


Ubiquitous Power Law Scaling in Nonlinear Self-Excited Hawkes Processes

Kiyoshi Kanazawa^{*}*Faculty of Engineering, Information and Systems, University of Tsukuba, Tennodai, Tsukuba, Ibaraki 305-8573, Japan and JST, PRESTO, 4-1-8 Honcho, Kawaguchi, Saitama 332-0012, Japan*Didier Sornette[†]*Department of Management, Technology and Economics, ETH Zurich, Zurich 8092, Switzerland and Institute of Risk Analysis, Prediction, and Management (Risks-X), Academy for Advanced Interdisciplinary Studies, Southern University of Science and Technology (SUSTech), Shenzhen 518055, China* (Received 15 February 2021; revised 10 August 2021; accepted 6 October 2021; published 28 October 2021)

The origin(s) of the ubiquity of probability distribution functions with power law tails is still a matter of fascination and investigation in many scientific fields from linguistic, social, economic, computer sciences to essentially all natural sciences. In parallel, self-excited dynamics is a prevalent characteristic of many systems, from the physics of shot noise and intermittent processes, to seismicity, financial and social systems. Motivated by activation processes of the Arrhenius form, we bring the two threads together by introducing a general class of nonlinear self-excited point processes with fast-accelerating intensities as a function of “tension.” Solving the corresponding master equations, we find that a wide class of such nonlinear Hawkes processes have the probability distribution functions of their intensities described by a power law on the condition that (i) the intensity is a fast-accelerating function of tension, (ii) the distribution of marks is two sided with nonpositive mean, and (iii) it has fast-decaying tails. In particular, Zipf’s scaling is obtained in the limit where the average mark is vanishing. This unearths a novel mechanism for power laws including Zipf’s law, providing a new understanding of their ubiquity.

DOI: 10.1103/PhysRevLett.127.188301

Introduction.—Many different types of data in the natural and social sciences exhibit power law density distributions of the size or frequencies of their characteristic variables. Namely, the probability density function (PDF) $P(S)$ of a variable S is given by $P(S) \sim 1/S^{1+\alpha}$ for large S values, with $\alpha > 0$. Many mechanisms have been proposed to rationalize it [1–4], such as proportional growth with additional conditions [5], family transformation of the Bose-Einstein distribution [6], least-effort principles [7], optimization between efficiency and faithfulness of self-reproduction [8] and so on.

Self-excited point processes assume that past events strongly influence the occurrence of future events. The Hawkes process [9] is the simplest such process, where the intensity (probability per unit time that a new event occurs) is linear in the sum of the triggering influence of all past events. In the last decade, the Hawkes process and generalizations have enjoyed an explosive growth in the investigation of their properties and in a large set of applications in all fields of knowledge [10–13].

Theoretically challenging, nonlinear self-excited processes have been scarcely investigated [14,15] except for a few special cases [16], even if they are *a priori* more suited to represent the interplay between stochasticity and nonlinear dynamics in many complex systems. Here, we study a class of nonlinear Hawkes processes characterized by fast-accelerating intensities as a function of an auxiliary field called the “tension,” and report the first explicit solution that is applicable to a wide class of nonlinear Hawkes processes. We find that this class of nonlinear Hawkes family universally exhibits intensity distributions with power law tails. In particular, Zipf’s scaling naturally appears when the distribution of marks is symmetric. A weaker condition is that the average mark is vanishing. These models are motivated by activation processes of the Arrhenius form, which are relevant in many applications in physics and also in seismicity and finance modeling as explained below.

Model.—The key ingredients of the nonlinear self-exciting Hawkes process considered here are the *intensity* $\lambda(t)$ and *tension* $\nu(t)$. Let us introduce a time series $\{t_i\}_i$, representing the time stamps of events, such as earthquakes, retweets on Twitter, or neural discharges in a brain. The intensity $\lambda(t)$ fully characterizes the statistics of occurrence of events, such that an event occurs with probability $\lambda(t)dt$ during the interval $[t, t + dt)$. We assume

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article’s title, journal citation, and DOI.

