PHYSICAL REVIEW E **75**, 020101(R) (2007)

Central limit theorem for anomalous scaling due to correlations

Fulvio Baldovin* and Attilio L. Stella†
Dipartimento di Fisica and Sezione INFN, Universitá di Padova, Via Marzolo 8, I-35131 Padova, Italy
(Received 27 September 2005; revised manuscript received 31 October 2006; published 20 February 2007)

We derive a central limit theorem for the probability distribution of the sum of many critically correlated random variables. The theorem characterizes a variety of different processes sharing the same asymptotic form of anomalous scaling and is based on a correspondence with the Lévy-Gnedenko uncorrelated case. In particular, correlated anomalous diffusion is mapped onto Lévy diffusion. Under suitable assumptions, the nonstandard multiplicative structure used for constructing the characteristic function of the total sum allows us to determine correlations of partial sums exclusively on the basis of the global anomalous scaling.

DOI: 10.1103/PhysRevE.75.020101 PACS number(s): 02.50.−r, 05.40.Fb, 89.75.Da

The central limit theorem (CLT) for sums of independent random variables [1] plays a fundamental role in statistical physics. This theorem is essential for the construction of equilibrium statistical mechanics [2], underlies the description of Brownian diffusion [3], and provides justification to numerical approaches like Monte Carlo methods. According to it, the probability density functions (PDF's) of the sums are Gaussian when the single-variable PDF's have finite second moment. If on the other hand these PDF's have long-range tails determining the divergence of the second moment, the Lévy-Gnedenko limit theorem states that the sums are Lévy distributed [1,4]. There are, however, many situations, like critical phenomena in statistical systems [5], financial time series [6], anomalous transport [7], and protein dynamics [8], where the presence of strong correlations leads to non-Gaussian PDF's obeying anomalous scaling with finite second moment [9]. Understanding how correlations determine anomalous scaling and universality is still a challenge in general, at least outside equilibrium statistical mechanics. Indeed, in this context renormalization group (RG) methods opened the way to a probabilistic interpretation of scaling and universality in critical phenomena [10]. RG transformations for effective Hamiltonians provide a framework for the discussion of critical scaling in cases when, due to the lack of independence, the simple factorization of individual variable characteristic functions (CF's), on which the CLT is based, does not hold. This framework requires new and more complicated forms of limit theorems and stability criteria [10]. In view of the key role played by the CLT in many fields, it is legitimate to ask if some form of CF factorization helps in discussing the correlated case. This could allow us to establish parallels between the treatments of independent and strongly dependent variables.

In this Rapid Communication we show that a nonstandard factorization of summand variable CF's allows one to construct the CF of the sum consistent with the assumption of asymptotic anomalous scaling. This factorization is at the basis of a novel CLT and, under further conditions, allows one to reconstruct the correlations of the asymptotic process.

In many physical situations, as one considers the sum $X \equiv \sum_{i=1}^{N} X_i$ of stochastic variables X_i with values x_i in the real axis, it is observed that for large N the PDF of X, $p_X(x)$, asymptotically obeys a simple scaling:

$$N^D p_X(x) \sim g\left(\frac{x}{N^D}\right), \tag{1}$$

where g is a scaling function and D is a scaling exponent. The scaling is anomalous if g is not a Gaussian function or $D \neq 1/2$. As appropriate in most physical applications, we consider cases in which p_X has finite second moment. The X_i's and X could be, respectively, the spins and the total magnetization of a critical ferromagnetic system. They could also represent the hour-by-hour increments and the total variation of a return in a financial time series which is sampled on intervals of N hours. Self-similarity is implied by Eq. (1) since plots of $N^D p_X$ vs x/N^D at different N asymptotically collapse onto the same curve g. To make this idea more precise, one can consider the normalized sum $Y \equiv \sum_{i=1}^{N} X_i/N^D$ and its PDF $p_Y(y) \equiv p_N(y)$. From Eq. (1) follows then, in the limit $N \to \infty$,

