

# AGES OF RECORDS IN RANDOM WALKS

RÉKA SZABÓ AND BÁLINT VETŐ

*Dedicated to Bálint Tóth on the occasion of his 60th birthday*

ABSTRACT. We consider random walks with continuous and symmetric step distributions. We prove universal asymptotics for the average proportion of the age of the  $k$ th longest lasting record for  $k = 1, 2, \dots$  and for the probability that the record of the  $k$ th longest age is broken at step  $n$ . Due to the relation to the Chinese restaurant process, the ranked sequence of proportions of ages converges to the Poisson–Dirichlet distribution.

## 1. INTRODUCTION

Recently there has been an increasing interest in the study of record statistics, since this field has found several applications in natural sciences, in particular in finance. If the discrete time series  $x_1, x_2, \dots$  is considered, then we say that a record event happens at time  $k$ , if  $x_k$  is larger than all the previous values in the series. The record statistics of a time series can be characterized with several variables, such as the time up to which a record survives, the number of records that occur up to time  $n$ , the age of the longest lasting record, or the probability that a record event occurs at any given time. Depending on the exact definition of the model, these variables exhibit interesting behavior.

In this paper we study the record statistics of the random walk where the jumps form an i.i.d. sequence of random variables drawn from a continuous distribution. A remarkable feature of the record statistics of this model is their universal behaviour, i.e. they do not depend on the details of the jump distribution as long as it is continuous and symmetric [12]. All these results have their roots in the Sparre Andersen theorem [1].

The statistics of the longest lasting record after  $n$  steps were first examined in [12] in the case of random walks with arbitrary symmetric and continuous distribution of the jumps. The asymptotic expected proportion of the longest lasting record was computed. Later in [10] apart from the expected value of the age of the longest lasting record another variable was considered: the probability, that the last record is the longest lasting one. It was observed that the statistics of the records is sensitive to the exact definition of the length of the last record event. Three different natural ways were given to define the length of the last record up to time  $n$ , and the universal behaviour of the longest lasting record was described in each case. Some of the constants characterizing the record statistics appeared before in the literature. The constant in the first case appeared first in [20] as the expected first coordinate of the Poisson–Dirichlet distribution and in [12] in the context of record ages. The constant in the second case appeared first in the context of the excursion lengths of a standard Brownian motion [22]. These constants also appeared in the study of fractional Brownian motion later in [9, 7].

The present paper is an extension of the work presented in [10]. We consider the full sequence of the  $k$ th longest lasting records for  $k = 1, 2, \dots$  and by introducing new constants, we describe their average asymptotic proportions and the limit of the probability that the record of the  $k$ th longest age is broken at step  $n$  for each  $k$  as  $n \rightarrow \infty$ . Some

---

2010 *Mathematics Subject Classification.* 60K05, 60G50.

*Key words and phrases.* longest lasting records, random walk, asymptotic average proportion, Laplace transform, renewal process, Poisson–Dirichlet distribution.

of these constants appeared before in [22] and in [6] where the probability that the last excursion is the  $k$ th longest lasting one in a standard Brownian motion was computed. These results were later extended for general renewal processes [11]. The corresponding observables for random walk excursions were studied in the recent work [8].

The theory of exchangeable partitions was developed in a series of papers by Pitman and coauthors, see also [18] and references therein. Based on this theory, the partition generated by the lengths of the record ages in one of the cases studied in the present paper is equal in law to the Chinese restaurant process at time  $n$ . As a consequence, the ranked sequence of ages of records rescaled by the total length converges to the Poisson–Dirichlet distribution [14, 15, 19] with parameter  $\alpha = 1/2$ . This distribution is also related to the ranked lengths of excursions in diffusion processes [20] and to renewal processes [2]. The Poisson–Dirichlet distribution with parameter  $\alpha = 1/2$  was proved to be the limiting distribution of ranked excursion lengths up to time  $n$  of the simple symmetric random walk [3].

The paper is organized as follows. In Section 2, we state our main result on the asymptotic behaviour of the Laplace transforms of two types of quantities: the average proportions of the ages of the longest lasting records and the probabilities that the record of the  $k$ th longest age is broken at step  $n$ . In two of the three cases, we can turn these asymptotics into a proof of convergence of the average proportions and in one case to the convergence of the record breaking probabilities by specific arguments. Then based on the relation to the Chinese restaurant process, we conclude a limit theorem about the rescaled ranked sequence of ages. The proof of the asymptotics of Laplace transforms is postponed to Section 3.

## 2. RESULTS

Let the sequence  $\eta_n$  for  $n = 1, 2, \dots$  consist of i.i.d. random variables from a common symmetric and absolutely continuous distribution. We consider the random walk  $X_0 = 0$  and  $X_n = X_{n-1} + \eta_n$  for  $n = 1, 2, \dots$ . A record event occurs at step  $n$ , if the random walk takes a value at time  $n$  which is larger than all the previous values of the random walk, that is,  $X_n > \max(X_0, \dots, X_{n-1})$  with the convention that the first record occurs at time 0.

Let  $t_1 = 0$  and recursively define

$$(1) \quad t_k = t_{k-1} + \min\{\tau > 0 : X_{t_{k-1}+\tau} > X_{t_{k-1}}\}$$

for  $k = 2, 3, \dots$  which gives the time when the  $k$ th record occurs. The age of the  $k$ th record is

$$(2) \quad \tau_k = t_{k+1} - t_k$$

for  $k = 1, 2, \dots$  which form a sequence of i.i.d. random variables by construction. Let  $R_n$  be the number of records until time  $n$ , i.e. define  $R_n = m$  if and only if  $t_m \leq n < t_{m+1}$ . By the i.i.d. structure of the ages of records  $(\tau_k)_{k=1}^\infty$ ,  $R_n$  becomes a renewal process. Let  $A_n$  denote the current age of the renewal process at time  $n$ , that is,  $A_n = n - t_m$  if  $R_n = m$ . For an illustration, see Figure 1.

There are three different natural choices for the set of ages which one considers. Given that  $R_n = m$ , the sequences of ages after  $n$  steps in the different cases are given by

$$(3) \quad \mathcal{A}_{n,m}^{(1)} = (\tau_1, \tau_2, \dots, \tau_{m-1}, A_n),$$

$$(4) \quad \mathcal{A}_{n,m}^{(2)} = (\tau_1, \tau_2, \dots, \tau_m),$$

$$(5) \quad \mathcal{A}_{n,m}^{(3)} = (\tau_1, \tau_2, \dots, \tau_{m-1})$$

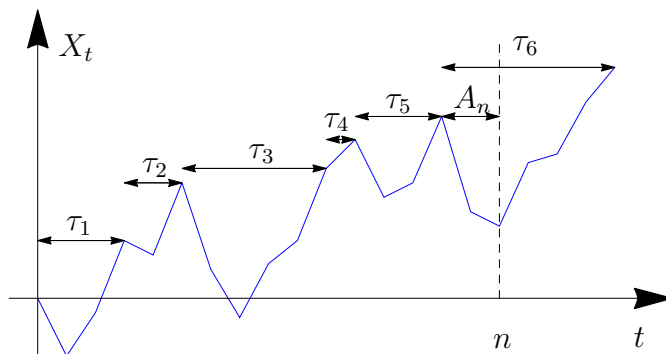


FIGURE 1. A random walk trajectory with  $R_n = 6$ , the age of the last record after  $n$  steps is  $A_n$ ,  $\tau_6$  or  $\tau_5$ .