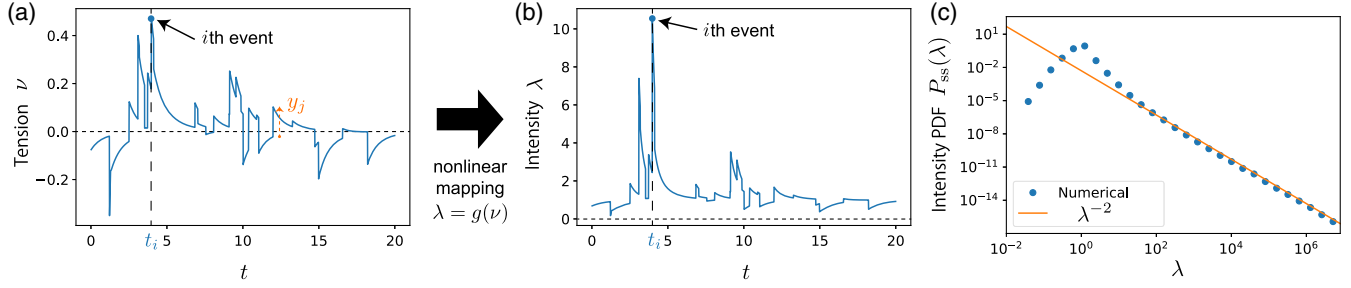


FIG. 1. (a) Sample tension trajectory $\{\nu(t)\}_t$ and (b) the corresponding intensity trajectory $\{\lambda(t)\}_t$ generated by the nonlinear Hawkes process (1) with (2). The tension trajectory $\{\nu(t)\}_t$ exhibits random jumps with size y_i distributed according to $\rho(y)$ and the corresponding intensity λ is given by $\lambda(t) = g(\nu(t))$. (c) Steady intensity distribution $P_{SS}(\lambda)$ for the exponential fast-accelerating intensity $g(\nu) \propto e^{\beta\nu}$ in the case where the mean mark is zero: $m = 0$, exhibiting Zipf's law $\propto \lambda^{-2}$. These figures are based on Monte Carlo simulations of the nonlinear Hawkes process with $\lambda = g(\nu) = \lambda_0 e^{\beta\nu}$, $h(t) = \sum_{k=1}^K \tilde{h}_k e^{-t/\tau_k}$, and $\rho(y) = (1/\sqrt{2\pi\sigma^2})e^{-y^2/(2\sigma^2)}$ with $K = 3$, $(\tau_1, \tau_2, \tau_3) = (1, 0.5, 2)$, $(\tilde{h}_1, \tilde{h}_2, \tilde{h}_3) = (0.5, 0.6, 0.1)$, $\lambda_0 = 1$, $\beta = 5$, and $\sigma = 0.1$ (see Supplemental Material [17] for the detailed numerical scheme).

that the intensity is a nonlinear positive and monotonically increasing function of the system tension $\nu(t)$,

$$\lambda(t) = g[\nu(t)], \quad (1)$$

where $g(\nu)$ is called the tension-intensity map. The tension $\nu(t)$ quantifies the total stress due to historical events, such as resulting from elastic deformations of the crust induced by earthquakes. In finance, $\lambda(t)$ can represent the rate of volatility jumps and $\nu(t)$ is the rate of financial returns whose amplitude exceeds some threshold. The tension at a given time is obtained as the sum of perturbations over all past events [see Fig. 1(a) for a realization], such that

$$\nu(t) = \sum_{i=1}^{N(t)} y_i h(t - t_i), \quad (2)$$

where each event i has a mark y_i distributed according to the PDF $\rho(y)$ and $N(t)$ is the number of events during $[0, t)$. Combining the relation (1) between tension and intensity and (2), we obtain a nonlinear version of the Hawkes process: $\lambda(t) = g[\sum_{i=1}^{N(t)} y_i h(t - t_i)]$. To represent that high tension promotes future events, we assume that the tension-intensity map is a nondecreasing function. For an affine function $g(\nu) = \nu_0 + \nu$, the model (1) with (2) reduces to the original linear Hawkes process and y_i can be interpreted as the average number of events of first generation triggered by event i and is thus called the fertility of event i , by imposing $\int_0^\infty h(t) dt = 1$. The memory function $h(t) \geq 0$ controls the distribution in time of the triggered events and decays to zero for large t .

