Scaling exponents for ordered maxima

E. Ben-Naim, P. L. Krapivsky, and N. W. Lemons

Phys. Rev. E **92**, 062139 — Published 22 December 2015

DOI: 10.1103/PhysRevE.92.062139

Scaling Exponents for Ordered Maxima

E. Ben-Naim,[1] P. L. Krapivsky,[2,3] and N. W. Lemons[1]

[1] *Theoretical Division and Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA*

[2] *Department of Physics, Boston University, Boston, Massachusetts 02215, USA*

[3] *Institut de Physique Théorique, Université Paris-Saclay, CEA and CNRS, 91191 Gif-sur-Yvette, France*

We study extreme value statistics of multiple sequences of random variables. For each sequence with N variables, independently drawn from the same distribution, the running maximum is defined as the largest variable to date. We compare the running maxima of m independent sequences, and investigate the probability S_N that the maxima are perfectly ordered, that is, the running maximum of the first sequence is always larger than that of the second sequence, which is always larger than the running maximum of the third sequence, and so on. The probability S_N is universal: it does not depend on the distribution from which the random variables are drawn. For two sequences, $S_N \sim N^{-1/2}$, and in general, the decay is algebraic, $S_N \sim N^{-\sigma_m}$, for large N. We analytically obtain the exponent $\sigma_3 \cong 1.302931$ as root of a transcendental equation. Furthermore, the exponents σ_m grow with m, and we show that $\sigma_m \sim m$ for large m.

PACS numbers: 02.50.-r, 05.40.-a, 05.45.Tp

I. INTRODUCTION

The theory of extreme values is a well-developed area of statistics and probability theory [1–6]. Extreme values such as the maximal and the minimal data points are important features of a dataset. Statistics of extreme events play a key role in a host of data rich subjects including climate science [7–9], geophysics [10–13], and economics [14–17].

The running maximum, defined as the largest variable to date in a sequence of variables, is a central quantity in extreme-value statistics. This quantity evolves rather slowly: the number of times it changes typically grows logarithmically with the number of random variables [5, 6]. Consequently, persistence [18–20] or first-passage properties [21] involving the running maximum often exhibit power-law dependence on the number of variables [22–24]. Methods and concepts from statistical physics provide a powerful tool for obtaining the nontrivial scaling exponents that characterize such power-law behaviors [25].

This investigation is motivated by a recent letter [26] concerning maximal positions of random walks. It was reported that the probability that the maxima of multiple random walks remain perfectly ordered decays as a power law with the number of steps [26, 27], and that the corresponding decay exponents are generally nontrivial.

Here, we study the same probability for *uncorrelated* random variables. Figure 1a shows three sequences of random variables and figure 1b shows the corresponding running maxima. The running maxima form three staircases. We are interested in the probability that these running maxima remain perfectly ordered, or equivalently, that the staircases do not intersect even once. When the random variables are uncorrelated, this probability is universal: it is completely independent of the distribution from which the variables are drawn.

The probability S_N that the three staircases are ordered decays algebraically with the number of random variables. Interestingly, the decay exponent is nontrivial:

$$S_N \sim N^{-\sigma}, \quad \text{with} \quad \sigma = 1.302931. \tag{1}$$

We obtain this exponent analytically. In the general case, we obtain upper and lower bounds implying that the exponent grows linearly with the number of independent sequences. Interestingly, this family of exponents gives a good approximation for the first-passage exponents found for maxima of random walks.

The rest of this paper is organized as follows. In section II, we investigate statistics of perfectly ordered maxima. We start with two sequences for which the analysis is straightforward. We then present theoretical results for the nontrivial case of three sequences. For the general case, we obtain the leading asymptotic behavior of the

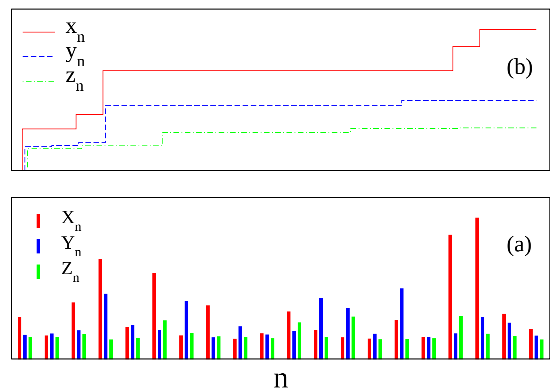

FIG. 1: (Color online) Illustration of perfectly ordered maxima. Panel (a) shows three sequences of random variables (bars), and the panel (b) displays the corresponding running maxima (staircases). The latter sequences are perfectly ordered. The leading sequence is shifted slightly to the left and the trailing sequence is shifted slightly to the right.

scaling exponents. In section III, we treat a related question concerning partially ordered maxima, and show that the exponents obtained in section II are part of an infinite set of families of scaling exponents. We conclude with a summary and a discussion of related open problems (Section IV). The appendix details technical derivations used in the three-sequence case.

II. PERFECT ORDER

We study extreme values of multiple sequences of N uncorrelated random variables. Let us first consider one sequence of random variables

$$\{X_1, X_2, \ldots, X_N\}, \tag{2}$$

where each of the variables X_i is drawn from the same distribution. We restrict our attention to continuous distributions for which there are no ties: $X_i \neq X_j$ for all $i \neq j$. The running maximum x_n is defined as the largest variable to date

$$x_n = \max\{X_1, X_2, \ldots, X_n\}\,, \tag{3}$$

with $n = 1, 2, \ldots, N$. Overall there are N maxima. These maxima are monotonically increasing $x_{n+1} \geq x_n$ for all n, and form the sequence $\{x_1, x_2, \ldots, x_N\}$. Figure 1 illustrates that maxima are *correlated* stochastic variables: by the definition (3), a running maxima involves memory of all preceding random variables.

In this study, we consider m independent sequences of random variables such as (2) and their corresponding maxima defined by (3) and ask: what is the probability that the maxima remain perfectly ordered?

A. Two Sequences

We start with two sequences. The second sequence of random variables, $\{Y_1, Y_2, \ldots, Y_N\}$, is drawn from the same distribution as the first sequence (2). The running maxima for the second sequence are again defined by $y_n = \max\{Y_1, Y_2, \ldots, Y_n\}$ for $n = 1, 2, \ldots, N$. We are interested in the probability S_N that the first set of maxima is always larger then the second set:

$$x_n > y_n, \qquad n = 1, 2, \ldots, N. \tag{4}$$

We term S_N the "survival" probability since the condition (4) defines a first-passage process [21].

The survival probability S_N obeys the *closed* recursion relation

$$S_N = S_{N-1}\left(1 - \frac{1}{2N}\right), \tag{5}$$

subject to the "initial" condition $S_0 = 1$. To appreciate (5) let us combine the two sets of N random variables into a larger set of $2N$ variables. Since all of these random variables are drawn from the same distribution, each variable is equally likely to be the largest. In particular, the variable Y_N is the largest with probability $\frac{1}{2N}$. The probability that the leading sequence remains in the lead at the Nth step is therefore $1 - \frac{1}{2N}$.