$$p_N(y) \sim g(y). \tag{2}$$

For the CF of p_N, $\tilde{p}_N(k) \equiv \int_{-\infty}^{+\infty} \exp(iky) p_N(y) dx$, Eq. (2) reads

$$\tilde{p}_N(k) \sim \tilde{g}(k), \tag{3}$$

where $\tilde{g}$ is the Fourier transform (FT) of g $[\tilde{p}_N(0)=\tilde{g}(0)=1]$. We assume here that p_N is even in y, so that $\tilde{p}_N(-k)=\tilde{p}_N^*(k)=\tilde{p}_N(k)$. Furthermore, $\tilde{p}_N(k)=1-\sigma^2 k^2/2+O(k^4)$, where the coefficient of k^2 is twice the second moment of p_N. Below, we choose units such that $\sigma^2/2=1$.

If we consider independent and identically distributed X_i's, the CLT accounts for the asymptotic scaling in Eqs. (2) and (3) stating that it is not anomalous; i.e., it has $D=1/2$ *and* g Gaussian. Let us call p_1 the PDF of any individual X_i and $\tilde{p}_1$ the corresponding CF. By N-times convolution, one gets

$$\tilde{p}_N(k) = [\tilde{p}_1(k/N^{1/2})]^N. \tag{4}$$

One can prove [1] that $\tilde{p}_N$ becomes Gaussian $[\sim \exp(-k^2)]$ at large N for any p_1 with finite variance $\int_{-\infty}^{+\infty} p_1(x) x^2 dx = 2$. Via inverse FT this implies a Gaussian form for the asymptotic p_N and $D=1/2$. A key concept here is that the limit PDF of the sum is stable; i.e., the sum of two independent Gaussian distributed variables is still Gaussian distributed. This stabil-

*Electronic address: baldovin@pd.infn.it
†Electronic address: stella@pd.infn.it

ity can be represented, e.g., by an invariance of $\tilde{g}$ under multiplication:

$$\tilde{g}(k/2^{1/2})\tilde{g}(k/2^{1/2}) = \tilde{g}(k). \quad (5)$$

This functional relation directly follows in the large-N limit from Eqs. (3) and (4) and has $\tilde{g}(k)=\exp(-k^2)$ as the only possible solution.

Here we investigate the possibility of a generalization of the multiplicative structure in Eq. (5) through the following steps: (i) We assume the existence of a $\tilde{g}$ and a D characterizing a given form of asymptotic anomalous scaling. (ii) We then introduce, in terms of $\tilde{g}$ and D themselves, a generalized multiplication $\otimes$ such that the identity

$$\tilde{g}(k/2^D) \otimes \tilde{g}(k/2^D) = \tilde{g}(k) \quad (6)$$

holds. (iii) Eventually, we apply this generalized multiplication $\otimes$ to CF's different from $\tilde{g}$ in order to prove a CLT, implying the existence of a wide class of correlated processes asymptotically behaving consistently with the scaling specified by $\tilde{g}$ and D.

We consider scaling functions with the additional property of $\tilde{g}$ being strictly monotonic in $[0,+\infty)$ [11], which in our context also implies $0<\tilde{g}(k)\leq 1 \forall\ k\in\mathbb{R}$. For $a_1,a_2\in(0,1]$, the generalized multiplication allowing satisfaction of Eq. (6) is

$$a_1 \otimes a_2 \equiv \tilde{g}(\{[\tilde{g}^{-1}(a_1)]^{1/D} + [\tilde{g}^{-1}(a_2)]^{1/D}\}^D), \quad (7)$$

where $\tilde{g}^{-1}$ is the inverse of $\tilde{g}$ in $[0,+\infty)$. One can easily verify that $a_1\otimes a_2\in(0,1]$ and $a_1\otimes 1=a_1$ and that $\otimes$ is associative and commutative [13]. Equation (6) is recovered by putting $a_1=a_2=\tilde{g}(k/2^D)$. It is important to remark that if $\tilde{g}$ is Gaussian and $D=1/2$, the $\otimes$ multiplication reduces to the ordinary one. The consideration of $a_1\neq a_2$ in Eq. (7) is clearly not needed to recover Eq. (6), but becomes of crucial importance to determine joint probabilities for partial sums of the X_i's compatible with the anomalous scaling of the total sum [12].