Conditions.—There is a large variety of nonlinear Hawkes processes defined via the pair of functions $g(\nu)$ and $\rho(y)$. Here, we focus on the wide class of nonlinear Hawkes processes that obey the three following conditions: (i) the tension-intensity map $g(\nu)$ is a fast-accelerating intensity, defined to diverge faster than any second-order

polynomial: $g(\nu) > O(\nu^2)$ for large ν ; (ii) the mark distribution is two sided with nonpositive mean, such that $\int_{-\infty}^0 \rho(y) dy \neq 0$, $\int_0^\infty \rho(y) dy \neq 0$, and $m := \int_{-\infty}^\infty y \rho(y) dy \leq 0$; and (iii) the mark distribution has fast-decaying tails, such that $\Phi(x) := \int_{-\infty}^\infty dy \rho(y) (e^{xy} - 1)$ exists. $\Phi(x) = 0$ has only two roots, where one is zero and the other is $c^* \geq 0$.

Here, the Bachmann-Landau-like inequality notation $a(x) > O[b(x)]$ means $\lim_{x \rightarrow \infty} a(x)/b(x) = \infty$. Also, the condition (iii) essentially means that fat-tail mark distributions, such as power law distributions, are out of scope in this Letter. Remarkably, all nonlinear Hawkes processes satisfying these three conditions have their steady-state intensity PDFs obeying the universal power law scaling, as we show below. Typical analytical forms satisfying condition (i) include $g(\nu) \propto \nu^n$ with $n > 2$ and

$$g(\nu) = \lambda_0 e^{\beta\nu}, \quad (3)$$

which is motivated by the physics of rupture [22] and earthquakes [23,24], modeled as activated processes following an Arrhenius law. Indeed, we assume that the tension ν is proportional to the seismic energy E (itself proportional to the total mechanical stress in the Earth crust), and let us assume that an earthquake happens if the system's state jumps over an energy barrier E_0 from a metastable state to another. According to the Arrhenius law, the escape rate is proportional to $e^{-\beta(E_0 - E)} \propto e^{\beta\nu}$ with a disordered-enhanced effective inverse temperature $\sim \beta$ [22], consistently with Eq. (3). This exponential dependence (3) also encompasses the class of multifractal processes emerging from the interplay between exponential activation and long memory [25], which have been shown to be relevant to model financial volatility [26]. By construction, the tension is dependent on all the marks of previous events, while future marks are drawn independently of the past history. This is consistent with the empirical unpredictability of earthquake magnitudes.

Condition (ii) guarantees the stationarity of the model as a result of the cumulative contribution of the negative marks $y_i < 0$, which prevent ν from diverging. An event with a negative (positive) mark y_i is likely to inhibit (induce) future events. The coexistence of events that inhibit and of events that promote future activity in our nonlinear Hawkes model is a fundamental extension to the general class of Hawkes processes. This allows us to realistically account for a ubiquitous inhibitory effect in real complex systems, such as the random mechanical stress relaxation after earthquake in seismology, or inhibitory synaptic potentials in neural networks. Note that, in contrast, the standard Hawkes process and many other versions only have positive marks, corresponding to taking into account excitations exclusively.

Power law intensity PDF.—When conditions (i)–(iii) are satisfied, the steady-state PDF of the intensity λ is analytically given by

$$P_{\text{SS}}(\lambda) \propto \lambda^{-1} \left[e^{-a\nu} \left\{ \frac{dg}{d\nu} \right\}^{-1} \right]_{\nu=g^{-1}(\lambda)}, \quad a := \frac{c^*}{h(0)}, \quad (4)$$

with c^* being the nonnegative root of $\Phi(c^*) = 0$ [see Supplemental Material (SM) [17]]. This formula readily reduces to various power law asymptotic forms, such as

$$P_{\text{SS}}(\lambda) \propto \begin{cases} \lambda^{-2-\beta^{-1}a} & [\text{for } g(\nu) \simeq \lambda_0 e^{\beta\nu}, \beta > 0], \\ \lambda^{-2+\frac{1}{n}} e^{-a(\lambda/\lambda_0)^{1/n}} & [\text{for } g(\nu) \simeq \lambda_0 \nu^n, n > 2]. \end{cases} \quad (5)$$

Beyond conditions (i)–(iii), no other properties or details, including the shape of the memory function, change the robust classes given by expressions (5).