where multiple occurrences are considered with multiplicities. Now we define the two main quantities of interest. Given the sequence  $\mathcal{A}_{n,m}^{(\beta)}$  for  $\beta = 1, 2, 3$ , let  $L_k^{(\beta)}(n)$  denote the  $k$ th largest element of  $\mathcal{A}_{n,m}^{(\beta)}$ . Further, let  $p_k^{(\beta)}(n)$  denote the probability that the last element of  $\mathcal{A}_{n,m}^{(\beta)}$  is the  $k$ th longest one of the sequence. In particular,

$$(6) \quad p_1^{(1)}(n) = \mathbf{P}(A_n \geq \max(\tau_1, \tau_2, \dots, \tau_{m-1})),$$

$$(7) \quad p_1^{(2)}(n) = \mathbf{P}(\tau_m \geq \max(\tau_1, \tau_2, \dots, \tau_{m-1})),$$

$$(8) \quad p_1^{(3)}(n) = \mathbf{P}(\tau_{m-1} \geq \max(\tau_1, \tau_2, \dots, \tau_{m-2})).$$

On the example of Figure 1, the sequence of the ages of records in the three cases are  $\mathcal{A}_{16,6}^{(1)} = (3, 2, 5, 1, 3, 2)$ ,  $\mathcal{A}_{16,6}^{(2)} = (3, 2, 5, 1, 3, 6)$  and  $\mathcal{A}_{16,6}^{(3)} = (3, 2, 5, 1, 3)$ . The corresponding values of  $L_k^{(\beta)}(16)$  are  $L_1^{(1)}(16) = 5$ ,  $L_2^{(1)}(16) = 3$ ,  $L_3^{(1)}(16) = 3$ ,  $L_4^{(1)}(16) = 2$ ,  $L_5^{(1)}(16) = 2$  and  $L_6^{(1)}(16) = 1$  for  $\beta = 1$ .

We introduce the distribution of the ages of records

$$(9) \quad f(k) = \mathbf{P}(\tau_1 = k) = \mathbf{P}(X_k > X_0 \geq \max(X_1, \dots, X_{k-1}))$$

for  $k = 1, 2, \dots$  and  $f(0) = 0$ , i.e.  $f(k)$  is the probability that the random walk crosses its starting point between steps  $k-1$  and  $k$ . Further, let

$$(10) \quad q(k) = \mathbf{P}(\tau_1 > k) = \mathbf{P}(\max(X_1, \dots, X_k) < X_0)$$

for  $k = 0, 1, \dots$ , in particular  $q(0) = 1$ , that is the probability that there is no record until time  $k$ . One clearly has

$$(11) \quad f(k) = q(k-1) - q(k)$$

by definition. The distribution of the ages of records is known by the classical Sparre Andersen theorem [1] which can also be found in Section XII.7 of [5].

**Theorem 1.** *Suppose that the step distribution of the random walk  $X_n$  is continuous and symmetric. Then the distribution and the generating function of the ages of records are given by*

$$(12) \quad q(k) = \frac{1}{2^{2k}} \binom{2k}{k}, \quad f(k) = \frac{1}{2^{2k-1}k} \binom{2k-2}{k-1},$$

$$(13) \quad \tilde{q}(z) = \sum_{k \geq 0} q(k)z^k = \frac{1}{\sqrt{1-z}}, \quad \tilde{f}(z) = \sum_{k \geq 1} f(k)z^k = 1 - \sqrt{1-z}.$$

As a consequence, one has the following asymptotics by the Stirling formula:

$$(14) \quad q(k) \sim \frac{1}{\sqrt{\pi k}}, \quad f(k) \sim \frac{1}{2\sqrt{\pi k}^{3/2}}$$

where the symbol  $\sim$  means that the ratio of the two sides converges to 1 as  $k \rightarrow \infty$ .

**Remark.** *By Theorem 1, the distribution of the ages of records does not depend on the step distribution of the random walk as long as the latter is continuous and symmetric. Therefore, instead of the random walk description, the problem of the record age statistics can be posed as a question on renewal processes where the holding times are distributed according to (12).*

*Furthermore, since only the sizes and not the ordering of the renewal intervals matters in what follows, it is enough to consider the partition of  $n$  generated by the ages of records in the random walk up to time  $n$ . This partition in the first case which corresponds to (3) equals in law to the Chinese restaurant process at time  $n$ , see Subsection 2.2.*

One naturally expects also based on numerical evidence that the asymptotic proportions of the ages of the longest records  $\mathbf{E}(L_k^{(\beta)}(n))/n$  and the probabilities  $p_k^{(\beta)}(n)$  converge as  $n \rightarrow \infty$  for fixed  $k$  and  $\beta$ . We can only prove the convergence of  $\mathbf{E}(L_k^{(1)}(n))/n$ ,  $\mathbf{E}(L_k^{(2)}(n))/n$  and  $p_k^{(1)}(n)$ , but we identify the limits in all the cases under the assumption of convergence.

**2.1. Limiting rank of the last interval and limiting proportions of ages of records.** In order to state the main result of the present paper, we define the constants

$$(15) \quad p_k^{(1)} = \frac{1}{2^{k-1}} \int_0^\infty \frac{x^{-1/2} e^{-x} \Gamma(-1/2, x)^{k-1}}{(x^{-1/2} e^{-x} + \gamma(1/2, x))^k} dx,$$

$$(16) \quad p_k^{(2)} = \frac{1}{2^k} \int_0^\infty \frac{x^{-3/2} (1 - e^{-x}) \Gamma(-1/2, x)^{k-1}}{(x^{-1/2} e^{-x} + \gamma(1/2, x))^k} dx,$$

$$(17) \quad C_k^{(1)} = p_k^{(1)},$$

$$(18) \quad C_k^{(2)} = \frac{1}{2^{k-1}} \int_0^\infty \frac{x^{-1/2} \Gamma(-1/2, x)^{k-1}}{(x^{-1/2} e^{-x} + \gamma(1/2, x))^k} dx,$$

$$(19) \quad C_k^{(3)} = \frac{1}{2^k} \int_0^\infty \frac{\Gamma(-1/2, x)^k}{(x^{-1/2} e^{-x} + \gamma(1/2, x))^k} dx$$

where

$$(20) \quad \Gamma(\nu, x) = \int_x^\infty e^{-t} t^{\nu-1} dt,$$

$$(21) \quad \gamma(\nu, x) = \int_0^x e^{-t} t^{\nu-1} dt$$

are the upper and lower incomplete gamma functions.

**Theorem 2.** *For the Laplace transforms of the probabilities  $p_k^{(\beta)}$ , for  $\beta = 1, 2$ , one has*

$$(22) \quad \sum_{n \geq 0} p_k^{(\beta)}(n) e^{-sn} \sim \frac{p_k^{(\beta)}}{s}$$

and

$$(23) \quad \sum_{n \geq 0} p_k^{(3)}(n) e^{-sn} \sim \frac{1}{\sqrt{s}} \ln \frac{1}{\sqrt{s}}$$

$\mathbf{k}$	$p_k^{(1)} = C_k^{(1)}$	$p_k^{(2)}$	$C_k^{(2)}$	$C_k^{(3)}$
1	0.62651 ...	0.80031 ...	$\infty$	0.24174 ...
2	0.14301 ...	0.08125 ...	0.18685 ...	0.07999 ...
3	0.06302 ...	0.03342 ...	0.07107 ...	0.04105 ...
4	0.03565 ...	0.01846 ...	0.03826 ...	0.02528 ...
5	0.02300 ...	0.01178 ...	0.02412 ...	0.01724 ...
6	0.01610 ...	0.00819 ...	0.01666 ...	0.01255 ...

