EVS for correlated variables

I. INTRODUCTION

Up to now, we considered the EVS of i.i.d. variables. Most of the time, however, there are correlations among the measured quantities (e.g. among the temperature or windspeed measurements at a given geographic location as a function of time). These correlations are often obvious but sometimes too weak to be noticable (see the series of signals in Fig.1). As we shall see, it will be important to distinguish the weakly and strongly correlated variables. We shall start by discussing a general characterization of correlations.

II. WEAK AND STRONG CORRELATIONS FROM A GENERAL POINT OF VIEW

Consider N variables y_1, y_2, ..., y_N and let's try to find a quantity to characterize the correlations among the ys. To do this, let us consider the fluctuations of the sum of the variables $Y = \sum_{i=1}^{N} y_i$

$$\langle (Y - \langle Y \rangle)^2 \rangle = \langle Y^2 \rangle - \langle Y \rangle^2 = \sum_{i,j=1}^{N} \left[\langle y_i y_j \rangle - \langle y_i \rangle \langle y_j \rangle \right] \quad (1)$$

For i.i.d. variables, there are no correlations among y_i and y_j for $i \neq j$ and it follows that the quantity we shall call the correlation function

$$C(i,j) = \langle y_i y_j \rangle - \langle y_i \rangle \langle y_j \rangle \quad (2)$$

is equal to zero, $C(i \neq j, j) = 0$, and the fluctuations in Y simplify to

$$\langle (\delta Y)^2 \rangle = \sum_{i=1}^{N} \left[\langle y_i^2 \rangle - \langle y_i \rangle^2 \right] = N \langle (\delta y_1)^2 \rangle \, . \quad (3)$$

In this case we observe that if the average value of y_i is of the order of 1 and, consequently, $\langle Y \rangle \sim N$, then the relative fluctuations of the sum diminishes in the large-N limit

$$\frac{\sqrt{\langle (\delta Y)^2 \rangle}}{\langle Y \rangle} \sim \frac{1}{\sqrt{N}} \, . \quad (4)$$

Let us now consider the case when all the measurements are correlated equally $C(i \neq j, j) = Q$. Then the fluctuations of the sum can be written as

$$\langle (\delta Y)^2 \rangle = N \langle (\delta y_1)^2 \rangle + N(N-1)Q \, . \quad (5)$$

For large N, the fluctuation is dominated by the correlations and we find

$$\langle (\delta Y)^2 \rangle = N^2 Q \, . \quad (6)$$

This means that the relative fluctuations remain finite in the $N \to \infty$ limit

$$\frac{\sqrt{\langle (\delta Y)^2 \rangle}}{\langle Y \rangle} \sim O(1) \, . \quad (7)$$

Clearly, the correlation dominated case is drastically different from the uncorrelated one. The question is what is the meaning of weak correlations and where is the dividing line between weak and strong correlations.

In order to clarify the situation, let us imagine that the i indexes have the meaning of spatial coordinates. Then $C(i,j)$ gives us the correlations between the spatial points i and j. Let us also assume that we have a homogeneous system, i.e. the correlations depend only on the distance between the spatial points $C(i,j) = \Phi(|i-j|)$. Then the fluctuations in Y can be written (after changing the two summation variables from i and j to $i+j$ and $i-j$, and using a one dimensional periodic boundary conditions for i)

$$\langle (\delta Y)^2 \rangle = N \langle (\delta y_1)^2 \rangle + 2N \sum_{n=1}^{N/2} \Phi(n) \, . \quad (8)$$

We can see that if the correlations are short ranged [e.g. extend only to $n = 1$ or they decay exponentially, $\Phi(n) \sim \exp(-\alpha n)$] then the sum in (8) is finite (converges for $N \to \infty$) and we can write

$$\langle (\delta Y)^2 \rangle = N \left(\langle (\delta y_1)^2 \rangle + 2 \sum_{n=1}^{\infty} \Phi(n) \right) \sim N \, . \quad (9)$$

This means now that the relative fluctuations $\sqrt{\langle (\delta Y)^2 \rangle} / \langle Y \rangle \sim 1/\sqrt{N}$ diminish in the same way as in an uncorrelated system. This is the usual thermodynamical scaling which is one of the basic element underlying the thermodynamic theories. This is why we shall call systems obeying (9) as weakly correlated systems.

