Supplemental Material: Ubiquitous power law scaling law in nonlinear self-excited Hawkes processes

Kiyoshi Kanazawa*
Faculty of Engineering, Information and Systems, University of Tsukuba, Tennodai, Tsukuba, Ibaraki 305-8573, Japan
JST, PRESTO, 4-1-8 Honcho, Kawaguchi, Saitama 332-0012, Japan

Didier Sornette†
ETH Zurich, Department of Management, Technology, and Economics, Zurich 8092, Switzerland
Institute of Risk Analysis, Prediction, and Management (Risks-X), Academy for Advanced Interdisciplinary Studies, Southern University of Science and Technology (SUSTech), Shenzhen 518055, China
(Dated: October 5, 2021)

I. METHODS

A. Markov embedding (discrete sum of exponentials)

Let us first focus on the case of a superposition of exponentials:

$$h(t) = \sum_{k=1}^{K} \tilde{h}_k e^{-t/\tau_k}. \tag{S1}$$

For this case, Eq. (2) can be converted into the following Markovian dynamics,

$$\nu(t) = \sum_{k=1}^{K} z_k(t), \quad \frac{dz_k}{dt} = -\frac{z_k}{\tau_k} + \tilde{h}_k \xi^{\mathrm{P}}_{\rho(y);\lambda(t)} \tag{S2}$$

with the state-dependent Poisson noise $\xi^{\mathrm{P}}_{\rho(y);\lambda(t)}$, defined by

$$\xi^{\mathrm{P}}_{\rho(y);\lambda(t)} = \sum_{i=1}^{N(t)} y_i \delta(t - t_i), \tag{S3}$$

where t_i is the ith event time and y_i is a random number obeying a given distribution $\rho(y)$. Note that the probability that an event occurs within interval $[t, t+dt)$ is given by

$$\lambda(t)dt = g(\nu(t))dt. \tag{S4}$$

This technique [1, 2] is called Markov embedding, where low-dimensional non-Markovian dynamics is converted onto higher-dimensional Markovian dynamics.

We note that this Markov embedding framework is sufficiently general since any memory kernel $h(t)$ can be written as a continuous sum of exponentials (via the Laplace transformation), which can be approximated by the discrete-sum formula (S1), such that

$$h(t) = \int_0^\infty dx \tilde{h}(x) e^{-t/x} \approx \sum_{k=1}^{K} \tilde{h}_k e^{-t/\tau_k}. \tag{S5}$$

In this sense, the discrete representation $\tilde{h}_k$ corresponds to the continuous function representation $\tilde{h}(x)$ via this relationship. This method can be formally generalised for the general continuous sum of exponentials as shown in Sec. II D.

---

* kiyoshi@sk.tsukuba.ac.jp
† dsornette@ethz.ch

B. Numerical scheme

We have numerically studied Eq. (2) based on the Monte Carlo simulations of Eq. (S2) for Fig. 1. Let us introduce a discretised time series,

$$0 = s_0 < s_1 < \cdots < s_N = T, \;\; \Delta s_i = s_{i+1} - s_i. \tag{S6}$$

Equation (S2) reads

$$z_k(s_{i+1}) - z_k(s_i) = -\frac{z_k(s_i)}{\tau_k}\Delta s_i + \begin{cases} 0 & (\text{Probability} = 1 - \lambda(s_i)\Delta s_i) \\ \tilde{h}_k y_k & (\text{Probability} = \lambda(s_i)\Delta s_i) \end{cases} \tag{S7}$$

for $i = 0, \dots, N-1$. The mark sequence $\{y_k\}_k$ obeys the normal distribution

$$P(y_k) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y_k - m)^2/(2\sigma^2)} \tag{S8}$$

with mean m and variance σ^2. The time step Δs_k in Eq. (S7) must be sufficiently small, such that $\lambda(s_i)\Delta s_i \ll 1$. We therefore employ an adaptive scheme

$$\Delta s_i = \min\left\{\Delta t^{(1)}_{\max}, \frac{\Delta t^{(2)}_{\max}}{\lambda(s_i)}\right\} \tag{S9}$$

with $\Delta t^{(1)}_{\max}$ and $\Delta t^{(2)}_{\max}$. In addition, we introduce a finite cutoff for the tension-intensity map,

$$\lambda(t) = g(\nu(t)) = \min\left\{\lambda_0 e^{\beta\nu(t)}, \lambda_{\max}\right\} \tag{S10}$$

with $\lambda_{\max} = 10^7$, to control rounding error.

For the numerical trajectory generated by Eq. (S7), we obtain the empirical steady intensity distribution as

$$P_{\rm ss}(\lambda) = \langle \delta(\lambda - \lambda(t)) \rangle = \lim_{T\to\infty} \frac{1}{T} \int_0^T \delta(\lambda - g(\nu(s)))ds \approx \frac{1}{T} \sum_{i=0}^{N-1} \delta(\lambda - g(\nu(s_i)))\Delta s_i \tag{S11}$$

under the assumption of ergodicity. In addition, we have applied a parallel computing method,

$$P_{\rm ss}(\lambda) = \left\langle \lim_{T\to\infty} \frac{1}{T} \int_0^T \delta(\lambda - g(\nu(s)))ds \right\rangle \approx \frac{1}{N_{\rm PC}} \sum_{j=1}^{N_{\rm PC}} \frac{1}{T} \sum_{i=0}^{N-1} \delta(\lambda - g(\nu(s_i)))\Delta s_i \tag{S12}$$

with a total number $N_{\rm PC}$ of parallel threads.

1. *Numerical simulation for the negative mean mark case* $m < 0$

Since the PDF obeying Zipf's law is shown for the zero mean mark case in the main text as Fig. 1c, here we provide numerical simulations for the cases with negative mean marks $m \le 0$, where the exponent deviates from Zipf's scalings. Our theory predicts the relation

$$P_{\rm ss}(\lambda) \propto \lambda^{-2-\beta^{-1}a}, \;\; a = -\frac{2m}{\sigma^2 h(0)} \tag{S13}$$

for the Gaussian mark distribution $\rho(y) = e^{-(y-m)^2/(2\sigma^2)}/\sqrt{2\pi\sigma^2}$ (see Sec. VI D 3 for the derivation), which agrees with the numerical intensity PDFs as shown in Fig. S1.

2. *Parameters*

In our work, we set the parameters summarised in Table I for numerical simulations. See also Numerical Code S1.pdf and Numerical Code S3.pdf for the details.

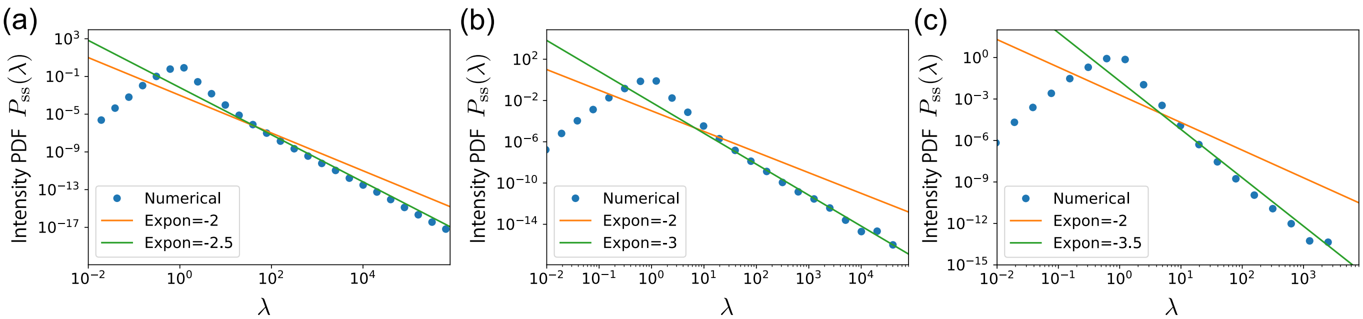


FIG. S1. Numerical simulations for the negative mean mark cases, where the deviations from Zipf's scaling are observed, such that $P_{\rm ss}(\lambda) \propto \lambda^{-2-a}$ with (a) $a = 0.5$, (b) $a = 1$, and (c) $a = 1.5$. The parameters are summarised in Table I.

parameter	Figure 1c	Figure S1a	Figure S1b	Figure S1c
$\Delta t^{(1)}_{\max}$	0.1			
$\Delta t^{(2)}_{\max}$	0.01			
K	3			
$\{\tau_i\}_{i=1,\dots,K}$	$(1.0, 0.5, 2.0)$			
$\{\tilde{h}_i\}_{i=1,\dots,K}$	$(0.5, 0.6, 0.1)$			
λ_0	1			
a	0 (Zipf)	0.5	1	1.5
β	5	1		
m	0	$-\dfrac{a\sigma^2 h(0)}{2}$		
σ	0.1	0.5		
T	5×10^4	5×10^6		
$N_{\rm PC}$	8			

TABLE I. Summary table of the parameters used in the numerical simulations.

