On the asymptotic behavior of the length of the longest increasing subsequences of random walks

J. Ricardo G. Mendonça,[1, 2, *] Hendrik Schawe,[3, †] and Alexander K. Hartmann[3, ‡]

[1] *Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, 03828-000 São Paulo, SP, Brazil*
[2] *LPTMS, CNRS, Université Paris-Sud, Université Paris-Saclay, 91405 Orsay CEDEX, France*
[3] *Institut für Physik, Universität Oldenburg, 26111 Oldenburg, Germany*

We numerically estimate the leading asymptotic behavior of the length L_n of the longest increasing subsequence of random walks with step increments following Student's t-distribution with parameter in the range $\frac{1}{2} \leq \nu \leq 5$. We found that the expected value $\mathbb{E}(L_n) \sim n^{\theta}$ with θ decreasing from $\theta(\nu = \frac{1}{2}) \sim 0.70$ to $\theta(\nu \geq 3) \sim 0.57$. For random walks with distribution of step increments of finite variance ($\nu > 2$) we found that $\mathbb{E}(L_n) \sim \sqrt{n}\ln n$ to leading order, confirming previous observations. We note that this asymptotic behavior (including the subleading term) resembles that of the largest part of random integer partitions under the uniform measure and that, curiously, both random variables seem to follow Gumbel statistics. We also provide more refined estimates for the asymptotic behavior of $\mathbb{E}(L_n)$ for random walks with step increments of finite variance.

I. INTRODUCTION

The longest increasing subsequence (LIS) problem is to find an increasing subsequence of maximum length of a given finite sequence of n elements taken from a partially ordered set. Let $\mathscr{X}_n = (X_1, \dots, X_n)$ be such a sequence—say, of real numbers. The longest (weakly) increasing subsequence of $\mathscr{X}_n$ is the longest subsequence $X_{i_1} \leq X_{i_2} \leq \cdots \leq X_{i_L}$ of $\mathscr{X}_n$ such that $1 \leq i_1 < i_2 < \cdots < i_L \leq n$, with L the length of the LIS. There may be more than one "longest" increasing subsequence for a given $\mathscr{X}_n$, with different elements but of the same maximum length. Algorithmically, it takes $O(n \log\log n)$ time to find one LIS of a given sequence of n elements [1].

The classical LIS problem is that of determining the LIS of a random permutation. The problem seems to have been first considered by Stanislaw Ulam in the early 1960s [2]. The resolution of the LIS problem for random permutations culminated with the exact characterization of L_n as a random variable distributed like $L_n \sim 2\sqrt{n} + \sqrt[6]{n}\,\chi$ with $\mathbb{P}(\chi \leq s) = F_2(s)$, the Tracy-Widom distribution for the fluctuations of the largest eigenvalue of a Gaussian unitary random matrix ensemble about its soft edge [3, 4]. Comprehensive expositions on the LIS problem for random permutations appear in [5, 6].

Recently, another version of the LIS problem has been posed [7, 8]: what is the behavior of the LIS of a random walk? Let $\mathscr{X}_n = (X_1, \dots, X_n)$ be the sequence of terms of a random walk given by

$$X_0 = 0, \quad X_t = X_{t-1} + \xi_t, \quad t = 1, \dots, n, \tag{1}$$

with the ξ_t, $t = 1, \dots, n$, independent random variables identically distributed according to some zero-mean, symmetric probability distribution $\phi(\xi)$. The sequence $\mathscr{X}_n$ constitutes a time-series of correlated random variables—if the expectation $\mathbb{E}(\xi^2) \neq 0$, then $\mathbb{E}(X_t X_s) \neq 0$. In [7], the authors showed that when $\phi(\xi)$ has finite positive variance, then for all $\varepsilon > 0$ and large enough n the length L_n of the LIS of $\mathscr{X}_n$ observes

$$c\sqrt{n} \leq \mathbb{E}(L_n) \leq n^{\frac{1}{2}+\varepsilon} \tag{2}$$

for some positive constant c. The upper bound in Eq. (2) does not rule out a logarithmic term, and can actually be read like

$$\mathbb{E}(L_n) \leq \sqrt{n}\,(\ln n)^a \tag{3}$$

for some $a \geq 0$. In [8], the authors further proved that the expected length of the LIS of a particular random walk with step lengths of ultra-heavy distribution without any finite (integer or fractional) moment scales with the length of the walk as

$$n^{\beta_0 - o(1)} \leq \mathbb{E}(L_n) \leq n^{\beta_1 + o(1)}, \tag{4}$$

with non-sharp $\beta_0 \simeq 0.690$ and $\beta_1 \simeq 0.815$. Besides the bounds (2)–(4), little is known rigorously about the LIS of random walks. Figure 1 displays two random walks of 300 steps distributed according to a Student's t-distribution (see Sec. II), one with parameter $\nu = 1$ (the same as the Cauchy distribution) and the other with parameter $\nu = 4$, together with one of their LIS each.