Clearly, the condition of weak scaling is the convergence of the sum $\sum_{n=1}^{\infty} \Phi(n)$. Otherwise, if e.g. we assume that $\Phi(n)$ has a power-law decay

$$\Phi(n) \sim 1/n^{-1+\eta} \quad ; \quad \eta > 0 \, , \quad (10)$$

then the large-N scaling of $\langle (\delta Y)^2 \rangle$ changes since the main contribution comes from the correlations and one has

$$\langle (\delta Y)^2 \rangle \sim N^{1+\eta} \, . \quad (11)$$

Correlations obeying (10) usually emerge in systems at their critical points and it is usual to refer to critical systems as strongly correlated. We can thus classify the correlations weak or strong according to the convergence or divergence of the sum $\sum_{n=1}^{\infty} \Phi(n)$. As we shall see, this classification is relevant in EVS, as well.

III. $1/f^\alpha$ NOISE: A MODEL WHERE THE CORRELATIONS CAN BE TUNED

In order to develop an intuition about the effect of correlations, we shall consider here the EVS of periodic signals displaying Gaussian fluctuations with $1/f^\alpha$ power spectra. While we shall use the terminology of time signals, one-dimensional stationary interfaces may equally be imagined, with the same spatial spectrum, and periodic boundary conditions. One should note that systems with $1/f^\alpha$ type fluctuations are abundant in nature, with examples ranging from voltage fluctuations in resistors [1], through temperature fluctuations in the oceans [2], to climatological temperature records [3], and to the number of stocks traded daily [4].

The $1/f^\alpha$ signals are rather simple since they decompose into independent modes in the Fourier space. The modes are not identically distributed, however, giving rise to correlations, which can be tuned by α. They yield signals with no correlations ($\alpha = 0$), with weak ($0 < \alpha < 1$), and strong correlations ($1 \leq \alpha < \infty$). Thus studying the EVS of $1/f^\alpha$ processes allows the studying of the effect of correlations on extreme events.

The $1/f^\alpha$ signals are defined as follows. Gaussian periodic signals $h(t) = h(t + T)$ of length T are considered with the probability density functional of $h(t)$ given by

$$\mathcal{P}[h(t)] \sim e^{-S[h(t)]}, \tag{12}$$

where the effective action S can be formally defined in real space but, in practice, is defined through its Fourier representation. First, define the Fourier transform of the signal

$$h(t) = \sum_{n=-N/2+1}^{N/2} c_n e^{2\pi i n t/T}\,, \tag{13}$$

where $c_n^* = c_{-n}$ and their phases (for $n \neq N/2$) are independent random variables uniformly distributed in the interval $[0, 2\pi]$, while $c_{N/2}$ is real. Since c_0 will not appear in the action we can set the average of the signal to zero, i.e. $c_0 = 0$. Note that the cutoff introduced by N means that the timescale is not resolved below

$$\tau = T/N \tag{14}$$

and thus a measurement of $h(t)$ yields effectively N data points.

Now, the action can be written in terms of the Fourier amplitudes as

$$S[c_n; \alpha] = 2\lambda T^{1-\alpha} \sum_{n=1}^{N/2} n^\alpha |c_n|^2\,, \tag{15}$$

where λ is a stiffness parameter which is set to $(2\pi)^\alpha/2$ hereafter.

As one can see from Eqs. (12) and (15), the amplitudes of the Fourier modes are independent, Gaussian distributed variables – but they are not identically distributed. Indeed, the fluctuations increase with decreasing wavenumber, with power spectrum

$$\langle |c_n|^2 \rangle \propto \frac{1}{n^\alpha}, \tag{16}$$

hence the name of $1/f^\alpha$ signal (in ususal notation $\langle |c_\omega|^2 \rangle \propto 1/\omega^\alpha$).