II. MASTER EQUATION

Since Eq. (S2) is Markovian, we can derive the corresponding master equation (i.e., the time-evolution equation for the probability density function (PDF)). Let us consider the phase point $\boldsymbol{z} := (z_1, \dots, z_K)$ and its PDF $P_t(\boldsymbol{z})$. Indeed, the PDF satisfies the following master equation

$$\frac{\partial P_t(\boldsymbol{z})}{\partial t} = \sum_{k=1}^{K} \frac{\partial}{\partial z_k} \frac{z_k}{\tau_k} P_t(\boldsymbol{z}) + \int_{-\infty}^{\infty} dy \rho(y) \left\{ G(\boldsymbol{z} - y\tilde{\boldsymbol{h}}) P_t(\boldsymbol{z} - y\tilde{\boldsymbol{h}}) - G(\boldsymbol{z}) P_t(\boldsymbol{z}) \right\} \tag{S14}$$

where we introduce $G(\boldsymbol{z}) := g\left(\sum_{k=1}^{K} z_k\right)$ and $\tilde{\boldsymbol{h}} := \left(\tilde{h}_1, \dots, \tilde{h}_K\right)$. In the following, we focus on the case with non-positive mean mark:

$$m := \int_{-\infty}^{\infty} y\rho(y) dy \le 0. \tag{S15}$$

Remarkably, the solution explodes for $m > 0$ (see Sec. VI C for an intuitve discussion on the condition of the explosive solutions).

A. Derivation

Equation (S14) can be derived as follows. Let us introduce an arbitrary function $f(\boldsymbol{z})$. The time-evolution of $f(\boldsymbol{z})$ is given by

$$df(\boldsymbol{z}) = \begin{cases} -\sum_{k=1}^{K} \frac{z_k}{\tau_k} \frac{\partial f(\boldsymbol{z})}{\partial z_k} dt, & \text{(No jump during } [t, t+dt) : \text{prob.} = 1 - \lambda(t)dt) \\ f(\boldsymbol{z} + y\tilde{\boldsymbol{h}}) - f(\boldsymbol{z}), & \text{(Jump in } [t, t+dt) : \text{prob.} = \lambda(t)\rho(y)dtdy) \end{cases} \tag{S16}$$

with jump size y obeying a given PDF $\rho(y)$. We take an ensemble average to obtain

$$\left\langle \frac{df}{dt} \right\rangle = \left\langle -\sum_{k=1}^{K} \frac{z_k}{\tau_k} \frac{\partial f(\boldsymbol{z})}{\partial z_k} + \int_{-\infty}^{\infty} dy \rho(y) G(\boldsymbol{z}) \left[f(\boldsymbol{z} + y\tilde{\boldsymbol{h}}) - f(\boldsymbol{z}) \right] \right\rangle. \tag{S17}$$

Here we integrate by part to obtain

$$-\int_{-\infty}^{\infty} d\boldsymbol{z} P_t(\boldsymbol{z}) \frac{z_k}{\tau_k} \frac{\partial f(\boldsymbol{z})}{\partial z_k} = \int_{-\infty}^{\infty} d\boldsymbol{z} f(\boldsymbol{z}) \frac{\partial}{\partial z_k} \frac{z_k}{\tau_k} P_t(\boldsymbol{z}). \tag{S18}$$

We also apply a variable transformation $\boldsymbol{z} + y\tilde{\boldsymbol{h}} \to \boldsymbol{z}$ to obtain

$$\int_{-\infty}^{\infty} d\boldsymbol{z} P_t(\boldsymbol{z}) G(\boldsymbol{z}) f(\boldsymbol{z} + y\tilde{\boldsymbol{h}}) = \int_{-\infty}^{\infty} d\boldsymbol{z} f(\boldsymbol{z}) G(\boldsymbol{z} - y\tilde{\boldsymbol{h}}) P_t(\boldsymbol{z} - y\tilde{\boldsymbol{h}}). \tag{S19}$$

This yields the identity

$$\int_{-\infty}^{\infty} d\boldsymbol{z} f(\boldsymbol{z}) \left\{ \frac{\partial P_t(\boldsymbol{z})}{\partial t} - \sum_{k=1}^{K} \frac{\partial}{\partial z_k} \frac{z_k}{\tau_k} P_t(\boldsymbol{z}) - \int_{-\infty}^{\infty} dy \rho(y) \left[P_t(\boldsymbol{z} - y\tilde{\boldsymbol{h}}) G(\boldsymbol{z} - y\tilde{\boldsymbol{h}}) - P_t(\boldsymbol{z}) G(\boldsymbol{z}) \right] \right\} = 0. \tag{S20}$$

Since this identity holds for an arbitrary function $f(\boldsymbol{z})$, we obtain Eq. (S14).

B. Solution for zero-mean mark distributions

The asymptotic solution of the master equation (S14) can be obtained as follows. Let us define the steady PDF $P_{\rm ss}(\boldsymbol{z}) := \lim_{t\to\infty} P_t(\boldsymbol{z})$ and $\phi(\boldsymbol{z}) := G(\boldsymbol{z}) P_{\rm ss}(\boldsymbol{z})$ to rewrite Eq. (S14) as

$$\sum_{k=1}^{K} \frac{1}{\tau_k} \frac{\partial}{\partial z_k} \left(\frac{z_k}{G(\boldsymbol{z})} \phi(\boldsymbol{z}) \right) + \int_{-\infty}^{\infty} dy \rho(y) \phi\left(\boldsymbol{z} - y\tilde{\boldsymbol{h}} \right) - \phi(\boldsymbol{z}) = 0 \tag{S21}$$

for $t \to \infty$. For large $\boldsymbol{z}$, the first term is negligibly small for the fast-accelerating intensity $G(\boldsymbol{z}) \gg \left(\sum_{k=1}^{K} z_k \right)^2$. We thus obtain

$$\int_{-\infty}^{\infty} dy \rho(y) \phi(\boldsymbol{z} - y\tilde{\boldsymbol{h}}) - \phi(\boldsymbol{z}) \approx 0 \quad \text{for large } \boldsymbol{z}. \tag{S22}$$

Let us apply a variable transformation from $\boldsymbol{z} = (z_1, \dots,\ z_K)$ to $\boldsymbol{Z} := (W, Z_2, \dots Z_K)$ with

$$z_1 = \tilde{h}_1 W, \quad z_2 = \tilde{h}_2 W + Z_2, \quad z_3 = \tilde{h}_3 W + Z_3, \quad \dots \quad , z_K = \tilde{h}_K W + Z_K, \tag{S23}$$

which leads to

$$\phi(\boldsymbol{z} - y\tilde{\boldsymbol{h}}) = \phi\left(\tilde{h}_1(W-y), \tilde{h}_2(W-y) + Z_2, \dots, \tilde{h}_K(W-y) + Z_k \right). \tag{S24}$$

By defining

$$\psi(W-y; Z_2, \dots, Z_K) = \psi(W-y; \boldsymbol{Z}') := \phi\left(\tilde{h}_1(W-y), \tilde{h}_2(W-y) + Z_2, \dots, \tilde{h}_K(W-y) + Z_k \right) \tag{S25}$$

with $\boldsymbol{Z}' := (Z_2, \dots, Z_K)$, we can rewrite Eq. (S22) as

$$\int_{-\infty}^{\infty} dy \rho(y) \psi\left(W - y; \boldsymbol{Z}' \right) - \psi\left(W; \boldsymbol{Z}' \right) \approx 0. \tag{S26}$$

This form of the integral equation is useful because the dependence on y disappears for $\boldsymbol{Z}'$, and can be regarded as an effectively one-dimensional integral equation.