Using Eq. (5) one gets $S_1 = \frac{1}{2}$, $S_2 = \frac{3}{8}$, $S_3 = \frac{5}{16}$, and generally

$$S_N = \binom{2N}{N} 2^{-2N}. \tag{6}$$

Importantly, this probability is *universal* as it holds regardless of the distribution from which the random variables are drawn. There are two requirements for equation (6) to hold: (i) all $2N$ variables must be independent and identically distributed, and (ii) the probability distribution that governs these random variables is continuous so that there are no ties.

To obtain the asymptotic behavior for large N, we use the Stirling formula $N! \simeq (2\pi N)^{-1/2}(N/e)^N$ and equation (6). We thus find that the survival probability decays algebraically,

$$S_N \simeq \pi^{-1/2}\, N^{-1/2} \tag{7}$$

for large N. As shown below, the survival probability has a similar algebraic decay in the general case, except that the decay exponent depends on the number of sequences.

The same probability distribution (6) arises in the context of discrete time random walks. Consider two walks that start at the origin. The probability S_N that the positions remain ordered, $x_1(n) > x_2(n)$ for $1 \leq n \leq N$, is given by Eq. (6). This result, known as the Sparre Andersen theorem [28–30], remains valid *regardless* of the step distribution, e.g., it holds for Levy walks with diverging average step length [31].

We also remark that the random variables X_n are uncorrelated, e.g. $\langle X_i X_j\rangle = \langle X_i\rangle\langle X_j\rangle$ for all $i \neq j$; further, there are no correlations between the sequences X_n and Y_n. As a result, the probability Π_N that the actual random variables are always ordered, $X_n > Y_n$ for all n, is purely exponential $\Pi_N = 2^{-N}$. In view of the much slower algebraic decay (7), we conclude that ordered sequences of random variables are much less likely than ordered sequences of maxima.

B. Three Sequences

We now consider three sequences. We denote the third sequence of random variables as $\{Z_1, Z_2, \ldots, Z_N\}$ and the corresponding sequence of maxima as $\{z_1, z_2, \ldots, z_N\}$. We are interested in the probability S_N that the three sequences remain perfectly ordered as illustrated in Fig. 1:

$$x_n > y_n > z_n, \qquad n = 1, 2, \ldots, N. \tag{8}$$

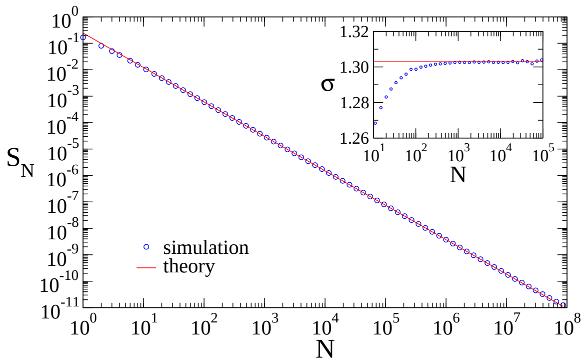

FIG. 2: (Color online) The probability S_N versus N for three sequences. Shown are results of Monte Carlo simulation and the theoretical prediction (25). The numerical simulation results were obtained from roughly 10^{13} independent runs. The inset shows the local slope $\sigma \equiv -d \ln S/d \ln N$.

One immediately finds $S_1 = \frac{1}{3!} = \frac{1}{6}$, but it is challenging to derive S_N for $N \geq 2$ (see Table I). Our numerical simulations (see Fig. 2) show that S_N decays algebraically,

$$S_N \sim N^{-\sigma}, \tag{9}$$

for $N \gg 1$ with $\sigma = 1.3028 \pm 0.0002$.

When there are three sequences, the survival probability S_N no longer obeys a closed recursion equation such as (5). There are, however, closed recursion equations for the probability $P_{N,j}$ that: (i) the maxima are ordered, that is, the condition (8) holds, and (ii) the number of variables from the first sequence that are larger than the intermediate maximum y_N equals j (with $1 \leq j \leq N$) as follows

$$\underbrace{OO\cdots OO}_{3N-1-j}\, Y\, \underbrace{XX\cdots XX}_{j}. \tag{10}$$

In this schematic representation, the $3N$ variables are ordered from smallest (on the left) to largest (on the right), and moreover, the labels are ignored. Further, the symbol O stands for a variable from either one of the three sequences. Importantly, in all the configurations that satisfy the requirement (8), the intermediate maximum is always from the second sequence. Hence, there is no need to keep track of the location of the maximum of the third sequence, and we merely need to ensure that the maximum z_N does not overtake the two other maxima x_N and y_N.

In the configuration (10), the number of variables from the first, second, or third sequence are all equal to N. By definition $P_{N,j}$ is the probability that the maxima are ordered and the number of variables from the first sequence $\{X_1, X_2, \ldots, X_N\}$ that are larger than the intermediate maximum y_N equals j. There are at most N such variables, and hence the the survival probability is the sum of the probabilities $P_{N,j}$:

$$S_N = \sum_{j=1}^{N} P_{N,j}. \tag{11}$$

N	S_N	$(3N)!\,S_N$
1	$\frac{1}{6}$	1
2	$\frac{29}{360}$	58
3	$\frac{4\,597}{90\,720}$	18 388
4	$\frac{5\,393}{149\,688}$	17 257 600
5	$\frac{179\,828\,183}{6\,538\,371\,840}$	35 965 636 600
6	$\frac{352\,052\,449\,513}{16\,005\,934\,264\,320}$	140 820 979 805 200

TABLE I: The survival probability S_N for $N = 1, 2, \ldots, 6$. Also listed are the integers $(3N)!S_N$. These are the number of possible ways to order the $3N$ random variables X_n, Y_n, and Z_n such that the condition (8) holds.

The probability $P_{N,j}$ obeys the recursion equation

$$\begin{aligned} P_{N+1,j} &= \frac{3N+2-j}{3N+3}\,\frac{3N+1-j}{3N+2}\,\frac{3N-j}{3N+1}\,P_{N,j} \\ &+ \frac{3N+2-j}{3N+3}\,\frac{3N+1-j}{3N+2}\,\frac{j}{3N+1}\,P_{N,j-1} \\ &+ \frac{3N+2-j}{(3N+3)(3N+2)(3N+1)}\sum_{k=j}^{N+1}(3N-k)P_{N,k} \\ &+ \frac{3N+2-j}{(3N+3)(3N+2)(3N+1)}\sum_{k=j}^{N+1}k\,P_{N,k-1}. \end{aligned} \tag{12}$$

This recursion equation evaluates the probability that the configuration (10) persists even after three new variables (X_{N+1}, Y_{N+1}, and Z_{N+1}) are added. A step-by-step derivation of (12) is detailed in Appendix A.

