

# Large deviations of the length of the longest increasing subsequence of random permutations and random walks

Jörn Börjes,<sup>1,\*</sup> Hendrik Schawe,<sup>1,†</sup> and Alexander K. Hartmann<sup>1,‡</sup>

<sup>1</sup>*Institut für Physik, Universität Oldenburg, 26111 Oldenburg, Germany*

(Dated: January 16, 2019)

We study numerically the distributions of the length  $L$  of the longest increasing subsequence (LIS) for the two cases of random permutations and of one-dimensional random walks. Using sophisticated large-deviation algorithms, we are able to obtain very large parts of the distribution, especially also covering probabilities smaller than  $P(L) = 10^{-1000}$ . This enables us to verify for the length of the LIS of random permutations the analytically known asymptotics of the rate function and even the whole Tracy-Widom distribution, to which we observe a rather fast convergence in the larger than typical part. For the length  $L$  of LIS of random walks, where no analytical results are known to us, we test a proposed scaling law and observe convergence of the tails into a collapse for increasing system size. Further, we obtain estimates for the leading order behavior of the rate functions of both tails.

## I. INTRODUCTION

We study the distribution of the length  $L$  of the *longest increasing subsequence* (LIS) [1] of different ensembles of random sequences. Here, a subsequence of a given sequence is obtained by removing arbitrary entries and keeping the order of the remaining entries. In particular, the remaining entries are not necessarily neighbors in the given sequence. For a LIS it is required that the remaining entries are increasing from left to right and the number of remaining elements is maximal. An application of the LIS is for aligning whole genomes [2]. The first mention of this problem seems to be from Stanislaw Ulam [3], and is therefore also known as “Ulam’s problem”. In his study the mean length  $L$  of LIS on random permutations (RP) of  $n$  integers were scrutinized by means of Monte Carlo simulations and it was conjectured that in the limit of large  $n$ , the length converges to  $L = c\sqrt{n}$ , with some constant  $c$ , which was later proven to be  $c = 2$  [4]. In the following years much work was published scrutinizing the large deviation behavior of this problem and explicit expressions for both the left (lower) and right (upper) tail were derived rigorously [5–7]. Interestingly, for the LIS of the random permutation it was shown that the distribution  $P(L)$  of its length is a Tracy-Widom distribution [8]. The Tracy-Widom distribution was at that time only known from random matrix theory, where it described the distribution of the largest eigenvalues of the *Gaussian unitary ensemble* (GUE), an ensemble of Hermitian random matrices. In physics it came into focus after an explicit mapping of a 1+1 dimensional polynuclear growth model [9]. Subsequently other mappings of 1+1 dimensional growth models belonging to the Kardar-Parisi-Zhang universality like an anisotropic ballistic deposition [10] were found. Other models in which

the Tracy-Widom distribution appears, include the totally asymmetric exclusion process [11] and directed polymers [12]. For a pedagogical overview about the relations of different models exhibiting a Tracy-Widom distribution, we recommend Ref. [13]. Fluctuations in growth processes following the Tracy-Widom distribution could also be observed in experiments, e.g., from growing liquid crystals where the Tracy-Widom distribution of the GUE appears for circular growth and of the *Gaussian orthogonal ensemble* (GOE) for growth from a flat surface [14, 15].

The Tracy-Widom distribution seems to occur always together with a *third order phase transition* between a *strongly-interacting* phase in the left tail and a *weakly-interacting* phase in the right tail, whose crossover is characterized by the Tracy-Widom distribution [16]. For these third order phase transitions, the probability density function behaves in the left tail as  $P(x) \approx e^{-n\Phi_-}$  with the role of the free energy played by the *rate function*  $\Phi_-(x) \sim (a-x)^3$  for  $x \rightarrow a$  from the left, where  $a$  is the critical point of the transition, i.e., the scaled mean value. Here,  $n$  is some large parameter, e.g, the system size. The  $O(x^3)$  leading order behavior of  $\Phi_-$  generally leads to a discontinuity in the third derivative of the free energy and therefore to a third order phase transition. This seems to be a characteristic sign predicting the main region of the distribution to follow a Tracy-Widom distribution. Therefore the behavior of the far tails of problems of this universality are of great interest to understand this connection better. Consequently the *large deviations* of some of these models were studied thoroughly [16, 17].