In order to improve our knowledge about the LIS of random walks, we ran Monte Carlo simulations to estimate the scaling of L_n for several different distributions of step lengths [9]. The simulations showed that

$$\mathbb{E}(L_n) \sim n^{\theta} \tag{5}$$

with a non-universal scaling exponent $0.60 \lesssim \theta \lesssim 0.69$ for the heavy-tailed distributions of step lengths examined, with θ increasing as the distribution of step lengths becomes more heavy-tailed. For distributions of finite variance, assuming the validity of Eq. (5), we found a value $\theta \simeq 0.57$, irrespective of the particular distribution. A closer look into the data for random walks with step lengths of finite variance led us to conjecture that the asymptotic behavior of $\mathbb{E}(L_n)$ in these cases is given by

$$\mathbb{E}(L_n) \sim \frac{1}{e}\sqrt{n}\,\ln n + \frac{1}{2}\sqrt{n} \tag{6}$$

plus lower order terms, although the constants are presumed based on least squares adjustements. Moreover, we found that the empirical distribution of L_n seems to be of the form

$$f(L_n) = \frac{1}{\mathbb{E}(L_n)}\, g\!\left(\frac{L_n}{\mathbb{E}(L_n)}\right), \tag{7}$$

* Email: jricardo@usp.br
† Email: hendrik.schawe@uni-oldenburg.de
‡ Email: a.hartmann@uni-oldenburg.de

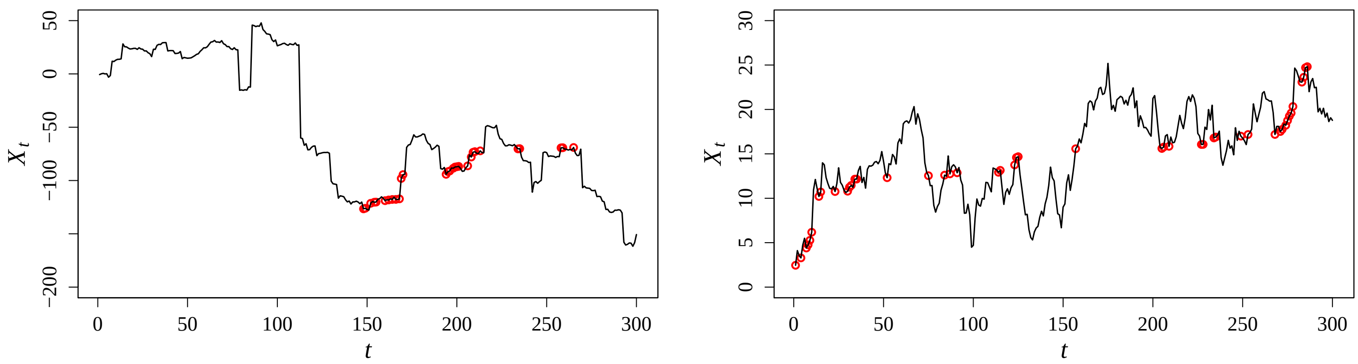

FIG. 1. Student's t-random walks of 300 steps each with parameters $\nu = 1$ (left panel) and $\nu = 4$ (right panel) together with one of their LIS (small circles). Note the different vertical scales.

with $\mathbb{E}(L_n)$ given by Eq. (5) or (6) depending on whether $\phi(\xi)$ has finite second moment or not. Accordingly, when the step lengths are of finite variance $g(z)$ should be universal. Plots of $g(z)$ for different distributions of step lengths appear in Ref. [9]; see also Fig. 4. The form (7) has been further tested in Ref. [10], that probed the distribution of L_n into regions of very small probabilities for random walks with uniform increments $\xi \sim U(-1,1)$. The authors found very good agreement between Eqs. (6)–(7) and their data. They also estimated that the large deviation rate function $\Phi(L)$ associated with the distribution of L_n by

$$f(L_n) \asymp \exp(-n\Phi(L_n)) \tag{8}$$

behaves asymptotically, in the limit of large $n \to \infty$, like $\Phi(L) \sim L^{-1.6}$ in the left tail and $\Phi(L) \sim L^{2.9}$ in the right tail. Despite this characterization, the distribution $g(z)$ remains unknown. It is tempting to conjecture that the actual exponents in $\Phi(L)$ are, respectively, $L^{-3/2}$ and L^3, in which case they would be the same as those of the large deviation rate function of the Tracy-Widom F_2 distribution with the sides (left $\leftrightarrow$ right) and signs flipped [11]. Note, however, that unlike in Eq. (8), the large deviation rate function for the Tracy-Widom distribution in the right tail is defined by a relationship of the form $\exp(-\sqrt{n}\Phi(L))$, i. e., with an unusual $\sqrt{n}$ scaling.

In this paper we provide an updated account on the LIS problem for random walks. While in previous studies of the problem only specific distributions $\phi(\xi)$ of step increments were considered, in Section II we employ a parametrized distribution, namely, the Student's t-distribution, that allows us to investigate the dependence of the scaling exponent θ in (5) with the heavy-tail index of $\phi(\xi)$. In Sec. III, we further verify the proposed scaling form Eqs. (6)–(7) with new and independent data, this time taking the full distribution of the data into account. We also set down some remarks on the resemblance between the statistics of the LIS problem for random walks of finite variance and the random partition problem under the uniform measure. Section IV concludes the paper with some perspectives for further study along the lines explored here.