With the condition that the mark distribution has zero mean $m := \int_{-\infty}^{\infty} y\rho(y)dy = 0$ with fast-decaying tail, the solution of this integral equation is given by

$$\psi\left(W;\boldsymbol{Z}'\right) = C_0\left(\boldsymbol{Z}'\right) + WC_1\left(\boldsymbol{Z}'\right) \tag{S27}$$

with arbitrary functions $C_0(\boldsymbol{Z}')$ and $C_1(\boldsymbol{Z}')$ that do not have W as an argument (see Sec. VI D for the derivation). As confirmed soon later, the natural boundary condition requires $C_1(\boldsymbol{Z}') = 0$ and thus the general solution is finally given by

$$\psi\left(W;\boldsymbol{Z}'\right) = C_0\left(\boldsymbol{Z}'\right). \tag{S28}$$

The tension distribution in the steady state $P_{\rm ss}(\nu) := \lim_{t\to\infty}\langle\delta(\nu-\nu(t))\rangle$ is given by marginalisation of the full distribution as

$$P_{\rm ss}(\nu) := \int_{-\infty}^{\infty} d\boldsymbol{z} P_{\rm ss}(\boldsymbol{z})\delta\left(\nu - \sum_{k=1}^{K} z_k\right) \propto \frac{1}{g(\nu)} \quad \text{for large } \nu, \tag{S29}$$

assuming that $\int_{-\infty}^{\infty} C_0(z_2',\dots,z_K')\Pi_{j=2}^{K}dz_j'$ is finite (see also Sec. VI E for the detailed calculation below). This implies Eq. (S72) in the main text. This implies that the steady intensity PDF is given by

$$P_{\rm ss}(\lambda) \propto \lambda^{-1}\left\{\frac{dg(\nu)}{d\nu}\right\}^{-1}\Bigg|_{\nu=g^{-1}(\lambda)} \tag{S30}$$

where we have used the Jacobian relationship $P_{\rm ss}(\lambda) = |d\nu/d\lambda|P_{\rm ss}(\nu)$, representing the conservation of probability under a change of variable.

1. *Consistency with the natural boundary condition.*

Technically, the steady state solution of the master equation should satisfiy the natural boundary condition, requiring a vanishing probability current at $\boldsymbol{z}\to\boldsymbol{\infty}$. Here we impose this condition on the steady state solution (S27). For large $\boldsymbol{z}$, the master equation is asymptotically given by

$$\frac{\partial P_t(\boldsymbol{z})}{\partial t} \simeq \int_{-\infty}^{\infty} dy\rho(y)\left\{G(\boldsymbol{z}-y\tilde{\boldsymbol{h}})P_t(\boldsymbol{z}-y\tilde{\boldsymbol{h}}) - G(\boldsymbol{z})P_t(\boldsymbol{z})\right\}, \tag{S31}$$

which is equivalent to

$$\frac{\partial P_t(W;\boldsymbol{Z}')}{\partial t} \simeq \int_{-\infty}^{\infty} dy\rho(y)\left\{G(W-y;\boldsymbol{Z}')P_t(W-y;\boldsymbol{Z}') - G(W;\boldsymbol{Z}')P_t(W;\boldsymbol{Z}')\right\}, \tag{S32}$$

after the variable transformation $\boldsymbol{z}\to\boldsymbol{Z} := (W;\boldsymbol{Z}')$ defined by Eq. (S23). The probability current is defined by the Kramers-Moyal expansion:

$$\frac{\partial P_t(W;\boldsymbol{Z}')}{\partial t} \simeq -\frac{\partial}{\partial W}J_t(W;\boldsymbol{Z}'), \quad J_t(W;\boldsymbol{Z}') := \sum_{n=1}^{\infty}\frac{(-1)^{n-1}\alpha_n}{n!}\frac{\partial^{n-1}}{\partial W^{n-1}}G(W;\boldsymbol{Z}')P_t(W;\boldsymbol{Z}'), \quad \alpha_n := \int_{-\infty}^{\infty} dy y^n\rho(y). \tag{S33}$$

The natural boundary condition requires the vanishing probability current at $W = +\infty$ as

$$\lim_{W\to+\infty} J_{\rm ss}(W;\boldsymbol{Z}') = 0. \tag{S34}$$

Since $\alpha_1 = m = 0$ for the zero-mean mark $m = 0$, in the steady state, the natural boundary condition is given by

$$J_{\rm ss}(W;\boldsymbol{Z}') = \sum_{n=2}^{\infty}\frac{(-1)^{n-1}\alpha_n}{n!}\frac{\partial^{n-1}}{\partial W^{n-1}}\left(C_0(\boldsymbol{Z}') + WC_1(\boldsymbol{Z}')\right) = -\frac{\alpha_2}{2}C_1(\boldsymbol{Z}') = 0, \tag{S35}$$

which requires that $C_1(\boldsymbol{Z}') = 0$.

2. Polynomial intensity case: $g(\nu) \propto \nu^n$ for some $n > 2$

We next study the power law forms of the intensity, by assuming various tension-intensity maps. Let us first consider the polynomial case of $g(\nu) = \lambda_0 + \lambda_1 \nu^n$ for some $n > 2$. This means that

$$\frac{dg}{d\nu} = n\lambda_1 \nu^{n-1}, \quad g^{-1}(\lambda) = \left(\frac{\lambda - \lambda_0}{\lambda_1}\right)^{1/n} \propto \lambda^{1/n} \quad \text{for large } \lambda. \tag{S36}$$

This means that the asymptotic form is given by the quasi-Zipf law:

$$P_{\rm ss}(\lambda) \propto \lambda^{-2+1/n} = \lambda^{-1-a}, \quad a := 1 - \frac{1}{n}. \tag{S37}$$

3. Superpolynomial intensity case $g(\nu) > O(\nu^n)$ with any $n > 2$

To develop some intuition, let us consider two typical cases. One typical case is given by $g(\nu) = \lambda_0 e^{\beta\nu}$, implying that

$$\frac{1}{g(\nu)}\frac{dg}{d\nu} = \beta, \quad g^{-1}(\lambda) = \frac{1}{\beta}\log\frac{\lambda}{\lambda_0} \quad \Longrightarrow \quad P_{\rm ss}(\lambda) \propto \lambda^{-2}. \tag{S38}$$

Another typical case is given by the super-exponential case $g(\nu) = \lambda_0 e^{\beta\nu^n}$ with $n > 1$. This case implies

$$\frac{1}{g(\nu)}\frac{dg}{d\nu} = n\beta\nu^{n-1}, \quad g^{-1}(\lambda) = \left(\frac{1}{\beta}\log\frac{\lambda}{\lambda_0}\right)^{1/n}, \quad \Longrightarrow \quad P_{\rm ss}(\lambda) \propto \lambda^{-2}\left(\log\lambda\right)^{-1+1/n} \quad \text{for large } \lambda. \tag{S39}$$

This means that Zipf's law holds up to the minor logarithmic factor.

Let us generalise these Zipf's law for general superpolynomial cases as follows. Considering the relation

$$\frac{1}{g(\nu)}\frac{dg}{d\nu} = \frac{d}{d\nu}\log g(\nu), \tag{S40}$$

we obtain

$$P_{\rm ss}(\lambda) \propto \lambda^{-1}\left\{\frac{dg}{d\nu}\right\}^{-1} = \lambda^{-1}\frac{1}{g(\nu)}\left\{\frac{d}{d\nu}\log g(\nu)\right\}^{-1} = \lambda^{-2}\left\{\frac{d}{d\nu}\log g(\nu)\right\}^{-1} = \lambda^{-2}\left\{\frac{d}{d\nu}\log\lambda(\nu)\right\}^{-1}. \tag{S41}$$

For most of physically motivated functions $\lambda = g(\nu)$, the logarithmic contribution from $\{\log\lambda\}^{-1}$ is subleading compared with Zipf's part λ^{-2} except for the polynomial intensity. Indeed, by assuming the asymptotic balance between the logarithmic and the power law parts as $(d/d\nu)\log\lambda(\nu) \simeq C\lambda^a$ with some real numbers $a \neq 0$ and $C < \infty$, we deduce the polynomial intensity $\lambda(\nu) = g(\nu) \propto \nu^{-1/a}$ as the corresponding exception. Thus, we find that the logarithmic factor $\{(d/d\nu)\log\lambda(\nu)\}^{-1}$ is a minor correction term for the superpolynomial cases.