For the distribution of the length of the LIS of random permutations there are also analytical results for the large deviations, i.e., the behavior for large values of  $n$  including the far tails [5–8], which also show the characteristic behavior of the above mentioned left-tail rate function. For the case of the length of the LIS of random walks, bounds for the behavior of the mean are known [18] and there is also numerical work which is concerned with the distribution in the typical region [19], i.e., those

---

\*Electronic address: joern.boerjes@uni-oldenburg.de

†Electronic address: hendrik.schawe@uni-oldenburg.de

‡Electronic address: a.hartmann@uni-oldenburg.de

LIS which occur with a high enough probability of about  $\geq 10^{-6}$ . We deem it worthwhile to look also for this system closer at the tails of the distribution for finite systems.

For the purpose of studying the large deviations of this problem numerically, we will utilize sophisticated large deviation sampling methods to observe the distribution of the length  $L$  for two ensembles of random sequences. This way we can observe directly the far tails of the Tracy-Widom distribution for the random permutation case [8] and can confirm the known large  $n$  asymptotics [7]. The second ensemble are one-dimensional random walks with increments from a uniform distribution. While we can observe the scaling proposed in Ref. [19] for the main region, the tails are subject to considerable finite-size effects. Nevertheless the distributions collapse over larger regions for larger sizes  $n$ . Also, we give estimates for the leading order behavior of the rate functions governing the left and right tail of the distribution  $P(L)$ .

This study will first introduce the different ensembles of interest and the algorithms used to obtain the distribution of the length in Sec. II. In Sec. III we will show the results we gathered and interpret them. We conclude this study in Sec. IV.

## II. MODELS AND METHODS

To define the longest increasing subsequence (LIS), we have to define a subsequence first. Given some sequence  $S = (S_1, S_2, \dots, S_n)$  a *subsequence* of length  $L$  is a sequence  $s = (S_{i_1}, S_{i_2}, \dots, S_{i_L})$  ( $1 \leq i_j \leq n, i_j < i_{j+1}$  for all  $j = 1, \dots, L$ ) containing only elements present in  $S$  in the same order as in  $S$ , though possibly with gaps. An *increasing subsequence* has elements such that every element in  $s$  is smaller than its predecessor, i.e.,  $S_{i_j} < S_{i_{j+1}}$  for  $j = 1, \dots, L - 1$ . The LIS is consequently the longest, i.e., the one with the highest number  $L$  of elements, of all possible increasing subsequences. Note that the LIS is not uniquely defined, but by definition its length is unique. As an example two different LIS are marked by overlines and underlines in the following sequence:  $S = (\underline{3}, 9, \underline{4}, \bar{1}, \bar{2}, \underline{7}, \bar{6}, \bar{8}, 0, 5)$

In this study the sequence  $S$  is either drawn from the ensemble of *random permutations* of  $n$  consecutive integers or from the ensemble of *random walks* with increments  $\delta_j$  ( $j = 1, \dots, n$ ) from a uniform distribution  $\delta_j \sim U(-1, 1)$ , such that

$$S_i = \sum_{j=1}^i \delta_j. \quad (1)$$

An example of each sequence with the corresponding LIS marked is shown in Fig. 1. Here the typical difference between the random permutation and random walks are visible: The entries of the random walk are strongly correlated such that the random walk typically consists of runs with downward or upward trends, such that the LIS

is typically confined in an upward trend and its entries therefore are close together. The random permutation, on the other hand, typically shows LIS with entries over the whole range. Therefore it is plausible that the distributions of the length of the LIS for these two ensembles differ [19].

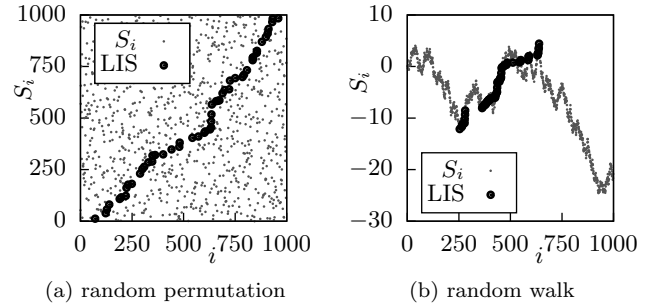


FIG. 1: Visualization of random sequences of length  $n = 1000$  where the value is plotted over the corresponding index. Marked with circles are the entries of one possible LIS. (a) random permutation, (b) random walk.

To find the LIS of any given sequence, we use the *patience sort algorithm*, which is originally a sorting algorithm. This choice is mainly motivated by the simplicity of the algorithm if one is only interested in the length of the LIS. We will only introduce the very simple version to obtain the length, but a comprehensive review of the connection of patience sort with the LIS can be found in Ref. [4]. In short, the patience sort algorithm works as follows: We iterate over the  $n$  entries  $S_i$  and place each into an initially empty stack (or pile)  $a_j$  on the smallest  $j$  such that for the top entry  $\text{top}(a_j) > S_i$  holds. Note that this will always ensure that the top entries of  $a$  are ascendingly sorted, such that we can determine  $j$  by a binary search in  $\mathcal{O}(\ln n)$ . Finally, the number of non-empty stacks  $a_j$  is equal to the length  $L$  of the LIS.

