# *Lengths and heights of random walk excursions*


Endre Csáki[1 †] and Yueyun Hu[2]

[1] *A. Rényi Institute of Mathematics, Hungarian Academy of Sciences, P.O.B. 127, Budapest, H-1364 Hungary*
`csaki@renyi.hu`
[2] *Laboratoire de Probabilités, CNRS-UMR 7599, Université Paris VI, 4 Place Jussieu, F-75252 Paris Cedex 05, France*
`hu@ccr.jussieu.fr`



Consider a simple symmetric random walk on the line. The parts of the random walk between consecutive returns to the origin are called excursions. The heights and lengths of these excursions can be arranged in decreasing order. In this paper we give the exact and limiting distributions of these ranked quantities. These results are analogues of the corresponding results of Pitman and Yor [10, 11, 12] for Brownian motion.



**Keywords:** simple random walk, excursion length, excursion height


## Contents



## 1 Introduction

Let $X_1, X_2, \ldots$ be independent random variables with distribution

$$P(X_i = +1) = P(X_i = -1) = \frac{1}{2}.$$

Put $S_0 = 0$, $S_i = X_1 + \ldots X_i$, $i = 1, 2, \ldots$. Then the sequence $\{S_i\}_{i=0}^{\infty}$ is called a simple symmetric random walk on the line. Consider the return times defined by $\rho_0 = 0$,

$$\rho_i = \min\{k > \rho_{i-1} : S_k = 0\}, \quad i = 1, 2, \ldots$$


[†] Research supported by the Hungarian National Foundation for Scientific Research, Grant Nos. T 037886 and T 043037.

Further, let

$$\xi(n) = \#\{k:\, 0 < k \le n, S_k = 0\}$$

be the local time of the random walk at zero, i.e. the number of returns to the origin up to time $n$.

The parts $(S_{\rho_{i-1}}, \dots, S_{\rho_i - 1})$, $\quad i = 1, 2, \dots$ between consecutive returns are called excursions. Consider the lengths $\tau_i = \rho_i - \rho_{i-1}$, and heights

$$\mu_i = \max_{\rho_{i-1} \le k \le \rho_i - 1} |S_k|, \qquad \mu_i^+ = \max_{\rho_{i-1} \le k \le \rho_i - 1} S_k$$

of $i$-th excursion.

Clearly, the random walk does not change sign within an excursion. We may call the excursion positive (negative) if the random walk assumes positive (negative) values within this excursion. If the $i$-th excursion is negative, then $\mu_i^+ = 0$.

In this paper we consider the ranked lengths and heights of excursions up to time $n$. In general, however the (fixed) time $n$ need not be an excursion endpoint, and we include the length and height of this last, possibly incomplete, excursion as well. Consider the sequences

$$L^{(1)}(n) \ge L^{(2)}(n) \ge \dots, \qquad M^{(1)}(n) \ge M^{(2)}(n) \ge \dots, \qquad \text{and} \qquad M_+^{(1)}(n) \ge M_+^{(2)}(n) \ge \dots,$$

where $L^{(j)}(n)$ is the $j$-th largest in the sequence

$$(\tau_1, \tau_2, \dots, \tau_{\xi(n)}, n - \rho_{\xi(n)}),$$

$M^{(j)}(n)$ is the $j$-th largest in the sequence

$$(\mu_1, \mu_2, \dots, \mu_{\xi(n)}, \max_{\rho_{\xi(n)} \le k \le n} |S_k|),$$

while $M_+^{(j)}(n)$ is the $j$-th largest in the sequence

$$(\mu_1^+, \mu_2^+, \dots, \mu_{\xi(n)}^+, \max_{\rho_{\xi(n)} \le k \le n} S_k).$$

We define $M_+^{(j)}(n) = M^{(j)}(n) = L^{(j)}(n) = 0$ if $j > \xi(n) + 1$.

Similar quantities for Brownian motion were investigated by Knight [9], Pitman and Yor [10, 11, 12], Wendel [13] and their strong limit properties were studied in [7, 8], [4, 5]. In [3] the properties of $L^{(1)}$ were investigated.

In this paper we present exact and limiting distribution results for ranked lengths and heights of random walk excursions.

## 2 Exact distributions

First we give the exact distributions for $M_+$. In the sequel the binomial coefficient $\binom{n}{k}$ is meant to be zero if $k > n$, $k < 0$ or $k$ is not an integer.