II. LIS OF HEAVY-TAILED RANDOM WALKS

We investigate the behavior of the scaling exponent appearing in the relation $L_n \sim n^\theta$ for random walks with heavy-tailed distribution of step increments as a function of their characteristic index α, defined by

$$\phi_\alpha(|\xi| \gg 1) \sim |\xi|^{-1-\alpha}. \tag{9}$$

We want to check whether there exists a well defined relationship between θ and α. In order to access a range of values of $\alpha \leq 2$ (such that $\mathbb{E}(\xi^2) = \infty$), we employ Student's t-distribution [12]

$$\phi_\nu(\xi) = \frac{\Gamma\left[\frac{1}{2}(\nu+1)\right]}{\sqrt{\nu\pi}\,\Gamma\left(\frac{1}{2}\nu\right)}\left(1+\frac{\xi^2}{\nu}\right)^{-\frac{1}{2}(\nu+1)}, \tag{10}$$

where $\Gamma(z)$ is the usual gamma function and ν is a parameter. This well-known distribution appears in inference problems about unknown parameters (mean or variance or both) of a normal population. In statistical applications $\nu \geq 1$ is a natural number, but for general modeling purposes ν can be taken a real positive number. When $\nu < \infty$, Student's t-distribution displays a heavy tail $\phi_\nu(|\xi| \gg 1) \sim |\xi|^{-1-\nu}$, with infinite variance if $\nu \leq 2$ and finite variance $\nu/(\nu-2)$ for $\nu > 2$. We see that ν plays the role of the tail index α in Eq. (9). Student's t-distribution becomes the Gaussian distribution in the limit $\nu \to \infty$.

For each parameter ν and walk length n, we generate 10^4 realizations of $\mathscr{X}_n$, compute their L_n and estimate the empirical average $\langle L_n \rangle$ and variance $\langle L_n^2 \rangle - \langle L_n \rangle^2$. In our simulations $10^4 \leq n \leq 10^7$ and $\frac{1}{2} \leq \nu \leq 5$. Whenever $\mathbb{E}(\xi^2)$ is finite (i. e., $\nu > 2$), we use normalized random variables $\xi/\sqrt{\mathbb{E}(\xi^2)}$ for the step increments. We found that these quantities scale like

$$\langle L_n \rangle \sim n^\theta \quad \text{and} \quad \langle L_n^2 \rangle - \langle L_n \rangle^2 \sim n^\gamma \tag{11}$$

over the three decades range of n investigated; data from [9, 10] indicate that Eq. (11) actually holds over much larger intervals. Figure 2 displays log-log plots of $\langle L_n \rangle$ and $\langle L_n^2 \rangle - \langle L_n \rangle^2$ for $\nu = \frac{2}{3}$ for illustration. Least-squares fits provide estimates for θ and γ; see Table I. In all cases we obtained

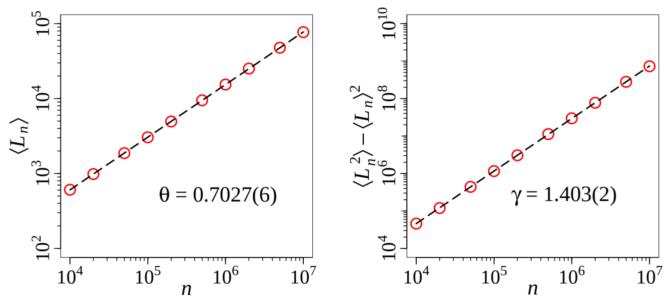

FIG. 2. Log-log plot of the empirical mean (left panel) and empirical variance (right panel) of L_n for the Student's t-random walk with parameter $\nu = 2/3$ together with the least-squares fits (dashed lines). The fact that the curves have virtually the same slope in the different vertical scales on the graphs suggests that $\gamma \simeq 2\theta$. Each point was obtained from an average over 10^4 sample random walks.

$\gamma \simeq 2\theta$ to a very good precision, in agreement with the picture provided by Fig. 2, suggesting that the p. d. f. of L_n indeed follows the form (7), since then the kth moment of L_n becomes

$$\mathbb{E}(L_n^k) = \int L_n^k f(L_n)\, dL_n = = \mathbb{E}(L_n)^k \int z_n^k g(z_n)\, dz_n = c_{n,k}\,\mathbb{E}(L_n)^k, \quad (12)$$

with $z_n = L_n/\mathbb{E}(L_n)$ and $c_{n,k}$ the kth moment of the distribution $g(z_n) = \mathbb{E}(L_n) f(\mathbb{E}(L_n)\, z_n)$. We see that, with $f(L_n)$ like in Eq. (7), all moments $\mathbb{E}(L_n^k) \propto \mathbb{E}(L_n)^k$, as our data for $k = 1$ and 2 do. Note that, as a consequence,

$$\mathrm{Var}(L_n) = \mathbb{E}(L_n^2) - \mathbb{E}(L_n)^2 = (c_{n,2} - c_{n,1}^2)\,\mathbb{E}(L_n)^2, \quad (13)$$

and the random variable L_n cannot possibly be self-averaging unless $(c_{n,2} - c_{n,1}^2) \xrightarrow{n} 0$, i. e., unless $g(z_n)$ becomes increasingly more concentrated with n,

$$g(z_n) \xrightarrow{n} \delta(z - c_1). \quad (14)$$

Our data indicate, however, that $g(z_n)$ remains broad irrespective of how large n gets.