Zipf's law.—Result (5) implies that Zipf's scaling appears as an important subclass of the nonlinear Hawkes processes as a special case $a = 0$:

$$P_{\text{SS}}(\lambda) \propto \begin{cases} \lambda^{-2} & (\text{for } g(\nu) > O(\nu^n) \text{ for any } n) \\ \lambda^{-2+\frac{1}{n}} & (\text{for } g(\nu) \simeq \lambda_0 \nu^n, n > 2), \end{cases} \quad (6)$$

except for minor logarithmic corrections. The condition $a = 0$ is realized exactly when $c^* = 0$, which holds for zero-mean mark $m = 0$, implying $a = c^*/h(0) = 0$. This is for instance realized for symmetric mark distribution $\rho(y) = \rho(-y)$. Approximate Zipf's distributions are obtained when $a = c^*/h(0)$ is small, which occurs for large $h(0)$.

Symmetric mark distributions are realized in the physics of earthquakes as discussed in [24]. Indeed, the stress perturbations induced by a (small) earthquake correspond to the stress field of a double couple, which can be simply represented by a concentrated set of four forces of the same norm, summing to zero (zero total force) and with total torque also equal to zero. A large earthquake is just a set of double-couple sources placed along its fault surface.

The stress induced by a double couple has a nice butterfly symmetry with four lobes, two positive and two negative ones, and is perfectly symmetric. With the correspondence that the tension ν is proportional to stress, and is given by (2), and that the exponential intensity function (3) derives from the physics of earthquake nucleation with Arrhenius law with an effective temperature [24], this justifies the symmetric property of the distribution of marks for earthquakes.

The results of numerical simulations for $m = 0$ (zero-mean marks) are presented in Fig. 1. Panel (a) shows a typical realization of ν for case (3), while panel (b) shows the derived temporal evolution of λ . Panel (c) shows the corresponding steady-state PDF of λ obeying Zipf's law (see also SM [17] for numerical simulations for the negative-mean cases $m < 0$).

Field-master equation.—Our general result for a large class of memory functions can be derived using our recently introduced field-master-equation framework [27,28] [see also Supplemental Material (SM) [17] for the technical detail]. The main idea is to convert the original low-dimensional non-Markovian stochastic process onto a high-dimensional Markovian field dynamics. This technique is called Markov embedding and has been applied for some special cases, such as memory functions composed of discrete sums of exponentials (see Refs. [29,30] for the generalized Langevin equation and Refs. [31,32] for Hawkes processes).

The Markov embedding scheme can be formulated for the nonlinear Hawkes process (1) with (2) as follows. Let us decompose the memory kernel $h(t)$ as a continuous sum of exponentials. This amounts to representing $h(t)$ as a Laplace-like transform of another function $\tilde{h}(x)$ of the auxiliary variable $x \in (0, \infty)$:

$$h(t) = \int_0^\infty dx \tilde{h}(x) e^{-t/x}. \quad (7)$$

Based on this decomposition, the original process (1) with (2) is equivalent to a Markovian stochastic partial differential equation (SPDE) for the excess tension $\{z(t, x)\}_{x \in \mathbb{R}^+}$

$$\frac{\partial z(t, x)}{\partial t} = -\frac{z(t, x)}{x} + \tilde{h}(x) \xi_{\rho(y); \lambda(t)}^P \quad (8)$$

with the total tension $\nu(t) = \int_0^\infty dx z(t, x)$ [see Figs. 2(a) and 2(b) for schematics of the Markov embedding scheme] and the compound Poisson noise $\xi_{\rho(y); \lambda(t)}^P = \sum_{i=1}^{N(t)} y_i \delta(t - t_i)$. Remarkably, while the original process is non-Markovian in a one-dimensional space $\nu(t)$, the field dynamics is Markovian in the infinite-dimensional space $\{z(t, x)\}_{x \in \mathbb{R}^+}$.

The equivalence between the original nonlinear Hawkes process (1) with (2) and the SPDE (8) can be shown as follows: the formal solution of the SPDE (8) is given by $z(t, x) = \int_{-\infty}^t \tilde{h}(x) e^{-(t-s)/x} \xi_{\rho(y); \lambda}^P(s) ds = \sum_{i=1}^{N(t)} \tilde{h}(x) y_i e^{-(t-t_i)/x}$.