TABLE 1. The numerical values of  $p_k^{(1)}$ ,  $p_k^{(2)}$  and  $C_k^{(\beta)}$  for  $\beta = 1, 2, 3$  for  $1 \leq k \leq 6$

as  $s \rightarrow 0$ . As a consequence,  $p_k^{(1)}(n) \rightarrow p_k^{(1)}$  and if  $p_k^{(2)}(n)$  converges as  $n \rightarrow \infty$ , then the limit is  $p_k^{(2)}$ . If  $\sqrt{n}p_k^{(3)}(n)/\ln n$  converges, then the limit is  $1/(2\sqrt{\pi})$ , that is

$$(24) \quad p_k^{(3)}(n) \sim \frac{\ln n}{2\sqrt{\pi n}}.$$

Further, one has the following asymptotics for the Laplace transforms of the expected ages of the longest lasting records

$$(25) \quad \sum_{n \geq 0} \mathbf{E}(L_k^{(\beta)}(n))e^{-sn} \sim \frac{C_k^{(\beta)}}{s^2}$$

as  $s \rightarrow 0$  with the constants (17)–(19) for  $k = 1, 2, \dots$  and  $\beta = 1, 2, 3$ . Consequently

$$(26) \quad \frac{\mathbf{E}(L_k^{(\beta)}(n))}{n} \rightarrow C_k^{(\beta)}$$

as  $n \rightarrow \infty$  for  $k = 1, 2, \dots$  and for  $\beta = 1, 2$ . If the sequence  $\mathbf{E}(L_k^{(3)}(n))/n$  converges as  $n \rightarrow \infty$ , then the limit is  $C_k^{(3)}$ .

Table 1 contains the first few numerical values of  $p_k^{(\beta)}$  and  $C_k^{(\beta)}$ . The constants  $p_k^{(1)}$  first appeared in [20] as the expected  $k$ th coordinate of the Poisson–Dirichlet distribution,  $p_k^{(2)}$  first appeared in [22] as the asymptotic probabilities that the last excursion is the  $k$ th longest lasting one in a standard Brownian motion, see also [6]. The constants  $C_k^{(2)}$  and  $C_k^{(3)}$  were previously unknown in the literature except for  $C_1^{(3)}$  which appeared in [10].  $C_1^{(2)} = \infty$  since the expectation of holding times in the corresponding renewal process is infinite, see (14).

**Remark.** The following heuristic argument explains why  $\ln n/\sqrt{n}$  is the order of magnitude of  $p_k^{(3)}(n)$ . Let  $\lambda(n)$  be the random time of the last record until time  $n$ . Since the rescaled spent waiting time converges in distribution to the arcsine law (see e.g. Section XIV.3 of [5]), one has

$$(27) \quad \mathbf{P}(\lambda(n) = l) \asymp \frac{1}{\sqrt{ln}}$$

for  $l \leq n/2$  as  $n \rightarrow \infty$  where  $\asymp$  means that the ratio of the two sides are bounded from below and from above by non-trivial constants. On the other hand, conditionally given that there was a record at time  $l$ , the number of renewal intervals before  $l$  equals  $\asymp \sqrt{l}$ , and these intervals follow each other in a uniform random order. Hence

$$(28) \quad \mathbf{P}(\text{the last renewal interval in } [0, l] \text{ is the } k\text{th longest one} \mid \lambda(n) = l) \asymp \frac{1}{\sqrt{l}}$$

for  $l$  large enough. By the law of total probability, if we sum the product of (27) and (28) for  $l$ , then the result is asymptotically  $\ln n/\sqrt{n}$ .

**2.2. Relation to the Chinese restaurant process.** The Chinese restaurant process is a two-parameter stochastic process that can be described as a seating arrangement in a restaurant, see [16] and [18]. Let  $\alpha \in [0, 1]$  and  $\theta > -\alpha$  denote the parameters and consider an initially empty restaurant with an infinite number of circular tables, each with infinite capacity. The first customer is seated at the first table. Suppose that  $n_1, \dots, n_k$  people are sitting at the tables where  $\sum_{i=1}^k n_i = n$ . Then the new customer chooses table  $i$  for  $1 \leq i \leq k$  with probability  $(n_i - \alpha)/(n + \theta)$ , or sits to a new table with probability  $(\theta + k\alpha)/(n + \theta)$ .

Let  $N_i(n)$  denote the  $i$ th largest element of the random seating arrangement  $n_1, \dots, n_k$ , that is the size of the  $i$ th largest table after the arrival of the  $n$ th customer. Then the size-biased permutation of the limit of the frequencies

$$(29) \quad \left( \frac{N_1(n)}{n}, \frac{N_2(n)}{n}, \dots \right)$$

can be represented as  $(\tau_1, \tau_2, \dots)$ ,

$$(30) \quad \tau_i = (1 - U_i) \prod_{j=1}^{i-1} U_j \quad i \geq 1$$

where the sequence of  $U_i$ 's is independent with beta distribution of parameter  $(\theta + i\alpha, 1 - \alpha)$ . This limit is the Poisson–Dirichlet distribution with parameters  $(\alpha, \theta)$  described first in [13] for  $\theta = 0$ .

**Lemma 3.** *The distribution of the partition generated by the ages of the records in the first case which corresponds to (3) has the same distribution as the Chinese restaurant process with parameters  $(\alpha, \theta) = (1/2, 0)$  at time step  $n$ .*

*Proof.* The equality in law can be seen in two steps. According to Lemma 7 of [17], the Brownian excursion partition up to  $n$  has the same distribution as the discrete renewal process with holding time distribution (12). The Brownian excursion partition is a partition of integers obtained by the random equivalence relation  $i \sim j$  iff  $U_i$  and  $U_j$  fall in the same excursion interval of a standard Brownian motion  $B$  away from 0 where  $U_i$  are uniformly distributed on  $[0, 1]$  independent of each other and of  $B$ .

On the other hand, by Section 4.5 of [18] based on [19] and [16], the Brownian excursion partition has the distribution of the Chinese restaurant process with parameters  $(\alpha, \theta) = (1/2, 0)$ . This finishes the proof.  $\square$

**Theorem 4.** *Let us consider a random walk with arbitrary symmetric and continuous distribution of the jumps. Let  $(\tau_1, \tau_2, \dots, \tau_{m-1}, A_n)$  be the sequence of the ages of its records up to time  $n$ . Then the relative sizes of the ranked ages of records jointly converge*

$$(31) \quad \left( \frac{L_1^{(1)}(n)}{n}, \frac{L_2^{(1)}(n)}{n}, \dots \right) \xrightarrow{d} \text{PD}(1/2, 0)$$

*in distribution in the vague sense.*

Similar results were already shown for Brownian motion [20] and for the simple symmetric random walk [3]: the limiting distribution of ranked excursion lengths up to time  $n$  converges in both cases to the Poisson–Dirichlet distribution with parameter  $\alpha = 1/2$ .

