Universal Order Statistics of Random Walks

Grégory Schehr*
Université Paris-Sud, CNRS, LPT, 91405 Orsay Cedex, France

Satya N. Majumdar†
Université Paris-Sud, CNRS, LPTMS, 91405 Orsay Cedex, France

(Received 15 November 2011; published 26 January 2012)

We study analytically the order statistics of a time series generated by the positions of a symmetric random walk of \( n \) steps with step lengths of finite variance \( \sigma^2 \). We show that the statistics of the gap \( d_{k,n} = M_{k,n} - M_{k+1,n} \) between the \( k \)th and the \((k + 1)\)th maximum of the time series becomes stationary, i.e., independent of \( n \) as \( n \to \infty \) and exhibits a rich, universal behavior. The mean stationary gap exhibits a universal algebraic decay for large \( k \), \( (d_{k,\infty})/\sigma \sim 1/\sqrt{2\pi k} \), independent of the details of the jump distribution. Moreover, the probability density (pdf) of the stationary gap exhibits scaling, \( P(d_{k,\infty} = \delta) = (\sqrt{\delta}/\sigma)P(\sqrt{\delta}/\sigma) \), in the regime \( \delta \sim (d_{k,\infty}) \). The scaling function \( P(x) \) is universal and has an unexpected power law tail, \( P(x) \sim x^{-4} \) for large \( x \). For \( \delta \gg (d_{k,\infty}) \) the scaling breaks down and the pdf gets cut off in a nonuniversal way. Consequently, the moments of the gap exhibit an unusual multiscaling behavior.

DOI: 10.1103/PhysRevLett.108.040601

PACS numbers: 05.40.Fb, 02.50.Cw, 05.45.Tp

During the last 50 years, extreme value statistics (EVS), the statistics of the maximum or the minimum of a set of random variables, have found many applications, ranging from engineering [1] to environmental sciences [2] or finance [3,4], where rare and extreme events may have drastic consequences. It was demonstrated [5] that EVS also plays a major role in the physics of complex and disordered systems. Therefore finding the distribution of the maximum \( x_{\text{max}} \) (or the minimum \( x_{\text{min}} \)) of a set of \( n + 1 \) random variables \( \{x_0, x_1, x_2, \ldots, x_n\} \) has been the subject of intense activity not just for independent and identically distributed (iid) random variables [1], but also recently for strongly correlated random variables [6–14] that are often more relevant in physical contexts. While the statistics of the extremum \( x_{\text{max}} \) (or \( x_{\text{min}} \)) is important, they concern the fluctuations of a single value among a typically large sample and a natural question is then the following: are these extremal values isolated, i.e., far away from the others, or are there many other events close to them? Such questions have led to the study of the density of states of near-extreme events [15,16].

This is, for instance, a crucial question in disordered systems, where the low temperature properties are governed by excited states close to the ground state. A natural way to characterize this phenomenon of crowding of near-extreme events is via the order statistics, i.e., arranging the random variables \( x_m \)'s in decreasing order of magnitude \( M_{1,n} > \cdots > M_{k,n} > \cdots > M_{n+1,n} \) where \( M_{k,n} \) denotes the \( k \)th maximum of the set \( \{x_0, x_1, \ldots, x_n\} \). Evidently, \( x_{\text{max}} = M_{1,n} \), while \( x_{\text{min}} = M_{n+1,n} \). A set of useful observables that are naturally sensitive to the crowding of extremum are the gaps between the consecutive ordered maxima: \( d_{k,n} = M_{k,n} - M_{k+1,n} \) denoting the \( k \)th gap.

While the study of order (or gap) statistics has received considerable interest in statistics literature, e.g., in the context of system reliability [17], the available results are restricted only to iid variables. In contrast, there hardly exist analytical results for the gap statistics for strongly correlated random variables. The importance of order statistics for such correlated variables came up recently in several physical contexts, notably in the study of the branching Brownian motion [18] and also for \( 1/f \) signals [19] with an application to the statistical analysis of cosmological observations [20]. Any solvable model for the order statistics for correlated variables would thus be welcome and this Letter takes a step in that direction.