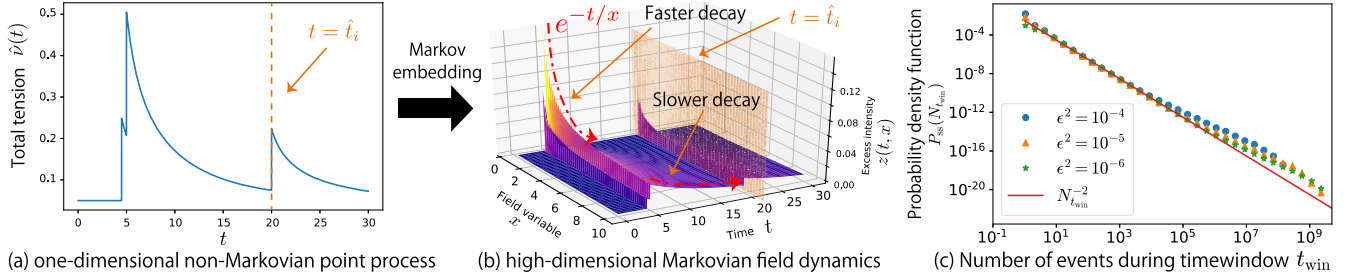


FIG. 2. (a),(b) Schematic of the Markov embedding from the one-dimensional non-Markovian process $\nu(t)$ to the infinite-dimensional Markovian field dynamics $\{z(t, x)\}_{x \in \mathbf{R}^+}$. The original process (a) is non-Markovian because its time evolution (1) with (2) depends on all the history $\{\nu(s)\}_{s \leq t}$. On the other hand, the field dynamics (b) is Markovian because its time evolution (8) depends only on the current configuration of the field variable $\{z(t, x)\}_{x \in \mathbf{R}^+}$. Note that the auxiliary field variable $x \in \mathbf{R}^+$ is introduced according to Eqs. (7) and (8) and is interpreted as a “position” at which the field is evaluated. The decay speed is faster for smaller x , while it is slower for larger x according to Eq. (8). (c) Statistics for the number of events $N_{t_{win}}$ occurring in short time windows of size t_{win} in the diffusive scaling limit (14) for various ϵ . We observe a Zipf law up to the upper cutoff $N_{cut} = O(\epsilon^2)$. Beyond the cutoff, a fatter tail is numerically observed (see SM [17]).

The total tension is then given by $\nu(t) := \int_0^\infty dx z(t, x) = \sum_{i=1}^{N(t)} y_i \int_0^\infty dx \tilde{h}(x) e^{-(t-t_i)/x} = \sum_{i=1}^{N(t)} y_i h(t - t_i)$. This is equivalent to (2). We thus find that the Markovian SPDE (8) is a correct representation after Markov embedding.

The SPDE (8) can be regarded as the “physical dynamics” of the field variable $\{z(t, x)\}_{x \in \mathbf{R}^+}$, since x can be considered as the “physical position” in $\mathbf{R}^+ := (0, \infty)$ on which the field is evaluated. This interpretation has the advantage that the functional methods for various SPDEs of stochastic field dynamics are available for advanced analytics (e.g., the functional Fokker-Planck equations for the reaction-diffusion equations [33]).

Since the SPDE (8) is Markovian, we can obtain the corresponding master equation. By introducing the probability density functional (PDF) $P_t[z] := P_t[\{z(t, x)\}_{\mathbf{R}^+}]$, the field-master equation is given by

$$\frac{\partial P_t[z]}{\partial t} = (\mathcal{L}_A + \mathcal{L}_J)P_t[z] \quad (9)$$

with the advective and jump Liouville operators \mathcal{L}_A and \mathcal{L}_J , respectively, defined by

$$\mathcal{L}_A P_t := \int_0^\infty dx \frac{\delta}{\delta z(x)} \frac{z(x)}{x} P_t[z], \quad (10a)$$

$$\mathcal{L}_J P_t := \int_{-\infty}^\infty dy \rho(y) G[z - y\tilde{h}] P_t[z - y\tilde{h}] - G[z] P_t[z], \quad (10b)$$

with $G[z] := g[\int_0^\infty z(t, x) dx]$ and $\rho(y)$ is the mark distribution.

Note that $P_t[z]$ is a path probability measure: the probability is given by $P_t[z] \mathcal{D}z$ that the configuration of the field variable is nearly equal to $\{z(t, x)\}_{\mathbf{R}^+}$, where $\mathcal{D}z := \prod_{x \in \mathbf{R}^+} dz(x)$ is the path-integral volume element. In addition, the ensemble average $\langle A \rangle$ is given by the path

integral $\langle A \rangle := \int AP_t[z] \mathcal{D}z$. Technically, the field-master equation (9) should be interpreted as a formal limit from discrete underlying descriptions according to the standard convention (see Ref. [33] and SM [17]). The steady-state solution $P_{SS}[z]$ is related to the steady-state intensity PDF $P_{SS}(\lambda)$ as $P_{SS}(\lambda) = \int \mathcal{D}z \delta(\lambda - \int_0^\infty dx z(t, x)) P_{SS}[z]$.