**Remark.** *It follows from Lemma 3 that the size of the age of the last record up to  $n$  is a size biased sample from the Chinese restaurant process. Since the relative cluster sizes*



in the latter converge to  $\text{PD}(1/2, 0)$ , one can conclude that the probability  $p_k^{(1)}(n)$  that the last record is the  $k$ th longest lasting one also converges. The limit is the expectation of the  $k$ th coordinate in the  $\text{PD}(1/2, 0)$  distribution which is exactly  $p_k^{(1)}$ , see Proposition 17 in [20]. This also proves (26) for  $\beta = 1$ .

**2.3. Convergence of average limiting proportions.** The next two propositions imply that the knowledge on the Laplace transforms in Theorem 2 can be turned into a proof of convergence of  $n^{-1}\mathbf{E}(L_k^{(\beta)}(n))$  for  $\beta = 1, 2$ .

**Proposition 5.** *For any fixed  $n$ , the following relation holds:*

$$(32) \quad \mathbf{E}(L_k^{(1)}(n+1)) = \mathbf{E}(L_k^{(1)}(n)) + p_k^{(1)}(n),$$

or equivalently,

$$(33) \quad \mathbf{E}(L_k^{(1)}(n)) = \sum_{j=0}^n p_k^{(1)}(j)$$

with the convention  $p_1^{(1)}(0) = 1$  and  $p_k^{(1)}(0) = 0$  for  $k = 2, 3, \dots$ . As a consequence, the equality of the constants  $p_k^{(1)} = C_k^{(1)}$  for  $k = 1, 2, \dots$  follows without using Theorem 2. Furthermore, the asymptotic equivalence (22) implies the convergence (26) for  $\beta = 1$ .

**Proposition 6.** *For any fixed  $k$ , the sequence  $\left(\sum_{j=1}^k \mathbf{E}(L_j^{(2)}(n))\right)_n$  is subadditive, that is,*

$$(34) \quad \sum_{j=1}^k \mathbf{E}(L_j^{(2)}(n+m)) \leq \sum_{j=1}^k \mathbf{E}(L_j^{(2)}(n)) + \sum_{j=1}^k \mathbf{E}(L_j^{(2)}(m)).$$

Consequently, the proportions  $\mathbf{E}(L_k^{(2)}(n))/n$  converge as  $n \rightarrow \infty$ .

*Proof of Proposition 5.* Note that the age of the  $k$ th longest record can only increase from time  $n$  to time  $n+1$ , if the last interval is the  $k$ th longest one at time  $n$ . This observation means that the difference  $L_k(n+1) - L_k(n)$  is a Bernoulli random variable with parameter  $p_k^{(1)}(n)$  which proves (32), then (33) follows immediately. To prove (26) for  $\beta = 1$ , the Hardy–Littlewood Tauberian theorem can be used (see e.g. Theorem 2 in XIII.5 of [5]). The theorem yields that (22) for  $\beta = 1$  implies that the right-hand side of (33) is asymptotically equivalent to  $np_k^{(1)}$ , i.e. (26) holds for  $\beta = 1$  with  $C_k^{(1)} = p_k^{(1)}$ .  $\square$

*Proof of Proposition 6.* We define the shifted random walk  $\tilde{X}_l = X_{n+l} - X_n$ . Note that if  $r$  is the first record time of the random walk  $X_l$  after time  $n$ , then  $r - n$  is also a record time of the shifted random walk  $\tilde{X}_l$ . Furthermore, all the record times and ages of records for  $X_l$  after time  $r$  correspond to those of  $\tilde{X}_l$  after time  $r - n$ .

Let  $\mathcal{A}_n^{(2)}$  denote the list of ages of records in the original random walk which started not later than  $n$ . Similarly, let  $\tilde{\mathcal{A}}_m^{(2)}$  be the list of ages of records in the shifted random walk which started not later than  $m$ . Now we consider the list  $\mathcal{A}_{n+m}^{(2)}$ . All the ages of records which started not later than  $n$  are contained in the list  $\mathcal{A}_n^{(2)}$ . Ages of records in the original random walk started after time  $n$  are contained in  $\tilde{\mathcal{A}}_m^{(2)}$  by the previous observation on the ages of records after time  $r$ . Hence the  $k$  largest elements of the list  $\mathcal{A}_{n+m}^{(2)}$  can be found either in  $\mathcal{A}_n^{(2)}$  or in  $\tilde{\mathcal{A}}_m^{(2)}$ . Since  $\mathcal{A}_m^{(2)}$  and  $\tilde{\mathcal{A}}_m^{(2)}$  are equal in distribution, for the expectations, the subadditivity relation (34) follows. By the subadditive lemma, the sequence  $\sum_{j=1}^k \mathbf{E}(L_j^{(2)}(n))/n$  converges for all  $k$  as  $n \rightarrow \infty$  which finishes the proof of the proposition.  $\square$

## 2.4. Summation identities.

**Proposition 7.** *The sum of the constants  $p_k^{(\beta)}$  and  $C_k^{(\beta)}$  are the following:*

$$(35) \quad \sum_{k=1}^{\infty} p_k^{(1)} = \sum_{k=1}^{\infty} p_k^{(2)} = \sum_{k=1}^{\infty} C_k^{(1)} = 1,$$

$$(36) \quad \sum_{k=2}^{\infty} C_k^{(2)} = \frac{1}{2\sqrt{\pi}} \int_0^{\infty} \frac{x^{-1/2} \Gamma(-1/2, x)}{x^{-1/2} + \gamma(1/2, x)} dx \simeq 0.43067 \dots,$$

$$(37) \quad \sum_{k=1}^{\infty} C_k^{(3)} = \frac{1}{2}.$$

The proof of the summation identities (35)–(37) in Proposition 7 are straightforward by summing the geometric series under the integral sign in (15)–(19) and by the integration by parts formula

$$(38) \quad \begin{aligned} \Gamma(-1/2, x) &= 2x^{-1/2}e^{-x} - 2\Gamma(1/2, x) \\ &= 2x^{-1/2}e^{-x} + 2\gamma(1/2, x) - 2\sqrt{\pi} \end{aligned}$$

where the fact  $\Gamma(1/2) = \sqrt{\pi}$  is used.

Next we explain why it is natural to expect the identities (35) and (37) to hold. Clearly, for any finite  $n$ ,

$$\sum_{k=1}^{\infty} p_k^{(1)}(n) = \sum_{k=1}^{\infty} p_k^{(2)}(n) = \sum_{k=1}^{\infty} \frac{\mathbf{E}(L_k^{(1)}(n))}{n} = 1$$

holds, hence by Fatou's lemma, the sums in (35) are at most 1. The reason for these sums not being less than 1 is the following. One expects an invariance principle for the ages of the records of the random walk  $X_n$  to hold, i.e. ages of records of the walk should converge to the ages of records of Brownian motion. Let us consider the trajectories of the random walk  $X_n$  as continuous functions by linear interpolation. By reflecting a trajectory of  $X_n$  downwards off 0 in the Skorokhod sense (see e.g. Section VI.2 in [21]), the record ages of the original trajectory correspond to the excursion lengths of the reflected trajectory (up to an error which is at most 1). Invariance principle for the excursions of the random walk was proved in [4] for the simple random walk, but it is expected for other step distributions as well. Provided that such an invariance principle is proved for the continuous and symmetric step distributions which are considered in the present paper, then the equality in (35) readily follows.