One can further establish a precise correspondence between this generalized multiplication and the ordinary one. Once fixed $\tilde{g}$ and D, let us consider the mapping $\mathcal{M}_{\tilde{g},D}:(0,1]\to(0,1]$ defined as $\mathcal{M}_{\tilde{g},D}(\cdot)\equiv\exp\{-[\tilde{g}^{-1}(\cdot)]^{1/D}\}$ and its inverse $\mathcal{M}^{-1}_{\tilde{g},D}(\cdot)\equiv\tilde{g}([-\ln(\cdot)]^D)$. Equation (6) can then be rewritten as

$$\mathcal{M}^{-1}_{\tilde{g},D}\{\mathcal{M}_{\tilde{g},D}[\tilde{g}(k/2^D)]\mathcal{M}_{\tilde{g},D}[\tilde{g}(k/2^D)]\} = \mathcal{M}^{-1}_{\tilde{g},D}\{\mathcal{M}_{\tilde{g},D}[\tilde{g}(k)]\}, \quad (8)$$

which exemplifies the fact that $\mathcal{M}_{\tilde{g},D}$ establishes an isomorphism between the generalized and ordinary multiplications. A key consequence is that $\hat{g}\equiv\mathcal{M}_{\tilde{g},D}(\tilde{g})$ obeys a condition of the form (5) with the exponent $1/2$ replaced by D. This is the well-known Lévy-Gnedenko stability condition for independent random variables, which has the singular Lévy CF $\exp(-|k|^{1/D})$ as solution [1,4]. Consistently, of course, $\hat{g}(k)=\exp(-|k|^{1/D})$. Notice that the Lévy-stable $\hat{g}$ loses the meaning of CF for $D<1/2$, because the corresponding PDF ceases to be positive definite. Here this limitation does not apply, since the inverse FT of $\hat{g}$ does not represent a PDF.

According to the Lévy-Gnedenko limit theorem [1], the Lévy-stable CF is approached in the $N\to\infty$ limit for the sum of N independent variables whose individual CF has the same leading singularity $\sim|k|^{1/D}$ at $k=0$. This circumstance and the above mapping suggest to look at the counterpart of such convergence process in the space of correlated PDF's. In analogy with the independent case [Eq. (4)], we can indeed construct the CF of the sum of N correlated variables, starting from a single-variable CF $\tilde{p}_1$, but replacing the ordinary multiplication with the generalized one, as specified by the chosen $\tilde{g}$ and D. As before, $\tilde{p}_1$ is assumed to be regular and to generate a finite second moment, but in general will not coincide with $\tilde{g}$. If we pose $\hat{p}_1\equiv\mathcal{M}_{\tilde{g},D}(\tilde{p}_1)$, this function is singular at $k=0$: $\hat{p}_1=1-|k|^{1/D}+O(|k|^{2/D})$. Hence, by the Lévy-Gnedenko limit theorem, $[\hat{p}_1(k/N^D)]^N\sim\hat{g}(k)=\exp(-|k|^{1/D})$ for $N\to\infty$ [1,4]. The above isomorphism guarantees then that [14]

$$\left[\tilde{p}_1\left(\frac{k}{N^D}\right)\right]^{\otimes N} \equiv \underbrace{\tilde{p}_1\left(\frac{k}{N^D}\right)\otimes\cdots\otimes\tilde{p}_1\left(\frac{k}{N^D}\right)}_{N\text{ terms}} \sim \tilde{g}(k) \quad (9)$$

for $N\to\infty$ and for any p_1 with finite variance $\sigma^2=2$. Equation (9) follows from the fact that $\mathcal{M}^{-1}_{\tilde{g},D}(\hat{g})=\tilde{g}$ and expresses a CLT for general $\tilde{g}$ and D. Starting from a single-variable PDF p_1, the iterated generalized multiplication of its CF yields the CF for the sum of the variables in a process where the X_i's are correlated. In force of the CLT, this process leads asymptotically to the universal anomalous scaling specified by $\tilde{g}$ and D.