C. Solution for negative-mean mark distributions

The above calculation can be generalised by assuming that the mean mark is negative and that the probability of the positive marks is nonzero:

$$m := \int_{-\infty}^{\infty} y\rho(y)dy < 0, \quad m_+ := \int_0^{\infty} y\rho(y)dy > 0. \tag{S42}$$

For large z, the steady master equation can be rewritten as

$$\int_{-\infty}^{\infty} dy\rho(y)\psi\left(W - y; \boldsymbol{Z}'\right) - \psi\left(W; \boldsymbol{Z}'\right) \approx 0 \tag{S43}$$

with $\psi(W;\boldsymbol{Z}') := G(\boldsymbol{z})P_{\rm ss}(\boldsymbol{z})$ after the variable transformaton $\boldsymbol{z} \to \boldsymbol{Z} := (W;\boldsymbol{Z}')$ defined by Eq. (S23). Under the condition (S42), the general solution is given by

$$\psi\left(W;\boldsymbol{Z}'\right) = C_0(\boldsymbol{Z}')e^{-c^*W} + C_1(\boldsymbol{Z}'), \tag{S44}$$

where $C_0(\boldsymbol{Z}')$ and $C_1(\boldsymbol{Z}')$ are arbitrary functions without W as an argument (see Sec. VI D for the derivation) and $c^* > 0$ is the unique positive root of $\Phi(c^*) = 0$ for the moment-generating function defined by

$$\Phi(x) := \int_{-\infty}^{\infty} \rho(y)(e^{xy} - 1)dy. \tag{S45}$$

We can prove that $\Phi(c) = 0$ has only two roots at $c = 0$ and $c = c^* > 0$ as shown in Sec. VI D. The outline of the proof is as follows: since the second order derivative of $\Phi(x)$ is always positive (see Eq. (S112) below), the first order derivative is an increasing function of x. If the mean of y is zero, at $x = 0$, the first order derivative is equal to zero, and thus positive for $x > 0$. This proves that $c^* = 0$ for $m = 0$. If the mean of y is negative, the first order derivative is negative at $x = 0$ but it increases and passes positive for larger x. Thus $\Phi(x)$ first decreases below 0 and then crosses it again at some c^*, which is the solution.

Finally, the natural boundary condition requires that $C_1(\boldsymbol{Z}')$ must be zero as shown later soon: $C_1(\boldsymbol{Z}') = 0$. We then obtain the general solution

$$\psi\left(W; \boldsymbol{Z}'\right) = C_0(\boldsymbol{Z}')e^{-c^*W}. \tag{S46}$$

The tension distribution in the steady state $P_{\rm ss}(\nu) := \lim_{t\to\infty}\langle\delta(\nu - \nu(t))\rangle$ is given by marginalisation of the full distribution as

$$P_{\rm ss}(\nu) := \int_{-\infty}^{\infty} d\boldsymbol{z} P_{\rm ss}(\boldsymbol{z})\delta\left(\nu - \sum_{k=1}^{K} z_k\right) \propto \frac{1}{g(\nu)} \exp\left(-\frac{c^*}{\tilde{h}_{\rm tot}}\nu\right) \quad \text{for large } \nu \tag{S47}$$

with

$$\tilde{h}_{\rm tot} := \sum_{k=1}^{K} \tilde{h}_k = h(t = 0), \tag{S48}$$

which is derived in Sec. VI F. This implies that the steady intensity PDF is given by

$$P_{\rm ss}(\lambda) \propto \lambda^{-1} \left[\exp\left(-\frac{c^*}{h(0)}\nu\right) \left\{\frac{dg(\nu)}{d\nu}\right\}^{-1}\right]_{\nu = g^{-1}(\lambda)} \tag{S49}$$

where we have used the Jacobian relationship $P_{\rm ss}(\lambda) = |d\nu/d\lambda| P_{\rm ss}(\nu)$, representating the probability conservation.

1. *Consistency with the natural boundary condition.*

Let us confirm the consistency of the general solution (S44) for the natural boundary condition. By substituting the general solution (S44) in the probability current $J_{\rm ss}(W; \boldsymbol{Z}')$, we obtain

$$J_{\rm ss}(W; \boldsymbol{Z}') = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}\alpha_n}{n!} \frac{\partial^{n-1}}{\partial W^{n-1}} \left(C_1(\boldsymbol{Z}') + e^{-c^*W} C_0(\boldsymbol{Z}')\right) = mC_1(\boldsymbol{Z}') + \frac{C_0(\boldsymbol{Z}')}{c^*} \Phi(c^*) e^{-c^*W}, \tag{S50}$$

where we have used the expansion of the generating function $\Phi(x) = \sum_{n=1}^{\infty} (\alpha_n/n!)x^n$. Since $\Phi(c^*) = 0$, we obtain $J_{\rm ss}(W; \boldsymbol{Z}') = mC_1(\boldsymbol{Z}')$. Because the natural boundary condition requires $\lim_{W\to\infty} J_{\rm ss}(W; \boldsymbol{Z}') = 0$ for any $\boldsymbol{Z}'$, the function $C_1(\boldsymbol{Z}')$ must be zero.

2. *Exponential intensity case: $g(\nu) \simeq \lambda_0 e^{\beta\nu}$*

For simplicity, let us consider the exponential intensity

$$\lambda = g(\nu) = \lambda_0 e^{\beta\nu}, \quad \beta > 0. \tag{S51}$$

Using formula (S49), the steady state PDF is given by the power law intensity distribution

$$P_{\rm ss}(\lambda) \propto \lambda^{-2-\beta^{-1}a}, \quad a := \frac{c^*}{h(0)}, \tag{S52}$$

suggesting a thinner tail for the negative-mean mark distributions $m < 0$.

3. Super-exponential intensity case: $g(\nu) \simeq \lambda_0 e^{\beta\nu^n}$ with $n > 1$

We next consider the super-exponential intensity

$$\lambda = g(\nu) = \lambda_0 e^{\beta\nu^n}, \quad \beta > 0, \quad n > 1. \tag{S53}$$

Using formula (S49), the steady PDF is given by

$$\begin{aligned} P_{\rm ss}(\lambda) &\propto \lambda^{-2}\left(\log\frac{\lambda}{\lambda_0}\right)^{-1+1/n}\exp\left(-\frac{c^*}{h(0)}\left(\frac{1}{\beta}\log\frac{\lambda}{\lambda}\right)^{1/n}\right) \\ &= \exp\left[-2\log\lambda - \frac{c^*}{h(0)}\left(\frac{1}{\beta}\log\frac{\lambda}{\lambda_0}\right)^{1/n} + \left(-1+\frac{1}{n}\right)\log\log\frac{\lambda}{\lambda_0}\right] \quad \text{for large } \lambda. \end{aligned} \tag{S54}$$

Here we can drop the sub-dominant double-logarithmic correction, since

$$\lim_{\lambda\to\infty}\frac{1}{\log\lambda}\log\log\frac{\lambda}{\lambda_0} = 0. \tag{S55}$$

We thus obtain the asymptotic formula for the super-exponential cases to leading order:

$$P_{\rm ss}(\lambda) \propto \lambda^{-2-\beta^{-1}a\left(\beta^{-1}\log\frac{\lambda}{\lambda_0}\right)^{-1+1/n}} \quad \text{for large } \lambda, \quad a := \frac{c^*}{h(0)}, \tag{S56}$$

which obeys quasi-Zipf's scaling with the correction in the exponent due to the asymmetry of the mark distribution.

One can notice that the correction term $(\log\lambda/\lambda_0)^{1/n}$ in Eq. (S54) can be also regarded as subleading in the sense that $\lim_{\lambda\to\infty}\frac{1}{\log\lambda}\left(\beta^{-1}\log\frac{\lambda}{\lambda_0}\right)^{1/n} = 0$, deducing the Zipf's law $P_{\rm ss}(\lambda) \propto \lambda^{-2}$ for large λ. However, one should be cautious in dropping this term for practical analyses, since the convergence speed is slow.

4. Polynomial intensity case: $g(\nu) \simeq \lambda_0\nu^n$ with $n > 2$

We also consider the polynomial intensity

$$\lambda = g(\nu) = \lambda_0\nu^n, \quad n > 2. \tag{S57}$$

Using formula (S49), the steady state PDF is given by the power law intensity distribution with the streched-exponential truncation:

$$P_{\rm ss}(\lambda) \propto \lambda^{-2+1/n}\exp\left[-a\left(\frac{\lambda}{\lambda_0}\right)^{1/n}\right] \quad \text{for large } \lambda, \quad a := \frac{c^*}{h(0)}. \tag{S58}$$

D. Markov embedding and field master equation for continuous sum of exponentials

We have formulated the Markov embedding method for the nonliear Hawkes process with the discrete sum of exponentials (S1) and have derived the corresponding master equation (S14). Here we formally generalise this methodology for the most general case of continuous sum of exponentials:

$$h(t) = \int_0^\infty dx\tilde{h}(x)e^{-t/x}. \tag{S59}$$

On the basis of this decomposition, the original nonlinear Hawkes process is converted into a Markovian stochastic partial differential equation (SPDE).

$$\nu(t) = \int_0^\infty dx z(t,x), \quad \frac{\partial z(t,x)}{\partial t} = -\frac{z(t,x)}{x} + \tilde{h}(x)\xi^{\rm P}_{\rho(y);\lambda(t)}. \tag{S60}$$

This conversion implies that the original one-dimensional non-Markovian process $\nu(t)$ is equivalent to an infinite-dimensional Markovian process described by $\{z(t,x)\}_{x\in(0,\infty)}$. Here $x\in(0,\infty)$ is the label of the auxiliary variables $\{z(t,x)\}_{x\in(0,\infty)}$ distributed on the auxiliary field $(0,\infty)$.