In this Letter, we present exact analytical results for the order statistics and the gap distribution of a time series \( \{x_0, x_1, \ldots, x_n\} \) where \( x_m \) represents the position of a random walker at discrete time \( m \). The walker starts at \( x_0 = 0 \) at time \( 0 \) and at each discrete step evolves via \( x_m = x_{m-1} + \eta_m \), where the noise \( \eta_m \)'s are iid jump lengths, each drawn from a symmetric and continuous distribution \( f(\eta) \) with zero mean and a finite variance \( \sigma^2 = \int_{-\infty}^{\infty} \eta^2 f(\eta) d\eta \). Even though the jump lengths are uncorrelated, the entries \( x_m \)'s are clearly correlated and represent perhaps the simplest, yet most ubiquitous correlated time series (discrete-time Brownian motion) with a large variety of applications [21,22], including for instance in queuing theory [23]—where \( x_m \) represents the length of a single server queue at time \( m \)—or in finance where \( x_m \) represents the logarithm of the price of a stock at time \( m \) [24]. Even for this relatively simple correlated time series, we show that the gap distribution exhibits a rather rich and universal behavior.

It is useful to summarize our main results. For large \( n \), one finds that \( (M_{k,n})/\sigma = \sqrt{2n/\pi} + O(1) \), independent of \( k \). Thus the property of the crowding of extremum
for the moments of the stationary gap: becomes universal for large $n$.

The initial value $x$ retains a rich, nontrivial $k$ dependence which becomes universal for large $k$, i.e., independent of the details of the jump distribution $f(\eta)$. We compute the stationary mean gap $\bar{d}_k = \langle d_{k,\infty} \rangle$ exactly for all $k$, for arbitrary $f(\eta)$ and show that, when expressed in units of $\sigma$, it has a universal algebraic tail, $\bar{d}_k/\sigma \approx 1/\sqrt{2\pi k}$ for large $k$. Next, we compute exactly the full pdf of the stationary gap $p_k(\delta) = \Pr(d_{k,\infty} = \delta)$ for the exponential jump distribution, $f(\eta) = b^{-1}\exp(-|\eta|/b)$ and show that for large $k$, there is a scaling regime when $\delta \sim \langle d_{k,\infty} \rangle \approx \sigma/\sqrt{2\pi k}$ where the pdf scales as, $p_k(\delta) \approx (\sqrt{k}/\sigma)\sqrt{\delta} \sqrt{2\pi k}$, with a nontrivial scaling function

$$P(x) = 4\left[ \sqrt{\frac{2}{\pi}}(1 + 2x^2) - e^{2x^2}x(4x^2 + 3)\text{erfc}(\sqrt{2x}) \right].$$

where erfc$(z) = (2/\sqrt{\pi}) \int_z^\infty e^{-t^2} dt$ is the complementary error function. While we were unable to compute the gap pdf for arbitrary $f(\eta)$, our numerical simulations provide strong evidence that the scaling function $P(x)$ in Eq. (1) is actually universal, i.e., independent of $f(\eta)$. Somewhat unexpectedly, we find that this universal scaling function has an algebraic tail $P(x) \sim x^{-4}$ for large $x$. For $\delta > \langle d_{k,\infty} \rangle \approx \sigma/\sqrt{2\pi k}$, the pdf gets cut off in a nonuniversal fashion. This is shown to have interesting consequences for the moments of the stationary gap: $\langle d_{k,\infty}^p \rangle \sim k^{-p/2}$ for $p < 3$, while $\langle d_{k,\infty}^p \rangle \sim k^{-3/2}$ for $p > 3$.

We start with the statistics of the $k$th maximum $M_{k,n}$ of the random walk $x_m = x_{m-1} + \eta_m$ of $n$ steps, starting from the initial value $x_0 = 0$. The goal is to write down an evolution equation for the cumulative distribution of the $k$th maximum $F_{k,n}(x) = \Pr[M_{k,n} \leq x]$. The event $M_{k,n} \leq x$ means that we have at most $(k-1)$ points above the level $x$ between step 1 and $n$. To keep track of this event, it is convenient first to define an auxiliary quantity $q_{k,n}(x)$ denoting the probability that the random walk, starting at $x_0 = x$, has $k$ points below 0 from step 1 to step $n$. It is then easy to see that $F_{k,n}(x)$ can be expressed as the sum

$$F_{k,n}(x) = \left\{ \begin{array}{ll}
\sum_{m=0}^{k-1} q_{m,n}(x), & x > 0, \\
\sum_{m=0}^{-k} q_{n-m,n}(-x), & x < 0,
\end{array} \right.$$

where we used that $f(\eta)$ is symmetric and continuous.