The recursion (12) is subject to the "initial" condition $P_{1,k} = \frac{1}{6}\delta_{k,1}$. The first iteration of (12) yields $P_{2,1} = \frac{1}{18}$, $P_{2,2} = \frac{1}{40}$, and hence, $S_2 = \frac{29}{360}$. Table I lists the next few values of the survival probability, obtained from iteration of the recursion equation (12).

We are primarily interested in the $N \to \infty$ asymptotic behavior. In this limit, the probabilities $P_{N,j}$ simplify greatly because the variables N and j become uncorrelated! Numerical evaluation of the recursion (12) shows that $P_{N,j}$ factorizes

$$P_{N,j} \simeq S_N\, p_j. \tag{13}$$

The quantity p_j is the limiting rank distribution of the intermediate maximum, i.e., the probability that there are j variables exceeding the intermediate maximum. From (11) and (13) we confirm that the distribution p_j is properly normalized

$$\sum_{j=1}^{\infty} p_j = 1. \tag{14}$$

We now substitute (13) along with (9) into Eq. (12) and divide both sides by S_N. The leading terms on the right- and left-hand sides [these are of order $O(1)$] cancel. Evaluating corrections to the leading behavior [these are

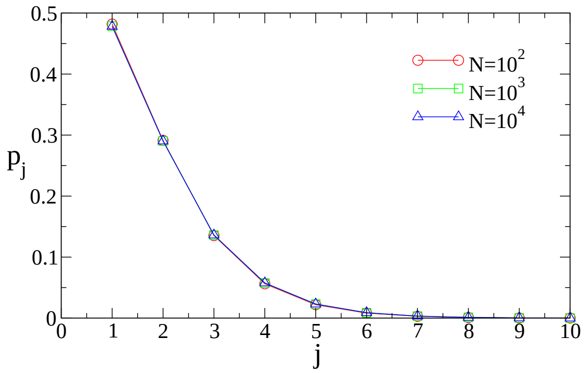

FIG. 3: (Color online) The rank distribution p_j versus j. Shown are results of numerical evaluation of the recursion (12) with increasing values of N.

of order $O(N^{-1})$], we arrive at the remarkably simple recursion relation for the rank distribution

$$\sigma\, p_j = (j+1)\, p_j - \frac{j}{3} p_{j-1} - \frac{1}{3}\sum_{k=j}^{\infty} p_k. \tag{15}$$

In deriving (15), we replaced S_{N+1}/S_N by $1-\sigma\, N^{-1}$. The first, second, and third terms on the right-hand side arise from the first, second, and third terms in (12). The recurrence (15) has to be solved subject to the normalization (14), and the exponent σ is essentially an eigenvalue.

First, we mention a few straightforward results. Summing (15) we express the average rank $\langle j\rangle = \sum_j j\, p_j$ through the exponent σ,

$$\langle j\rangle = 3\sigma - 2. \tag{16}$$

Next, we iterate (15) to obtain the first few probabilities,

$$\begin{aligned} p_1 &= \frac{1}{3(2-\sigma)},\\ p_2 &= \frac{7-3\sigma}{3^2(2-\sigma)(3-\sigma)},\\ p_3 &= \frac{59-48\sigma+9\sigma^2}{3^3(2-\sigma)(3-\sigma)(4-\sigma)}. \end{aligned} \tag{17}$$

The rank distribution is a rapidly decreasing function (Fig. 3). To the leading order, the distribution decays exponentially, $p_j \propto 3^{-j}$, as seen by comparing the leading terms $j\, p_j = (j/3)p_{j-1}$. The algebraic correction to this leading asymptotic behavior can be easily extracted from (15). Indeed, one writes $p_j = 3^{-j} f_j$ and recasts (15) into $\left(\sigma+\frac{1}{2}\right) f_j = (j+1) f_j - j f_{j-1}$ for $j \gg 1$, from which we get $f_j \sim j^{\sigma-1/2}$ and hence

$$p_j \simeq b\, j^{\sigma-1/2} 3^{-j}. \tag{18}$$

The prefactor $b = 1.58063$ is derived in the Appendix (see Appendices B and C).

To determine the eigenvalue σ, we employ the generating function technique. By multiplying (15) by z^{j+1} and summing over all $j \geq 1$, we recast the infinite system of equations (15) into the first-order differential equation

$$(3-z)\frac{dP(z)}{dz} + P(z)\left(\frac{1}{1-z} - \frac{3\sigma}{z}\right) = \frac{z}{1-z}, \tag{19}$$

for the generating function

$$P(z) = \sum_{j\geq 1} p_j\, z^{j+1}\,. \tag{20}$$

The normalization (14) yields $P(1) = 1$ and the average rank (16) gives $P'(1) = 3\sigma - 1$. In addition, we have $P(0) = 0$. The solution to Eq. (19) subject to the above boundary conditions reads

$$\begin{aligned} P(z) &= \sqrt{\frac{1-z}{3-z}}\left(\frac{z}{3-z}\right)^{\sigma} U(z),\\ U(z) &= \int_0^z \frac{du}{(1-u)^{3/2}}\,\frac{(3-u)^{\sigma-1/2}}{u^{\sigma-1}}\,. \end{aligned} \tag{21}$$

To obtain the eigenvalue σ, we evaluate the behavior of this solution in the vicinity of $z = 1$. The solution consists of a regular term $P_{\rm reg}(z)$ and a singular term $P_{\rm sing}(z)$, that is $P(z) = P_{\rm reg}(z) + P_{\rm sing}(z)$. We already know that $P(1) = 1$ and $P'(1) = 3\sigma - 1$, and hence, $P_{\rm reg}(z) = 1 - (3\sigma-1)(1-z) + \cdots$ as $z \to 1$. On the other hand, the leading behavior of the singular term is (see Appendix B)

$$P_{\rm sing}(z) \simeq F(\sigma)\sqrt{1-z}. \tag{22}$$

The amplitude $F(\sigma)$ can be expressed in terms of the Euler gamma function the hypergeometric function

$$F(\sigma) = -\sqrt{\pi}\,\frac{\Gamma(2-\sigma)}{\Gamma\!\left(\frac{3}{2}-\sigma\right)}\; {}_2F_1\!\left(-\tfrac{1}{2}, \tfrac{1}{2}-\sigma; \tfrac{3}{2}-\sigma; -\tfrac{1}{2}\right). \tag{23}$$

The quantity $P'(1)$ must be finite and hence, the leading term of the singular component of the solution must vanish, $F(\sigma) = 0$. Consequently, the exponent σ is a root of the following equation involving hypergeometric function

$${}_2F_1\!\left(-\tfrac{1}{2}, \tfrac{1}{2}-\sigma; \tfrac{3}{2}-\sigma; -\tfrac{1}{2}\right) = 0. \tag{24}$$

The quantity σ is a transcendental number,

$$\sigma = 1.302931\ldots \tag{25}$$

Results of Monte Carlo simulations are in excellent agreement with this theoretical prediction (Fig. 2). In contrast with the behavior (7) where the decay exponent is rational, we see that for three sequences the exponent σ governing the behavior (9) is apparently irrational.