The master equation corresponding to the SPDE (S60) can be formally written with the formalism of functional calculus. Indeed, by introducing the probability density functional $P[z]:=P_t[\{z(x)\}_x]$ and the intensity functional $G[z]:=g\left(\int_0^\infty dxz(t,x)\right)$, the field master equation [3, 4] is given by

$$\frac{\partial P_t[z]}{\partial t}=\int_0^\infty dx\frac{\delta}{\delta z(x)}\frac{z(x)}{x}P_t[z]+\int_{-\infty}^\infty dy\rho(y)\left\{G[z-y\tilde{h}]P_t[z-y\tilde{h}]-G[z]P_t[z]\right\}, \tag{S61}$$

which should be interpreted as a formal limit from the discrete representation (S14) according to the standard convention [5], and thus has the same asymptotic solution (S29).

The functional description for the field master equation (S61) is formally introduced as follows. Let us discuss the nonlinear Hawkes process (S2) for the discrete sum of exponentials (S1), whose master equation is given by Eq. (S14). Here we introduce a lattice for x with interval dx, such that

$$\tau_k=x_k=kdx,\quad \tilde{h}(x_k)dx=\tilde{h}_k,\quad z_k(t)=z(t,x_k)dx \tag{S62}$$

for the non-negative integer $k=1,\dots,K$. Equation (S2) is then rewritten as

$$\nu(t)=\sum_{k=1}^K z(t,x_k),\quad \frac{\partial z(t,x_k)}{\partial t}=-\frac{z(t,x_k)}{x_k}+\tilde{h}_k\xi^{\mathrm{P}}_{\rho(y);\lambda(t)},\quad h(t)=\int_0^\infty dx\tilde{h}(x)e^{-t/x}\approx\sum_k\tilde{h}_ke^{-t/x}. \tag{S63}$$

The corresponding master equation is given by

$$\frac{\partial P_t(\boldsymbol{z})}{\partial t}=\sum_{k=1}^K dx\frac{1}{dx}\frac{\partial}{\partial z(x_k)}\frac{z(x_k)}{x_k}P_t(\boldsymbol{z})+\int_{-\infty}^\infty dy\rho(y)\left\{G(\boldsymbol{z}-y\tilde{\boldsymbol{h}})P_t(\boldsymbol{z}-y\tilde{\boldsymbol{h}})-G(\boldsymbol{z})P_t(\boldsymbol{z})\right\}, \tag{S64}$$

by introducing a vector $\tilde{\boldsymbol{h}}:=(\tilde{h}_1,\dots,\tilde{h}_K)$. Let us take the formal limit $K\to\infty$ and $dx\downarrow 0$ to deduce the field master equation (S61) by replacement

$$\frac{\delta}{\delta z(x)}[...]:=\lim_{dx\downarrow 0}\lim_{K\to\infty}\frac{1}{dx}\frac{\partial}{\partial z(x_k)}[...],\quad \int_0^\infty dx[...]=\lim_{dx\downarrow 0}\lim_{K\to\infty}\sum_{k=1}^K dx[...], \tag{S65}$$

which follows the convention [5]. We note that the rigorous foundation for the functional description has not been established yet [5], and constitutes a problem out of scope of our paper.

III. ILLUSTRATIVE CASE

As an appendix, let us focus on the illustrative case where the memory function $h(t)=\tilde{h}e^{-t/\tau}$ is a single exponential and the distribution of marks is symmetric: $\rho(y)=\rho(-y)$, in order to provide an intuitive understanding of the underlying generating mechanism of the Zipf and quasi-Zipf laws. In this case, the original model can be converted into a simple process obeying the stochastic differential equation (SDE)

$$\frac{d\nu}{dt}=-\frac{\nu}{\tau}+\tilde{h}\xi^{\mathrm{P}}_{\rho(y);\lambda} \tag{S66}$$

in terms of the compound Poisson process $\xi^{\mathrm{P}}_{\rho(y);\lambda}$ with jump-size distribution $\rho(y)$ and corresponding intensity $\lambda=g(\nu)$. We apply the diffusive approximation: $\xi^{\mathrm{P}}_{\rho(y);\lambda(t)}\approx\sqrt{2Dg(\nu)}\xi^{\mathrm{G}}$, with the standard white Gaussian noise ξ^{G} and $D:=(\tilde{h}^2/2)\int_{-\infty}^\infty y^2\rho(y)dy$, which requires that the second-order moment of the PDF $\rho(y)$ exists. This can be shown by the Kramers-Moyal (KM) expantion and truncating its series up to the second order. Indeed, the KM expantion of the master equation is given by

$$\begin{aligned}\frac{\partial P_t(\nu)}{\partial t}&=\frac{1}{\tau}\frac{\partial}{\partial\nu}\nu P_t(\nu)+\int_{-\infty}^\infty dy\rho(y)[g(\nu-\tilde{h}y)P_t(\nu-\tilde{h}y)-g(\nu)P_t(\nu)]\\&=\frac{1}{\tau}\frac{\partial}{\partial\nu}\nu P_t(\nu)+\sum_{k=1}^\infty\frac{\tilde{h}^{2k}\alpha_{2k}}{(2k)!}\frac{\partial^{2k}}{\partial\nu^{2k}}[g(\nu)P_t(\nu)]\end{aligned} \tag{S67a}$$

with the $k$the-order KM coefficient defined by

$$\alpha_k := \int_{-\infty}^{\infty} y^k \rho(y) dy. \tag{S67b}$$

By truncating the KM expansion up to the second-order, we obtain

$$\frac{\partial P_t(\nu)}{\partial t} \approx \frac{1}{\tau}\frac{\partial}{\partial \nu}\nu P_t(\nu) + D\frac{\partial^2}{\partial \nu^2} g(\nu) P_t(\nu), \tag{S67c}$$

which is equivalent to Eq. (S66). Since $\sqrt{2Dg(\nu)} \gg \nu$ for fast-accelerating intensities, we obtain a ν-dependent diffusion process

$$\frac{d\nu}{dt} \approx -\frac{\nu}{\tau} + \sqrt{2Dg(\nu)}\xi^{\mathrm{G}} \quad (\text{for large } \nu) \ . \tag{S68}$$

We note that the truncation of the KM expansion can be proved by making assumption of the diffusive scaling using the system-size expansion (see Sec. IV for the mathematical detail).

A. Assuming the fast-acceralating intensities

Let us consider the case of fast-accelerating intensities $g(\nu) > O(\nu^2)$. For this case, the linear part $-\nu/\tau$ becomes smaller than the diffusive term $\sqrt{2Dg(\nu)}\xi^{\mathrm{G}}$ at large ν:

$$\sqrt{2Dg(\nu)}\xi^{\mathrm{G}} > O(\nu) = O(-\nu/\tau). \tag{S69}$$

This means that the diffusive model (S68) can be regarded as the inhomogeneous diffusive process for large ν without relaxation term:

$$\frac{d\nu}{dt} \approx \sqrt{2Dg(\nu)}\xi^{\mathrm{G}} \quad (\text{for large } \nu) \ , \tag{S70}$$

which corresponds to

$$\frac{\partial P_t(\nu)}{\partial t} \approx D\frac{\partial^2}{\partial \nu^2} g(\nu) P_t(\nu) \quad (\text{for large } \nu) \ . \tag{S71}$$

This means that the corresponding steady state PDF is asymptotically given by

$$P_{\mathrm{ss}}(\nu) \approx \frac{1}{g(\nu)} = o(\nu^{-2}) \quad (\text{for large } \nu), \tag{S72}$$

which is consistent with the normalisation condition $\int_{-\infty}^{\infty} d\nu P_{\mathrm{ss}}(\nu) = 1$ under the assumption of fast-accelerating intensities $g(\nu) > O(\nu^2)$. Since this asymptotic solution is consistent with the normalisation condition, the solution (S72) is the correct asymptotic form for $P_{\mathrm{ss}}(\nu)$. Since $\lambda = g(\nu)$, the identity $P_{\mathrm{ss}}(\nu)d\nu = P_{\mathrm{ss}}(\lambda)d\lambda$ expressing the conservation of probability under a change of variable leads to the following expression for the PDF of λ:

$$P_{\mathrm{ss}}(\lambda) \propto \frac{1}{\lambda}\left\{\frac{dg}{d\nu}\left(g^{-1}(\lambda)\right)\right\}^{-1}. \tag{S73}$$

This recovers the universal and quasi-Zipf's laws (6) for both superpolynomial and polynomial fast-accelerating intensity as shown in Sec. II B.