**Theorem 2.1.** *For $a > 0$, $\ell$ integers, we have*

$$P(M_+^{(j)}(n) \geq a, S_n = \ell) = P(S_n = 2ja - \ell) = \frac{1}{2^n}\binom{n}{ja + \frac{n-\ell}{2}}, \qquad \ell < a \tag{2.1}$$

$$P(M_+^{(j)}(n) \geq a, S_n = \ell) = P(S_n = 2(j-1)a + \ell) = \frac{1}{2^n}\binom{n}{(j-1)a + \frac{n+\ell}{2}}, \qquad \ell \geq a \tag{2.2}$$

*and*

$$\begin{aligned} P(M_+^{(j)}(n) \geq a) &= P(S_n = (2j-1)a) + 2P(S_n > (2j-1)a) \\ &= \frac{1}{2^n}\left(\binom{n}{ja + \frac{n-a}{2}} + 2\sum_{\ell > (2j-1)a}\binom{n}{\frac{n+\ell}{2}}\right). \end{aligned} \tag{2.3}$$

*Proof.* These results are well-known for $j = 1$, their usual proofs being based on the reflection principle of André [1]. Our present proofs are also based on the reflection principle, applied successively. Now consider the case $j = 2$. The event $\{M_+^{(2)}(n) \geq a\}$ means that there are at least two upcrossings from 0 to $a$ by the random walk up to time $n$. Let $\tau_1$ be the first hitting time of the level $a$, $\theta_1$ be the first return time to zero after $\tau_1$ and let $\tau_2$ be the first hitting time of $a$ after $\theta_1$. Reflect the part between $\tau_1$ and $\theta_1$ through $a$, then $\theta_1$ becomes the first hitting time of the level $2a$ for the new path. Put the part between $\theta_1$ and $n$ continuously to the new path. If $\ell \geq a$, we are done, and reversing this procedure, we can see that there is a bijection between the set of paths $\{M_+^{(2)}(n) \geq a, S_n = \ell\}$ and $\{S_n = 2a + \ell\}$. This proves (2.2) for $j = 2$. If $\ell < a$, then $\tau_2$ becomes the first hitting time of $3a$ for the new path. Reflect the part between $\tau_2$ and $n$ through $3a$, we get a path with $S_n = 4a - \ell$. By reversing this procedure, we can see that there is a bijection between $\{M_+^{(2)}(n) \geq a, S_n = \ell\}$ and $\{S_n = 4a - \ell\}$, proving (2.1) for $j = 2$. By repeating the above procedure, we get (2.1) and (2.2) for $j \geq 3$ as well. Now (2.3) follows from (2.1) and (2.2) by summing up for $\ell$. □

Concerning the distributions of $M$ we present

**Theorem 2.2.** *For $a > 0$, integer we have*

$$\begin{aligned} P(M^{(j)}(n) \geq a, S_n = 0) &= 2^{j-n}\sum_{k=0}^{\infty}(-1)^k\binom{k+j-1}{k}P(S_n = 2(k+j)a) \\ &= 2^{j-n}\sum_{k=0}^{\infty}(-1)^k\binom{k+j-1}{k}\binom{n}{(k+j)a + \frac{n}{2}} \end{aligned} \tag{2.4}$$

*and*

$$\begin{aligned} &P(M^{(j)}(n) \geq a) \\ &= 2^{j-n}\sum_{k=0}^{\infty}(-1)^k\binom{k+j-1}{k}\left(P(S_n = (2k+2j-1)a) + 2P(S_n > (2k+2j-1)a)\right) \\ &= 2^{j-n}\sum_{k=0}^{\infty}(-1)^k\binom{k+j-1}{k}\left(\binom{n}{(j+k)a + \frac{n-a}{2}} + 2\sum_{\ell > (2k+2j-1)a}\binom{n}{\frac{n+\ell}{2}}\right). \end{aligned} \tag{2.5}$$

*Proof.* We follow the proof presented in Pitman and Yor [12] for the case of Brownian bridge. In fact, we show that Theorem 2.2 is a consequence of Theorem 2.1.

