(t)}{B_{\geq 0}(t)}, \quad (\text{A7})$$

where we have used Eq. (A4). It remains to divide both sides of Eq. (A7) by  $t$  and integrate over  $t$  from  $t = 0$  to  $t = s$ , with  $B_{\geq 0}(0) = 1$ . One obtains

$$B_{\geq 0}(s) = \exp \left( \sum_{n \geq 1} \text{Prob}(x_n = 0) \frac{s^n}{n} \right). \quad (\text{A8})$$

Finally, from  $Z(s) = \sqrt{B_{\geq 0}(s)}$ , the relation  $\text{Prob}(x_n = 0) = (\alpha/2\pi) \int_{-\pi/\alpha}^{\pi/\alpha} \hat{f}(k)^n dk$ , and  $\hat{f}(k) = \hat{f}(-k)$ , one obtains

$$\begin{aligned} Z(s) &= \exp \left( \frac{1}{2} \sum_{n \geq 1} \text{Prob}(x_n = 0) \frac{s^n}{n} \right) \\ &= \exp \left( -\frac{\alpha}{2\pi} \int_0^{\pi/\alpha} \ln [1 - s \hat{f}(k)] dk \right). \end{aligned} \quad (\text{A9})$$

## Appendix B: Derivation of the scaling form of $E(R_n)$ in the general case

In this appendix, we derive the scaling form given in Eq. (82) for any discrete jump distribution with Lévy index  $0 < \mu \leq 2$ . Write  $P(h = 0) = p$ ,  $P(h = 1) = p_1$ , and  $u_m = P(h = m)/p_1$ . From  $\hat{f}(k) = p + 2p_1 \sum_{m \geq 1} u_m \cos(m\alpha k)$  and  $p + 2p_1 \sum_{m \geq 1} u_m = 1$ , one gets

$$\hat{f}(k) = 1 - \frac{(1-p)}{\sum_{m \geq 1} u_m} \sum_{m \geq 1} u_m [1 - \cos(m\alpha k)]. \quad (\text{B1})$$

Introducing  $\hat{f}_0(k)$  the Fourier transform of the jump distribution corresponding to  $p = 0$  with fixed  $u_m$  ( $m \geq 1$ ), we rewrite Eq. (B1) in the simpler form

$$\hat{f}(k) = 1 - (1-p)[1 - \hat{f}_0(k)]. \quad (\text{B2})$$

Using the expression (B2) of  $\hat{f}(k)$  on the right-hand side of Eq. (57) and considering the scaling regime  $s, p \rightarrow 1$  with fixed  $(1-p)/(1-s)$ , one obtains

$$Z(s) \sim \frac{Z_0(1)}{\sqrt{1-p}} \exp \left( -\frac{\alpha}{2\pi} \int_0^{\pi/\alpha} \ln \left[ 1 + \frac{(1-s)/(1-p)}{1 - \hat{f}_0(k)} \right] dk \right), \quad (\text{B3})$$



where  $Z_0(1)$  is given by Eq. (43) with  $s = 1$  and  $\hat{f}_0(k)$  instead of  $\hat{f}(k)$ . Now, inverting the generating function in Eq. (50), one gets the integral representation

$$E(R_n) = \frac{1}{2i\pi} \oint \frac{ds}{s^{n+1}(1-s)^{3/2}Z(s)}. \quad (\text{B4})$$

Making the change of variable  $s = \exp(-\lambda/n)$  in Eq. (B4), using the fact that, in the  $n \rightarrow +\infty$  limit, only the vicinity of  $s = 1$  contributes to the  $s$ -integral, and taking for  $Z(s)$  the asymptotic expression in Eq. (B3), one obtains the following scaling form for  $E(R_n)$ ,

$$E(R_n) \sim \mathcal{G}[(1-p)n], \quad (\text{B5})$$

valid for  $n \rightarrow +\infty$ ,  $p \rightarrow 1$ , and fixed  $(1-p)n = O(1)$ , with the scaling function

$$\mathcal{G}(x) = \frac{\sqrt{x}}{2i\pi Z_0(1)} \int_{\mathcal{L}} \exp\left(\frac{\alpha}{2\pi} \int_0^{\pi/\alpha} \ln\left[1 + \frac{\lambda/x}{1 - \hat{f}_0(k)}\right] dk\right) \frac{e^\lambda}{\lambda^{3/2}} d\lambda. \quad (\text{B6})$$