The validity of Eq. (9) does not require $D>1/2$ because again the inverse FT of $\hat{p}_1$ is not constrained to remain positive. However, other positivity requirements can pose limits on the choice of p_1. Indeed, there is no guarantee that, if $\tilde{p}_1$ is a CF, $\tilde{p}_1\otimes\tilde{p}_1$ will also be, in general. Since positivity control is a hard mathematical issue [1,15], we addressed it numerically by analyzing the convergence process in Eq. (9) for several $\tilde{g}$'s and $\tilde{p}_1$'s. We verified that as long as $\tilde{p}_1$ has the same general properties assumed for $\tilde{g}$, $p_N(y)\equiv(1/2\pi)\int_{-\infty}^{+\infty}\exp(-iky)[\tilde{p}_1(k/N^D)]^{\otimes N}dk$ remains positive definite for any N. For illustration we report the results for the case $\tilde{g}(k)=1/(1+k^2)$—i.e., $g(y)=\exp(-|y|)/2$ and $\tilde{g}^{-1}(a)=-\sqrt{1/a-1}$ for $a\in(0,1]$. Figures 1 and 2 show the evolution of $p_X(x)$ under the generalized multiplications of the single-variable CF for a Gaussian $p_1(x)=\exp(-x^2/4)/\sqrt{4\pi}$ and, respectively, $D=0.9$ and $D=0.25$. In general, larger D's imply faster convergence to the fixed point. However, after a sufficient number of iterations, all the collapses we checked are almost perfect. One may wonder if Eq. (9) remains valid for more general forms of p_1. A first extension of the above results can be obtained by considering single-variable PDF's with two symmetric peaks, which, e.g., could be relevant for magnetic or diffusive phenomena. In this case $\tilde{p}_1$ is not strictly positive anymore, so that a continuation of the generalized multiplication to negative values is required. One can indeed find a continuation that preserves the isomorphism with the ordinary multiplication [12]. We verified [12] that while Eq. (9) remains valid as-