C. General Case

We now discuss the general case where there are m sequences, each containing N random variables. All $m\,N$

m	σ_m
1	0
2	1/2
3	$1.302931\ldots$
4	2.255 ± 0.015
5	3.24 ± 0.03
6	4.2 ± 0.1
7	5.2 ± 0.2

TABLE II: The exponent σ_m versus the number of sequences m. The numerical simulation results represent an average over roughly 10^{13} independent realizations.

random variables are independently drawn from the same distribution function. Based on the results for two and three sequences, we expect that the probability S_N that the m maxima remain ordered is universal, being independent of the details of the distribution function from which the variables are drawn. Moreover, we anticipate that the survival probability decays algebraically,

$$S_N \sim N^{-\sigma_m} \tag{26}$$

for large N. Henceforth, the dependence of S_N on m is left implicit. The decay exponent σ_m depends on the number of sequences m. We already know the values $\sigma_1 = 0$, $\sigma_2 = 1/2$, and $\sigma_3 \cong 1.302931$.

In principle, the recursive equation (12) for the case $m = 3$ can be generalized to higher values of m. However, such a description involves the positions of all $m-2$ intermediate maxima, and it is tedious. Instead, we use Monte Carlo simulations. Table II lists the numerically obtained values for $4 \leq m \leq 7$, while figure 4 shows the very same data points. These results suggest that the exponent grows linearly with the number of sequences. Below, we derive upper and lower bounds for the exponent and from these two bounds, we deduce the linear growth analytically.

It is straightforward to establish a lower bound for the probability S_N and consequently, an upper bound for the exponent σ_m. Let us consider the special case where: (i) the m maxima are properly ordered on the very first step, and (ii) all m maxima remain constant. The probability for the first event is $1/m!$ and the probability for the second scenario is N^{-m}. Hence, we have the lower bound

$$S_N \geq \frac{1}{m!}\, N^{-m}\,. \tag{27}$$

This simple argument gives the upper bound $\sigma_m \leq m$.

To establish an upper bound for the probability S_N and a consequent lower bound for the exponent σ_m, we introduce a natural generalization of the survival probability S_N for the case $m = 2$. Equation (6) corresponds to the situation where there are two sequences of random variables and the first set of maxima is always larger. When there are m sequences, we can similarly ask: what is the probability A_N that the first set of maxima is always ahead? For example, when $m = 3$ this is the probability that $x_n > y_n$ and $x_n > z_n$ for all $n = 1, 2, \ldots, N$. The probability A_N satisfies a straightforward generalization of the recursion (5)

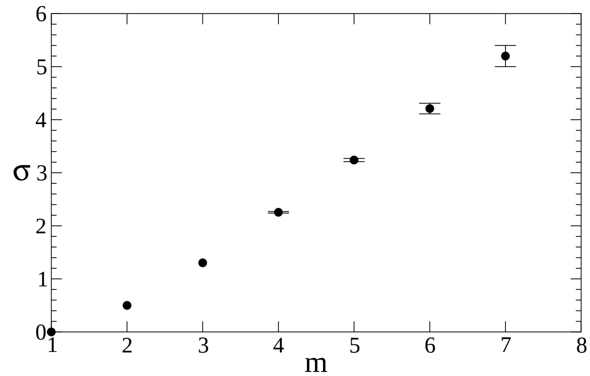

FIG. 4: The exponent σ versus m. Shown are results of numerical simulations, listed in Table I.

$$A_N = A_{N-1}\left(1 - \frac{m-1}{m\,N}\right). \tag{28}$$

When m new variables are added, the probability that a new "global" maximum is set equals $1/N$ and the probability that this new maximum belongs to one of the $m-1$ trailing sequences is simply $\frac{m-1}{m}$. Starting from $A_0 = 1$, the recursion equation (28) gives

$$A_N = \frac{\Gamma(N + \frac{1}{m})}{\Gamma(\frac{1}{m})\,\Gamma(N+1)}, \tag{29}$$

where $\Gamma(x)$ is the Euler Gamma function. When $m = 2$, we recover S_N given in (6). Using the asymptotic relation, $\Gamma(N+a)/\Gamma(N) \simeq N^a$ as $N \to \infty$, we obtain the power-law decay

$$A_N \simeq \frac{1}{\Gamma(\frac{1}{m})}\, N^{-\alpha_m}, \qquad \alpha_m = 1 - \frac{1}{m}, \tag{30}$$

for large N. The family of exponents α_m approaches a constant $\alpha \to 1$ in the limit $m \to \infty$,

We are now in a position to construct an upper bound for S_N. The probability that the first maximum is always larger than all other $m-1$ maxima is given by A_N in (29). The probability that the second maximum is always larger then $m-2$ remaining maxima can be estimated by replacing m with $m-1$ in (29). Similarly, the probability that the third maximum is always larger than the next $m-3$ maxima is approximated by replacing m with $m-2$ in (29), and so on. The product of these $m-1$ probabilities constitutes an upper bound

$$S_N \leq \prod_{k=2}^{m} \frac{\Gamma(N + \frac{1}{k})}{\Gamma(\frac{1}{k})\,\Gamma(N+1)}. \tag{31}$$

If the maxima are always perfectly ordered, then every one of the $m-1$ conditions mentioned above is always

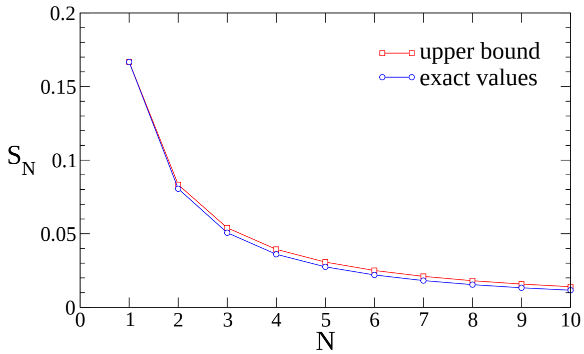

FIG. 5: (Color online) The survival probability S_N for $m = 3$. The upper bound (31) shown in squares is compared with the exact values (circles), obtained from the exact recursion (12).

satisfied. However, since the opposite is not necessarily true, the product of $m-1$ probabilities overestimates the probability S_N. Figure 5 demonstrates that this upper bound gives a very good approximation for the probability S_N. For example, when $m = 3$, the upper bound is $\frac{1}{12} = \frac{30}{360}$ whereas the exact value is $S_2 = \frac{29}{360}$ (see Table I). The lower bound (27) is a much poorer approximation, in comparison.