Figure 3 displays a log-log plot of θ against ν. The plot does not suggest any clear functional relationship between θ and ν—we were hoping for something like $\theta \sim \nu^z$ in the interval $\nu \le 2$. Otherwise, θ saturates at $\theta \simeq 0.57$ for distributions of step lengths of finite variance ($\nu > 2$), with a "transient" behavior in the interval $2 < \nu \le 3$ that we attribute to the finite length of the random walks ($n \le 10^7$ steps). If we repeat the analysis of Ref. [9] for the Student's t-random walks, we obtain that for random walks with step lengths of infinite variance L_n follows (5) with a nonuniversal exponent θ, as we can see from Table I, while for random walks with step lengths of finite variance L_n follows (6), confirming previous results employing other distributions [9, 10]. In Section III B, however, we revisit the estimation of the constants appearing in (6).

Figure 4 displays data collapse for our LIS data for some selected ν employing expressions (5) or (6), depending whether $\nu < 2$ or $\nu \ge 2$, respectively. The curve resulting from the data collapse corresponds to the empirical distribution $g(z)$ in Eq. (7). We see very good data collapse, all virtually with the same form for $g(z_n)$. We also see that it is definitely not the case that $g(z_n) \xrightarrow{n} \delta(z)$ (cf. Eqs. (12)–(14)).

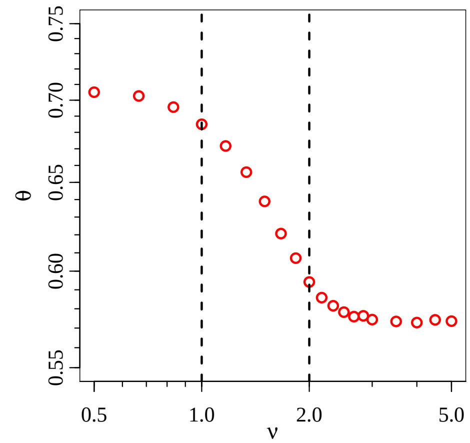

FIG. 3. Log-log plot of the scaling exponent θ against the tail index ν. The dashed vertical lines delimit the intervals in which $\phi_\nu(\xi)$ (Eq. (10)) does not have finite integer moments ($\nu \le 1$), has only finite mean ($1 < \nu \le 2$), and has finite mean and variance ($\nu > 2$).

III. REMARKS ON THE LIS OF RANDOM WALKS OF FINITE VARIANCE

A. A wishful (but unlikely) connection with integer partitions

Recall that a partition of a natural number n is a sequence of integers $\lambda_1 \ge \cdots \ge \lambda_k > 0$ such that $\lambda_1 + \cdots + \lambda_k = n$. For example, $(5,4,3)$ and $(4,4,2,1,1)$ are two partitions of $n = 12$. No closed-form expression for the number $p(n)$ of partitions of n is known. Asymptotically, for $n \to \infty$ we have the Hardy-Ramanujan formula $p(n) \sim \exp(\pi\sqrt{2n/3})/4\sqrt{3}\,n$ [13].

TABLE I. Exponents θ and γ according to (11) for selected values of tail index ν. The ratio $\gamma/\theta \simeq 2$ suggests form (7) for the p. d. f. of L_n. The numbers between parentheses indicate the uncertainty in the last digit of the data.

ν	θ	γ	γ/θ
1/2	0.7051(6)	1.419(2)	2.01
1	0.6850(5)	1.372(3)	2.00
3/2	0.639(1)	1.282(3)	2.01
2	0.594(2)	1.198(5)	2.01
5/2	0.578(2)	1.161(6)	2.01
3	0.574(2)	1.151(5)	2.00
7/2	0.573(2)	1.151(5)	2.01
4	0.573(2)	1.148(5)	2.00
5	0.574(2)	1.151(5)	2.01

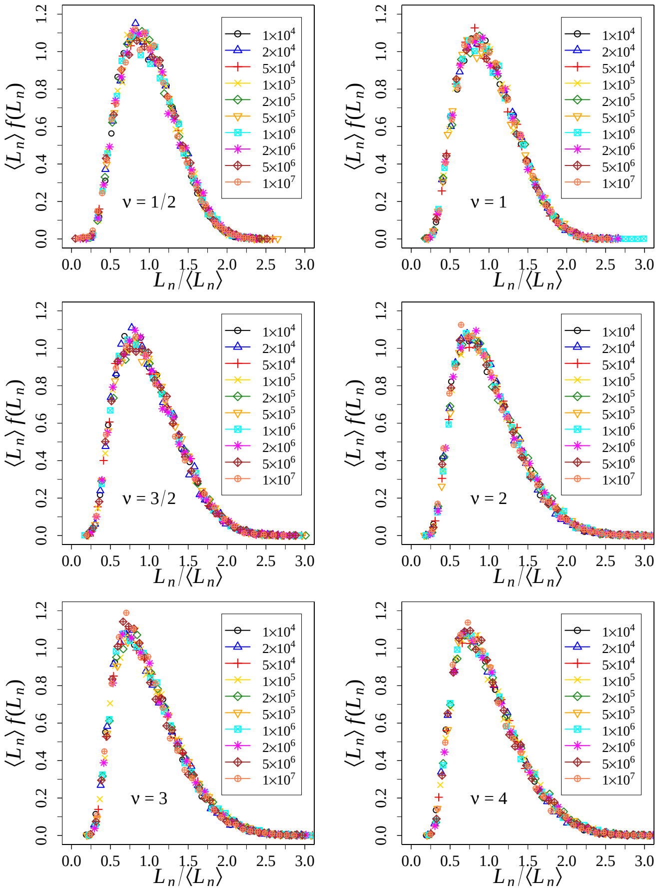

FIG. 4. Data collapse for some LIS data according to expressions (5)–(6). The upper panels (first row) display data for the LIS of random walks with step lengths with both infinite mean and variance ($\nu \leq 1$), the middle panels (second row) display data when the step increments have finite mean but infinite variance ($1 < \nu \leq 2$), and the bottom panels (last row) display data obtained from random walks with step increments with both mean and variance finite ($\nu > 2$).