### A. Large-deviation Sampling

To be able to gather statistics of the large-deviation regime numerically [20], we need to apply a sophisticated sampling scheme. Therefore we use a well tested [21–23] Markov chain Monte Carlo sampling which treats the system as a physical system at some artificial *temperature* with the observable of interest as its *energy*. Since the algorithm has been presented comprehensively in the literature, we here only state the details specific to the current application. In our case, we identify the state of the system with the sequence, the length  $L$  with the energy and sample the equilibrium state at temperature  $\Theta$  using the Metropolis algorithm [24, 25]. Controlling the temperature allows us to direct the sampling to different regimes of the distributions, to eventually cover the distributions over a large part of the support. To evolve our Markov chain of sequences, we have to introduce change

moves, which modify a sequence and consequently the energy  $L$ . For the random permutation we swap two random entries and for the random walk we replace one of the increments  $\delta_j$  (cf. Eq. (1)) by a new random number drawn from the same uniform distribution. These changes are accepted according to the Metropolis acceptance ratio

$$P_{\text{acc}} = \min(1, e^{-\Delta L/\Theta}), \quad (2)$$

where  $\Delta L$  is the change in energy due to the change move. This Markov chain of sequence realizations will converge to an equilibrium state. As usual with Markov chain Monte Carlo simulations, we need to ensure equilibration and that the samples are decorrelated [25].

As should be intuitively plausible, in equilibrium the realizations will generally have a lower than typical energy for low temperatures and typical energies for high temperatures. We can also introduce negative temperatures for larger than typical energies. This way the temperature can be tuned to guide the simulation towards realizations within a specific range of energies  $L$ . Since we know the equilibrium distribution  $Q_\Theta(S)$  at temperature  $\Theta$  of realizations, i.e., sequences  $S$ , to be

$$Q_\Theta(S) = \frac{1}{Z_\Theta} e^{-L(S)/\Theta} Q(S), \quad (3)$$

with the natural distribution  $Q(S)$ , we can later correct for the bias introduced by the temperature and arrive at the unbiased distribution  $P(L)$  with good statistics also in the regions unreachable by simple sampling. Therefore consider the sampled equilibrium distributions  $P_\Theta(L)$ . To connect them to the distribution of realizations  $Q_\Theta(S)$ , we can sum all realizations with the same value of  $L$ , leading to

$$P_\Theta(L) = \sum_{\{S|L(S)=L\}} Q_\Theta(S) \quad (4)$$

$$= \sum_{\{S|L(S)=L\}} \frac{1}{Z_\Theta} e^{-L(S)/\Theta} Q(S) \quad (5)$$

$$= \frac{1}{Z_\Theta} e^{-L/\Theta} P(L). \quad (6)$$

Solving this equation for  $P(L)$  allows to correct for the bias introduced by the temperature. An intermediate snapshot of this process is shown in Fig. 2.

The constants  $Z_\Theta$  can be obtained by enforcing continuity of the distribution, i.e.,

$$P_{\Theta_j}(L) e^{L/\Theta_j} Z_{\Theta_j} = P_{\Theta_i}(L) e^{L/\Theta_i} Z_{\Theta_i} \quad (7)$$

for pairs of  $i, j$  for which the gathered data  $P_{\Theta_i}(L)$  overlaps with  $P_{\Theta_j}(L)$ . While this can be used to approximate the ratios of pairwise  $Z_{\Theta_i}$ , the absolute value can then be obtained by normalization of the whole distribution. This procedure requires a clever choice of temperatures, since gaps in the sampled range of  $L$  would make it impossible to find a ratio of  $Z_{\Theta_i}$  on the left and right side