By using the upper bound (31) and the asymptotic behavior (30) we arrive at the lower bound,

$$\sigma_m \geq \alpha_2 + \alpha_3 + \cdots \alpha_m \,. \tag{32}$$

Hence, we conclude that the exponent σ_m is bounded from above and from below as follows,

$$m - \left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{m}\right) \leq \sigma_m \leq m\,. \tag{33}$$

For example, for $m = 3$ we have the bounds $\frac{7}{6} \leq \sigma_3 \leq 3$, and indeed, the lower bound is much tighter compared with the upper bound. Most significantly, the two bounds establish the linear growth (see figure 4)

$$\sigma_m \simeq m. \tag{34}$$

The asymptotic behavior $1 + \frac{1}{2} + \cdots + \frac{1}{m} \simeq \ln m + \gamma$ (here $\gamma \cong 0.577215$ is the Euler constant) shows that the deviation from linear growth is at most logarithmic. The linear growth (34) is in contrast with the quadratic $m(m-1)/4$ growth of the exponent characterizing ordering of random walks [32–36]. Finally, we mention a numerical observation: the empirical formula $\tilde{\sigma}_m = (m-1)^2/m$, which is exact for $m = 1$ and $m = 2$, yields an excellent approximation for the values listed in Table I.

The exponents σ_m provide an excellent approximation to an analogous set of exponents that characterize random walks. Let s_N be the probability that the maxima of the positions of m independent random walks, each consisting of N steps, are ordered. This quantity decays algebraically $s_N \sim N^{-\nu_m}$. For two random walks, it was established analytically that $\nu_2 = \frac{1}{4}$ [26]. In view of the identity $\nu_2 = \sigma_2/2$, we compare the values $\sigma_m/2$ with the values of ν_m obtained using numerical simulations for $m = 3, 4, 5, 6$ [26]. Table III shows that the ordering exponents, obtained in the present study for uncorrelated random variables, provide an excellent approximation for the ordering exponents that characterize the maximal positions of random walks, at least for small values of m. Random walk positions are certainly correlated random variables, and therefore, the results above, which are strictly valid for uncorrelated random variables, may in practice provide useful approximations for certain classes of correlated random variables.

m	ν_m	$\sigma_m/2$
2	1/4	1/4
3	0.653	0.651465
4	1.13	1.128
5	1.60	1.62
6	2.01	2.10

TABLE III: The exponent ν_m for ordering of random walk maxima [26] versus the values $\sigma_m/2$, see Table II.

We also mention that the quantity s_N is not universal as it depends, albeit rather weakly, on the step length distribution [37]. Nevertheless, the decay exponents ν_m listed in Table III are universal, namely they are valid as long as the step length distribution is symmetric and has finite variance [26].

III. PARTIAL ORDER

The exponents σ_m characterizing the statistics of perfectly ordered maxima are part of a broader family of exponents. We have already obtained one such family of exponents, $\alpha_m = 1 - \frac{1}{m}$, that characterize the probability A_N that the first sequence of maxima is always larger than all other $m-1$ maxima. For example, when $m = 2$ the requirement that x_n is largest is equivalent to the requirement $x_n > y_n$ for all $n = 1, 2, \ldots, N$. Therefore, $S_N = A_N$ for $m = 2$, and hence $\alpha_2 = \sigma_2$. Another trivial identity is $\alpha_1 = \sigma_1$.

Similarly, we can introduce the probability B_N that the first two sets of maxima remain ordered and exceed the other $m-2$ maxima. In other words, the sequence of maxima $\{x_1, x_2, \ldots, x_N\}$ is always the largest, and the sequence $\{y_1, y_2, \ldots, y_N\}$ is always the second largest. Therefore, the probabilities B_N and S_N are identical when there are two or three sequences. We expect the algebraic decay

$$B_N \sim N^{-\beta_m}, \tag{35}$$

with $\beta_2 = \sigma_2$ and $\beta_3 = \sigma_3$.

The recursion equation (12) is straightforward to generalize from three to $m \geq 2$ sequences, thereby giving a recursive calculation of the probability B_N. By repeating the analysis leading to (12) we find that the rank

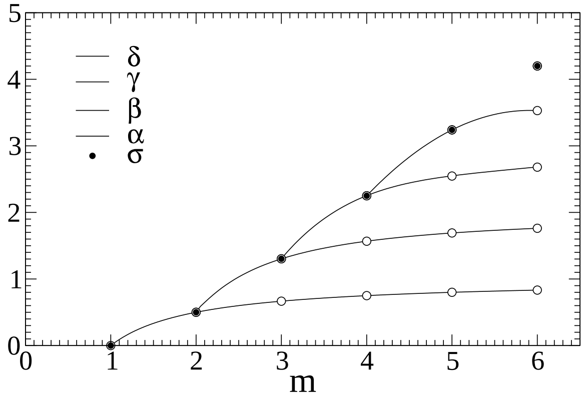

FIG. 6: Families of ordering exponents. Shown are the first four families of exponents; the values of α_m, β_m, γ_m, and δ_m are listed in Table III (circles). The ordering exponents σ_m describing perfectly ordered maxima (bullets) are also shown. The analytic curves for α and β are given in equations (30) and (37). The curves for γ and δ represent a fourth-order polynomial best fit to the simulation results.

distribution p_j obeys the recursion equation

$$\beta\, p_j = (j+1)\, p_j - \frac{j}{m} p_{j-1} - \frac{1}{m}\sum_{k=j}^{\infty} p_k \tag{36}$$

and the normalization condition (14). The exponent $\beta \equiv \beta_m$ is again an eigenvalue, and the average rank can be expressed through this exponent, $\langle j\rangle = [1 + m(\beta-1)]/(m-2)$. Moreover, the rank distribution has the following tail $p_j \sim j^{\beta-(m-2)/(m-1)}\, m^{-j}$. By following the analysis in Sec. II B, it is straightforward to show that exponent β is a root of a transcendental equation involving the hypergeometric function,

$${}_2F_1(-\mu, 1-\mu-\beta; 2-\mu-\beta; -\mu) = 0, \tag{37}$$

with the shorthand notation $\mu = 1/(m-1)$. When $m = 3$ this equation coincides with (24) and hence $\beta_3 = \sigma_3$. The next three values are $\beta_4 = 1.56479$, $\beta_5 = 1.69144$, and $\beta_6 = 1.76164$. One can also deduce the asymptotic behavior, $2 - \beta_m \simeq m^{-1}$, when $m \gg 1$. Just like the family of exponents α_m, the curve β_m saturates in the large m limit: $\beta_m \to 2$. While the parameter m is discrete, the solution to the transcendental equation (37) can be evaluated for all $m \geq 2$ and the resulting *continuous* curve is shown in Fig. 6.