Let $\{\varepsilon_i\}_{i=1}^{\infty}$ be a sequence of i.i.d. random variables with $P(\varepsilon_i = 0) = P(\varepsilon_i = 1) = 1/2$, and let $g_j$ denote the index of the $j$-th 1 in the above sequence. It is well-known that $g_j$ has negative binomial distribution

$$P(g_j = k + j) = \frac{1}{2^{k+j}}\binom{k+j-1}{k}, \qquad k = 0, 1, \ldots$$

Since heights and signs of excursions are independent, we have

$$\left\{M_+^{(j)}(n)\right\}_{j=1}^{\infty} \stackrel{\mathcal{D}}{=} \left\{M^{(g_j)}(n)\right\}_{j=1}^{\infty},$$

where $\stackrel{\mathcal{D}}{=}$ denotes equality in distribution and the sequences $\{\varepsilon_i\}_{i=1}^{\infty}$ and $\{M^{(j)}(n)\}_{j=1}^{\infty}$ are assumed to be independent. Hence

$$Ef\left(M_+^{(j)}(n)\right) = Ef\left(M^{(g_j)}(n)\right) = \sum_{k=0}^{\infty} \frac{1}{2^{k+j}}\binom{k+j-1}{k} Ef\left(M^{(k+j)}(n)\right) \tag{2.6}$$

and using Lemma 9 in Pitman and Yor [12] one can invert (2.6) as

$$Ef\left(M^{(j)}(n)\right) = 2^j \sum_{k=0}^{\infty} (-1)^k \binom{k+j-1}{k} Ef\left(M_+^{(k+j)}(n)\right)$$

On choosing $f(x) = I\{x \geq a\}$, where $I$ denotes indicator of the event in the brackets, we get (2.5) from (2.3). The proof of (2.4) is similar. □

Concerning the lengths $L^{(j)}(n)$ no explicit formula is known for their distributions. In [3] a recurrence formula is given for $j = 1$ as

$$P(L^{(1)}(n) \leq z) = \sum_{k=1}^{z} b_k P(L^{(1)}(n-k) \leq z), \tag{2.7}$$

where

$$b_k = P(\rho_1 = k) = \frac{1}{k2^{k-2}}\binom{k-2}{\frac{k-2}{2}}$$

is the probability that the first return time to zero is equal to $k$. (2.7) is equivalent to

$$P(L^{(1)}(n) > z) = \sum_{k=1}^{z} b_k P(L^{(1)}(n-k) > z) + \sum_{k=z+1}^{\infty} b_k \tag{2.8}$$

for $z < n$ and $P(L^{(1)}(n) > z) = 0$ if $z \geq n$.

This can be generalized as follows.

**Theorem 2.3.** *For $j > 1$, $n \geq 2j - 1$, $z \geq 1$ integers we have*

$$P(L^{(j)}(n) > z) = \sum_{k=1}^{z} b_k P(L^{(j)}(n-k) > z) + \sum_{k=z+1}^{n} b_k P(L^{(j-1)}(n-k) > z). \tag{2.9}$$

*Proof.* The proof of (2.9) is straightforward by considering that the first return to zero occurs at time $k$. If $k > z$, it means that the length of the first excursion is larger than $z$, so in the remaining part we need only $j-1$ excursions whose lengths are greater than $z$. □

The generating function of the above distributions can be given as follows. Let $0 < u < 1$ and define

$$\phi_j(u,z) = \sum_{n=1}^{\infty} u^n P(L^{(j)}(n) > z).$$

By recurrence formula (2.9), we obtain that for $j > 1$,

$$\begin{aligned}
\phi_j(u,z) &= \sum_{k=1}^{z} \sum_{n=k+z}^{\infty} b_k u^n P(L^{(j)}(n-k) > z) + \sum_{k=z+1}^{\infty} b_k \sum_{n=k+z}^{\infty} u^n P(L^{(j-1)}(n-k) > z) \\
&= \sum_{k=1}^{z} b_k u^k \sum_{m=z}^{\infty} u^m P(L^{(j)}(m) > z) + \sum_{k=z+1}^{\infty} b_k u^k \sum_{m=z}^{\infty} u^m P(L^{(j-1)}(m) > z) \\
&= \sum_{k=1}^{z} b_k u^k \phi_j(u,z) + \sum_{k=z+1}^{\infty} b_k u^k \phi_{j-1}(u,z),
\end{aligned}$$

since $P(L^{(j-1)}(m) > z) = 0$ when $m < z$. For $j = 1$, we get from (2.8) that

$$\phi_1(u,z) = \sum_{k=1}^{z} b_k u^k \phi_1(u,z) + \sum_{k=z+1}^{\infty} b_k \frac{u^{z+1}}{1-u}.$$