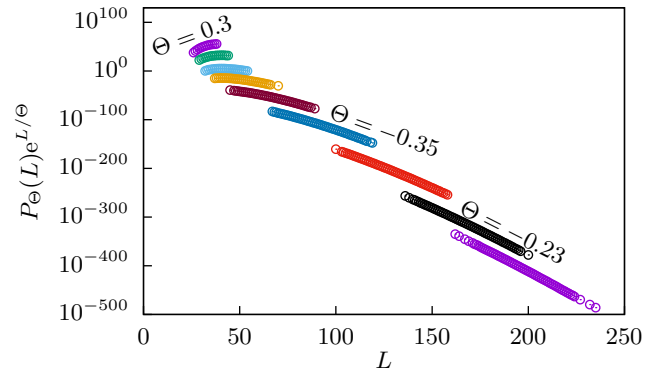


FIG. 2: Intermediate step after correction with Eq. 6 but before determination of the values  $Z_{\Theta_i}$  (i.e., all  $Z_{\Theta_i} = 1$ ). The data is gathered for random walk sequences of length  $n = 512$ . Each shade of gray (color) is sampled at a different temperature  $\Theta$ , for three datasets the corresponding temperatures are annotated. (For clarity some evaluated temperatures are omitted.)

of the gap. We used in the order of 100 distinct temperatures. In general, the larger the size  $n$ , the more temperatures are needed.

### III. RESULTS

We applied the temperature-based sampling scheme to obtain the probability distributions of the length of the LIS for the two cases of random permutations and of random walks with uniform increments. In both cases, we studied five different system sizes  $n$  up to  $n = 4096$  each.

#### A. Random Permutations

First, we will look at the distribution of the length of the LIS of random permutations. For this case there are already a lot of properties known in the limit of  $n \rightarrow \infty$ .

It is known that the distribution should converge to a suitably rescaled Tracy-Widom distribution  $\chi$  of the GUE ensemble [8] for large values of  $n$  as

$$P_n \left( (L - 2\sqrt{n}) n^{-1/6} \right) = \chi \left( (L - 2\sqrt{n}) n^{-1/6} \right). \quad (8)$$

Rescaled to accommodate this leading behavior, our results are shown in Fig. 3. By using the large-deviation approach, we are able to measure probabilities as small as  $10^{-1000}$  and below, allowing us to go beyond the first numerical work [19] on the distribution of LIS. We can observe a very good collapse up to probabilities of  $10^{-200}$  of our data onto the Tracy-Widom distribution given in the tables of Ref. [26].

Also note that the collapse works very well in the intermediate right tail but converges a bit slower in the left

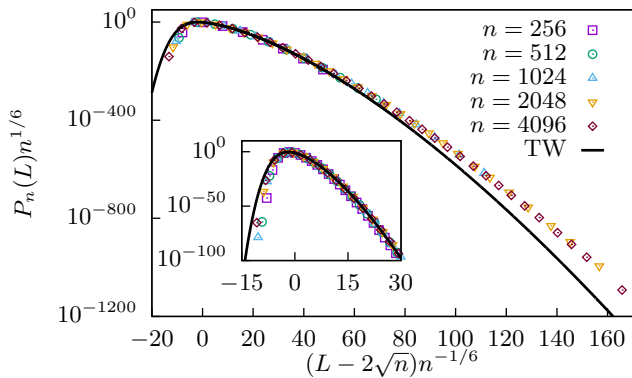


FIG. 3: Numerically obtained distributions for different system sizes  $n$  rescaled according to Eq. (8). The Tracy-Widom distribution is drawn as a black line [26] and is expected to be the curve all distributions collapse onto. The inset shows a zoom on the intermediate tails. On the left the tendency of our data towards the Tracy-Widom distribution with increasing system size  $n$  is visible. (For clarity some datapoints are discarded to show the same density of symbols for every system size.)

tail and far slower in the far-right tail. The inset zooms into the intermediate tail of the probability density function  $P > 10^{-100}$ , where the collapse fits very well to the expected Tracy-Widom distribution. In the far tails we observe considerable deviations, from the tabulated data, which are at least in part caused by finite-size effects due to the relatively small sizes  $n$  of our sequences. For a more extensive study of these finite-size effects, one could obtain the empirical distribution for more sizes, and extrapolate the finite-size effects to  $n \rightarrow \infty$ , as done in [27]. Nevertheless, our numerically obtained tails fit very well to another expected form which will be explained later, such that we assume a stronger influence of finite-size effects in the far tails for this scaling instead of systematic errors.

Also note that while we can sample a very large part of the distribution of the length of longest subsequences of random permutations even including events with a probability less than  $10^{-1000}$  for the largest permutations, we can not reach across the whole range of possible values and would possibly need to modify our sampling algorithm by either switching to a better change move or consider a different sampling like Wang-Landau's method [28].