Derivation.—We now provide an outline of the theoretical derivation of the solution of the field-master equation (see SM [17] for more detailed calculations and an illustrative-case study with the exponential memory). Let us introduce $\phi[z] := G[z] P_{SS}[z]$ to rewrite Eq. (9) in the steady state as

$$0 = \int_0^\infty dx \frac{\delta}{\delta z} \left(\frac{z\phi[z]}{xG[z]} \right) + \int_{-\infty}^\infty dy \rho(y) \phi[z - y\tilde{h}] - \phi[z]. \quad (11)$$

Since the first term is negligible for large z assuming the condition (i), the asymptotic solution satisfies

$$\int_{-\infty}^\infty dy \rho(y) \phi[z - y\tilde{h}] - \phi[z] \approx 0 \quad \text{for large } z. \quad (12)$$

Under conditions (ii) and (iii), its solution is given by $\phi[z] \approx C_0[Z'] e^{-c^* W}$ with $W := z(x^*)/h(x^*)$ and $Z'(x) := z(x) - h(x)z(x^*)/h(x^*)$ for $\mathbf{R}^+ := \mathbf{R}^+ \setminus \{x^*\}$ by selecting an appropriate number $x^* \in \mathbf{R}$. Here, C_0 is an arbitrary functional without W as an argument. After a path integral for marginalization, we obtain

$$P_{SS}(\nu) := \int \mathcal{D}z P_{SS}[z] \delta\left(\nu - \int_0^\infty dx z(x)\right) \approx \frac{e^{-c^* \nu/h(0)}}{g(\nu)}. \quad (13)$$

Equation (4) then follows by the change of variable $\nu \rightarrow \lambda$.

Intuition.—Let us consider the case of an exponential growing intensity (3) and a simple exponential memory

kernel $h(t) = (n/\tau)e^{-t/\tau}$, where the integral of the memory, $n = \int_0^\infty h(t)dt > 0$, would be interpreted as the branching ratio in the linear case. Suppose that the initial tension is zero and thus the initial intensity is λ_0 . Naively, one could infer that the typical waiting time till the next event, the expected event interval (EEI), is given by $(1/\lambda_0)$. Choosing the parameters such that $\tau \ll (1/\lambda_0)$ would imply that the influence of an event in triggering future events is extremely localized temporally and one should expect no intermittency, no power laws and a rather trivial behavior. This reasoning is wrong as it neglects the nonlinear nature of the model with strong feedback loops. Indeed, defining the small parameter $\eta := \lambda_0\tau \ll 1$, after one event occurs with positive mark $y > 0$, the intensity is given instantaneously by $\lambda(t) = \lambda_0 e^{\beta\lambda_0 y n/\eta}$ and the corresponding EEI is of the order of $[1/\lambda(t)] = (1/\lambda_0)e^{-\beta\lambda_0 y n/\eta}$. Paradoxically, as the memory τ of the event is vanishingly smaller than the naive characteristic timescale $(1/\lambda_0)$, the time needed for the next event to be triggered becomes exceedingly smaller, since $(1/\lambda_0)e^{-\beta\lambda_0 y n/\eta} \ll \tau \ll (1/\lambda_0)$ for sufficiently small η such that $\eta \ln(1/\eta) \ll \beta\lambda_0 y n$. Hence, in contradiction with the naive view, a very short memory enhances triggering and creates a very rich bursty dynamics of events. Readily generalized to multiple events, this reasoning gives an intuition on the basic source of the scale-free nature of the power law intensity PDF (5), suggesting the absence of both characteristic intensity and EEI.