By scanning through α, systems of wide interest may be generated. For example, $\alpha = 0, 1, 2, 4$ correspond respectively to white-noise, $1/f$-noise [5], an Edwards-Wilkinson interface [6] or Brownian curve, and a Mullins-Herring interface [7, 8] or random acceleration process [9].

For $1/f^\alpha$ signals, the correlations may be tuned as one can see in Fig.1. An α-scan leads us from the absence of correlations ($\alpha = 0$, white noise) to the limit of a deterministic signal ($\alpha = \infty$). In order to see where is the dividing line between weak and strong correlations, let us consider mean-square fluctuations of the signal (similar to the fluctuations of $\langle (\delta Y)^2 \rangle / N$ discussed before) and is called the roughness of the signal (of the interface)

$$w_2 = \frac{1}{T} \int_0^T dt \left[h(t) - \overline{h} \right]^2 = 2 \sum_{n=1}^{N/2} |c_n|^2\,, \tag{17}$$

where the overbar indicates an average over t, and the second equality shows that w_2 is the integrated power-spectrum of the system.

Using (16), it is easy to show that $\langle w_2 \rangle$ has the following asymptote for large system sizes ($T \to \infty$)

$$\langle w_2 \rangle \sim \begin{cases} T^{\alpha-1} & \text{for } 1 < \alpha \leq \infty \\ \ln T/\tau & \text{for } \alpha = 1 \\ \tau^{\alpha-1} & \text{for } 0 \leq \alpha < 1\,. \end{cases} \tag{18}$$

Thus the fluctuations diverge with system size for $1 \leq \alpha < \infty$ in contrast to the finite fluctuations in the $0 \leq \alpha < 1$ regime. Since diverging fluctuations are the sign of strong correlations, this gives a reason for separating the $0 \leq \alpha < 1$ and $1 \leq \alpha < \infty$ regions and attaching the name of weak and strong correlations to each, respectively.

IV. EVS FOR $1/f^\alpha$ SIGNALS

The EVS of $1/f^\alpha$ signals is concerned with is the maximum relative height (MRH). This is the highest peak of a signal over a given time interval T, measured from the average level. Specifically, for each realization of the signal, $h(t)$, the MRH is

$$h_m := \max_t h(t) - \overline{h(t)}\,. \tag{19}$$

where $\max_t h(t)$ is the peak of the signal and $\overline{h(t)}$ is its time average. The MRH, h_m, varies from realization to realization, and is therefore a random variable whose probability density function (PDF), denoted by $P(h_m)$, we would like to determine. The physical significance of h_m is obvious. For instance, in a corroding surface it gives the maximal depth of damage or, in general, it is the maximal peak of a surface. To name another example, when natural water level fluctuations are considered, it is related to the necessary dam height.

It is relatively easy to carry out simulations for finding $P(h_m)$. Exact results, however, are available only for special values of $\alpha = 0,\, 1,\, 2,\, \infty$.

V. WEAK CORRELATION REGIME ($0 \leq \alpha < 1$)

For $\alpha = 0$, we have Gaussian distributed i.i.d. variables, so we have the Fisher-Tippett-Gumbel (FTG) distribution as the limit distribution in the large-size ($L \to \infty$) limit. As can be seen from Fig.2, the limit distribution is approached very slowly with increasing L. This is in agreement with our results for finite-size scaling (previous lecture) where we showed that the first shape correction has an amplitude proportional to $1/\ln L$.

Since the fluctuations do not diverge for $\alpha < 1$ (weak correlations), we expect that the limit distribution remains the FTG distribution. This is indeed observed for $\alpha = 0$ (upper left), 0.25 (upper right), 0.5 (lower left), 0.75 (lower right panel) in Fig.2. Of course, simulations are not proofs and the slow convergence gives reasons to doubt this conjecture. There are, however, mathematical works [10] where it is poved rigorously that the limit distribution for $\alpha < 1$ is the FTG distribution.