The next step is to write a backward recurrence equation for $q_{k,n}(x)$ by considering the stochastic jump $x \to x'$ at the first step (see Fig. 2) and then subsequently using the Markov property of the evolution. One gets, for $n \geq 1$,

$$q_{k,n}(x) = \int_0^x q_{k,n-1}(x')f(x' - x)dx' + \int_{-x}^0 q_{n-k,n-1}(-x')f(x' - x)dx'.$$

starting from $q_{0,0}(x) = 1$. The first term corresponds to a jump from $x > 0$ to $x' > 0$ while the second term corresponds to a jump from $x > 0$ to $x' < 0$ (see Fig. 2).

The integral equation (3) is of the Wiener-Hopf type which are generically hard to solve for arbitrary jump distribution $f(x)$. However, for the special case $f(x) = \frac{1}{2\sigma}\exp(-|x|/b)$ (with $b = \sigma/\sqrt{2}$), using the useful property, $f''(x) = b^2f(x) - b^2\delta(x)$, we were able to reduce this integral equation into a differential recurrence equation which can subsequently be solved by generating function method. Skipping details [25], we get

$$q(z, s, x) = \sum_{n=0}^\infty \sum_{k=0}^n s^k z^k q_{k,n}(x)$$

$$= \frac{1}{1 - s} + \left( \frac{1}{\sqrt{(1-s)(1-zs)}} - \frac{1}{1 - s} \right) \times \exp\left( -\sqrt{2(1-s)}\frac{x}{\sigma} \right).$$

from which, using Eq. (2), the $p$th moment of the $k$th maximum $\langle M_{n,k}^p \rangle$ can be extracted. In particular, for $p = 1$ we get [25]

$$q_{k,n}(x) = \begin{array}{ll}
\frac{1}{1 - s} + \left( \frac{1}{\sqrt{(1-s)(1-zs)}} - \frac{1}{1 - s} \right) \times \exp\left( -\sqrt{2(1-s)}\frac{x}{\sigma} \right), & x > 0, \\
0, & x < 0.
\end{array}$$

FIG. 1. A realization of a random walk of $n = 6$ steps. We denote by $M_{k,6}$ the $k$th maximum and focus, in particular, on the gaps $d_{k,n} = M_{k,n} - M_{k+1,n}$. Note that $x_0$ is taken into account in the statistics.

FIG. 2. Illustration of the backward equation in Eq. (3).
\[
\langle M_{k,n} \rangle = \frac{\sigma}{\sqrt{2\pi}} \left( \frac{\Gamma(k + \frac{1}{2})}{\Gamma(k + 1)} \right).
\]

It follows from (5) that for large \( n \), \( \langle M_{k,n} \rangle / \sigma \sim \sqrt{2n/\pi} \), independently of \( k \), while the average gap \( \langle d_{k,n} \rangle \) is given by
\[
\langle d_{k,n} \rangle = \frac{\Gamma(k + \frac{1}{2})}{\sqrt{2\pi}} \left( \frac{\Gamma(k - n + \frac{1}{2})}{\Gamma(n + k + 1)} + \frac{\Gamma(k + \frac{1}{2})}{\sqrt{2\pi}} \right).
\]

Note that \( \langle d_{k,n} \rangle = \langle d_{n-k+1,n} \rangle \) reflecting the up-down (max-min) symmetry of the walk. Interestingly, as \( n \to \infty \), \( \langle d_{k,n} \rangle \) approaches a finite value
\[
\lim_{n \to \infty} \langle d_{k,n} \rangle = \frac{\Gamma(k + \frac{1}{2})}{\sqrt{2\pi}k!}.
\]

In addition, for large \( k \),
\[
\lim_{n \to \infty} \langle d_{k,n} \rangle = \frac{1}{\sqrt{2\pi k}} + O(k^{-1}).
\]

Next we show that the result (8) is actually universal and holds for arbitrary symmetric and continuous jump distribution \( f(x) \). To make progress for general \( f(x) \), we came across a very useful combinatorial identity known as Pollaczek-Wendel identity [26,27]. Using this identity and a few manipulations [25], we were able to derive the following exact result
\[
\lim_{n \to \infty} \langle d_{k,n} \rangle = \tilde{d}(k) = \frac{\sigma}{\sqrt{2\pi}} \left( \frac{\Gamma(k + \frac{1}{2})}{\Gamma(k + 1)} \right)
- \frac{1}{\sqrt{2\pi k}} \int_0^\infty dq \frac{[\hat{f}(q)]^k}{q^2} - \frac{1}{(1 + q^2)^k},
\]

where \( \hat{f}(q) = \int_0^\infty f(x) e^{i\eta x} dx \) is the Fourier transform of the jump distribution. The asymptotic analysis of (9) gives the universal result [25]
\[
\tilde{d}(k) / \sigma \sim (2\pi k)^{-1/2}, \quad k \gg 1,
\]

independent of \( f(\eta) \). This \( k^{-1/2} \) dependence of \( \tilde{d}(k) \) (10) was actually noticed in the numerical study of periodic random walks in Ref. [19] and was also conjectured to be exact, based on scaling arguments.