There is an infinite set of probabilities generalizing A_N and B_N: The probability C_N that the first three maxima (out of a total of $m \geq 3$) remain ordered $x_n > y_n > z_n$ and exceed all others for all $1 \leq n \leq N$; the probability D_N that the first four maxima (out of total of $m \geq 4$) remain ordered and exceed all others, and so on. These probabilities are characterized by three families of exponents,

$$C_N \sim N^{-\gamma_m}, \quad D_N \sim N^{-\delta_m}. \tag{38}$$

m	α	β	γ	δ
1	0			
2	1/2	1/2		
3	2/3	1.302931	1.302931	
4	3/4	1.56479	2.255	2.255
5	4/5	1.69144	2.547	3.24
6	5/6	1.76164	2.680	3.53

TABLE IV: The first few families of exponents. The exponents α_m and β_m are obtained analytically. When $m \geq 4$, the exponents γ_m and δ_m are from numerical simulations.

We have $\beta_3 = \gamma_3 = \sigma_3$, $\gamma_4 = \delta_4 = \sigma_4$, and $\delta_5 = \sigma_5$. The first four families of exponents are shown in Fig. 6. These families of exponents form an intriguing structure that resembles a scallop. The smallest two integer exponents in the mth family coincide with the exponents σ_m and σ_{m+1}. An interesting question for future research is whether the families of ordering exponents adhere to a universal scaling curve when the number of conditions such as (8) becomes very large.

IV. CONCLUSIONS

We have shown that the probability that the running maxima of independent sets of random variables are ordered decays algebraically with the number of variables. The scaling exponents that characterize this decay are in general nontrivial. When there are three sequences, the scaling exponent is eigenvalue of a recurrence equation, and it is also a root of a transcendental equation. The scaling exponents grow linearly with the number of independent sequences. We have also seen that ordering exponents for uncorrelated random variables provide an excellent approximation for the corresponding set of exponents that characterize maximal positions of random walks.

The key observation that allowed us to treat the three-variable case analytically is that the rank of the intermediate maximum decouples from the sequence length in the asymptotic regime. This observation, combined with the power-law decay of the overall survival probability reduces the complexity of the underlying combinatorial problem: enumerating the number of ways to order the random variables such that the respective maxima remain ordered.

One can study the probability that the running maxima of the first k sequences are ordered and exceed the maxima of remaining $m-k$ sequences. We have examined such probabilities for $k = 1, 2, 3, 4$ and derived analytic expressions for the exponents for the case with persistent leader ($k = 1$) and the case with persistent leader and persistent second leader. We have seen that there are families of exponents, in some cases equivalent to eigenvalues, that form an intriguing structure. For uncorrelated random variables, the intersection points of these eigenvalue families mark a linear envelope. It would therefore be interesting to investigate the corresponding exponent families for ordered maxima of Brownian trajectories, where that envelope shows quadratic growth [32–36].

Furthermore, there are many similar survival probabilities, for example, the probability L_N that the running maxima of one sequence are never the smallest. Numerically, we observe the algebraic decay $L_N \sim N^{-\lambda_m}$. The first exponent is obvious, $\lambda_2 = 1/2$. Our numerical simulations yield a slowly decreasing set of exponents: $\lambda_3 = 0.3801$, $\lambda_4 = 0.3145$, $\lambda_5 = 0.2726$, and $\lambda_6 = 0.2430$.

We emphasize that the probability that the actual random variables remain perfectly ordered decays exponentially while the probability that the running maxima maintain perfect order decays much more slowly, namely algebraically, with sequence length. Hence, it is far more likely to observe ordered maxima. In several contexts such as temperature records [9] or stock market time series [15], record highs or record lows are followed very closely to see for example if one year is the hottest or if one stock is consistently outperforming its peers. Hence, we expect that the questions we investigated theoretically in this study may be of practical relevance in analysis of time series. Moreover, consistently ordered extreme values provide a natural way to quantify persistent upward or downward trends in the data.

We anticipate that the actual survival probabilities, particularly the values of the scaling exponents, may be measurable in empirical data, including those with correlated random variables. One such example is inter-event times for earthquakes where a series of recent studies demonstrate how "persistence" properties of maxima of uncorrelated variables provide excellent predictions for empirical observations [11, 22, 23].

The concept of ordered maxima can also be employed in analysis of physical systems. For example, in disordered materials such as ensembles granular particles, one may be able to probe density fluctuations in different regions and in particular, extreme values of the packing fraction versus time. Measurement of the evolution of extreme values as described in the current investigation can be used to identify "hot-spots" or regions of persisting high density or high effective temperatures [39–41].

We acknowledge financial support through US-DOE grant DE-AC52-06NA25396 for support (EB & NWL).

[1] M. R. Leadbetter, G. Lindgren, H. Rootzen, *Extremes and Related Properties of Random Sequences and Processes*, (Springer, Berlin, 1983).
[2] E. I. Gumbel, *Statistics of Extremes* (Dover, New York 2004).
[3] R. S. Ellis, *Entropy, Large Deviations, and Statistical Mechanics* (Springer, Berlin 2005).
[4] S. I. Resnick, *Extreme Values, Regular Variation and Point Processes* (Springer, Berlin, 2007).
[5] B. C. Arnold, N. Balakrishnan and H. N. Nagraja, *Records* (Wiley-Interscience, 1998).
[6] V. B. Nevzorov, *Records: Mathematical Theory*, Translation of Mathematical Monographs **194** (American Mathematical Society, Providence, RI, 2001).
[7] S. Redner and M. R. Petersen, Phys. Rev. E **74**, 061114 (2006)
[8] W. I. Newman, B. D. Malamud, and D. L. Turcotte, Phys. Rev. E **82**, 066111 (2010).
[9] G. Wergen, A. Hense, and J. Krug, Climate Dynamics **42**, 1275 (2014).
[10] R. M. Vogel, A. Zafirakou-Koulouris, and N. C. Matalas, Water Resources Research **37**, 1723 (2001).
[11] W. I. Newman, Phys. Rev. E **89**, 062113 (2013).
[12] R. Shcherbakov, J. Davidsen, and K. F. Tiampo, Phys. Rev. E **87**, 052811 (2013).
[13] Q. H. Wen, X. Zhang, Y. Xu, and B. Wang, Geophys. Res. Lett. **40**, 1171 (2013).
[14] P. Embrechts, G. Klüppelberg and T. Mikosch, *Modelling Extremal Events for Insurance and Finance* (Springer-Verlag, Berlin, 1997).
[15] J. -P. Bouchaud and M. Potters, *Theory of Financial Risk and Derivative Pricing* (Cambridge University Press, Cambridge 2003).
[16] S. Y. Novak, *Extreme Value Methods with Applications to Finance* (Chapman & Hall/CRC Press, London, 2011).
[17] G. Wergen, M. Bogner, and J. Krug, Phys. Rev. E **83**, 051109 (2011).
[18] S. N. Majumdar, C. Sire, A. J. Bray, and S. J. Cornell, Phys. Rev. Lett. **77**, 2867 (1996).
[19] B. Derrida, V. Hakim, R. Zeitak, Phys. Rev. Lett. **77**, 2871 (1996).
[20] A. J. Bray, S. N. Majumdar, and G. Schehr, Adv. Phys. **62**, 225 (2013).
[21] S. Redner, *A Guide to First-Passage Processes* (Cambridge University Press, Cambridge, 2001).
[22] E. Ben-Naim and P. L. Krapivsky, Phys. Rev. E **88**, 022145 (2013).
[23] P. W. Miller and E. Ben-Naim, J. Stat. Mech. P10025 (2013).
[24] E. Ben-Naim and P. L. Krapivsky, J. Phys. A **47**, 255002 (2014).
[25] P. L. Krapivsky, S. Redner and E. Ben-Naim, *A Kinetic View of Statistical Physics* (Cambridge University Press, Cambridge, UK, 2010).
[26] E. Ben-Naim and P. L. Krapivsky, Phys. Rev. Lett. **113**, 030604 (2014).
[27] J. Randon-Furling, EPL **109**, 40015 (2015).
[28] E. Sparre Andersen, Math. Scand. **1**, 263 (1953); *ibid.* **2**, 195 (1954).
[29] W. Feller, *An Introduction to Probability Theory and its Applications*, Vol. I (Wiley, New York, 1968).
[30] S. N. Majumdar, Physica A **389**, 4299 (2010).
[31] P. Lévy, *Processus Stochastiques et Mouvement Brownien* (Gauthier-Villars, Paris, 1948).
[32] M. E. Fisher, J. Stat. Phys. **34**, 667 (1984).
[33] D. A. Huse and M. E. Fisher, Phys. Rev. B **29**, 239