VI. STRONG CORRELATION REGIME ($\alpha \geq 1$)

As we enter the strong correlation regime, the average value of the extremum and its fluctuations diverge. Thus, it is essential that we plot the distributions in terms of $y = h_{max}/\langle h_{max}\rangle$ or $x = (h_{max} - \langle h_{max}\rangle)/\sigma$ where σ is the variance of the maximal value.

Fig.3 displays simulation results for $\alpha = 1$ and 2 and one can see that the convergence became much faster Indeed, for these values of αs, exact limit distributions are available [11, 12] which are indistinguishable from the curves shown. It is remarkable that the shape of the limit distribution does not change much with α in the strongly correlated regime (see Fig.4 where the $\alpha = 2, 4, \infty$ results are compared with the FTG distribution), all changes are absorbed in the divergence of the average and the variance.

VII. ANALYTIC CALCULATION AT $\alpha = 2$

The exact calculation of the EVS distribution for $\alpha = 2$ [11] and later for $\alpha = 1$ [12] were difficult excercises in calculating path integrals. Here, we shall carry out a claculation of the $\alpha = 2$ case for a slighly modified EVS. We shall consider the maximum value of the signal with respect to the initial value

$$h_m := \max_t [h(t) - h(0)]\,. \tag{20}$$

This quantity may be of interest for example when we begin to measure temperature at a given place, or we bought shares in a company and are interested in our expected profit.

Since $\alpha = 2$ is a random walk, we shall consider $h(t)$ as a random walk on the x axis and to make connection to this process in usual variable, we shall write $x(t) \equiv h(t)$. Accordingly, we shall denote the initial position of the random walker $h(0) \equiv x_0$. Then the quantity we are after is the probability $M(x < b|x_0; t)$ that x remains below b till time t if the random walker started from x_0 at $t = 0$. Once we calculated $M(x < b|x_0; t)$, the EVS distribution is obtained by taking its derivative by b.

In order to calculate $M(x < b|x_0; t)$ we shall use the following identity

$$M(x < b|x_0; t) = \int_{-\infty}^{b} \hat{P}_b(x, t|x_0, 0)dx \tag{21}$$

where $\hat{P}_b(x, t|x_0, 0)$ is the probability that, starting from x_0 at $t = 0$, a path of the random walker did not cross $x = b$ by time t.

In order to find $\hat{P}_b(x, t|x_0, 0)$, we have to solve the diffusion equation

$$\partial_t \hat{P}_b(x, t|x_0, 0) = D\partial_x^2 \hat{P}_b(x, t|x_0, 0) \tag{22}$$

with the initial condition (the random walker starts at x_0)

$$\hat{P}_b(x, 0|x_0, 0) = \delta(x - x_0)\,, \tag{23}$$

and with the boundary condition of a sink being placed at $x = b$

$$\hat{P}_b(b, t|x_0, 0) = 0\,. \tag{24}$$

The solution of the above problem consist of realizing that the Green function of the diffusion equation is the following Gaussian

$$\hat{P}_b(x, t|x_0, 0) = \frac{1}{\sqrt{4\pi Dt}} e^{-(x-x_0)^2/4Dt}\,, \tag{25}$$

and the boundary condition can be satisfied by the mirror image method. Thus the solution is given by

$$\hat{P}_b(x, t|x_0, 0) = \frac{e^{-(x-x_0)^2/4Dt} - e^{-(x+x_0-2b)^2/4Dt}}{\sqrt{4\pi Dt}}\,. \tag{26}$$

The above solution should be inserted into (21) and then the probability density $P(x = b|x_0; t)$ of the maximum being between b and $b+db$ is obtained from the derivative

$$P(x=b|x_0;t) = \frac{\partial M(x<b|x_0;t)}{\partial b} = \hat{P}_b(b,t|x_0,0) + \int_{-\infty}^{b} \frac{\partial \hat{P}_b(x,t|x_0;0)}{\partial b} dx\,. \quad (27)$$