The left tail asymptotic, i.e.,  $L/\sqrt{n} = x < 2$ , of the probability density function is given by the analytically known rate function [6, 7]

$$\lim_{n \rightarrow \infty} \frac{1}{n} \ln P_n(L) = -2H_0(x) \quad (9)$$

with

$$H_0(x) = -\frac{1}{2} + \frac{x^2}{8} + \ln \frac{x}{2} - \left(1 + \frac{x^2}{4}\right) \ln \left(\frac{2x^2}{4+x^2}\right); \quad (10)$$

the right tail asymptotic, i.e.,  $L/\sqrt{n} = x > 2$ , is given by [5, 7]

$$\lim_{n \rightarrow \infty} \frac{1}{\sqrt{n}} \ln P_n(L) = -U_0(x) \quad (11)$$

with

$$U_0(x) = 2x \cosh^{-1}(x/2) - 2\sqrt{x^2 - 4}. \quad (12)$$

Note that Eq. (11) behaves atypically for a rate function as the distribution behaves like  $P_n \propto e^{-\sqrt{n}U_0}$ , which according to the definition, e.g., given in [29], does therefore not fulfill the large deviation principle. Nevertheless, it describes the behavior of the distribution in leading order.

We use our sampled data to test these rate functions. If the data are suitably rescaled according to Eq. (9) and (11), in the corresponding tails we can observe a very nice convergence of the data to the rate functions. This is plotted in Fig. 4. This excellent agreement of analytical and numerical results over hundreds of decades in probability, gives us confidence that our approach works well and can be extended to cases where no analytical results are known. Also note that we can observe in our data the leading order behavior of the left tail rate function  $H_0$ , which goes with the exponent 3 characteristic for the third order phase transition confirming its connection with the Tracy-Widom distribution [16].

## B. Random Walks

The second class of sequences  $S$  we scrutinized are random walks. Here no analytical results are known to us, thus the distribution beyond the high-probability peak region seems to be unknown. Again, by applying the large-deviation approach, we sample basically the whole distribution, and can even compare the right tail of our distribution with the corner case of  $L = n$ , which only occurs if all increments  $\delta$  are positive and therefore with probability  $2^{-n}$ . This case is marked in Fig. 5 to emphasize the quality of our data. For the left tail, we can not sample so far, as the very steep decline of the distribution is difficult to handle for our sampling scheme.

For random walks with increments from a symmetric uniform distribution, indeed for increments from any symmetric distribution with finite variance, the scaling of the mean as  $\langle L \rangle \propto n^\theta$  and the variance as  $\sigma^2 \propto n^{2\theta}$  was observed in Ref. [19] with  $\theta = 0.5680(15)$ . More interestingly the same reference suggests that the whole distribution follows the scaling form

$$P_n(L) = n^{-\theta} g(n^{-\theta} L), \quad (13)$$

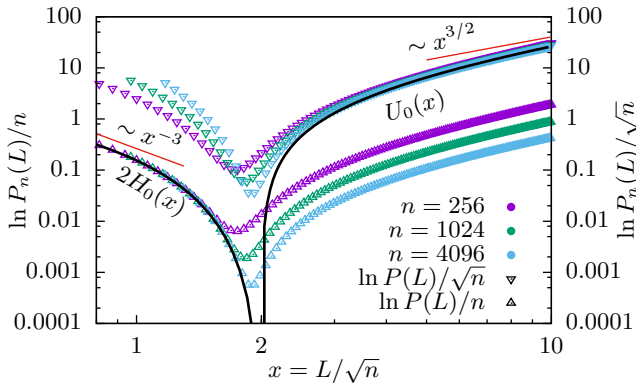


FIG. 4: Empirical rate functions for different system sizes  $n$ . On the top (triangles down) scaled as  $\ln P_n(L)/\sqrt{n}$  to emphasize the right tail behavior. On the bottom (triangles up) scaled as  $\ln P_n(L)/n$  to emphasize the left tail behavior. The analytically known rate functions for both tails  $2H_0$  and  $U_0$  are shown in the correspondingly scaled region and a convergence of the data to these functions is well visible. The leading order terms of the series expansion (cf. [5]) are also shown as straight lines next to the rate function.

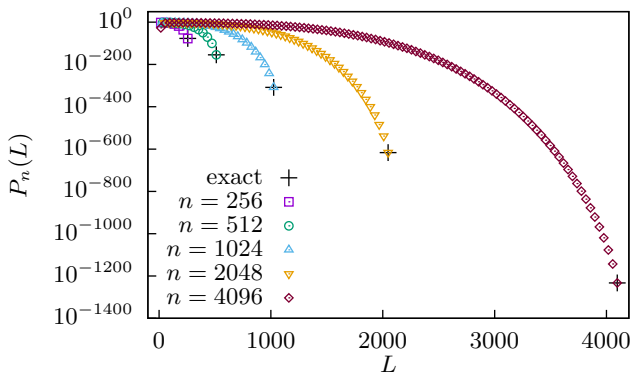


FIG. 5: Probability distributions  $P_n(L)$  of the length of LIS of random walks with exact extremevalues for the  $n = L$  case. (For clarity only every 40th bin is visualized, including the last bin.)