It follows that

$$\phi_1(u,z) = \frac{u^{z+1} \sum_{k=z+1}^{\infty} b_k}{(1-u)\left(1 - \sum_{k=1}^{z} b_k u^k\right)},$$

and for $j \geq 2$,

$$\phi_j(u,z) = \left( \frac{\sum_{k=z+1}^{\infty} b_k u^k}{1 - \sum_{k=1}^{z} b_k u^k} \right)^{j-1} \phi_1(u,z) = \frac{\left(\sum_{k=z+1}^{\infty} b_k u^k\right)^{j-1} u^{z+1} \sum_{k=z+1}^{\infty} b_k}{(1-u)\left(1 - \sum_{k=1}^{z} b_k u^k\right)^j}. \tag{2.10}$$

We note the following interpretation of the generating function $\phi_j$. Let $N$ be a random variable, independent of our random walk, having geometric distribution with parameter $1-u$, i.e.

$$P(N = n) = (1-u)u^{n-1}, \quad n = 1, 2, \ldots \tag{2.11}$$

Then

$$P(L^{(j)}(N) > z) = \frac{1-u}{u} \phi_j(u,z). \tag{2.12}$$

Some joint distributions can also be given. Consider e.g. the joint distribution of $M^{(1)}(n)$ and $L^{(1)}(n)$.

**Theorem 2.4.**

$$P(M^{(1)}(n) < a, L^{(1)}(n) \le z) = \sum_{k=1}^{z} c_k P(M^{(1)}(n-k) < a, L^{(1)}(n-k) \le z), \tag{2.13}$$

*where*

$$c_k = \frac{1}{2^{k-1}} \sum_{i=-\infty}^{\infty} \left( \binom{k-2}{\frac{k-2}{2}+ia} - \binom{k-2}{\frac{k}{2}+ia} \right).$$

*Proof.* Consider again that the first return time to zero is equal to $k$. The probability of the event

$$\{S_0 = 0, 0 < |S_i| < a, i = 1, 2, \ldots, k-1, S_k = 0\}$$

is exactly $c_k$ (cf., e.g. [6]). For the remaining path we need also that $M^{(1)} < a, L^{(1)} \le z$. Hence we have (2.13). $\square$

# 3 Limiting distributions

Now one can obtain the limiting distributions for the ranked heights from the exact formulae of Section 2 by the usual procedure: Stirling formula, etc. based on the asymptotic results

$$\lim_{n\to\infty} P(S_n > y\sqrt{n}) = 1 - \Phi(y)$$

and

$$\binom{2n}{n + [y\sqrt{2n}]} \sim \frac{2^{2n}}{\sqrt{\pi n}} e^{-2y^2}, \qquad n \to \infty, \tag{3.1}$$

where $\Phi(\cdot)$ is the standard normal distribution function.

**Theorem 3.1.**

$$\lim_{n\to\infty} P(M_+^{(j)}(n) \ge y\sqrt{n}) = 2(1 - \Phi((2j-1)y)) \tag{3.2}$$

*and*

$$\lim_{n\to\infty} P(M_+^{(j)}(2n) \ge y\sqrt{2n} \,|\, S_{2n} = 0) = e^{-2j^2y^2}. \tag{3.3}$$

**Theorem 3.2.**

$$\lim_{n\to\infty} P(M^{(j)}(n) \ge y\sqrt{n}) = 2^{j+1} \sum_{k=0}^{\infty} (-1)^k \binom{k+j-1}{k} (1 - \Phi((2k+2j-1)y)) \tag{3.4}$$

*and*

$$\lim_{n\to\infty} P(M^{(j)}(2n) \ge y\sqrt{2n} \,|\, S_{2n} = 0) = 2^j \sum_{k=0}^{\infty} (-1)^k \binom{k+j-1}{k} e^{-2(k+j)^2y^2}. \tag{3.5}$$

These limiting distributions are the same as the corresponding distributions in the case of Brownian motion. For (3.3) and (3.5) see [12] and for (3.2) see [5].

The limiting distribution of $L^{(j)}$ can be obtained in the following form.

**Theorem 3.3.**
$$\lim_{n\to\infty} P(L^{(j)}(n) < xn) = P(V_j < x),$$
*where $V_j$ is a random variable with*
$$E(\exp(-\lambda/V_j)) = \frac{e^{-\lambda}(G(\lambda))^{j-1}}{\left(\sqrt{\pi\lambda}+G(\lambda)\right)^j}, \tag{3.6}$$
*where*
$$G(\lambda) = \frac{1}{2}\int_1^\infty \frac{e^{-\lambda y}}{y^{3/2}}\,dy.$$

**Remark:** (3.6) is given in Wendel [13] and Pitman and Yor [10], $V_j$ being the ranked excursion length of Brownian motion.