The large argument behavior of  $\mathcal{G}(x)$  is obtained by doing the same analysis as the one for the behavior of Eq. (57) near  $s = 1$ . Making the substitutions  $(1-s) \rightarrow \lambda/x$  and  $\hat{F}(k) \rightarrow 1/(1 - \hat{f}_0(k))$  in the integral on the right-hand side of Eq. (57), and using the small  $k$  behavior  $\hat{f}_0(k) \sim 1 - (a_0 k)^\mu$ , with  $a_0 = a/(1-p)^{1/\mu}$  independent of  $p$ , one finds

$$\begin{aligned} \mathcal{G}(x) &\sim \frac{\sqrt{x}}{2i\pi Z_0(1)} \int_{\mathcal{L}} \left(1 + \frac{\alpha}{2a_0} \sqrt{\frac{\lambda}{x}} \delta_{\mu,2}\right) \frac{e^\lambda}{\lambda^{3/2}} d\lambda \\ &= \frac{2}{Z_0(1)} \sqrt{\frac{x}{\pi}} + \frac{\alpha}{2a_0 Z_0(1)} \delta_{\mu,2} \quad (x \rightarrow +\infty), \end{aligned} \quad (\text{B7})$$

where we have kept the terms surviving the  $x \rightarrow +\infty$  limit, only. Injecting this result onto the right-hand side of Eq. (B5) and using the relations  $Z_0(1) = Z(1)\sqrt{1-p}$  and  $a_0 Z_0(1) = a Z(1)$ , one can check that the scaling form (B5) coincides with the large  $n$  behavior (65) in the limit  $(1-p)n \gg 1$ .

On the other hand, the small argument behavior of  $\mathcal{G}(x)$  is obtained from the equation (B6) as

$$\begin{aligned} \mathcal{G}(x) &\sim \frac{\sqrt{x}}{2i\pi Z_0(1)} \int_{\mathcal{L}} \exp\left(\frac{\alpha}{2\pi} \int_0^{\pi/\alpha} \ln\left[\frac{\lambda/x}{1 - \hat{f}_0(k)}\right] dk\right) \frac{e^\lambda}{\lambda^{3/2}} d\lambda \\ &= \frac{1}{Z_0(1)} \exp\left(-\frac{\alpha}{2\pi} \int_0^{\pi/\alpha} \ln[1 - \hat{f}_0(k)] dk\right) = 1 \quad (x \rightarrow 0). \end{aligned} \quad (\text{B8})$$

Using this result on the right-hand side of Eq. (B5), one finds  $E(R_n) \sim 1$  in the opposite limit  $(1-p)n \ll 1$ . All these results hold for any discrete jump distribution with Lévy index  $0 < \mu \leq 2$ .