The first term on the second line is zero due to the boundary condition (24) while the second term can be calculated using the following identity

$$\frac{\partial \hat{P}_b(b,t|x_0;0)}{\partial b} = \frac{2}{\sqrt{\pi D t}} \frac{\partial}{\partial x} e^{-(x+x_0-2b)^2/4Dt}\,. \quad (28)$$

The final result is obtained as

$$P(x|x_0;t) = \frac{1}{\sqrt{\pi D t}} e^{-(b-x_0)^2/4Dt}\,, \quad b \geq x_0\,. \quad (29)$$

The above expression is valid, of course, only for $b \geq x_0$ since the maximum is measured with respect to the initial value, i.e. $b = x_0$ is the minimum value of the possible maxima. Note that the distribution (29) is normalized, thus there is no extra weight at $b = x_0$ coming from paths which live below x_0 in the time interval $[0,t]$.

VIII. HOMEWORK

Homework 9. Having calculated $M(x<b|x_0;t)$ through $\hat{P}_b(x,t|x_0,0)$ [see (21)] allows us to calculate the so called first passage time probability $\mathcal{P}_b(t|x,0)dt$ defined as the probability that the random walker, starting from x_0 at time $t=0$ passes the $x=b$ value for the first time between t and $t+dt$. It follows from the above definition that $M(x<b|x_0;t)$ can be expressed through $\mathcal{P}_b(t|x,0)dt$ as follows

$$M(x<b|x_0;t) = 1 - \int_0^t \mathcal{P}_b(t'|x,0)dt\,. \quad (30)$$

The meaning of the above equation is that the probability that the path lives below $x=b$ is equal to 1 minus the probability that the path has crossed b sometimes in the time interval $[0,t]$.

Use the relationship (30) to calculate $\mathcal{P}_b(t|x,0)dt$.

[1] A. V. Yakimov and F. N. Hooge, *A simple test of the Gaussian character of noise*, Physica B 291, 97–104 (2000).

[2] R. A. Monetti and S. Havlin and A. Bunde, *Long-term persistence in the sea surface temperature fluctuations*, Physica A 320, 581–589 (2002).

[3] R. Blender and K. Fraedrich and B. Hunt, *Millennial climate variability: GCM-simulation and Greenland ice cores*, Geophys. Res. Lett. 33, L04710 (2006).

[4] F. Lillo and R. N. Mantegna, *Variety and volatility in financial markets*, Phys. Rev. E 62, 6126-6134 (2000).

[5] M. B. Weissman, $1/f$ *noise and other slow, nonexponential kinetics in condensed matter*, Rev. Mod. Phys. 60, 537–571 (1988).

[6] S. F. Edwards and D. R. Wilkinson, *The Surface Statistics of a Granular Aggregate*, Philos. Trans. R. Soc. London A 381, 17–31 (1982).

[7] W. W. Mullins, *Theory of Thermal Grooving*, J. Appl. Phys. 28, 333–339 (1957).

[8] J. Villain, *Continuum models of crystal growth from atomic beams with and without desorption*, J. Phys. I France 1, 19–42 (1991).

[9] T. W. Burkhardt, *Semiflexible polymer in the half plane and statistics of the integral of a Brownian curve*, J. Phys. A: Math. Gen. 26, L1157 (1993).

[10] S. M. Berman, *Limit theorems for the maximum term in stationary sequences*, Ann. Math. Stat. 35, 502 (1964).

[11] S. N. Majumdar and A. Comtet, Phys. Rev. Lett. 92, 225501 (2004).

[12] Y. V. Fyodorov, P. Le Doussal, and A. Rosso, J. Stat. Mech.: Theory and Exp., P10005 (2009).