with a not-explicitly known function  $g$ . Note the similarity to the scaling for the random permutation case in Eq. 8, though here we do not need to subtract the mean value before rescaling with the standard deviation, since both happen to have the same exponent  $\theta$  in this case. Using our data for the tails of the distribution, we can test whether this scaling holds over the whole distribution or only in the main region. If we rescale the axis of the plot suitably, the distributions for different sizes  $n$  should collapse on the scaling function  $g$ , in the case that Eq. (13) holds. Since we only have results for comparatively small systems sizes in comparison to Ref. [19], we have to include the logarithmic corrections to scaling,

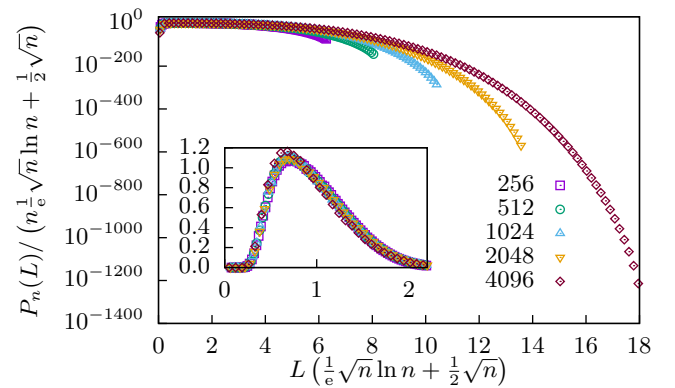


FIG. 6: Collapse of different system sizes on a common curve  $g$  from Eq. (13). Apparently the far tail shows severe finite-size effects, though for increasing sizes  $n$  a convergence to a common curve is visible. (For clarity not all datapoints are drawn.)

which are estimated by Ref. [19] to be

$$\langle L \rangle \approx \frac{1}{e} \sqrt{n} \ln n + \frac{1}{2} \sqrt{n}. \quad (14)$$

Note that this proposed scaling is proportional to  $\sqrt{n}$ , i.e.,  $\theta = 0.5$  instead of the slightly larger value measured without accounting for logarithmic corrections. Also note that this follows the analytically expected form [18]. In fact, using expression Eq. (14) instead of  $n^\theta$  in Eq. (13), leads to a better collapse of our data. This is shown in Fig. 6, where the collapse does seem to work except for the very far tails, which is an effect – at least partially – caused by finite-size effects, since the length of the LIS can for finite  $n$  never be longer than  $n$ . This pattern occurs often when looking at the far tails of discrete systems, e.g., for the convex hull of random walks on lattices in [30–33] or in a toy model for non interacting Fermions in a landscape with  $n$  random energy levels [27].

Since for the rate functions characterizing the distribution of the length of LIS of random walks there is no known result, we use our numerical data to give a rough estimate of the rate function. Therefore we look into the empirical rate function  $\Phi_n(L) = \frac{1}{n} \ln P_n(L)$ , which is plotted in Fig. 7 for the data already shown in Fig. 5.

Using the empirical rate function we can obtain the asymptotics of the rate function from our data. Note that to estimate the right tail rate function we use the intermediate tail and not the far tail, which is bending up due to finite-size effects. Since we are only interested in the leading order exponent of the rate function, i.e., assuming  $\Phi(L) \propto L^\kappa$  for very small and very large values of  $L$ , we can rescale the axes arbitrarily due to the scale invariance of power laws. For convenience we look at  $x = L/L_{\max}$  to limit the range to the interval  $[0, 1]$ . For the left tail we observe a leading order behavior of the rate function of approximately  $\Phi(L) \sim L^{-1.6}$  and for the right tail  $\Phi(L) \sim L^{2.9}$ , though larger values of the exponent are possible, if looking at a range farther right,

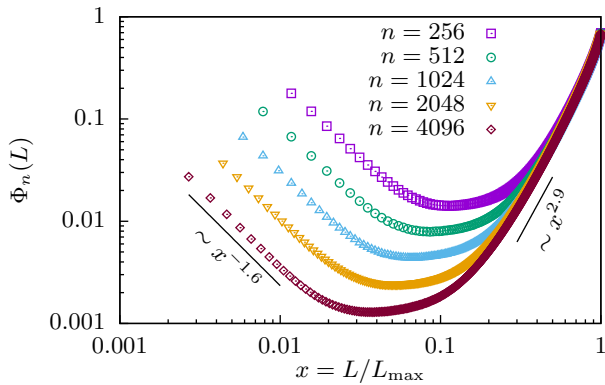


FIG. 7: Empirical rate function  $\Phi_n(L)$  for the length of the LIS of random walks. (For clarity not every datapoint is shown.)

which is probably caused by finite-size effects, as the very long LIS are suppressed by the hard limit of  $L \leq n$ . Note that the exponent of the left tail is clearly distinct from 3, such that it does not show signs of a third order phase transition. Also it does not show a Tracy-Widom distribution in the main region (also see [19]), which is consistent with the expectation that these two properties do occur together [16].