1. *Intuitive understanding*

An intuitive understanding that the SDE (S68) leads to a stationary solution and thus a bona fide PDF for ν is obtained by applying Ito's lemma on the change of variable $\nu \to \chi := e^{-(\beta/2)\nu}$ for the case (3) with $D = 1/2$ and $\lambda_0 = 1$, leading to

$$d\chi \approx \mu_\chi dt + \frac{\beta}{2} dB \ , \ \text{ with } \mu_\chi := \frac{\beta^2}{8\chi} - \frac{\chi \ln \chi}{\tau}, \tag{S74}$$

where we use the mathematical notation $dB := \xi^{\mathrm{G}} dt$ to represent the increment of the standard Brownian (or Wiener) process. Expression (S74) describes the motion of a Brownian particle in the potential $V(\chi) = -\int^{\chi} \mu_{\chi'} d\chi' = (1/2\tau)\chi^2 \ln(\chi e^{-2}) - (\beta^2/8)\ln\chi$, from which the drift force μ_χ derives. The behaviour of λ at large values is controlled by the dynamics of ν at large positive values, which corresponds to χ close to 0. As χ approaches 0, μ_χ diverges on the positive side and repeals χ from the origin. Thus ν and λ never diverges. When χ grows, μ_χ becomes negative and also diverge in amplitude, pushing it back to smaller values, thus preventing ν to become too negative and therefore stopping λ from being too small.

B. The case of the quadratic Hawkes process

As a marginal case, let us consider the case with a quadratic intensity

$$g(\nu) = \lambda_0 + \lambda_1 \nu^2 = O(\nu^2), \tag{S75}$$

which does not belong to the fast-accelerating intensity and is out of scope of our main manuscript. A part of this case is studied in Ref. [6], and this non-linear Hawkes process is called the quadratic Hawkes (QHawkes) processes exhibiting quite different behaviour (e.g., interested readers should see Ref. [6], where a special case of the QHawkes process is investigated by assuming the diffusive limit and the exponential memory kernel $h(t) = \tilde{h}e^{-t/\tau}$. Before our work, this was the only study where an explicit asymptotic solution of the nonlinear Hawkes processes was given for a specific setup).

The QHawkes process belongs to another class of nonlinear Hawkes processes, because the relaxation term $-\nu/\tau$ and the diffusive term $\sqrt{2Dg(\nu)}\xi^{\mathrm{G}} \propto \sqrt{2\lambda_1 D}\nu\xi^{\mathrm{G}}$ are of the same order for large ν:

$$O(-\nu/\tau) = O\left(\sqrt{2Dg(\nu)}\xi^{\mathrm{G}}\right) = O(\nu). \tag{S76}$$

Indeed, the QHawkes is essentially similar to the Kesten process [7] because the diffusive model (S68) can be asymptotically regarded as a continuous version of the Kesten process as

$$\nu(t+dt) \approx \left\{1 - \frac{dt}{\tau} + \sqrt{2\lambda_1 D} dB(t)\right\}\nu(t) \quad \text{(for large } \nu\text{)} \tag{S77}$$

with the increment of the standard Brownian process $dB := \xi^{\mathrm{G}} dt$. Since the solution of the Kesten processes are known to obey non-universal power laws (in the sense that the power law exponents continuously vary according to system parameters), the intensity distribution of the QHawkes process also obeys a power law relation,

$$P_{\mathrm{ss}}(\lambda) \propto \lambda^{-3/2-a} \tag{S78}$$

with a non-universal positive number a which can take any positive number according to system parameters, such as D, τ and λ_1. We note that this result can be confirmed by directly solving the Fokker-Planck equation (S67c), even without knowing the theoretical background of the linear Kesten processes. This is in contrast with the universal Zipf's law in our work under the symmetric assumption $\rho(y) = \rho(-y)$ with a fixed power law exponent independent of system parameters. Thus, the QHawkes process is essentially different from our nonlinear Hawkes models because the QHawkes process can be regarded as a linear-Kesten family member while our setup belongs to nonlinear Kesten families.

IV. DIFFUSIVE ASYMPTOTICS: THE SYSTEM-SIZE EXPANSION

Here we briefly review the system-size expansion, an established perturbative method for systems with small-noise or with weak-coupling. This method is relevant to the diffusive approximation in Sec. III and Zipf's law (15) for the number of events in the main text.

In Sec. III, we have used the diffusive approximation to derive multiplicative Gaussian noise terms $\sqrt{\lambda}\xi^{\mathrm{G}}$ from the Poisson noise terms $\xi^{\mathrm{P}}_{\rho(y);\lambda}$. This approximation is asymptotically correct by assuming the diffusive scaling

$$g(\nu) = \frac{1}{\epsilon^2}\bar{g}(\nu), \quad \rho(y) = \frac{1}{\epsilon}\bar{\rho}\left(\frac{y}{\epsilon}\right) \tag{S79}$$

with a small parameter $\epsilon > 0$. We also assume that $\bar{g}(\nu)$ and $\bar{\rho}(Y)$ are ϵ-independent with scaled mark $Y := y/\epsilon$. We note that this method is essentially equivalent to the system-size expansion (or the Ω expansion), invented by van Kampen [8]; historically ϵ is often written by $\epsilon := 1/\Omega$ with large parameter Ω, called the system size (see a recent related Letter [10] and a review [11] for more detail).

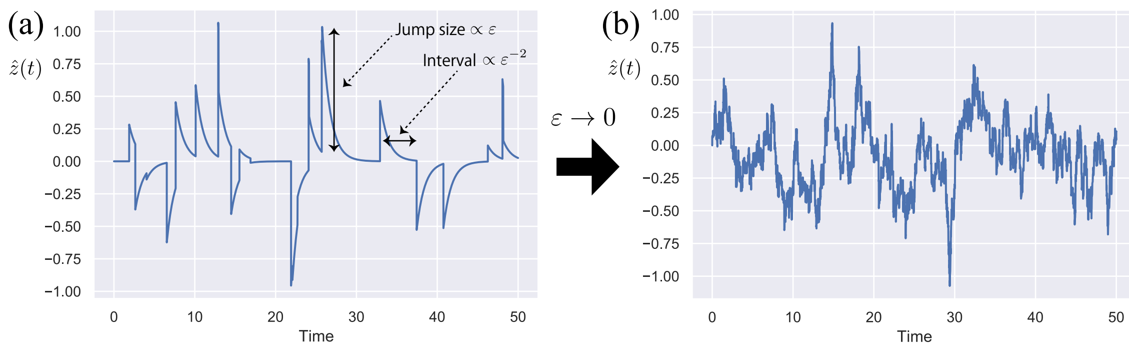


FIG. S2. Typical trajectories of the nonlinear Hawkes process for the diffusive limit (S81). Typically, the inter-events interval is proportional to ϵ^2 and the jump-size is proportional to ϵ. The original point process (S83) gradually reduces to the Langevin dynamics (S86) for small ϵ.

The intuitive explanation of this scaling (S79) is given as follows: the compound Poisson noise is given by $\xi^{\mathrm{P}}_{\rho(y);g(\nu)} = \sum_{i=1}^{N(t)} y_i\delta(t-t_i)$. Here we assume that the mark y_i is sufficiently small; y_i is proportional to a small parameter ϵ as

$$y_i := \epsilon Y_i, \quad \xi^{\mathrm{P}}_{\rho(y);g(\nu)} = \sum_{i=1}^{N(t)} \epsilon Y_i \delta(t-t_i) = \epsilon \xi^{\mathrm{P}}_{\bar{\rho}(y);g(\nu)} \tag{S80}$$

with ϵ-independent mark Y_i and mark distribution $\bar{\rho}(y)$. Considering the Jacobian relation associated with the preservation of probability, we obtain

$$\rho(y)dy = \bar{\rho}(Y)dY \quad \Longleftrightarrow \quad \rho(y) = \frac{1}{\epsilon}\bar{\rho}\left(\frac{y}{\epsilon}\right). \tag{S81}$$

In this sense, the scaling (S81) can be regarded as a small-noise limit or a weak-interaction limit. However, if we naively take the small noise limit $\epsilon \to 0$, the effect of the noise completely disappears. To keep the minimal effect of the noise, let us assume that the intensity is sufficiently large as

$$g(\nu) = \frac{1}{\epsilon^2}\bar{g}(\nu). \tag{S82}$$

We thus obtain the diffusive scaling (S79). In this sense, this scaling implies that the mark size is small but the frequency $\sim \epsilon^{-2}$ is sufficiently high. We thus obtain the following specific form of the nonlinear Hawkes process (see Fig. S2 for typical trajectories):

$$\lambda(t) = \frac{1}{\epsilon^2}\bar{g}\left(\epsilon \sum_{i=1}^{N(t)} Y_i h(t-t_i)\right). \tag{S83}$$