- 
- [1] Hoyt D V 1981 *Climatic Change* **3** 243
  - [2] Redner S and Petersen M R 2006 *Phys. Rev. E* **74** 061114
  - [3] Matalas N C 1997 *Climatic Change* **37** 89
  - [4] Vogel R M, Zafirakou-Koulouris A and Matalas N C 2001 *Water Resour. Res.* **37** 1723
  - [5] Wergen G, Bogner M and Krug J 2011 *Phys. Rev. E* **83** 051109
  - [6] Wergen G, Majumdar S N and Schehr G 2012 *Phys. Rev. E* **86** 011119
  - [7] Challet D 2015 preprint arXiv:1502.05367
  - [8] Barlevy G 2002 *Rev. Economic Stud.* **69** 65
  - [9] Barlevy G and Nagaraja H N 2006 *J. Appl. Probab.* **43** 1119
  - [10] Gembris D, Taylor J G and Suter D 2002 *Nature* **417** 406
  - [11] Gembris D, Taylor J G and Suter D 2007 *J. Appl. Stat.* **34** 529
  - [12] Glick N 1978 *Am. Math. Mon.* **85** 2
  - [13] Krug J 2007 *J. Stat. Mech.* P07001
  - [14] Ben-Naim E, Redner S and Vazquez F 2007 *Europhys. Lett.* **77** 30005
  - [15] Meehl G A, Tebaldi C, Walton G, Easterling D and McDaniel L 2009 *Geophys. Res. Lett.* **36** L23701
  - [16] Wergen G and Krug J 2010 *Europhys. Lett.* **92** 30008
  - [17] Anderson A and Kostinski A 2011 *J. Appl. Meteorol. Climatol.* **50** 1859
  - [18] Franke J, Wergen G and Krug J 2012 *Phys. Rev. Lett.* **108** 064101
  - [19] Wergen G, Hense A and Krug J 2014 *Clim. Dyn.* **42** 1275
  - [20] Wilks D S and Wilby R L 1999 *Prog. Phys. Geography* **23** 329
  - [21] Benestad R E 2006 *J. Climate* **19** 630
  - [22] Majumdar S N, von Bomhard P and Krug J (2019) *Phys. Rev. Lett.* **122**, 158702
  - [23] Krug J and Jain K 2005 *Physica A* **358** 1
  - [24] Oliveira L P, Jensen H J, Nicodemi M and Sibani P 2005 *Phys. Rev. B* **71** 104526
  - [25] Sibani P, Rodriguez G F and Kenning G G 2006 *Phys. Rev. B* **74** 224407

- [26] Godrèche C and Luck J M 2008 *J. Stat. Mech.* P11006
- [27] Chandler K N 1952 *J. R. Stat. Soc. Ser. B. Methodol.* **14** 220
- [28] Foster F G and Stuart A 1954 *J. R. Stat. Soc.* **16** 1
- [29] Nevzorov V B 1987 *Theory Probab. Appl.* **32** 201
- [30] Arnold B C, Balakrishnan N and Nagaraja H N 1998 *Records* (New York: Wiley)
- [31] Schmittmann B and Zia R K P 1999 *Am. J. Phys.* **67** 1269
- [32] Nevzorov V B 2001 *Records: Mathematical Theory* (Providence, RI: American Mathematical Society)
- [33] W. Feller, *Introduction to Probability Theory and Its Applications* (Wiley, New York, 1966), Vol. 2.
- [34] Majumdar S N and Ziff R M 2008 *Phys. Rev. Lett.* **101** 050601
- [35] Majumdar S N 2010 *Physica A*, **389**, 4299
- [36] Sparre Andersen E 1954 *Math. Scand.* **2** 195
- [37] Godrèche C, Majumdar S N and Schehr G 2017 *J. Phys. A: Math. Theor.* **50** 333001
- [38] Sabhapandit S 2011 *Europhys. Lett.* **94** 20003
- [39] Le Doussal P and Wiese K J 2009 *Phys. Rev. E* **79** 051105
- [40] Majumdar S N, Schehr G and Wergen G 2012 *J. Phys. A: Math. Theor.* **45** 355002
- [41] Edery Y, Kostinski A B, Majumdar S N and Berkowitz B 2013 *Phys. Rev. Lett.* **110**, 180602
- [42] Mori F, Le Doussal P, Majumdar S N and Schehr G 2020 *Phys. Rev. Lett.* **124**, 090603
- [43] Kearney M J 2020 *J. Stat. Mech.*, 023206
- [44] Abramowitz M and Stegun I A 1985 *Handbook of Mathematical Functions* (Washington: Applied mathematics series, National Bureau of Standards)
- [45] Vervaat M 1979 *Ann. of Probab.* **7**, 143