To explain (37), one should observe that

$$(39) \quad \sum_{k=1}^{\infty} \mathbf{E}(L_k^{(3)}(n)) = \frac{n}{2}$$

holds for any  $n$  since the left-hand side is the expected time of the last record before  $n$  which is the expected time of the maximal value of the random walk  $X_k$  for  $0 \leq k \leq n$ . By reverting the trajectory, the time of the maximal value of the random walk  $\widehat{X}_k = X_{n-k} - X_n$  for  $0 \leq k \leq n$  and that of the original walk add up to  $n$ . Since this time reversal preserves the measure, it explains (39). The same invariance principle for the ages of records of the random walk  $X_n$  which was used in the explanation of (35) could turn this argument into a proof of (37).



## 3. PROOF OF LAPLACE TRANSFORM ASYMPTOTICS

This section is devoted to the proof of Theorem 2 which is the main result of the paper. We start with an a priori bound on the constants  $C_k^{(\beta)}$ .

**Lemma 8.** *For any  $\beta = 1, 2, 3$  and for  $k \geq 2$ , we have*

$$(40) \quad \frac{L_k^{(\beta)}(n)}{n} \leq \frac{1}{k-1}$$

almost surely. As a consequence, we have

$$(41) \quad C_k^{(\beta)} \leq \frac{1}{k-1}.$$

*Proof.* The bound (40) is clear, since at least  $k-1$  of the  $k$  longest intervals are in  $[0, n]$ . Consequently,

$$(42) \quad \sum_{n \geq 0} \mathbf{E}(L_k^{(\beta)}(n))e^{-sn} \leq \frac{1}{k-1} \sum_{n \geq 0} ne^{-sn} = \frac{1}{k-1} \frac{e^{-s}}{(1-e^{-s})^2} \sim \frac{1}{k-1} \frac{1}{s^2}$$

which proves (41). □

*Proof of Theorem 2.* To compute  $L_k^{(\beta)}(n)$  and  $p_k^{(\beta)}(n)$ , first we have to determine the joint distribution of  $R_n$  and the sequence of ages up to time  $n$ . Assume that  $R_n = m$  and let  $\bar{l}^{(1)} = (l_1, l_2, \dots, l_{m-1}, a)$ ,  $\bar{l}^{(2)} = (l_1, l_2, \dots, l_m)$  and  $\bar{l}^{(3)} = (l_1, l_2, \dots, l_{m-1})$  be the possible realizations of ages in the three different cases. The probability of a realization is

$$(43) \quad \mathbf{P}(\bar{l}^{(\beta)}, n, m) = \mathbf{P}(\mathcal{A}_{n,m}^{(\beta)} = \bar{l}^{(\beta)}, R_n = m)$$

which can be expressed with  $q(k)$  and  $f(k)$  of (12) in the cases  $\beta = 1, 2, 3$  as

$$(44) \quad \mathbf{P}(\bar{l}^{(1)}, n, m) = f(l_1)f(l_2) \dots f(l_{m-1})q(a) \mathbb{1} \left( \sum_{k=1}^{m-1} l_k + a = n \right)$$

$$(45) \quad \mathbf{P}(\bar{l}^{(2)}, n, m) = f(l_1)f(l_2) \dots f(l_m) \mathbb{1} \left( \sum_{k=1}^{m-1} l_k < n \leq \sum_{k=1}^m l_k \right)$$

$$(46) \quad \mathbf{P}(\bar{l}^{(3)}, n, m) = f(l_1)f(l_2) \dots f(l_{m-1}) \sum_{a \geq 0} q(a) \mathbb{1} \left( \sum_{k=1}^{m-1} l_k + a = n \right)$$

where  $\mathbb{1}(\cdot)$  is the indicator function. This explicit expression for the joint distribution in case  $\beta = 1$  first appeared in [12].

Case  $\beta = 1$ . Let  $p_k^{(1)}(n, m)$  denote the probability that  $R_n = m$  and the last element of  $\mathcal{A}_{n,m}^{(1)}$  is the  $k$ st longest one of the sequence. In this case, there are exactly  $k-1$  longer intervals than  $A_n$  among  $\tau_1, \dots, \tau_{m-1}$ . This probability can be expressed as

$$(47) \quad \begin{aligned} p_k^{(1)}(n, m) &= \mathbf{P}(A_n = L_k^{(1)}(n), R_n = m) \\ &= \binom{m-1}{k-1} \sum_{a \geq 0} \sum_{l_1=a+1}^{\infty} \dots \sum_{l_{k-1}=a+1}^{\infty} \sum_{l_k=1}^a \dots \sum_{l_{m-1}=1}^a \mathbf{P}(\bar{l}^{(1)}, n, m). \end{aligned}$$

By summing over all  $m$ , we get exactly  $p_k^{(1)}(n)$ , the probability that the last element of  $\mathcal{A}_{n,m}^{(1)}$  is the  $k$ st longest one. For its generating function,

$$\begin{aligned}
\tilde{p}_k^{(1)}(z) &= \sum_{n \geq 0} p_k^{(1)}(n) z^n \\
(48) \quad &= \sum_{n \geq 0} \sum_{m \geq 1} p_k^{(1)}(n, m) z^n \\
&= \sum_{m \geq 1} \sum_{j \geq 0} q(j) z^j \binom{m-1}{k-1} \left( \sum_{l=1}^j f(l) z^l \right)^{m-k} \left( \sum_{l=j+1}^{\infty} f(l) z^l \right)^{k-1}.
\end{aligned}$$

Next we apply the identity

$$(49) \quad \sum_{m \geq 1} \binom{m-k}{k-1} x^{m-k} = \frac{1}{(1-x)^k}.$$

The expression in (48) for the generating function can be written as

$$(50) \quad \tilde{p}_k^{(1)}(z) = \sum_{j \geq 0} \frac{q(j) z^j \left( \sum_{l=j+1}^{\infty} f(l) z^l \right)^{k-1}}{\left( 1 - \sum_{l=1}^j f(l) z^l \right)^k}.$$

The following identity can be derived from (11)

$$(51) \quad 1 - \sum_{k=1}^t f(k) z^k = q(t) z^t + (1-z) \sum_{k=0}^{t-1} q(k) z^k.$$

Using this identity, the denominator can be rewritten as

$$(52) \quad \tilde{p}_k^{(1)}(z) = \sum_{j \geq 0} \frac{q(j) z^j \left( \sum_{l=j+1}^{\infty} f(l) z^l \right)^{k-1}}{\left( q(j) z^j + (1-z) \sum_{l=0}^{j-1} q(l) z^l \right)^k}.$$

To prove (22) for  $\beta = 1$ , one has to show

$$(53) \quad s \tilde{p}_k^{(1)}(e^{-s}) = s \sum_{j \geq 0} \frac{q(j) e^{-sj} \left( \sum_{l=j+1}^{\infty} f(l) e^{-sl} \right)^{k-1}}{\left( q(j) e^{-sj} + (1-e^{-s}) \sum_{l=0}^{j-1} q(l) e^{-sl} \right)^k} \rightarrow p_k^{(1)}$$

for any  $k$  fixed as  $s \rightarrow 0$ . Using (14), one can see that there is a  $C > 0$  such that for any  $x > 0$  large enough, the summand  $j = x/s$  in (53) is bounded by  $C e^{-x}$  uniformly in  $s$ . This bound can then be turned into a uniform exponential bound on the tail of the sum in (53). The uniform exponential bound justifies the exchange of limits and reduces (53) to

$$(54) \quad \lim_{K \rightarrow \infty} \lim_{s \rightarrow 0} \sum_{j=0}^{K/s} \frac{s q(j) e^{-sj} \left( \sum_{l=j+1}^{\infty} f(l) e^{-sl} \right)^{k-1}}{\left( q(j) e^{-sj} + (1-e^{-s}) \sum_{l=0}^{j-1} q(l) e^{-sl} \right)^k} = p_k^{(1)}.$$

where sum on the left-hand side is a Riemann sum for the integral in (15) on  $[0, K]$  instead of  $[0, \infty)$ . This verifies (22) for  $\beta = 1$ . The convergence  $p_k^{(1)}(n) \rightarrow p_k^{(1)}$  follows from the remark after Theorem 4.