As is well known, integer partitions play an important role in the solution of the LIS problem for random permutations [4–6]. In this case, the partitions carry the Plancherel measure given by $\mathbb{P}(\lambda) = (\dim \lambda)^2/n!$, where $\dim \lambda$ is the dimension of the irreducible representation of the symmetric group $\mathfrak{S}_n$ indexed by λ or, equivalently, the number of Young tableaux of shape λ. The correspondence between permutations and integer partitions (via the Robinson-Schensted-Knuth correspondence between permutations and Young tableaux) then allows one to identify the largest part of the partition λ with the length of the LIS of the original permutation.

Clearly, other probability measures for random integer partitions have also been considered. Of particular interest to us is the uniform measure given by $\mathbb{P}(\lambda) = 1/p(n)$. This is so because the expected size of the largest part λ_1 of a partition of a large integer n drawn from the set of all partitions of n uniformly at random is given asymptotically by [14–18]

$$\mathbb{E}(\lambda_1) = \sqrt{\frac{n}{4\zeta(2)}}\big(\ln n + 2\gamma_E - \ln \zeta(2)\big) + O(\ln n), \qquad (15)$$

where $\zeta(2) = \pi^2/6$ and $\gamma_E = 0.577215\cdots$ is Euler's constant. Equation (15) has, to leading and subleading order, the same functional form as the conjectured expression (6) for the asymptotic behavior of the length of the LIS of random walks with step lengths of finite variance. Moreover, the constant $1/\sqrt{4\zeta(2)} = 0.389848\cdots$ accompanying the leading term of Eq. (15) is close to the conjectured $1/e = 0.367879\cdots$ in (6). In the random partition model, however, the largest part λ_1 fluctuates, asymptotically for large n, like a Gumbel random variable with distribution

$$\mathbb{P}(\lambda_1 \leq \lambda) = F_G\left(\frac{\lambda - \sqrt{n/4\zeta(2)}\,\ln\big(n/4\zeta(2)\big)}{2\sqrt{n/4\zeta(2)}}\right), \qquad (16)$$

where $F_G(z) = \exp(-\exp(-z))$ [15, 17]. We can thus check whether Eq. (15) makes sense in the context of the LIS problem for random walks beyond mere coincidence by checking whether our LIS data follow a Gumbel distribution.

The mean and variance of a Gumbel random variable with p.d.f.

$$f_G(x;\mu,\beta) = \frac{1}{\beta}\exp\left[-\left(\frac{x-\mu}{\beta}\right) - \exp\left(-\frac{x-\mu}{\beta}\right)\right] \qquad (17)$$

are given by

$$\mathbb{E}(x) = \mu + \gamma_E \beta \quad \text{and} \quad \mathrm{Var}(x) = \frac{\pi^2}{6}\beta^2. \qquad (18)$$

If we substitute the sample mean $\langle L_n \rangle = 69946$ for $\mathbb{E}(x)$ and the sample variance $\langle L_n^2 \rangle - \langle L_n \rangle^2 \simeq 8.399 \times 10^8$ for $\mathrm{Var}(x)$ of a LIS dataset obtained from 10^4 Gaussian random walks of $n = 10^8$ steps each, we obtain the following simple estimation of the parameters μ and β for the data,

$$\hat{\mu} \simeq 56903 \quad \text{and} \quad \hat{\beta} \simeq 22597. \qquad (19)$$

We do not care about the uncertainties in $\hat{\mu}$ or $\hat{\beta}$ because they are relatively small and because the estimation procedure itself (the "method of moments") is only approximate. While $\hat{\mu}$ (as well as $\langle L_n \rangle$) is not very far from the respective factor in Eq. (16) with $n = 10^8$, to wit, $\mu = \sqrt{n/4\zeta(2)}\,\ln(n/4\zeta(2)) = 64468$, the value of $\hat{\beta}$ differs significantly from $\beta = 2\sqrt{n/4\zeta(2)} = 7797$. Figure 5 displays the histogram of the LIS data together with a plot of $f_G(z) = F_G'(z) = \exp(-z - \exp(-z))$. The fit looks good, but not excellent. In fact, the Gumbel distribution has a skewness of $12\sqrt{6}\,\zeta(3)/\pi^3 = 1.139\cdots$, while the data distribution has skewness $\simeq 0.976$ (irrespective of linear scaling). Whether this discrepancy is a finite-size effect is not clear at this moment. It should be remarked, however, that LIS data obtained from a uniform $U(-1,1)$ distribution of step increments with other values of walk length n provide the same overall picture as in Fig. 5 and seems to be nearly independent of n.

We leave the quantification of the "Gumbel hypothesis" to a future study employing more sophisticated density estimation techniques and hypothesis testing to tame uncontrolled magical thinking [19]. In Section III B, however, we will discover that Eq. (15) cannot be easily discarded as a possible scaling form for the length of the LIS of random walks with step lengths of finite variance.