Under this assumption, we can prove that the KM expansion (S67a) converges to the Fokker-Planck equation (S67c). Indeed, the KM coefficients (S67b) have the scalings

$$\alpha_k = \int_{-\infty}^{\infty} y^k \rho(y)dy = \epsilon^k \int_{-\infty}^{\infty} Y^k \bar{\rho}(Y)dY = \epsilon^k \bar{\alpha}_k, \quad \bar{\alpha}_k := \int_{-\infty}^{\infty} Y^k \bar{\rho}(Y)dY. \tag{S84}$$

The KM expansion (S67a) therefore can be transformed as

$$\begin{aligned}\frac{\partial P_t(\nu)}{\partial t} &= \frac{1}{\tau}\frac{\partial}{\partial \nu}\nu P_t(\nu) + \sum_{k=1}^{\infty} \epsilon^{2(k-1)} \frac{\tilde{h}^{2k}\bar{\alpha}_{2k}}{(2k)!}\frac{\partial^{2k}}{\partial \nu^{2k}}[\bar{g}(\nu)P_t(\nu)] \\ &= \frac{1}{\tau}\frac{\partial}{\partial \nu}\nu P_t(\nu) + \frac{\tilde{h}^2\bar{\alpha}_2}{2}\frac{\partial^2}{\partial \nu^2}\bar{g}(\nu)P_t(\nu) + O(\epsilon^2).\end{aligned} \tag{S85}$$

We thus asymptotically obtain the Fokker-Planck equation (S67c) for the diffusive scaling. In addition, this Fokker-Planck equation is equivalent to a multiplicative Langevin dynamics described by

$$\frac{d\nu}{dt} = -\frac{\nu}{\tau} + \sqrt{2D\bar{g}(\nu)}\xi^{\mathrm{G}}, \quad D := \frac{\tilde{h}^2\bar{\alpha}_2}{2} \tag{S86}$$

by using the Ito convention for the small ϵ limit. This methodology can be readily generalised for general memory kernel $h(t)$, by considering the system-size expansion for the field-master equation.

V. ZIPF'S LAW FOR THE EVENTS-NUMBER STATISTICS

We have studied Zipf's law for the steady distribution of intensity, $P_{\rm ss}(\lambda) \propto \lambda^{-1-a}$ by assuming the symmetry $\rho(y) = \rho(-y)$. Since the intensity is one of the fundamental characteristic quantities for point processes in general, our finding will be useful for understanding various Zipf's law even for other quantities. Here we discuss its application to the number of events occurring during a finite time window $t_{\rm win}$ as an example.

Let us consider a long-time interval $[0, T)$ and randomly select a time point $t^* \in [0, T)$. We then count the number of events during an interval $[t^*, t^* + t_{\rm win})$ as $N_{t_{\rm win}}(t^*)$ to observe the corresponding PDF $P_{t=t^*}(N_{t_{\rm win}})$. In the steady state, we can assume that $P_{t=t^*}(N_{t_{\rm win}})$ does not depend on the selection of t^* and therefore we write $P_{t=t^*}(N_{t_{\rm win}})$ by $P_{\rm ss}(N_{t_{\rm win}})$.

Here we derive Zipf's law for $P_{\rm ss}(N_{t_{\rm win}})$ using Zipf's law for $P_{\rm ss}(\lambda)$, by focusing on the exponential tension-intensity map

$$\lambda = g(\nu) = \lambda_0 e^{\beta\nu}. \tag{S87}$$

The basic idea of the derivation is to use the superposition of the Poisson distributions for a sufficiently short time window $t_{\rm win}$. For a short time window $t_{\rm win}$, let us assume that the intensity is approximately constant during $[t^*, t^* + t_{\rm win})$: $\lambda(t) \approx$ const. for $t \in [t^*, t^* + t_{\rm win})$. Under this assumption, the number of events obeys the Poisson distribution:

$$P(N_{t_{\rm win}}|\lambda) \approx \frac{(\lambda t_{\rm win})^{N_{t_{\rm win}}}}{N_{t_{\rm win}}!} e^{-\lambda t_{\rm win}}. \tag{S88}$$

Since t^* is randomly selected, the PDF of $\lambda(t^*)$ obeys Zipf's law (6) for $P_{\rm ss}(\lambda) \propto \lambda^{-2}$. We thus derive Zipf's law for the unconditional PDF of the number of events as

$$P_{\rm ss}(N_{t_{\rm win}}) = \int_0^\infty P(N_{t_{\rm win}}|\lambda) P_{\rm ss}(\lambda) d\lambda \propto N_{t_{\rm win}}^{-2}. \tag{S89}$$

This equation assumes that one can neglect the dependence between the realisations of λ's in subsequent windows. This assumption is likely incorrect for large realisations of λ, giving a clue as to the origin of the deviation from Zipf's law beyond the cut-off value $N_{\rm cut} = O(\epsilon^{-2})$. The study of this strong dependence regime is left for a future work.

While the idea of the superposition of the Poisson statistics works formally, the criteria for the sufficiently-short time window is ambigious for the case of nonlinear Hawkes processes, because the intensity obeys the scale-free distribution (i.e., Zipf's law) and has no clear characteristic timescales as the result of the intemittent properties of the nonlinear Hawkes processes. Because of this scale-free nature, the convergenve of the superposition technique (S89) is not uniform in terms of $N_{t_{\rm win}}$ and has a finite cutoff $N_{\rm cut}$.

To clarify this technical issue, let us consider the dimensional analysis of the nonlinear Hawkes processes by assuming the diffusive scaling limit (S79). Under this condition, the Markov-embedding representation (S2) of the nonlinear Hawkes process is approximated by

$$\frac{dz_k}{dt} = -\frac{z_k}{\tau_k} + \tilde{h}_k \xi^{\rm P}_{\rho(y);\lambda(t)} \approx -\frac{z_k}{\tau_k} + \sqrt{2\bar{D}_k \bar{g}(\nu)}\xi^{\rm G}, \quad \bar{D}_k := \frac{1}{2}\tilde{h}_k^2 \tilde{\alpha}_2 \tag{S90}$$

for small ϵ. Assuming that ν is approximately constant $\nu \approx \nu^* := \nu(t^*)$ during $[t^*, t^* + t_{\rm win})$, the typical displacement $\Delta z_{\rm diff}$ due to diffusion is given by

$$\Delta z_{\rm diff}^2 \approx 2\bar{D}_k \bar{g}(\nu^*) t_{\rm win}. \tag{S91}$$

Let us estimate the typical value of z. Since the steady PDF of tensions ν is given by $P_{\rm ss}(\nu) \sim e^{-\beta\nu}$, the typical value of ν is given by β^{-1}. Since ν is directly related to z_k as $\nu = \sum_{k=1}^K \nu_k$, the typical value of z_k is also the order of β^{-1}: $z_k^* = \beta^{-1}$. We thus estimate that the time window $t_{\rm win}$ is sufficiently short if the following condition is satisfied:

$$\Delta z_{\rm diff} \ll z_k^* \quad \Longleftrightarrow t_{\rm win} \ll \frac{1}{2\beta^2 \bar{D}_k \bar{g}(\nu^*)} = \frac{1}{\epsilon^2}\frac{1}{2\beta^2 \bar{D}_k \lambda(t^*)}. \tag{S92}$$

Remarkably, $\lambda(t^*)$ obeys a scale-free distribution $P_{\rm ss}(\lambda) \propto \lambda^{-2}$ and has no specific characteristic value. It means that this criteria has an explicit dependence on the value of $\lambda(t^*)$. Mathematically, this means that the short time window approximation does not uniformly hold in terms of λ if the time window is fixed. Since the number of events during $t_{\rm win}$ is estimated to be $N \approx \lambda t_{\rm win}$, the cutoff of the PDF derived from the superposition relation (S89) is estimated to be

$$N_{\rm cut} := \lambda(t^*) t_{\rm win} = \frac{1}{\epsilon^2}\frac{1}{2\beta^2 \bar{D}_k} = O(\epsilon^{-2}). \tag{S93}$$

In this sense, Zipf's law for $P_{\rm ss}(N_{t_{\rm win}})$ holds only up to the cutoff $N_{t_{\rm win}} \ll N_{\rm cut}$ as

$$P_{\rm ss}(N_{t_{\rm win}}) \propto N_{t_{\rm win}}^{-2} \quad (N_{t_{\rm win}} \ll N_{\rm cut}). \tag{S94}$$

Since the cutoff diverges as $\lim_{\epsilon\downarrow 0} N_{\rm cut} = \infty$, this asymptotic relation holds for a wide regime for small ϵ and should be regarded as an intermediate asymptotics [12] for the diffusive limit (see Fig. S3 for the numerical confirmation).