Number-of-events statistics.—The intensity PDF is a fundamental quantity to characterize temporal properties of point processes and allows one to derive various other quantities. One such variable that is directly observable is the total number of events $N_{t_{\text{win}}}$ occurring in a finite time window t_{win} . Assuming symmetric mark distributions $\rho(y) = \rho(-y)$, we show that Zipf's law also holds for the distribution of $N_{t_{\text{win}}}$. For simplicity, we consider the diffusive scaling limit (i.e., essentially equivalent to the system-size expansion [33], an established perturbative method invented by van Kampen [34] based on realistic scaling assumptions; see SM for a brief review) by introducing a small parameter ϵ :

$$g(\nu) = \frac{1}{\epsilon^2} \bar{g}(\nu), \quad \rho(y) = \frac{1}{\epsilon} \bar{\rho}\left(\frac{y}{\epsilon}\right) \quad (14)$$

with ϵ -independent functions \bar{g} and $\bar{\rho}$. We focus on the case with $\bar{g}(\nu) = \lambda_0 e^{\beta\nu}$. In this diffusive limit, corresponding to the mark size being typically much smaller in absolute value than the tension at any given time, the statistics of $N_{t_{\text{win}}}$ obeys Zipf's law for a sufficiently short time window t_{win} [see Fig. 2(c)]:

$$P_{\text{SS}}(N_{t_{\text{win}}}) \propto N_{t_{\text{win}}}^{-2} \quad \text{for } N_{t_{\text{win}}} < N_{\text{cut}} \quad (15)$$

as an intermediate asymptotics [35] with upper cut-off $N_{\text{cut}} = O(\epsilon^{-2})$.

This relation can be derived from a superposition of Poisson statistics. Let us consider a long time series in $[0, T)$ and then randomly select a time point $\tau \in [0, T)$. For a sufficiently short time window $[\tau, \tau + t_{\text{win}})$, we can assume that $\lambda(t)$ is constant and the number of events obeys the Poisson statistics $P(N_{t_{\text{win}}} | \lambda) = (\lambda t_{\text{win}})^{N_{t_{\text{win}}}} e^{-\lambda t_{\text{win}}} / N_{t_{\text{win}}}!$. Choosing τ randomly and neglecting dependences between count numbers across different windows, the unconditional distribution is given by the superposition of the Poisson distribution as

$$P_{\text{SS}}(N_{t_{\text{win}}}) \simeq \int P(N_{t_{\text{win}}} | \lambda) P_{\text{SS}}(\lambda) d\lambda \propto N_{t_{\text{win}}}^{-2}. \quad (16)$$

This example shows that Zipf's law (6) for the intensity PDF $P_{\text{SS}}(\lambda)$ is directly relevant to Zipf's laws for other observable quantities.

Theoretically, relation (15) is expected to hold only up to the cutoff N_{cut} (see SM [17]), which diverges as $\epsilon \rightarrow 0$, guaranteeing the robust universality of Zipf's law for $P_{\text{SS}}(N_{t_{\text{win}}})$ in the diffusive limit. Beyond the cutoff, we numerically observe a fatter tail stemming from dependences between count numbers in adjacent time windows, which becomes dominant at very high count numbers, as can be seen from its impact on ν given by (2).

Conclusion.—As power laws are widely observed in many complex systems, our theoretical finding suggests the nonlinear self-excited mechanism as an explanation for the universality of power laws. Intuitively, these properties emerge from the intricate interplay between a kind of multiplicative process, memory and endogeneity or reflexivity. Our new tools and results will be useful for data analysis of real complex systems. Interested readers are referred to Ref. [36] for more mathematical details.

This work was supported by (i) JST, PRESTO Grant No. JPMJPR20M2, Japan; (ii) the Japan Society for the Promotion of Science KAKENHI (Grant No. 20H05526); (iii) Intramural Research Promotion Program in the University of Tsukuba; and (iv) partially by the National Natural Science Foundation of China under Grant No. U2039202. The numerical computation for Fig. 2(c) was carried out at the Yukawa Institute Computer Facility.

*kiyoshi@sk.tsukuba.ac.jp

†dsornette@ethz.ch

- [1] D. M. W. Powers, NeMLaP3/CoNLL98: New methods in language processing and computational natural language learning, in *Applications and Explanations of Zipf's Law*, edited by D. M. W. Powers (ACL, 1998), pp 151–160, <https://aclanthology.org/W98-1218/>.
- [2] D. Sornette, *Critical Phenomena in Natural Sciences*, Springer Series in Synergetics (Springer, New York, 2004).
- [3] M. E. Newman, *Contemp. Phys.* **46**, 323 (2005).