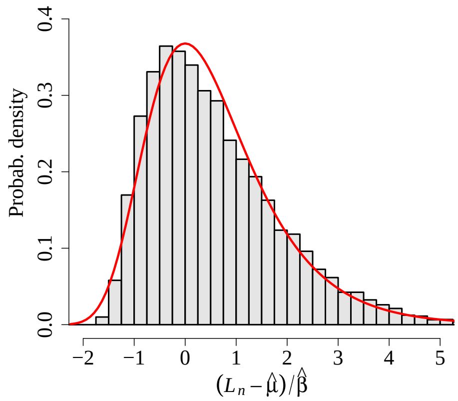

FIG. 5. Histogram of L_n obtained from 10^4 Gaussian random walks of 10^8 steps each together with the standard Gumbel p.d.f. (solid line). The LIS data is centered and normalized by the adjusted parameters $\hat{\mu}$ and $\hat{\beta}$, respectively; see Eqs. (17)–(19).

B. Constraining the range of the parameters in the asymptotic formula for $\mathbb{E}(L_n)$

The constants $1/e$ and $1/2$ appearing in the asymptotic formula (6) for the $\mathbb{E}(L_n)$ of random walks with distribution of step increments of finite variance were originally guessed in Ref. [9] based on linear least-squares fits to the data. Equation (15), with the same functional form, displays, respectively, the constants $1/\sqrt{4\zeta(2)} = 0.389848\cdots$ and $\big(2\gamma_E - \ln\zeta(2)\big)/\sqrt{4\zeta(2)} = 0.256025\cdots$. These led us to reassess the numerical values of the constants appearing in (6) by a more refined approach.

Assuming that Eq. (7) holds for the distribution of L_n and that $\mathbb{E}(L_n)$ behaves asymptotically like

$$\mathbb{E}(L_n) \sim a\sqrt{n}\ln n + b\sqrt{n}, \tag{20}$$

we can use the entire measured distributions to estimate the parameters a and b and give confidence intervals on their possible values. For a given pair (a,b) we scale the data points according to Eq. (20) and estimate the quality $S^{(c)}_{a,b}$ (superscript (c) for "collapse," see below) of the resulting data collapse by a method first introduced in [20] and refined in [21] in the context of the finite-size scaling analysis of phase transitions. The method works by estimating the best master curve on which the data points for different sizes n should collapse. The quality $S^{(c)}_{a,b}$ is defined as the mean-square distance of the data points to the master curve in units of the standard error, similar to a χ^2 test. If the data points are on average one standard error away from the estimated master curve, the data collapse will have a quality of $S^{(c)}_{a,b} = 1$. Values $S^{(c)}_{a,b} \ll 1$ indicate that the standard errors are overestimated; values $S^{(c)}_{a,b} \gg 1$ indicate that the data points do not collapse within error bars, i. e., that the quality of the data collapse is bad. A data collapse of bad quality might be due, besides the inevitable errors in the estimation of the master curve, also to finite-size effects in the data and corrections to the functional form (20) itself.

In our case, care should be exercised in the application of the method because the scaling function $g(z)$ (see Eq. (7)) is insensitive to the multiplication of L_n, and thus $\mathbb{E}(L_n)$, by a nonzero factor, i. e., to any rescaling $(a,b) \to (ra,rb)$ by some (real) $r \neq 0$. This leaves the determination of the optimal (a,b) ill-defined. To fix this we compare the average value of each data set with the values predicted by Eq. (20) for every pair (a,b) tested. We then apply the same method as before to compute the quality figure $S^{(m)}_{a,b}$ (superscript (m) for "mean") using Eq. (20) as the master curve with an added generous uncertainty of $\pm 0.05 \times \mathbb{E}(L_n)$ to account for finite-size effects and possible lower-order terms. Note that the minimum of $S^{(m)}_{a,b}$ corresponds to a standard least-squares fit.

The above mentioned analyses were performed on data obtained from random walks of $n = 2^{16}$, 2^{17}, 2^{18}, and 2^{19} steps distributed according to a uniform $U(-1,1)$ distribution; for each value of n, 10^6 sample random walks (and thus 10^6 data points L_n) are generated. Because data collapse is a matter of the form of a curve, we perform the collapse on $\ln f(L_n)$. This means that instead of the absolute standard errors σ of each data point we use the relative $\sigma/f(L_n)$ instead.