For the proof of (25), one can use Proposition 5 to get the relation

$$(55) \quad (1-z) \sum_{n \geq 0} \mathbf{E}(L_k^{(1)}(n)) z^n = z \sum_{n \geq 0} p_k^{(1)}(n) z^n.$$

By substituting  $z = e^{-s}$ , one readily gets (25) for  $\beta = 1$  with  $C_k^{(1)} = p_k^{(1)}$ . The convergence (26) for  $\beta = 1$  follows as a consequence of Proposition 5 and the Hardy–Littlewood Tauberian theorem.

Case  $\beta = 2$ . With similar notations as in the previous case, we have the following formula

$$\begin{aligned} p_k^{(2)}(n, m) &= \mathbf{P}(\tau_m = L_k^{(2)}(n), R_n = m) \\ &= \binom{m-1}{k-1} \sum_{j \geq 1} \sum_{l_1=j+1}^{\infty} \cdots \sum_{l_{k-1}=j+1}^{\infty} \sum_{l_k=1}^j \cdots \sum_{l_{m-1}=1}^j \mathbf{P}(\bar{l}^{(2)}, n, m). \end{aligned}$$

Using (45) and (49), the generating function of the sequence  $p_k^{(2)}(n)$  can be expressed as

$$\begin{aligned} \tilde{p}_k^{(2)}(z) &= \sum_{m \geq 1} \sum_{j \geq 0} f(j) \frac{1-z^j}{1-z} \binom{m-1}{k-1} \left( \sum_{l=1}^j f(l) z^l \right)^{m-k} \left( \sum_{l=j+1}^{\infty} f(l) z^l \right)^{k-1} \\ &= \sum_{j \geq 0} \frac{1}{1-z} \left( \frac{f(j)(1-z^j) \left( \sum_{l=j+1}^{\infty} f(l) z^l \right)^{k-1}}{\left( 1 - \sum_{l=1}^j f(l) z^l \right)^k} \right). \end{aligned}$$

Following the method presented in the previous case, we similarly get (22) for  $\beta = 2$  with  $p_k^{(2)}$  given in (16).

However in the  $\beta = 2$  case, there is no simple relation of  $\mathbf{E}(L_k^{(\beta)}(n))$  and  $p_k^{(\beta)}(n)$  like (32) for  $\beta = 1$ , hence we cannot deduce (25) from (22), but we proceed with direct computations. By (14), the expected total length of the last unfinished interval is infinite, therefore for  $\beta = 2$ , one has  $\mathbf{E}(L_1^{(2)}(n)) = \infty$ . In order to compute the expected interval lengths for  $k > 1$ , denote the distribution function of  $L_k^{(2)}(n)$  by  $F_k^{(2)}(t, n)$ . By the law of total probability, one can write

$$(56) \quad F_k^{(2)}(t, n) = \mathbf{P}(L_k^{(2)}(n) \leq t) = F_{k-1}^{(2)}(t, n) + \mathbf{P}(L_k^{(2)}(n) \leq t, L_{k-1}^{(2)}(n) > t).$$

We decompose the distribution function depending on the number of renewal intervals until  $n$  as

$$(57) \quad F_k^{(2)}(t, n) = \sum_{m \geq 1} F_k^{(2)}(t, n, m)$$

where

$$\begin{aligned} F_k^{(2)}(t, n, m) &= \mathbf{P}(L_k^{(2)}(n) \leq t, R_n = m) \\ &= F_{k-1}^{(2)}(t, n, m) + \mathbf{P}(L_k^{(2)}(n) \leq t, L_{k-1}^{(2)}(n) > t, R_n = m). \end{aligned}$$

In the last probability above, if the first  $k-1$  longest lasting time intervals are greater than  $t$ , we can distinguish two different cases: if these intervals are all among  $\tau_1, \dots, \tau_{m-1}$ , or if  $\tau_m$  is one of these intervals. Based on this observation,  $F_k^{(2)}(t, n, m)$  can be rewritten

as follows

$$\begin{aligned}
(58) \quad F_k^{(2)}(t, n, m) &= F_{k-1}^{(2)}(t, n, m) \\
&+ \binom{m-1}{k-1} \sum_{l_1=t+1}^{\infty} \cdots \sum_{l_{k-1}=t+1}^{\infty} \sum_{l_k=1}^t \cdots \sum_{l_{m-1}=1}^t \sum_{l_m=0}^t \mathbf{P}(\bar{l}^{(2)}, n, m) \\
&+ \binom{m-1}{k-2} \sum_{l_1=t+1}^{\infty} \cdots \sum_{l_{k-2}=t+1}^{\infty} \sum_{l_{k-1}=1}^t \cdots \sum_{l_{m-1}=1}^t \sum_{l_m=t+1}^{\infty} \mathbf{P}(\bar{l}^{(2)}, n, m).
\end{aligned}$$

The generating function

$$(59) \quad \tilde{F}_k^{(2)}(t, z) = \sum_{n \geq 0} F_k^{(2)}(t, n) z^n$$

can be decomposed based on (57), and using the recursion (58) for each term and by substituting the value of  $\mathbf{P}(\bar{l}^{(2)}, n, m)$  from (45), one obtains

$$\begin{aligned}
(60) \quad \tilde{F}_k^{(2)}(t, z) &= \tilde{F}_{k-1}^{(2)}(t, z) + \frac{1}{1-z} \left( \frac{\left( \sum_{j=t+1}^{\infty} f(j) z^j \right)^{k-1} \sum_{j=1}^t f(j) (1-z^j)}{\left( 1 - \sum_{j=1}^t f(j) z^j \right)^k} \right) \\
&+ \frac{1}{1-z} \left( \frac{\left( \sum_{j=t+1}^{\infty} f(j) z^j \right)^{k-2} \sum_{j=t+1}^{\infty} f(j) (1-z^j)}{\left( 1 - \sum_{j=1}^t f(j) z^j \right)^{k-1}} \right).
\end{aligned}$$