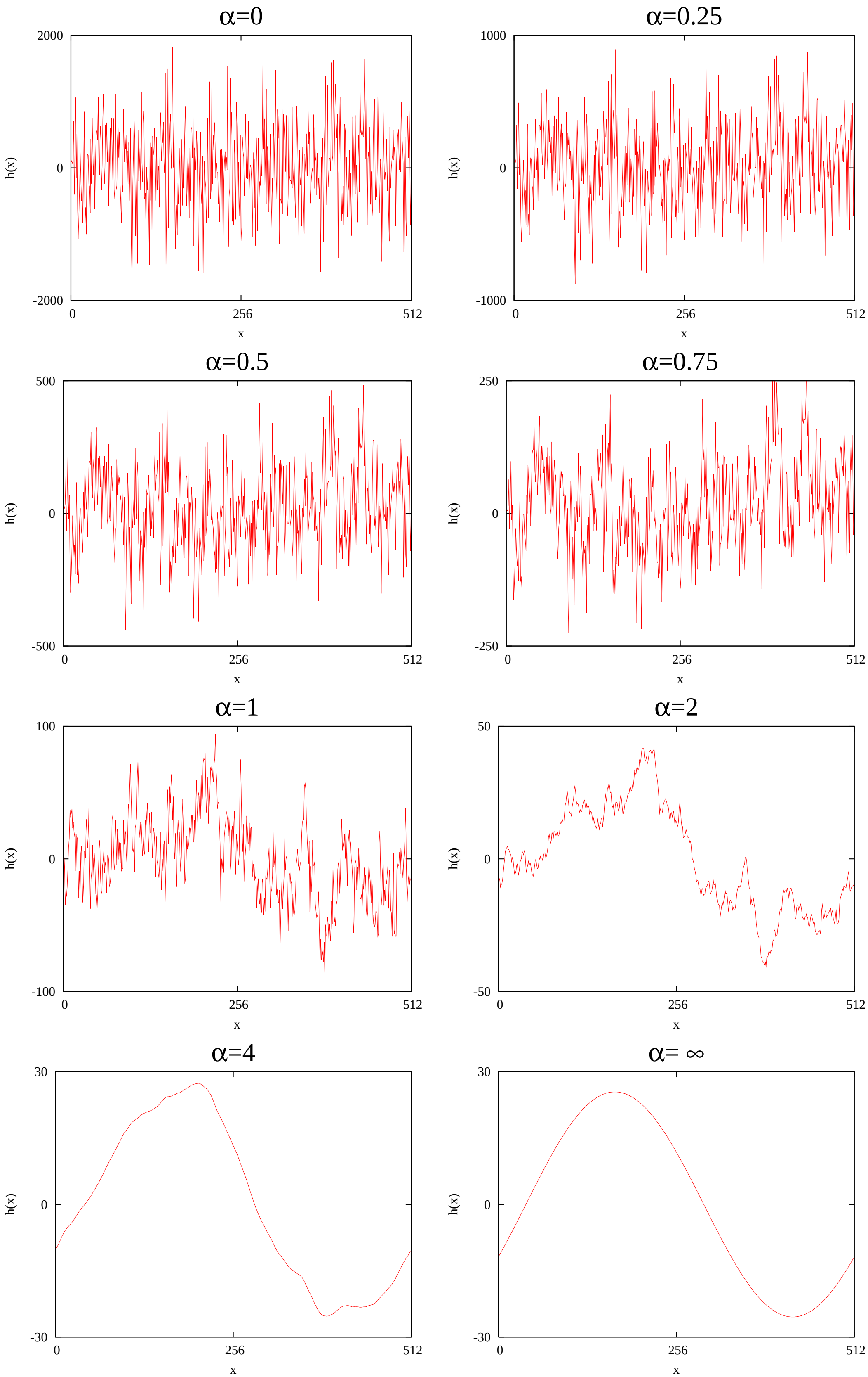

FIG. 1: Typical profiles of various $1/f^{\alpha}$ signals of length $N = T/\tau = 8192$. Note that, contrary to the visual illusion, the $0 \leq \alpha < 1$ surfaces are flat, while the $1 \leq \alpha < \infty$ signals are rough. In the former case, the amplitude of the signals is size independent, while in the latter case the amplitude diverges with system size. For ease of comparison, we have rescaled the signals to be approximately equal in height.