- [4] A. Saichev, Y. Malevergne, and D. Sornette, in *Theory of Zipf's Law and Beyond*, Lecture Notes in Economics and Mathematical Systems Vol. 632 (Springer, New York, 2009).
- [5] Y. Malevergne, A. Saichev, and D. Sornette, *J. Econ. Dyn. Control* **37**, 1195 (2013).
- [6] B. M. Hill and M. Woodroffe, *J. Am. Stat. Assoc.* **70**, 212 (1975).
- [7] R. Ferrer i Cancho and R. V. Solé, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 788 (2003).
- [8] C. Furusawa and K. Kaneko, *Phys. Rev. Lett.* **90**, 088102 (2003).
- [9] A. G. Hawkes, *Biometrika* **58**, 83 (1971).
- [10] D. Sornette and I. Osorio, Prediction, in *Epilepsy: The Intersection of Neurosciences, Biology, Mathematics, Physics and Engineering*, edited by I. Osorio, H. P. Zaveri, M. G. Frei, and S. Arthurs (CRC Press, Taylor & Francis Group, London, 2010), pp. 203–237.
- [11] M.-A. Rizoïu, Y. Lee, S. Mishra, and L. Xie, A tutorial on Hawkes processes for events in social media, in *Frontiers of Multimedia Research* (ACM, 2017), pp. 191–218.
- [12] A. G. Hawkes, *Quant. Finance* **18**, 193 (2018).
- [13] A. Reinhart, *Stat. Sci.* **33**, 299 (2018).
- [14] P. Brémaud and L. Massoulié, *Ann. Prob.* **24**, 1563 (1996).
- [15] F. Gao and L. Zhu, *Stoch. Proc. Appl.* **128**, 4051 (2018).
- [16] P. Blanc, J. Donier, and J.-P. Bouchaud, *Quant. Finance* **17**, 171 (2017).
- [17] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevLett.127.188301> for the details of the theoretical and numerical methods, which includes Refs. [18–21].
- [18] H. Kesten, *Acta Math.* **131**, 207 (1973).
- [19] A. S. Cherny and H.-J. Engelbert, *Singular Stochastic Differential Equations* (Springer-Verlag, Berlin, 2005).
- [20] K. Kanazawa, T. G. Sano, T. Sagawa, and H. Hayakawa, *Phys. Rev. Lett.* **114**, 090601 (2015).
- [21] K. Kanazawa, *Statistical Mechanics for Athermal Fluctuation: Non-Gaussian Noise in Physics* (Springer, Berlin, 2017), <https://dx.doi.org/10.1007/978-981-10-6332-9>.
- [22] A. Saichev and D. Sornette, *Phys. Rev. E* **71**, 016608 (2005).
- [23] G. Ouillon and D. Sornette, *J. Geophys. Res.* **110**, B04306 (2005).
- [24] D. Sornette and G. Ouillon, *Phys. Rev. Lett.* **94**, 038501 (2005).
- [25] V. A. Filimonov and D. Sornette, *Europhys. Lett.* **94**, 46003 (2011).
- [26] Z.-Q. Jiang, W.-J. Xie, W.-X. Zhou, and D. Sornette, *Rep. Prog. Phys.* **82**, 125901 (2019).
- [27] K. Kanazawa and D. Sornette, *Phys. Rev. Lett.* **125**, 138301 (2020).
- [28] K. Kanazawa and D. Sornette, *Phys. Rev. Research* **2**, 033442 (2020).
- [29] R. Zwanzig, *Nonequilibrium Statistical Mechanics* (Oxford University Press, New York, 2001).
- [30] R. Kupferman, *J. Stat. Phys.* **114**, 291 (2004).
- [31] J.-P. Bouchaud, J. Bonart, J. Donier, and M. Gould, *Trades, Quotes and Prices* (Cambridge University Press, Cambridge, 2018), <https://doi.org/10.1017/9781316659335>.
- [32] A. Dassios and H. Zhao, *Adv. Appl. Probab.* **43**, 814 (2011).
- [33] C. W. Gardiner, *Stochastic Methods*, 4th ed. (Springer, Berlin, 2009).
- [34] N. G. Van Kampen, *Stochastic Processes in Physics and Chemistry* (Elsevier, New York, 1992).
- [35] G. I. Barenblatt, *Scaling, Self-Similarity, and Intermediate Asymptotics* (Cambridge University Press, Cambridge, England, 1996).
- [36] K. Kanazawa and D. Sornette, [arXiv:2110.01523](https://arxiv.org/abs/2110.01523).