By computing the expected value using the formula

$$(61) \quad \mathbf{E}(L_k^{(2)}(n)) = \sum_{t \geq 0} \left( 1 - F_k^{(2)}(t, n) \right)$$

and by considering the generating functions of the two sides of (61), one can turn the recursion of generating functions (60) into the asymptotic recursion of Laplace transforms

$$\begin{aligned}
(62) \quad \sum_{n \geq 0} e^{-sn} \mathbf{E}(L_{k-1}^{(2)}(n)) &\sim \sum_{n \geq 0} e^{-sn} \mathbf{E}(L_k^{(2)}(n)) \\
&+ \frac{1}{s^2} \int_0^{\infty} \left( \frac{1}{2^{k-1}} \frac{\int_x^{\infty} y^{-3/2} (1-e^{-y}) dy (\Gamma(-1/2, x))^{k-2}}{(x^{-1/2} e^{-x} + \gamma(1/2, x))^{k-1}} \right. \\
&\quad \left. + \frac{1}{2^k} \frac{\int_0^x y^{-3/2} (1-e^{-y}) dy (\Gamma(-1/2, x))^{k-1}}{(x^{-1/2} e^{-x} + \gamma(1/2, x))^k} \right) dx
\end{aligned}$$

as  $s \rightarrow 0$  for  $k \geq 3$  by following the method presented for  $\beta = 1$ . Simplifying the integral above and using the integration by parts (38), one gets

$$(63) \quad \sum_{n \geq 0} e^{-sn} \mathbf{E}(L_{k-1}^{(2)}(n)) \sim \sum_{n \geq 0} e^{-sn} \mathbf{E}(L_k^{(2)}(n)) + \frac{1}{s^2} \frac{\sqrt{\pi}}{2^{k-2}} \int_0^{\infty} \frac{x^{-1/2} \Gamma(-1/2, x)^{k-2}}{(x^{-1/2} e^{-x} + \gamma(1/2, x))^k} dx$$

for  $k \geq 3$ . Having a  $k \rightarrow \infty$  decay bound on  $L_k^{(2)}(n)$  by Lemma 8, the asymptotic equivalence (25) follows for some sequence of constants  $C_k^{(2)}$  which satisfies the recursion

$$(64) \quad C_{k-1}^{(2)} = C_k^{(2)} + \frac{\sqrt{\pi}}{2^{k-2}} \int_0^{\infty} \frac{x^{-1/2} \Gamma(-1/2, x)^{k-2}}{(x^{-1/2} e^{-x} + \gamma(1/2, x))^k} dx$$

for  $k \geq 3$ . The solution for this recursion with  $\lim_{k \rightarrow \infty} C_k^{(2)} = 0$  is exactly (18). Hence (25) follows for  $\beta = 2$ . The sequence  $\mathbf{E}(L_k^{(2)}(n))/n$  converge as  $n \rightarrow \infty$  by Proposition 6, thus (25) together with the Hardy–Littlewood Tauberian theorem implies (26) for  $\beta = 2$ . Case  $\beta = 3$ . With the same notations as before, we have

$$\begin{aligned} p_k^{(3)}(n, m) &= \mathbf{P}(\tau_m = L_k^{(3)}(n), R_n = m) \\ &= \binom{m-1}{k-1} \sum_{j \geq 1} \sum_{l_1=j+1}^{\infty} \cdots \sum_{l_{k-1}=j+1}^{\infty} \sum_{l_k=1}^j \cdots \sum_{l_{m-2}=1}^j \sum_{a=0}^{\infty} \mathbf{P}(\bar{l}^{(3)}, n, m) \end{aligned}$$

and by using (49) for the generating function,

$$(65) \quad \tilde{p}_k^{(3)}(z) = \sum_{n \geq 0} p_k^{(3)}(n) z^n = \tilde{q}(z) \sum_{j \geq 0} \frac{f(j) z^j \left( \sum_{l=j+1}^{\infty} f(l) z^l \right)^{k-1}}{\left( 1 - \sum_{l=1}^j f(l) z^l \right)^k}$$

where  $\tilde{q}(z)$  is given by (13). Again by (13), we rewrite the denominator of (65) as

$$(66) \quad 1 - \sum_{l=1}^j f(l) z^l = \sqrt{1-z} + \sum_{l=j+1}^{\infty} f(l) z^l.$$

Since  $\tilde{q}(e^{-s}) \sim 1/\sqrt{s}$ , it is enough for the proof of (23) to show that

$$(67) \quad \sum_{j \geq 0} \frac{f(j) e^{-sj} \left( \sum_{l=j+1}^{\infty} f(l) e^{-sl} \right)^{k-1}}{\left( \sqrt{1-e^{-s}} - \sum_{l=j+1}^{\infty} f(l) e^{-sl} \right)^k} \sim \ln \frac{1}{\sqrt{s}}$$

as  $s \rightarrow 0$ .

To verify (67), we fix an  $\varepsilon > 0$  small and we divide the sum into a sum for  $j \leq \varepsilon/s$  and a sum for  $j > \varepsilon/s$ . For the sum for  $j > \varepsilon/s$ , we can use the argument which lead to (54) for the  $\beta = 1$  case to derive a uniform exponential tail bound. Hence the  $j > \varepsilon/s$  part of the sum in (67) can be considered as a Riemann sum for the integral

$$(68) \quad \int_{\varepsilon}^{\infty} \frac{x^{-3/2} \Gamma(-1/2, x)^{k-1}}{(2\sqrt{\pi} + \Gamma(-1/2, x))^k} dx.$$

Since  $\Gamma(-1/2, x) = (1 + \mathcal{O}(x))2/\sqrt{x}$  for  $x \rightarrow 0$ , the integrand above is equal to  $(1 + \mathcal{O}(\sqrt{x}))/2x$  for small  $x$ . Therefore, the integral in (68) is asymptotically  $-\frac{1}{2} \ln \varepsilon + \mathcal{O}(1)$  as  $\varepsilon \rightarrow 0$  where the error terms are uniform.

For the  $j \leq \varepsilon/s$  part of the sum, we again rewrite the left-hand side of (67) using (66) and we consider the sum

$$(69) \quad \sum_{j=0}^{\varepsilon/s} \frac{f(j) e^{-sj} \left( 1 - \sqrt{1-e^{-s}} - \sum_{l=1}^j f(l) e^{-sl} \right)^{k-1}}{\left( 1 - \sum_{l=1}^j f(l) e^{-sl} \right)^k}.$$

Next we compare the two series  $\sum_{l=1}^{\infty} f(l) = 1$  and  $\sum_{l=1}^{\infty} f(l) e^{-sl} = 1 - \sqrt{1-e^{-s}}$ . Note that both series have positive entries, and the second one is obtained from the first one by multiplying the summands by numbers from  $(0, 1)$ . Since the two sums differ by  $\sqrt{1-e^{-s}} = \mathcal{O}(\sqrt{s})$ , this remains an upper bound on the difference of their partial sums, i.e.

$$(70) \quad \left| \sum_{l=1}^j f(l) e^{-sl} - \sum_{l=1}^j f(l) \right| \sim \left| \sum_{l=1}^j f(l) e^{-sl} - \left( 1 - \frac{1}{\sqrt{\pi j}} \right) \right| = \mathcal{O}(\sqrt{s}).$$

Using this bound, the sum (69) can be written as

$$(71) \quad \sum_{j=0}^{\varepsilon/s} \frac{\frac{1}{2\sqrt{\pi j^{3/2}}}(1 + \mathcal{O}(\varepsilon)) \left(\frac{1}{\sqrt{\pi j}} + \mathcal{O}(\sqrt{s})\right)^{k-1}}{\left(\frac{1}{\sqrt{\pi j}} + \mathcal{O}(\sqrt{s})\right)^k}$$

where the  $\mathcal{O}(\varepsilon)$  and  $\mathcal{O}(\sqrt{s})$  errors are uniform. Since  $j \leq \varepsilon/s$ , we have  $1/\sqrt{j} \geq \sqrt{s/\varepsilon}$ , hence the  $\mathcal{O}(\sqrt{s})$  errors are negligible in (71) if  $\varepsilon$  is small enough. Therefore, (71) is asymptotically equal to

$$(72) \quad \frac{1}{2}(1 + \mathcal{O}(\varepsilon)) \sum_{j=1}^{\varepsilon/s} \frac{1}{j} \sim \frac{1}{2}(1 + \mathcal{O}(\varepsilon))(\ln \varepsilon - \ln s).$$

We have shown that as  $s \rightarrow 0$ , the left-hand side of (67) is asymptotically equal to the sum of the integral (68) and (72), i.e. it is equal to  $-\frac{1}{2}(1 + \mathcal{O}(\varepsilon)) \ln s + \mathcal{O}(\varepsilon \ln \varepsilon)$ . Since  $\varepsilon$  was arbitrary, (23) is proved.