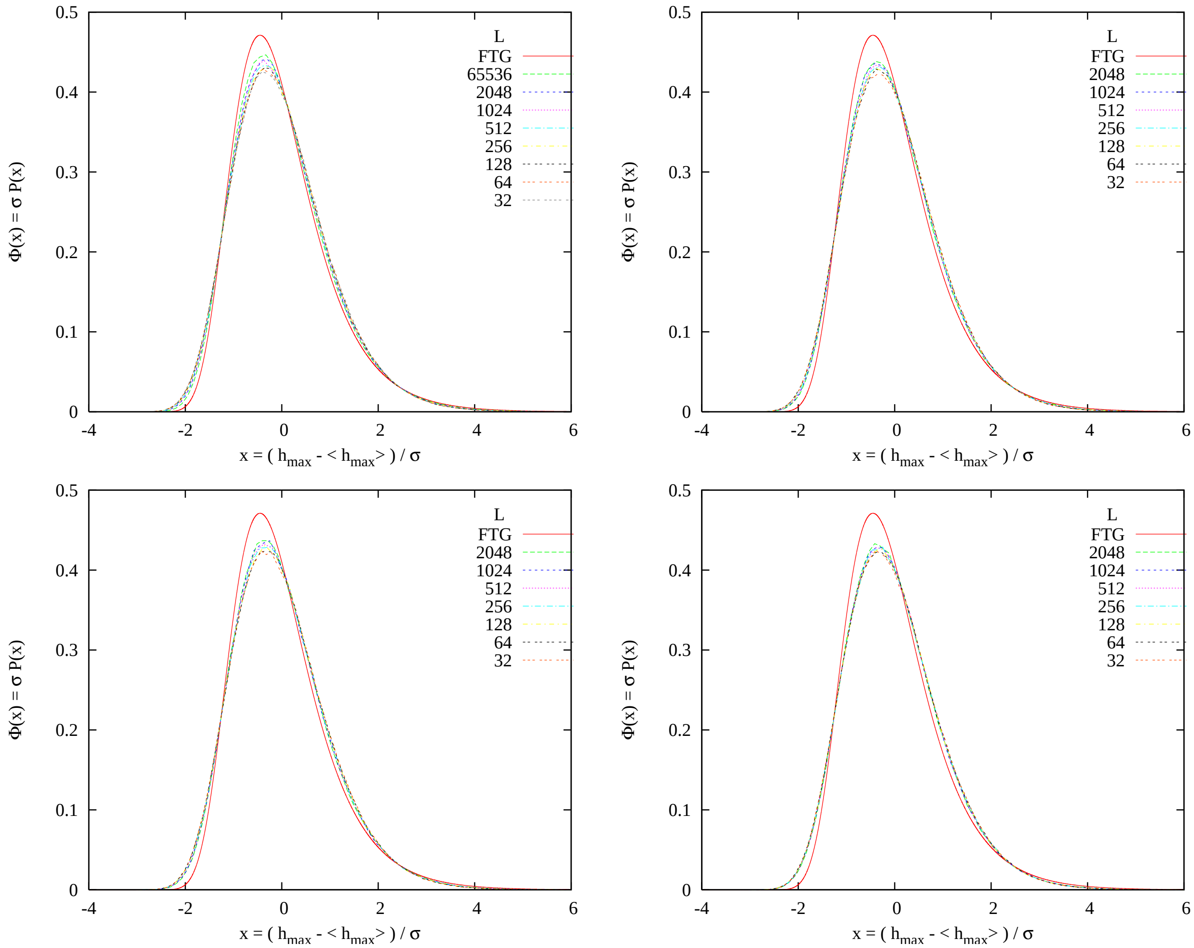

FIG. 2: EVS distributions for $\alpha = 0$ (upper left), 0.25 (upper right), 0.5 (lower left), 0.75 (lower right panel) using scaling variable $x = (h_{max} - \langle h_{max} \rangle)/\sigma$ where σ is the variance of the maximal value. The length of the signal is $L = T/\tau$ with τ being the resolution in the time signal. The upper solid red line is the FTG distribution which is known to be the exact limit distribution for $L \to \infty$.