This result naturally raises the question whether only the first moment of the gap is universal, or perhaps the universality extends even to the pdf of the gap, once it is scaled by the nonuniversal scale factor \( \sigma \). This led us next to investigate the full pdf of \( d_{k,n} \). It is convenient first to consider the joint cumulative distribution \( S_{k,n}(x,y) = \text{Pr}[M_{k,n} > y, M_{k+1,n} < x] \), with \( y > x \). If we can compute this, then the gap pdf \( P_k(n,d_{k,n} = \delta) \) can be obtained from the relation
\[
P_k(n, \delta) = - \int_{\mathbb{R}^2} \frac{\partial^2 S_{k,n}(x,y)}{\partial x \partial y} \theta(y - x) \delta(x + \delta - y) dx dy.
\]

To compute \( S_{k,n}(x,y) \), as before, it is convenient to first define an auxiliary quantity \( Q_{k,n}(x,\Delta) \) denoting the probability that a random walk of \( n \) steps, starting from \( x_0 = x \), has \( k \) points in the interval \( (-\infty, -\Delta) \) (with \( k \geq 1 \)) and \( n - k \) points on the positive side, hence with no point in the interval \( [-\Delta, 0] \). The joint distribution \( S_{k,n}(x,\Delta) \) can be expressed in terms of \( Q \) as
\[
S_{k,n}(x,\Delta) = \begin{cases} 
Q_{k,n}(x, y - x), & x > 0, \\
0, & x < 0 \text{ and } y > 0, \\
Q_{n-k+1,n}(-y, y - x), & x < 0 \text{ and } y < 0.
\end{cases}
\]

Following similar arguments leading to Eqs. (3), we derive a backward integral equation, for \( n \geq 1 \),
\[
Q_{k,n}(x,\Delta) = \int_0^\infty Q_{k,n-1}(x',\Delta)f(x - x')dx' + \int_0^\infty Q_{n-k,n-1}(-x',\Delta)f(x - x' + \Delta)dx',
\]

starting from \( Q_{0,0}(x,\Delta) = 1 \). As before, this integral equation can be reduced to a linear differential recurrence equation for the special case, \( f(x) = \frac{1}{\pi} \exp(-|x|/b) \) and subsequently solved via the generating function method [25]. We get
\[
\sum_{n=0}^\infty \sum_{k=0}^n z^k s^n Q_{k,n}(x,\Delta) = \frac{1}{1 - s} + A(z, s, \frac{\sqrt{2}\Delta}{\sigma}) \times \exp(-\sqrt{2(1-s)} \frac{x}{\sigma}).
\]

where \( A(z, s, \Delta) \) has a complicated expression [25] omitted here for clarity. From this result and using (11) and (12), we find [25] that as \( n \to \infty \), \( P_k(n,\delta) \to p_k(\delta) \) where
\[
\sum_{k=1}^\infty z^k p_k(\delta) = \frac{8z^2}{b} \frac{u(z) - v(z) e^{-2(\delta/b)}}{[u(z) + v(z) e^{-2(\delta/b)}]},
\]

with \( u(z) = \sqrt{1 - z + 1} \) and \( v(z) = \sqrt{1 - z - 1} \). Extracting \( p_k \) for all \( k \) from (15) is hard. However, one can easily extract the asymptotic behavior for large \( k \), by analyzing the \( z \to 1 \) limit of (15). This yields, for \( k \gg 1 \) and \( \delta \) fixed, \( p_k(\delta) \sim k^{-3/2} F(\delta) \) where \( F(\delta) \) decays exponentially for large \( \delta \) and represents the cutoff function. However, before the distribution gets cut off for large \( \delta \), there is a scaling regime \( \delta \sim \tilde{d}(k) \sim \sigma / \sqrt{2\pi k} \), with \( k \) large, where we anticipate a scaling form for the gap pdf
\[
p_k(\delta) \approx (\sqrt{k}/\sigma) P(\sqrt{k}\delta/\sigma),
\]
and we expect that the scaling function \( P(x) \) is independent of \( k \). Indeed, taking \( k \to \infty \) and \( \delta \to 0 \) limit in (15) while keeping the scaled variable \( \sqrt{k}\delta/\sigma \) fixed we find the scaling function \( P(x) \) satisfies
\[ \int_{0}^{\infty} e^{-x^4} \sqrt{P(x^2)} dx = (1 + \sqrt{2}/3)^{3/2}. \] (17)

This Laplace transform (17) can be inverted to yield finally the expression given in Eq. (1). The asymptotic behaviors of \( P(x) \) are given by

\[ P(x) \sim \begin{cases} 4\sqrt{2/\pi}x^{-4}, & x \to 0, \\ (3\sqrt{8/\pi})x^{-4}, & x \to \infty, \end{cases} \] (18)

which thus exhibits a surprising power law tail.