Following the same computation as in the  $\beta = 2$  case, the distribution function of  $L_k^{(3)}(n)$  restricted to the event  $R_n = m$  is

$$\begin{aligned} F_k^{(3)}(t, n, m) &= F_{k-1}^{(3)}(t, n, m) + \mathbf{P}(L_k^{(3)}(n) \leq t, L_{k-1}^{(3)}(n) > t, R_n = m) \\ &= F_{k-1}^{(3)}(t, n, m) + \binom{m-1}{k-1} \sum_{l_1=t+1}^{\infty} \cdots \sum_{l_{k-1}=t+1}^{\infty} \sum_{l_k=1}^t \cdots \sum_{l_{m-1}=1}^t \sum_{a=0}^{\infty} \mathbf{P}(\bar{l}^{(3)}, n, m) \end{aligned}$$

and the generating function of  $F_k^{(3)}(t, n) = \sum_{m \geq 1} F_k^{(3)}(t, n, m)$  is

$$\begin{aligned} (73) \quad \tilde{F}_k^{(3)}(t, z) &= \sum_{n \geq 0} F_k^{(3)}(t, n) z^n \\ &= \sum_{n \geq 0} \sum_{m \geq 1} F_k^{(3)}(t, n, m) z^n \\ &= \tilde{F}_{k-1}^{(3)}(t, z) + \tilde{q}(z) \frac{\left(\sum_{j=t+1}^{\infty} f(j) z^j\right)^{k-1}}{\left(1 - \sum_{j=1}^t f(j) z^j\right)^k}. \end{aligned}$$

The generating function of  $\mathbf{E}(L_k^{(3)}(n))$  can easily be expressed from the previous equation. For  $k \geq 2$ , we get the asymptotic recursion

$$(74) \quad \sum_{n \geq 0} e^{-sn} \mathbf{E}(L_k^{(3)}(n)) \sim \sum_{n \geq 0} e^{-sn} \mathbf{E}(L_{k-1}^{(3)}(n)) - \frac{1}{s^2} \int_0^{\infty} \frac{\sqrt{\pi}}{2^{k-1}} \frac{\Gamma(-1/2, x)^{k-1}}{(x^{-1/2} e^{-x} + \gamma(1/2, x))^k} dx.$$

As for the case  $\beta = 2$ , it readily implies a recursion for  $C_k^{(3)}$  which is satisfied by the right-hand side (19). Further the limit of  $C_k^{(3)}$  as  $k \rightarrow \infty$  is 0 by Lemma 8, which is shown by a dominated convergence argument for the right-hand side of (19). This finishes the proof.  $\square$

#### 4. CONCLUSION

The asymptotics for the Laplace transform of the probability that the last record is the  $k$ th longest lasting one and for the Laplace transform of the expected proportion of the  $k$ th longest lasting record is understood by Theorem 2. In some of the cases, these could rigorously be turned into the asymptotics for the sequences themselves by specific arguments. The rigorous proof of the asymptotics in other cases however remains open.



## ACKNOWLEDGEMENTS

The authors thank Balázs Ráth for discussions related to this paper, in particular for the remark after Theorem 2. They are grateful for the anonymous referees for their insightful suggestions and comments. This work was supported by OTKA (Hungarian National Research Fund) grant K100473. B. V. is grateful for the Postdoctoral Fellowship of the Hungarian Academy of Sciences and for the Bolyai Research Scholarship.

## REFERENCES

- [1] E. S. Andersen, *On the fluctuations of sums of random variables I*, Math. Scand. **1** (1953), 263–285
- [2] B. Barsak, *Limits of renewal processes and Pitman–Yor distribution*, Electron. Commun. Probab. **20** (2015), no. 51, 1–13
- [3] E. Csáki and Y. Hu, *Lengths and heights of random walk excursions*, Discrete Math. Theor. Comput. Sci. AC (2003), 45–52
- [4] E. Csáki and Y. Hu, *Invariance principles for ranked excursion lengths and heights*, Electron. Comm. Probab. **9** (2004), 14–21
- [5] W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. 2 Wiley, 2. edition, 1971
- [6] S. Finch, *Excursion durations*, unpublished manuscript (2008)
- [7] R. García-García, A. Rosso and G. Schehr, *The longest excursion of fractional Brownian motion: numerical evidence of the non-Markovian effects*, Phys. Rev. E **81** (2010), 010102(R)
- [8] C. Godrèche, *Longest interval between zeros of the tied-down random walk, the Brownian bridge and related renewal processes*, preprint, arXiv: 1611.01434
- [9] C. Godrèche, S. N. Majumdar and G. Schehr, *The longest excursion of stochastic processes in nonequilibrium systems*, Phys. Rev. Lett. **102** (2009), 240602
- [10] C. Godrèche, S. N. Majumdar and G. Schehr, *Universal statistics of longest lasting records of random walks and Lévy flights*, J. Phys. A: Math. Theor. **47** (2014), 255001
- [11] C. Godrèche, S. N. Majumdar and G. Schehr, *Statistics of the longest interval in renewal processes*, J. Stat. Mech. (2015), P03014
- [12] S. N. Majumdar and R. M. Ziff, *Universal record statistics of random walks and Lévy flights*, Phys. Rev. Lett. **101** (2008), 050601
- [13] M. Perman, *Random discrete distributions derived from subordinators*, PhD thesis, University of California, Berkeley, 1990
- [14] M. Perman, *Order statistics for jumps of normalised subordinators*, Stochastic Process. Appl. **46** (1993), 264–281
- [15] M. Perman, J. Pitman and M. Yor, *Size-biased sampling of Poisson point processes and excursions*, Probab. Theory Related Fields **92** (1992), 21–39
- [16] J. Pitman, *Exchangeable and partially exchangeable random partitions*, Probab. Theory Related Fields **102** (1995), 145–158
- [17] J. Pitman, *Partition structures derived from Brownian motion and stable subordinators*, Bernoulli **3** (1997), 79–96
- [18] J. Pitman, *Combinatorial stochastic processes*, Lecture notes for St. Flour course, 2002
- [19] J. Pitman and M. Yor, *Arcsine laws and interval partitions derived from a stable subordinator*, Proc. London Math. Soc. **65** (1992), 326–356
- [20] J. Pitman and M. Yor, *The two-parameter Poisson–Dirichlet distribution derived from a stable subordinator*, Ann. Probab. **25** (1997), 855–900
- [21] D. Revuz and M. Yor, *Continuous martingales and Brownian motion*, 3. edition, Springer, 1999
- [22] C. L. Scheffer, *The rank of the present excursion*, Stochastic Process. Appl. **55** (1995), 101–118

UNIVERSITY OF GRONINGEN, NIJENBORGH 9, 9747 AG GRONINGEN, THE NETHERLANDS  
*E-mail address:* r.szabo@rug.nl

MTA–BME STOCHASTICS RESEARCH GROUP, EGRY J. U. 1, 1111 BUDAPEST, HUNGARY  
*E-mail address:* vetob@math.bme.hu