(1984).
[34] D. J. Grabiner, Ann. Inst. Poincare: Prob. Stat. **35**, 177 (1999).
[35] P. L. Krapivsky and S. Redner, J. Phys. A **29**, 5347 (1996); D. ben-Avraham, B. M. Johnson, C. A. Monaco, P. L. Krapivsky, and S. Redner, J. Phys. A **36**, 1789 (2003).
[36] E. Ben-Naim and P. L. Krapivsky, J. Phys. A **43**, 495007 (2010); J. Phys. A **43**, 495008 (2010).
[37] E. Ben-Naim and P. L. Krapivsky, unpublished.
[38] I. M. Gelfand and G. E. Shilov, *Generalized Functions*, vol I. (Academic Press, Boston, 1964).
[39] J. B. Knight, C. G. Fandrich, C. N. Lau, H. M. Jaeger, and S. R. Nagel, Phys. Rev. E **51**, 3957 (1995).
[40] C. S. O'Hern, A. A. Langer, A. J. Liu, and S. R. Nagel, Phys. Rev. Lett. **88**, 075507 (2002).
[41] T. Aste, M. Saadatfar, T. J. Senden, Phys. Rev. E **71**, 061302 (2005).

Appendix A: Derivation of Eq. (12)

The quantity $P_{N,j}$ is the probability that: (i) the condition $x_n > y_n > z_n$ holds for all $n = 1, 2, \ldots, N$ and that (ii) there are exactly j variables from the sequence $\{X_1, X_2, \ldots, X_N\}$ that are larger than the intermediate maximum y_N. We introduce a related auxiliary quantity $Q_{N,j}$ which is the probability that: (i) the condition $x_n > y_n > z_n$ holds for all $n = 1, 2, \ldots, N$ and that (ii) there are exactly j variables from the sequence $\{X_1, X_2, \ldots, X_{N+1}\}$ that are larger than the intermediate maximum y_N. The probability $Q_{N,j}$ is directly related to the quantity $P_{N,j}$,

$$Q_{N,j} = \frac{3N-j}{3N+1}\, P_{N,j} + \frac{j}{3N+1}\, P_{N,j-1}. \tag{A1}$$

To obtain this recursion we consider how the configuration (10) changes with the addition of the variable X_{N+1}. The rank of the intermediate maximum remains the same if $X_{N+1} < y_N$. The probability of this event equals $\frac{3N-j}{3N+1}$ and hence the first term. The second term accounts for the complementary scenario, $X_{N+1} > y_N$, where the rank of the intermediate maximum y_N increases by one.

Next, we consider the auxiliary quantity $R_{N,j}$ which is the probability that: (i) the condition $x_n > y_n$ holds for all $n = 1, 2, \ldots, N+1$, (ii) the condition $y_n > z_n$ holds for all $n = 1, 2, \ldots, N$, and (iii) there are exactly j variables from the sequence $\{X_1, X_2, \ldots, X_{N+1}\}$ that are larger than the intermediate maximum y_{N+1}. The probability $R_{N,j}$ follows from the quantity $Q_{N,j}$,

$$R_{N,j} = \frac{3N+1-j}{3N+2}\, Q_{N,j} + \frac{1}{3N+2} \sum_{k=j}^{N+1} Q_{N,k}. \tag{A2}$$

Now we have to compare the random variable Y_{N+1} with the maximum y_N. If $Y_{N+1} < y_N$, the rank of the intermediate maximum does not change. The probability of this event is $\frac{3N+1-j}{3N+2}$ and hence, the first term. The second term accounts for the complementary situation in which the rank of the intermediate maximum increases.

Finally, the probability $P_{N+1,j}$ follows immediately from $R_{N,j}$

$$P_{N+1,j} = \frac{3N+2-j}{3N+3} R_{N,j}. \tag{A3}$$

To obtain this equation, we consider the addition of the variable Z_{N+1}. We must ensure that the maximum from the trailing sequence does not overtake the intermediate maximum, $Z_{N+1} < y_{N+1}$, and the probability for this event is simply $\frac{3N+2-j}{3N+3}$. By substituting (A1) into (A2) and then substituting (A2) into (A3) we obtain the recursion equation (12).