Figure 6 displays the contour plot of the composite quality factor $S_{a,b} = \frac{1}{2}\big(S^{(c)}_{a,b} + S^{(m)}_{a,b}\big)$ for our data, which are collected by a scan through the (a,b) space in discrete steps of $\Delta a = 0.003$ and $\Delta b = 0.01$ of the aforementioned procedure. The best quality $S^{(\min)}_{a,b} \approx 0.7$ was achieved for $(a,b) = (0.36, 0.36)$. The equi-quality lines at $S^{(\min)}_{a,b} + 1$ and $S^{(\min)}_{a,b} + 2$ can be roughly understood as 1σ and 2σ confidence intervals around the best quality $S^{(\min)}_{a,b}$ [20, 21]. Note that the precise shape of the confidence intervals depend on the weighting of the terms in the composite quality—for instance, a larger weight on $S^{(m)}_{a,b}$, say $\hat{S}_{a,b} = \frac{1}{4}\big(S^{(c)}_{a,b} + 3S^{(m)}_{a,b}\big)$, leads to a further elongation of the equi-quality loci. Figure 7 shows the data collapse as well as the least-squares fit which, in combination, yield the best quality. A further data collapse using high precision data from Ref. [10] in the far tails of the distribution shows a different picture, with $S^{(\min)}_{a,b} = 81$ at $(a,b) = (0.33, 0.68)$. This data collapse, although still including the pair $(1/e, 1/2)$ within the second interval, is clearly of bad quality ($S^{(\min)}_{a,b} \gg 1$), probably because of strong finite-size effects, since we only have data in the far tails of the distributions for small walk lengths $n \leq 4096$. This data collapse is therefore not shown.

We see from Fig. 6 that both pairs $(a,b) = (1/e, 1/2)$, which was proposed in [9] based on least-squares fits of mean values $\langle L_n \rangle$ for the LIS of Gaussian random walks, and $(a,b) = (0.389848\cdots, 0.256025\cdots)$ of Eq. (15) lie within the second confidence interval suggested by our method. Since the data sources are independent and even originate from different distributions of step increments, we see this as an argument in favor of the proposed scaling form (20).

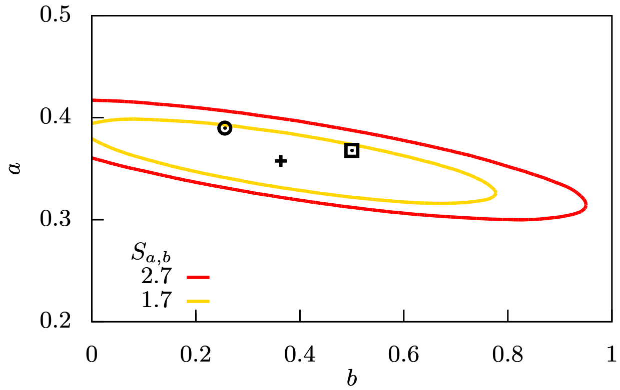

FIG. 6. Quality landscape $S_{a,b}$ obtained for random walks of $n = 2^{16}$, 2^{17}, 2^{18}, and 2^{19} uniformly, $U(-1,1)$ distributed steps; 10^6 random walks were generated for each n. The square symbol indicates the point $(1/e, 1/2)$ proposed in [9], the circle indicates the point $(0.389848\cdots, 0.256025\cdots)$ corresponding to Eq. (15), and the cross indicates the point $(0.36, 0.36)$ at which $S_{a,b}$ attains its minimum.

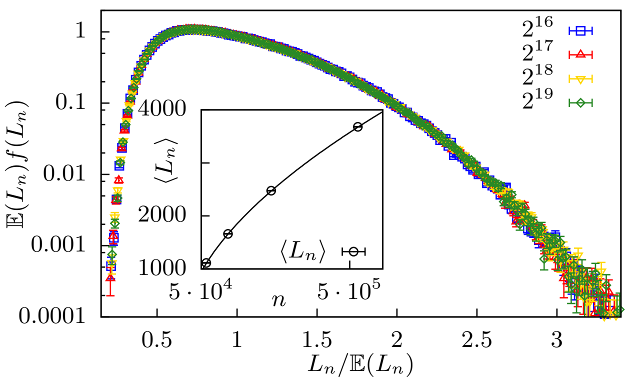

FIG. 7. Data collapse of the distributions with the optimal pair $(a,b) = (0.36, 0.36)$ that minimizes $S_{a,b}$. The inset shows the empirical averages $\langle L_n \rangle$ and their prediction according to Eq. (20) with the optimal constants $(a,b) = (0.36, 0.36)$.

IV. SUMMARY AND CONCLUSIONS

We have extended previous studies on the length L_n of the LIS of heavy-tailed random walks by considering Student's t-distributions with several different values of the parameter ν. We found that L_n scales like $\mathbb{E}(L_n) \sim n^\theta$ with a non-universal θ when $\phi(\xi)$ has infinite variance, but could not find a clear relationship $\theta = \theta(\nu)$ between these quantities besides the decreasing behavior $\theta'(\nu) < 0$. When $\phi(\xi)$ is of finite variance ($\nu > 2$), we recover the asymptotic behavior given by (6), but with newly estimated constants. Our best current estimate for the constants appearing in expression (20), based on a sophisticated combined consideration of the behavior of the mean and of the scaling of the full distribution, are $a = 0.36(3)$ and $b = 0.36(30)$.

We could not obtain data for $\nu < \frac{1}{2}$. The simulation of very heavy-tailed random walks is complicated by the fact that one needs to add numbers of widely different orders of magnitude while keeping their full significance. This can be done with numerical libraries that implement arbitrary precision arithmetic, but the efficiency of the simulations suffers enormously. It would be desirable to compute the LIS of heavy-tailed random walks in the $\nu \to 0$ limit to check how θ scales with ν in this limit and how it compares with the bounds in (4).