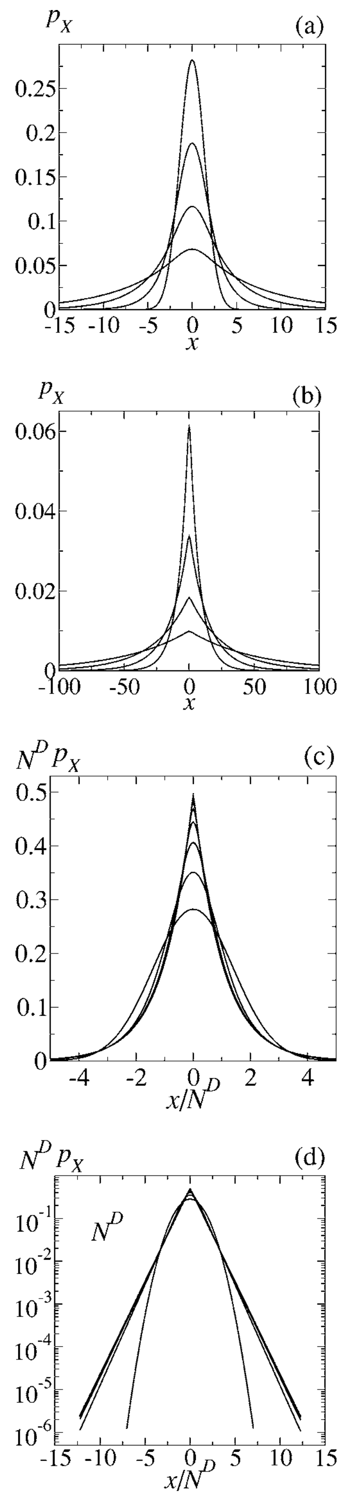

FIG. 1. Gaussian p_1 and $D=0.9$. Plot of p_X for $N=2^k$, $k=0,1,2,3$ (a) and $k=4,5,6,7$ (b). As N increases, the central peak of p_X decreases. The rescaled plots for $k=0,1,\ldots,7$ have increasing peaks and reveal convergence for large N: (c) and (d).

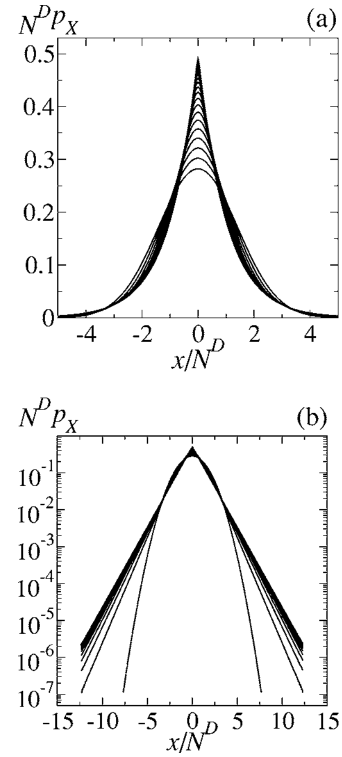

FIG. 2. Rescaled plots of p_X with $D=0.25$, $N=2^k$, $k=0,1,\ldots,22$: (a) and (b). Comparison with Fig. 1 indicates slower convergence when D is smaller.

ymptotically, for this new class of $\widetilde{p}_1$'s positivity problems of the iterated PDF's can arise during the initial stages of the convergence process.

In all examples, only the constraint $\sigma^2=2$ and, possibly, positivity requirements pose limitations on the domain of attraction of the stable PDF. This universality, typical of the CLT, is a consequence of the multiplicative structure in Eq. (9). Indeed, in Eq. (9) normalization $[\widetilde{p}_N(0)=1^{\otimes N}=1]$, centering $[\int_{-\infty}^{+\infty} y p_N(y)dy=0]$, and variance $[\int_{-\infty}^{+\infty} y^2 p_N(y)dy=\sigma^2=2]$ are conserved, as in the independent case. Thus, the trajectory described by p_N in function space differs substantially from a typical RG flow, in which the variance is not conserved and relevant scaling fields determine the critical surface [10]. Here, relevant fields are not present [16] and the result in Eq. (9) identifies at least a subset of the universality domain of the assumed asymptotic anomalous scaling specified by g and D. A further feature of our findings is that the choice of g does not imply a selection on admissible values of D and vice versa. This appears consistent with the variety of different anomalous scaling functions and exponents observed in natural phenomena [18].

A basic issue is that of identifying an explicit mechanism by which correlations are introduced by the $\otimes$ multiplication. Let us consider the normalized partial sums $Y_1=\Sigma_{i=1}^{N/2}X_i/(N/2)^D$ and $Y_2=\Sigma_{i=N/2+1}^{N}X_i/(N/2)^D$. The correlations between Y_1 and Y_2 are fully specified once their joint PDF $p_N^{(2)}(y_1,y_2)$ is given. Knowledge of p_N alone does not allow one to determine $p_N^{(2)}$ in general since many $p_N^{(2)}$ are such to satisfy the obvious condition $p_N(y)=\int_{-\infty}^{+\infty} p_N^{(2)}(y_1,y_2)\,\delta(y-y_1-y_2)dy_1dy_2$. Thus, many different correlation patterns are compatible with the anomalous scaling of p_N. Equation (9) asymptotically fixes $\widetilde{p}_N^{(2)}(k/2^D,k/2^D)\sim\widetilde{g}(k/2^D)\otimes g(k/2^D)$, where $\widetilde{p}_N^{(2)}$ is the FT of $p_N^{(2)}$. In the independent case this last result would hold with