Appendix B: The generating function $P(z)$ in the limit $z \to 1$

We evaluate the asymptotic behavior of the function $U(z)$ which appears in (21) as $z \to 1$ using the following steps:

$$\begin{aligned}
U(z) &= \int_0^z \frac{du}{(1-u)^{3/2}} \frac{(3-u)^{\sigma-1/2}}{u^{\sigma-1}} \qquad\qquad \text{(B1)}\\
&= \int_0^z \frac{du}{(1-u)^{3/2}} \left[\frac{(3-u)^{\sigma-1/2}}{u^{\sigma-1}} - 2^{\sigma-1/2}\right] - 2^{\sigma+1/2} + 2^{\sigma+1/2}(1-z)^{-1/2}\\
&= 2^{\sigma+1/2}\left(\frac{1}{\sqrt{1-z}} - 1\right) + \int_0^1 \frac{du}{(1-u)^{3/2}} \left[\frac{(3-u)^{\sigma-1/2}}{u^{\sigma-1}} - 2^{\sigma-1/2}\right] - \int_z^1 \frac{du}{(1-u)^{3/2}} \left[\frac{(3-u)^{\sigma-1/2}}{u^{\sigma-1}} - 2^{\sigma-1/2}\right]\\
&= 2^{\sigma+1/2}\left\{\frac{1}{\sqrt{1-z}} - 1 + \frac{1}{2}\int_0^1 \frac{du}{(1-u)^{3/2}} \left[u^{1-\sigma}\left(\frac{3-u}{2}\right)^{\sigma-1/2} - 1\right] - \frac{6\sigma-5}{4}\sqrt{1-z} + O\left[(1-z)^{3/2}\right]\right\}\\
&= 2^{\sigma+1/2}\left(\frac{1}{\sqrt{1-z}} + F(\sigma) - \frac{6\sigma-5}{4}\sqrt{1-z}\right) + O\left[(1-z)^{3/2}\right].
\end{aligned}$$

The quantity $F(\sigma)$ can be expressed in terms of the hypergeometric function,

$$\begin{aligned}
F(\sigma) &= \frac{1}{2}\int_0^1 \frac{du}{(1-u)^{3/2}}\left[u^{1-\sigma}\left(\frac{3-u}{2}\right)^{\sigma-1/2}-1\right]-1 \\
&= \frac{1}{2}\int_0^1 dv\, v^{-3/2}\Big[(1-v)^{1-\sigma}(1+\tfrac{1}{2}v)^{\sigma-1/2}-1\Big]-1 \\
&= \frac{1}{2}\frac{\Gamma(-\frac{1}{2})\,\Gamma(2-\sigma)}{\Gamma(\frac{3}{2}-\sigma)}\,{}_2F_1(-\tfrac{1}{2},\tfrac{1}{2}-\sigma;\tfrac{3}{2}-\sigma;-\tfrac{1}{2})-\frac{1}{2}\int_0^1 dv\, v^{-3/2}-1 \\
&= -\sqrt{\pi}\,\frac{\Gamma(2-\sigma)}{\Gamma\big(\frac{3}{2}-\sigma\big)}\;{}_2F_1(-\tfrac{1}{2},\tfrac{1}{2}-\sigma;\tfrac{3}{2}-\sigma;-\tfrac{1}{2}).
\end{aligned} \tag{B2}$$

First, we note that the integral which specifies $F(\sigma)$ is finite. In deriving the third line we used the integral representation of the hypergeometric function

$$\int_0^1 dv\, v^{b-1}(1-v)^{c-b-1}(1-zv)^{-a} = \frac{\Gamma(b)\,\Gamma(c-b)}{\Gamma(c)}\;{}_2F_1(a,b;c;z), \tag{B3}$$

and the identity ${}_2F_1(a,b;c;z) = {}_2F_1(b,a;c;z)$. In deriving the fourth line, we used the identity (valid in the sense of distributions, or generalized functions [38]) $\int_0^1 dv\, v^\lambda = \frac{1}{\lambda+1}$. This identity is certainly valid in the complex plane where $\mathrm{Re}(\lambda) > -1$, and then it may be analytically continued into the whole complex plane, in particular to the value $\lambda = -3/2$.

Appendix C: The generating function $P(z)$ in the limit $z \to 3$

The integral representation (21) implicitly assumes that $z < 1$. To obtain the asymptotic behavior of the generating function $P(z)$ in the limit $z \to 3$, we write the solution of Eq. (19) in a different form

$$\begin{aligned}
P(z) &= \sqrt{\frac{z-1}{3-z}}\left(\frac{z}{3-z}\right)^{\sigma}\tilde{U}(z)\,, \\
\tilde{U}(z) &= \tilde{U}(3) + \int_z^3 \frac{du}{(u-1)^{3/2}}\,\frac{(3-u)^{\sigma-1/2}}{u^{\sigma-1}}\,.
\end{aligned} \tag{C1}$$

To determine the constant $\tilde{U}(3)$ we transform $\tilde{U}(z)$ as follows:

$$\begin{aligned}
\tilde{U}(z) &= \tilde{U}(3) + \int_z^3 \frac{du}{(u-1)^{3/2}}\left[\frac{(3-u)^{\sigma-1/2}}{u^{\sigma-1}}-2^{\sigma-1/2}\right]+2^{\sigma-1/2}\int_z^3\frac{du}{(u-1)^{3/2}} \\
&= \tilde{U}(3) + \int_z^3 \frac{du}{(u-1)^{3/2}}\left[\frac{(3-u)^{\sigma-1/2}}{u^{\sigma-1}}-2^{\sigma-1/2}\right]+2^{\sigma+1/2}\left[\frac{1}{\sqrt{z-1}}-\frac{1}{\sqrt{2}}\right] \\
&= \frac{2^{\sigma+1/2}}{\sqrt{z-1}}+\tilde{U}(3)-2^{\sigma}+\int_1^3\frac{du}{(u-1)^{3/2}}\left[\frac{(3-u)^{\sigma-1/2}}{u^{\sigma-1}}-2^{\sigma-1/2}\right]-\int_1^z\frac{du}{(u-1)^{3/2}}\left[\frac{(3-u)^{\sigma-1/2}}{u^{\sigma-1}}-2^{\sigma-1/2}\right] \\
&= \frac{2^{\sigma+1/2}}{\sqrt{z-1}}-\int_1^z\frac{du}{(u-1)^{3/2}}\left[\frac{(3-u)^{\sigma-1/2}}{u^{\sigma-1}}-2^{\sigma-1/2}\right].
\end{aligned}$$

The leading singular contribution to the $P(z)$ equals the sum of the second, third, and fourth terms on the third line, and these terms cancel so that the generating function is regular at $z = 1$. In the limit $z \to 3$ we have $P(z) \simeq \sqrt{2}\,3^\sigma\,\tilde{U}(3)\,(3-z)^{-\sigma-1/2}$. In addition, we use the expansion $(1-x)^{-a} = \sum_{n\geq 0}\Gamma(a+n)\,x^n/[\Gamma(a)\Gamma(n+1)]$ to deduce the tail behavior (18) with

$$b = \sqrt{\frac{2}{3}}\;\frac{1}{3\,\Gamma(\sigma+1/2)}\left\{2^\sigma-\int_1^3\frac{du}{(u-1)^{3/2}}\left[\frac{(3-u)^{\sigma-1/2}}{u^{\sigma-1}}-2^{\sigma-1/2}\right]\right\}. \tag{C2}$$