The distribution \( P(x) \) describes the typical fluctuations of \( d_k \), which are of order \( O(k^{-1/2}) \) for large \( k \). Having derived it for the special case of exponential jump distribution, it is natural to wonder whether the same function \( P(x) \) appears for other jump distributions as well. Remarkably, our numerical simulations show that \( P(x) \) is indeed universal. In Fig. 3(a) we show a plot of the (scaled) pdf of the gaps \( P_{\text{pdf}}(\delta \sigma \kappa^{-1/2}) \) as a function of the scaling variable \( \delta \kappa^{1/2}/\sigma \) for three different jump distributions: exponential, Gaussian, and uniform. The data shown correspond to random walks of \( n = 10^5 \) and \( k = 90 \) and they have been obtained by averaging over \( 10^6 \) independent trajectories of the random walk. The dotted line corresponds to \( P(x) \) given in Eq. (1). The good collapse of these different curves, for \( \delta \kappa^{1/2}/\sigma \leq 1 \) indicate that the typical fluctuations of \( d_k \), of order \( O(k^{-1/2}) \), are universal—indeed universal, independent of \( f(\kappa) \)—and described by \( P(x) \) in Eq. (1).

In contrast to the typical fluctuations that are described by a universal scaling function, the atypically large fluctuations corresponding to \( \delta \gg \hat{d}(k) \sim k^{-1/2} \) are not universal. This pdf \( p_{\hat{d}}(\delta) \) for \( \delta \gg k^{-1/2} \) actually gets cut off in a nonuniversal way, as we have seen before for the exponential jump distribution. Thus there are two scales of \( \delta \): a typical fluctuation which is universal and large fluctuations that are nonuniversal. This has very interesting consequences on the behavior of the moments \( \langle d_k^p \rangle \) as a function of \( k \) (for large \( k \)). One conjectures, and this is corroborated by an exact calculation for the exponential distribution from Eq. (15),

\[ \frac{\langle d_k^p \rangle}{\sigma^p} \sim \begin{cases} \frac{1}{\sqrt{2\pi}} k^{-1/2}, & p = 1, \\ \frac{1}{2} k^{-1}, & p = 2, \\ D_p (\log k)^{-3/2}, & p = 3, \\ D_p k^{-3/2}, & p \geq 4, \end{cases} \] (19)

where the amplitudes for \( p < 3 \) are universal, while the amplitudes \( D_p \) are not. Our numerical data, shown in Fig. 3(b), are in agreement with these results (19). Indeed for \( p = 1 \) and \( p = 2 \) the value of the scaled moments \( \langle d_k^p \rangle/\sigma^p \) for different jump distributions do coincide and exhibit a power law decay with \( k \) in agreement with Eq. (19). The solid lines in Fig. 3 indicate the power law behavior expected from Eq. (19). On the other hand, for \( p = 4, 5 \) these scaled moments do not coincide and they exhibit a power law decay with, seemingly the same exponent, although a precise estimate of the exponent 3/2 for higher moments is quite difficult.

In conclusion, we have presented exact results for the gap statistics of symmetric random walks with a finite variance of steps \( \sigma^2 \) and found a rather rich and universal behavior independent of the details of the jump distribution. This presents an interesting and useful example of solvable order statistics in a correlated time series. In view of recent applications of random walks to fluctuating interfaces in 1 + 1 dimensions [9–11,16], it will be interesting to see if the universal gap statistics found here also holds for different boundary conditions of the interface. It would also be interesting to extend these results to cases where \( \sigma \) is infinite such as in Lévy flights and also to asymmetric jump distributions.

We warmly thank Zoltan Rácz for very stimulating discussions at the earliest stage of this work. We thank the Lorentz Center in Leiden for the hospitality where part of this work was accomplished. This research was support...
by ANR Grant No. 2011-BS04-013-01 WALKMAT and in part by the Indo-French Centre for the Promotion of Advanced Research under Project No. 4604-3.

*gregory.schehr@th.u-psud.fr
†majumdar@lptms.u-psud.fr