*Proof.* Let $N = N(z)$ be geometric random variable as in (2.11) with $u = \exp(-\lambda/z)$. First we show that
$$\lim_{z\to\infty} P(L^{(j)}(N) > z) = \frac{e^{-\lambda}(G(\lambda))^{j-1}}{\left(\sqrt{\pi\lambda}+G(\lambda)\right)^j}. \tag{3.7}$$

Using the asymptotic relation (3.1), we can see
$$b_{2r} \sim \frac{1}{2\sqrt{\pi}r^{3/2}}, \quad r\to\infty, \qquad \text{and} \qquad \sum_{k=z+1}^\infty b_k \sim \sqrt{\frac{2}{\pi z}}, \quad z\to\infty,$$
hence
$$\sum_{k=z+1}^\infty b_k \exp(-\lambda k/z) \sim \frac{1}{2\sqrt{\pi}}\sum_{2r>z+1}\frac{1}{r^{3/2}}\exp(-2\lambda r/z) \sim \frac{1}{2\sqrt{\pi}}\int_{z/2}^\infty \frac{\exp(-2\lambda v/z)}{v^{3/2}}\,dv = \sqrt{\frac{2}{\pi z}}G(\lambda),$$
as $z\to\infty$. Moreover, using the well-known generating function for $b_k$, we get
$$1-\sum_{k=1}^z b_k u^k = 1-\sum_{k=1}^\infty b_k u^k + \sum_{k=z+1}^\infty b_k u^k = \sqrt{1-u^2}+\sum_{k=z+1}^\infty b_k u^k \sim \sqrt{\frac{2\lambda}{z}}+\sqrt{\frac{2}{\pi z}}G(\lambda).$$
Now (3.7) follows by using (2.10), (2.12) and the above asymptotic relations.

Now let $H^{(j)}$ be the inverse of $L^{(j)}$, i.e. $H^{(j)}(z) = \inf\{n : L^{(j)}(n) > z\}$. It is then easy to see that
$$E\left(\exp\left(-\frac{\lambda H^{(j)}(z)}{z}\right)\right) = \exp(-\lambda/z)P(L^{(j)}(N) > z),$$
hence by (3.7) we conclude that the limiting distribution of $H^{(j)}(z)/z$ is the distribution of $1/V_j$, which in turn implies that the limiting distribution of $L^{(j)}(n)/n$ is the distribution of $V_j$. $\square$

## Acknowledgements

The authors are indebted to the referees for their useful remarks. Cooperation between the authors was supported by the joint French-Hungarian Intergovernmental Grant "Balaton" no. F-39/00.

## References

[1] André, D., Solution directe du problème résolu par M. Bertrand, *C. R. Acad. Sci. Paris* **105** (1887), 436-437.

[2] K.L. Chung and P. Erdős, On the application of the Borel-Cantelli lemma, *Trans. Amer. Math. Soc.* **72** (1952), 179–186.

[3] E. Csáki, P. Erdős and P. Révész, On the length of the longest excursion, *Z. Wahrsch. verw. Gebiete* **68** (1985), 365–382.

[4] E. Csáki and Y. Hu, Asymptotic properties of ranked heights in Brownian excursions, *J. Theoret. Probab.* **14**, (2001), 77–96.

[5] E. Csáki and Y. Hu, On the ranked excursions heights of a Kiefer process, preprint.

[6] E. Csáki and S.G. Mohanty, Excursion and meander in random walk, *Canad. J. Statist.* **9** (1981), 57–70.

[7] Y. Hu and Z. Shi, Extreme lengths in Brownian and Bessel excursions, *Bernoulli* **3** (1997), 387–402.

[8] Y. Hu and Z. Shi, Shortest excursion lengths, *Ann. Inst. H. Poincaré Probab. Statist.* **35** (1999), 103–120.

[9] F.B. Knight, On the duration of the longest excursion, *Seminar on stochastic processes, 1985 (Gainesville, Fla., 1985), Progr. Probab. Statist.*, **12**, Birkhuser Boston, Boston, MA, 1986, pp. 117–147.

[10] J. Pitman and M. Yor, The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator, *Ann. Probab.* **25** (1997), 855-900.

[11] J. Pitman and M. Yor, Ranked functionals of Brownian excursions, *C. R. Acad. Sci. Paris Ser. I* **326** (1998), 93–97.

[12] J. Pitman and M. Yor, On the distribution of ranked heights of excursions of a Brownian bridge, *Ann. Probab.* **29** (2001), 361–384.

[13] J.G. Wendel, Zero-free intervals of semi-stable Markov processes, *Math. Scand.* **14** (1964), 21–34.