Comparing this leading order behavior to the behavior of the random permutation case as visualized in Fig. 4, shows that the tails decay differently. For a direct comparison of our results, consider Fig. 8. While the right-tail exponent is larger in the random walk case, the probability density decays slower (cf. inset of Fig. 8). This apparent contradiction, is understandable when considering that the rate function of the random permutation case grows much faster near the minimum at  $\langle L \rangle$ , such that the rate function in the RP case has larger absolute values and the probability density decreases much faster. It is interesting that the empirical rate functions behave qualitatively so different for such closely related models, e.g., the branches left and right of the minimum show opposite curvature in the two cases. Generally, it is visible that the distribution  $P(L)$  is much broader in the RW case, especially towards quite large values of  $L$ .

#### IV. CONCLUSIONS

We obtained numerical data for the distribution of the length of the longest increasing subsequence for two cases of sequences of random numbers, namely, for random permutations and for one dimensional random walks. By applying sophisticated large-deviation algorithms, we are able to sample the distributions over literally hundreds of decades in probability. The case of random permutations is already well studied in the analytical literature and our results confirm, to our knowledge, for the first time these results. Since our data is gathered for finite

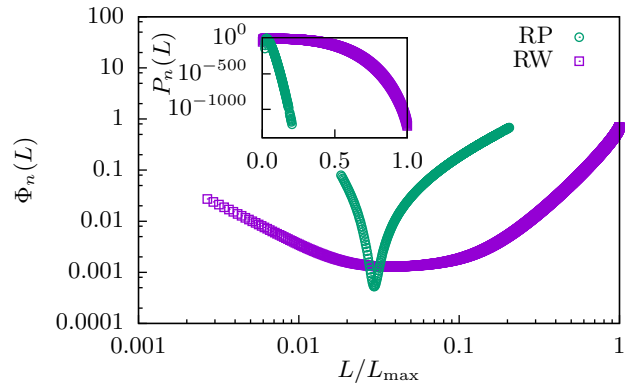


FIG. 8: Direct comparison of the distributions for both cases, the random permutation and the random walk. The main plot shows the empirical rate function  $\Phi_n(L)$ . The inset shows the probability density  $P_n(L)$ . Both show sequences of length  $n = 4096$ .

system sizes, we can observe a rather fast convergence to the analytical results valid in the  $n \rightarrow \infty$  limit. These results show also the validity and convergence of our simulations. For the case of random walks we can observe the leading order behavior of the rate function far into the tails. This result could be used to guide analytical work on this topic or at least to test future analytical results. A direct comparison of the empirical rate functions in the tails shows qualitatively very different behavior. While the rate function of the random walk seems to be a convex function, the random permutation case consists in principle of two concave parts.

A possible future direction extending this work would be an interpolation between the random permutation and random walk case, where one could observe the change of the exponents governing the rate function. Since a set of distinct random numbers  $\delta_j$  drawn uniformly, from  $[-1, 1]$  should show the same statistics for the longest increasing subsequence of a random permutation, we could introduce a parameter  $c$  governing the correlation length. The sequence would be constructed as  $S_i = \sum_{j=\max(0, i-c)}^i \delta_j$ . For  $c = 0$  this would correspond to a random permutation and for  $c = n$  to a random walk. In addition to this simple type of correlation, one could study power-law correlated random numbers or increments, leading possibly to even more complicated behavior.

#### Acknowledgments

We are indebted to Satya N. Majumdar who brought this problem to our attention and gave valuable feedback on a draft of this manuscript. Also we want to thank J. Ricardo G. Mendonça for interesting discussions about the problem and Christoph Norrenbrock for his advice during the preparation of the manuscript. We acknowledge 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 Cul-

ture (MWK) of the Lower Saxony State. HS acknowledges support by DFG grant HA 3169/8-1.