$\tilde{g}$ Gaussian, $D=1/2$ and the standard multiplication replacing $\otimes$. Furthermore, $\tilde{p}_N^{(2)}(k_1/2^{1/2},k_2/2^{1/2})\sim\tilde{g}(k_1/2^{1/2})\tilde{g}(k_2/2^{1/2})$ would clearly hold in that case also for $k_1\neq k_2$, so that $\tilde{p}_N^{(2)}$ would be fully specified in terms of $\tilde{g}$. It is natural to ask if, under suitable assumptions, an analogous factorization of $\tilde{p}_N^{(2)}$ with $k_1\neq k_2$ holds in terms of the $\otimes$ multiplication also in the correlated case. This property would imply the possibility of expressing the correlations determining the anomalous scaling in terms of p_N alone. It can be shown [12] that such a factorization is indeed possible if additional symmetries of $p_N^{(2)}$ are assumed, like the vanishing of linear correlations between Y_1 and Y_2, $\int_{-\infty}^{+\infty}y_1y_2p_N^{(2)}(y_1,y_2)dy_1dy_2=0$. This vanishing does not hold for other stochastic processes possessing anomalous scaling considered in the literature, like, for example, fractional Brownian motion [17]. Because of market efficiency, the vanishing of linear correlations characterizes, e.g., financial time series, where p_N is the PDF of the normalized return of an index in time N. For such series we were able to show [12] that the asymptotic form of $\tilde{p}_N^{(2)}$ can be uniquely determined starting from p_N and using a $\otimes$ multiplication. The agreement of the theoretical predictions with the empirically sampled $p_N^{(2)}$ is quite remarkable [12]. Thus, the generalized multiplication operation defined above is also a key for the full characterization of a relevant class of stochastic evolution processes.

One can also establish a connection between the present CLT and anomalous diffusion. Let us consider a single-variable PDF of the form $p_1(x)=[\delta(x-\Delta)+\delta(x+\Delta)]/2$ and define a time $t\equiv N\tau$. Here, releasing the condition $\sigma^2=2$, Δ and τ are, respectively, the space and time spans of random steps, and t is the time at which the Nth step occurs. In the continuum limit $N\to\infty$, $\tau\to 0$, $\Delta\to 0$, such that t and $\mathcal{D}^{1/2D}\equiv\Delta^{1/D}/\tau$ remain finite, one recovers [12] $\lim_{N\to\infty}[\tilde{p}_1(k)]^{\otimes N}\equiv\tilde{p}(k,t)=\mathcal{M}_{\tilde{g},D}^{-1}[\hat{p}(k,t)]$, where $\hat{p}(k,t)$ satisfies the standard Lévy diffusion equation [19]

$$\frac{\partial\hat{p}(k,t)}{\partial t}=-\frac{\mathcal{D}^{1/2D}|k|^{1/D}}{2^{1/2D}}\hat{p}(k,t). \qquad (10)$$

Assuming $p(x,0)=\delta(x)$, one gets the solution $\hat{p}(k,t)=\exp(-\mathcal{D}^{1/2D}|k|^{1/D}t/2^{1/2D})$, which corresponds to $\tilde{p}(k,t)=\tilde{g}(\mathcal{D}^{1/2}kt^D)=1-\mathcal{D}k^2t^{2D}/2+O(\mathcal{D}^2k^4t^{4D})$. Hence, $\langle x^2\rangle(t)=\mathcal{D}t^{2D}$ [12]. Thus, correlated sub- ($D<1/2$) and super- ($D>1/2$) diffusive solutions can be obtained through our mapping from the propagator of the uncorrelated Lévy diffusion equation (10). This enables the description of the evolution towards the asymptotic anomalous diffusion regime (analogous to Figs. 1 and 2) without introducing a broad distribution of waiting times elapsing between successive steps as is done in the continuous-time random walk approach [7,20].