While the Plancherel distribution of the largest part of an integer partition coincides with the distribution of the length of the LIS of a uniformly distributed random permutation [4–6], the similarity between (6) and Eq. (15) does not imply any obvious relationship between the LIS of random walks and random integer partitions under the uniform measure. Our best estimated constants a and b for expression (20), however, cannot rule out Eq. (15) as a good candidate scaling form for the length of the LIS of random walks with step lengths of finite variance, and whether the LIS of these random walks follows a Gumbel distribution is open to debate. In a further study we intend to apply more refined density estimation techniques in the selection of an empirical model for the data; knowledge of the tail behavior, as provided by [10], is a valuable piece of information in this regard.

Finally, we remark that the elucidation of a possible combinatorial structure behind the LIS of random walks remains a tantalizing issue.

ACKNOWLEDGMENTS

The authors thank Alexey Bufetov (U. Bonn) for having called their attention to random integer partitions under the uniform measure and Pavel L. Krapivsky (BU, USA) and Satya N. Majumdar (LPTMS, France) for useful conversations. They also acknowledge usage of the HPC facilities of the GWDG Göttingen and the CARL cluster in Oldenburg funded by the DFG (INST 184/157-1 FUGG) and the Ministry of Science and Culture (MWK) of the Lower Saxony State. JRGM thanks the LPTMS for kind hospitality during a sabbatical leave in France and FAPESP (Brazil) for partial support through grant no. 2017/22166-9. HS and AKH acknowledge support by DFG grant no. HA 3169/8-1.

[1] S. Bespamyatnikh and M. Segal, Enumerating longest increasing subsequences and patience sorting, Inf. Proc. Lett. **76** (1–2), 7–11 (2000).

[2] S. M. Ulam, Monte Carlo calculations in problems of mathematical physics, in *Modern Mathematics for the Engineer: Second Series*, edited by E. F. Beckenbach (McGraw-Hill, New York, 1961), pp. 261–281.

[3] C. A. Tracy and H. Widom, Level-spacing distributions and the Airy kernel, Commun. Math. Phys. **159** (1), 151–174 (1994).

[4] J. Baik, P. Deift, and K. Johansson, On the distribution of the length of the longest increasing subsequence of random permutations, J. Am. Math. Soc. **12** (4), 1119–1178 (1999).

[5] D. Aldous and P. Diaconis, Longest increasing subsequences: from patience sorting to the Baik-Deift-Johansson theorem, Bull. Am. Math. Soc. **36** (4), 413–432 (1999).

[6] D. Romik, *The Surprising Mathematics of Longest Increasing Subsequences* (Cambridge University Press, New York, 2015).

[7] O. Angel, R. Balka, and Y. Peres, Increasing subsequences of random walks, Math. Proc. Cambridge **163** (1), 173–185 (2017).

[8] R. Pemantle and Y. Peres, Non-universality for longest increasing subsequence of a random walk, ALEA Lat. Am. J. Probab. Math. Stat. **14**, 327–336 (2017).

[9] J. R. G. Mendonça, Empirical scaling of the length of the longest increasing subsequences of random walks, J. Phys. A: Math. Theor. **50** (8), 08LT02 (2017).

[10] J. Börjes, H. Schawe, and A. K. Hartmann, Large deviations of the length of the longest increasing subsequence of random permutations and random walks, Phys. Rev. E **99** (4), 042104 (2019).

[11] S. N. Majumdar and G. Schehr, Top eigenvalue of a random matrix: large deviations and third order phase transition, J. Stat. Mech.: Theory Exp. **(2014)**, P01012 (2014).

[12] Wikipedia contributors, Student's *t*-distribution — Wikipedia, The Free Encyclopedia. Accessed: April 15, 2019.

[13] G. H. Hardy and S. Ramanujan, Asymptotic formulæ in combinatory analysis, Proc. Lond. Math. Soc. (2) **17**, 75–115 (1918).

[14] K. Husimi, Partitio numerorum as occurring in a problem of nuclear physics, Proc. Phys.-Math. Soc. Japan (3rd Ser.) **20**, 912–925 (1938).

[15] P. Erdős and J. Lehner, The distribution of the number of summands in the partitions of a positive integer, Duke Math. J. **8** (2), 335–345 (1941).

[16] B. Fristedt, The structure of random partitions of large integers, Trans. Am. Math. Soc. **337** (2), 703–735 (1993).

[17] A. Comtet, P. Leboeuf, and S. N. Majumdar, Level density of a Bose gas and extreme value statistics, Phys. Rev. Lett. **98** (7), 070404 (2007); A. Comtet, S. N. Majumdar, and S. Ouvry, Integer partitions and exclusion statistics, J. Phys. A: Math. Theor. **40** (37), 11255–11269 (2007).

[18] P. J. Grabner, A. Knopfmacher, and S. Wagner, A general asymptotic scheme for the analysis of partition statistics, Combin. Probab. Comput. **23** (6), 1057–1086 (2014).

[19] P. Diaconis, Theories of data analysis: From magical thinking through classical statistics, in *Exploring Data Tables, Trends and Shapes*, edited by D. C. Hoaglin, F. Mosteller, and J. W. Tukey (Wiley, New York, 1985), pp. 1–36.

[20] N. Kawashima and N. Ito, Critical behavior of the three-dimensional $\pm J$ model in a magnetic field, J. Phys. Soc. Japan **62** (2), 435–438 (1993).

[21] J. Houdayer and A. K. Hartmann, Low-temperature behavior of two-dimensional Gaussian Ising spin glasses, Phys. Rev. B **70** (1), 014418 (2004).