- 
- [1] D. Romik, *The Surprising Mathematics of Longest Increasing Subsequences* (Cambridge University Press, USA, 2015).
- [2] A. L. Delcher, S. Kasif, R. D. Fleischmann, J. Peterson, O. White, and S. L. Salzberg, *Nucleic acids research* **27**, 2369 (1999).
- [3] S. M. Ulam, in *Modern Mathematics for the Engineer: Second Series*, edited by E. Beckenbach and M. Hestenes (Dover Publications, Incorporated, 2013), Dover Books on Engineering Series, chap. 11, pp. 261–281, ISBN 9780486497471.
- [4] D. Aldous and P. Diaconis, *Bulletin of the American Mathematical Society* **36**, 413 (1999).
- [5] T. Seppäläinen, *Probability Theory and Related Fields* **112**, 221 (1998), ISSN 1432-2064.
- [6] B. F. Logan and L. A. Shepp, in *Young Tableaux in Combinatorics, Invariant Theory, and Algebra* (Elsevier, 1977).
- [7] J.-D. Deuschel and O. Zeitouni, *Combinatorics, Probability and Computing* **8**, 247 (1999).
- [8] J. Baik, P. Deift, and K. Johansson, *Journal of the American Mathematical Society* **12**, 1119 (1999).
- [9] M. Prähofer and H. Spohn, *Phys. Rev. Lett.* **84**, 4882 (2000).
- [10] S. N. Majumdar and S. Nechaev, *Physical Review E* **69**, 011103 (2004).
- [11] K. Johansson, *Communications in Mathematical Physics* **209**, 437 (2000), ISSN 1432-0916.
- [12] J. Baik and E. M. Rains, *Journal of Statistical Physics* **100**, 523 (2000), ISSN 1572-9613.
- [13] S. N. Majumdar, in *Complex Systems: Lecture Notes of the Les Houches Summer School 2006*, edited by J. Bouchaud, M. Mézard, and J. Dalibard (Elsevier Science, 2006), Les Houches, chap. 4, ISBN 9780080550596, URL <https://arxiv.org/abs/cond-mat/0701193>.
- [14] K. A. Takeuchi and M. Sano, *Phys. Rev. Lett.* **104**, 230601 (2010).
- [15] K. A. Takeuchi, M. Sano, T. Sasamoto, and H. Spohn, *Scientific reports* **1**, 34 (2011).
- [16] S. N. Majumdar and G. Schehr, *Journal of Statistical Mechanics: Theory and Experiment* **2014**, P01012 (2014).
- [17] P. L. Doussal, S. N. Majumdar, and G. Schehr, *EPL (Europhysics Letters)* **113**, 60004 (2016).
- [18] O. Angel, R. Balka, and Y. Peres, *Mathematical Proceedings of the Cambridge Philosophical Society* **163**, 173 (2017).
- [19] J. R. G. Mendonça, *Journal of Physics A: Mathematical and Theoretical* **50**, 08LT02 (2017).
- [20] A. K. Hartmann, *Big Practical Guide to Computer Simulations* (World Scientific, Singapore, 2015).
- [21] A. K. Hartmann, *Phys. Rev. E* **65**, 056102 (2002).
- [22] A. K. Hartmann, *The European Physical Journal B* **84**, 627 (2011), ISSN 1434-6036.
- [23] A. K. Hartmann, *Phys. Rev. E* **89**, 052103 (2014).
- [24] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *The journal of chemical physics* **21**, 1087 (1953).
- [25] M. Newman and G. Barkema, *Monte carlo methods in statistical physics chapter 1-4* (Oxford University Press: New York, USA, 1999).
- [26] M. Prähofer and H. Spohn, *Journal of Statistical Physics* **115**, 255 (2004), ISSN 1572-9613, tables at <http://www-m5.ma.tum.de/KPZ>.
- [27] H. Schawe, A. K. Hartmann, S. N. Majumdar, and G. Schehr, *EPL (Europhysics Letters)* **124**, 40005 (2018).
- [28] F. Wang and D. P. Landau, *Phys. Rev. Lett.* **86**, 2050 (2001).
- [29] H. Touchette, *Physics Reports* **478**, 1 (2009), ISSN 0370-1573.
- [30] G. Claussen, A. K. Hartmann, and S. N. Majumdar, *Phys. Rev. E* **91**, 052104 (2015).
- [31] H. Schawe, A. K. Hartmann, and S. N. Majumdar, *Phys. Rev. E* **96**, 062101 (2017).
- [32] H. Schawe, A. K. Hartmann, and S. N. Majumdar, *Phys. Rev. E* **97**, 062159 (2018).
- [33] H. Schawe and A. K. Hartmann, arXiv preprint arXiv:1808.10698 (2018).