In summary, assuming anomalous scaling [Eqs. (1) and (2)] we have constructed a multiplicative functional identity for the CF of the asymptotic sum of strongly correlated random variables which allowed the definition of a generalized multiplication. An isomorphism between this multiplication and the ordinary one leads to establish a CLT in which the anomalous scaling represents the asymptotic limit. Thus, for a given asymptotic anomalous scaling form we have characterized a large class of processes falling in the corresponding universality domain. In particular cases our strategy also allows a full determination of the joint probabilities and thus of the correlations of partial sums of the random variables [12]. In the context of stochastic processes, the presence of correlations implies that past events have an influence on the future behavior. Knowledge of the joint probability of consecutive events (like Y_1 and Y_2) hence entails a predictive power which has been recently exploited in finance [12].

We thank M. Baiesi for collaboration and J.R. Banavar, H. Fogedby, G. Jona-Lasinio, F. Leyvraz, D. Mukamel, E. Orlandini, M. Paczuski, C. Tebaldi, and H. Touchette for discussions.

[1] See, e.g., B. V. Gnedenko and A. N. Kolmogorov, *Limit Distributions for Sums of Independent Random Variables* (Addison-Wesley, Reading, MA, 1954).
[2] A. I. Khinchin, *Mathematical Foundations of Statistical Mechanics* (Dover, New York, 1949).
[3] C. W. Gardiner, *Handbook of Stochastic Methods*, 3rd ed. (Springer-Verlag, Berlin, 2004).
[4] P. Lévy, *Théorie de l'Additions des Variables Aléatoires* (Gauthier Villars, Paris, 1954).
[5] L. P. Kadanoff, *Statistical Physics, Statics, Dynamics and Renormalization* (World Scientific, Singapore, 2005).
[6] B. Mandelbrot, J. Business **36**, 394 (1963); R. Mantegna and H. E. Stanley, *An Introduction to Econophysics* (Cambridge University Press, Cambridge, England, 2000).
[7] J. Bouchaud and A. Georges, Phys. Rep. **195**, 127 (1990); R. Metzler and J. Klafter, *ibid.* **339**, 1 (2000).
[8] S. C. Kou and X. S. Xie, Phys. Rev. Lett. **93**, 180603 (2004).
[9] We do not address weakly dependent random variables as, e.g., in B. M. Brown, Ann. Math. Stat. **42**, 59 (1971).
[10] G. Jona-Lasinio, Nuovo Cimento Soc. Ital. Fis., B **26**, 99 (1975); Phys. Rep. **352**, 439 (2001) and references therein.
[11] With additional technical effort [12] our results can be extended to more general cases, like the magnetization PDF of the critical 2D Ising model [10].
[12] F. Baldovin and A. L. Stella (unpublished).
[13] Similar multiplications have been considered, e.g., by G. Kaniadakis, Phys. Rev. E **66**, 056125 (2002).
[14] The rigorous conditions under which our correspondence becomes a homeomorphism are not discussed here.
[15] B. G. Giraud and R. Peschanski, Acta Phys. Pol. B **37**, 331 (2006).
[16] For Gaussian g and arbitrary D one can linearize the transformation in Eq. (9) at the asymptotic fixed point and compute its eigenvalues explicitly [12].
[17] See, e.g., B. B. Mandelbrot, *The Fractal Geometry of Nature* (Freeman, New York, 1982).
[18] This contrasts with the point of view expressed in C. Tsallis, Physica A **365**, 7 (2006), and references therein.
[19] B. D. Hughes, M. F. Shlesinger, and E. W. Montroll, Proc. Natl. Acad. Sci. U.S.A. **78**, 3287 (1981).
[20] M. F. Shlesinger, B. J. West, and J. Klafter, Phys. Rev. Lett. **58**, 1100 (1987).