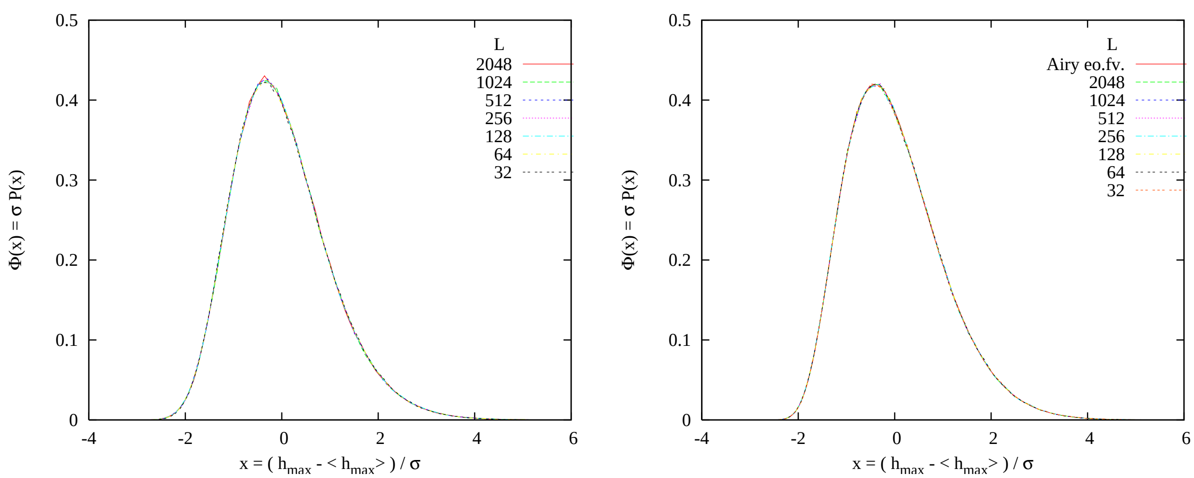

FIG. 3: EVS distributions for $\alpha = 1$ (left panel) and $\alpha = 2$ (right panel) using scaling variable $x = (h_{max} - \langle h_{max} \rangle)/\sigma$. The length of the signal is $L = T/\tau$ with τ being the resolution in the time signal. For $\alpha = 2$, the exact results (the so called Airy distribution) is also plotted.

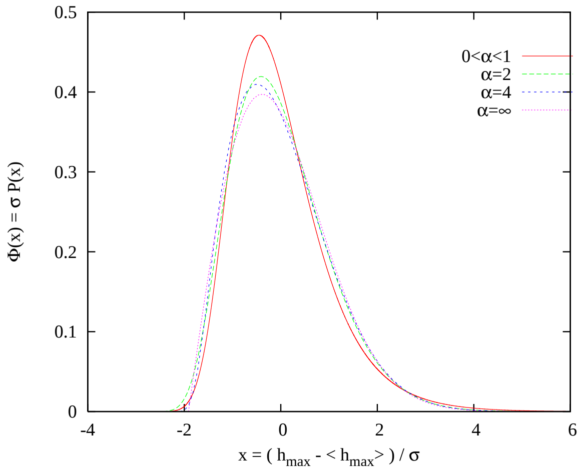

FIG. 4: EVS distributions for $\alpha = 2, 4, \infty$ are compared with the FTG distribution ($\alpha < 1$). The scaling variable is $x = (h_{max} - \langle h_{max} \rangle)/\sigma$. The length of the signal is $L = T/\tau$ with τ being the resolution in the time signal.