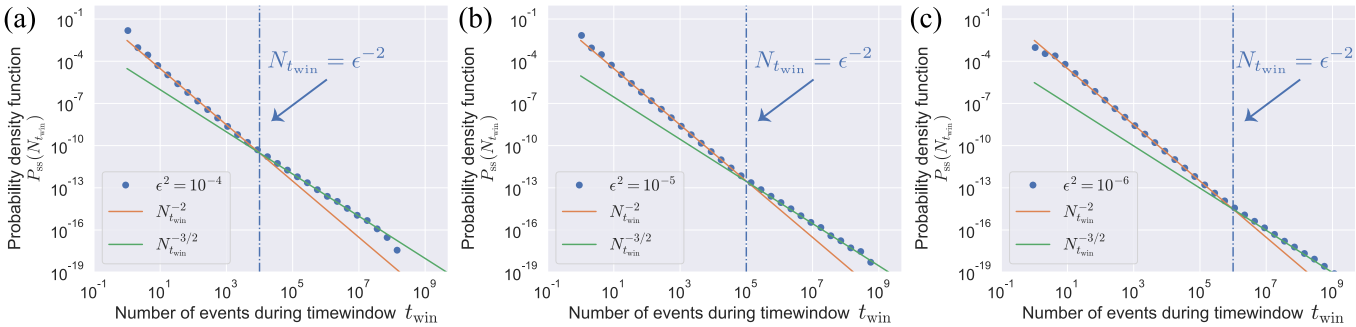


FIG. S3. Numerical confirmation of Zipf's law for the number of events $N_{t_{\rm win}}$ during a short time $t_{\rm win}$. By setting $t_{\rm win} = 10^{-5}$, we numerically observe the PDF of $N_{t_{\rm win}}$ for $\epsilon^2 = 10^{-4}, 10^{-5}, 10^{-6}$. Zipf's law $P_{\rm ss}(N_{t_{\rm win}}) \propto N_{t_{\rm win}}^{-2}$ was found to hold up to $N_{\rm cut} = O(\epsilon^{-2})$ as theoretically predicted, supporting that Zipf's law is indeed valid as intermediate asymptotics. Beyond the cutoff $N > N_{\rm cut}$, we numerically found a fatter tail characterised by a power law exponent $3/2$. This finding is an interesting issue, requiring further investigation.

Numerical scheme

For the numerical simulation of Figure 2c, we set $(\lambda_0, \beta, \sigma, \lambda_{\rm max}, t_{\rm win}, \Delta t_{\rm max}^{(1)}, \Delta t_{\rm max}^{(2)}) = (\epsilon^{-2}, 5, \epsilon, \infty, 10^{-5}, 10^{-6}, 10^{-2})$ for $g(\lambda) = \max\{\lambda_0 e^{\beta\nu}, \lambda_{\rm max}\}$ and $\rho(y) = e^{-y^2/(2\sigma^2)}/\sqrt{2\pi\sigma^2}$. The physical running time on the supercomputer of Kyoto University was bounded at 20 hours. The number of parallel threads was 112 and the total simulation times were $T_{\rm tot} \approx 38777$ for $\epsilon^2 = 10^{-4}$, $T_{\rm tot} \approx 38830$ for $\epsilon^2 = 10^{-5}$, and $T_{\rm tot} \approx 23552$ for $\epsilon^2 = 10^{-6}$, by defining $T_{\rm tot} := \sum_{i=1}^{112} T_i$ with simulation time on the ith thread. The other parameters are the same as those for Figure 1 (see Numerical Code S2.cpp for the detail).

VI. TECHNICAL NOTE ON CALCULATIONS

A. On the Bachmann-Landau-like inequality notation

In this Letter, the Bachmann-Landau equality and inequality notation is defined by

$$a(x) = O(b(x)) \iff \lim_{x\to\infty} \frac{a(x)}{b(x)} < \infty, \tag{S95a}$$

$$a(x) > O(b(x)) \iff \lim_{x\to\infty} \frac{a(x)}{b(x)} = \infty, \tag{S95b}$$

$$a(x) = o(b(x)) \iff \lim_{x\to\infty} \frac{a(x)}{b(x)} = 0. \tag{S95c}$$

Using this notation, we obtain

$$a(x) > O(b(x)) \iff 1/a(x) = o\left(1/b(x)\right) \iff b(x) = o(a(x)). \tag{S95d}$$

B. On the Laplace-like transformation

In the main text, we have introduced a Laplace-like transformation

$$h(t) = \int_0^\infty dx \tilde{h}(x) e^{-t/x}. \tag{S96}$$

This transformation is similar to the Laplace transformation. Indeed, by introducing the variable transformation $x := 1/s$, we obtain

$$h(t) = \int_0^\infty ds H(s) e^{-ts} = \int_0^\infty dx \frac{1}{x^2} H\left(\frac{1}{x}\right) e^{-t/x} \tag{S97}$$

with the Laplace representation $H(s)$. This calculation implies that $\tilde{h}(x)$ is equivalent to $x^{-2}H(x^{-1})$.

C. On the condition for explosive solutions

Here we present an intuitive discussion on the existence of solutions for nonpositive mean mark $m \leq 0$. To capture intuitively the nature of the dynamics, let us truncate the Kramers-Moyal expansion (S50) up to the second order:

$$\frac{\partial P_t(W;\boldsymbol{Z}')}{\partial t} \simeq \left[-m\frac{\partial}{\partial W}G(W;\boldsymbol{Z}') + \frac{\alpha_2}{2}\frac{\partial^2}{\partial W^2}G(W;\boldsymbol{Z}')\right]P_t(W;\boldsymbol{Z}), \quad \alpha_2 := \int_{-\infty}^{\infty} dy y^2\rho(y). \tag{S98}$$

Recall that the function G has been introduced in equation (S14) as $G(\boldsymbol{z}) := g\left(\sum_{k=1}^K z_k\right)$. This Fokker-Planck equation is equivalent to the following stochastic differential equation

$$dW \simeq mG(W;\boldsymbol{Z}')dt + \sqrt{G(W;\boldsymbol{Z}')}dB_t, \quad d\boldsymbol{Z}' = 0 \tag{S99}$$

with the standard Brownian motion B_t. For positive mean mark $m > 0$, the time-evolution is explosive. Indeed, the noise term is negligible for large $W \to \infty$,

$$|mG(W;\boldsymbol{Z}')| \gg \sqrt{G(W;\boldsymbol{Z}')} \tag{S100}$$

and the effectively deterministic dynamics

$$\frac{dW}{dt} \simeq mG(W;\boldsymbol{Z}') \tag{S101}$$

is obviously explosive as soon as G grows faster than linearly as a power law or exponential function. On the other hand, the negative-mean case $m < 0$ is not explosive.

This intuitive discussion is consistent with the rigorous mathematical results on singular stochastic processes [9]. In addition, the rigorous results in Ref. [9] guarantee that the solution is not explosive even for the marginal case $m = 0$. Indeed, according to [9], for an SDE

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t, \quad X_0 = x_0, \tag{S102}$$

the classification of the solution is based on the following quantities

$$\rho_{\mathrm{CE}}(x) := \exp\left(-\int_a^x \frac{2b(y)}{\sigma^2(y)}dy\right), \quad s_{\mathrm{CE}}(x) := -\int_x^{\infty}\rho(y)dy, \quad x \in [a,\infty) \tag{S103}$$

and

$$\int_a^{\infty}\rho_{\mathrm{CE}}(x)dx, \quad \int_a^{\infty}\frac{|s_{\mathrm{CE}}(x)|dx}{\rho_{\mathrm{CE}}(x)\sigma^2(x)} \tag{S104}$$

with some appropriate a. For the SDE (S99), $\rho_{\mathrm{CE}}(x)$ and $s_{\mathrm{CE}}(x)$ are given by

$$\rho_{\mathrm{CE}}(x) := \exp\left(-\int_a^x \frac{2mG}{G}dy\right) = e^{-2m(x-a)}, \tag{S105}$$

$$s_{\mathrm{CE}}(x) := -\int_x^{\infty} dy e^{-2m(y-a)} = \begin{cases} -\infty & (m \leq 0) \\ \frac{1}{2m}e^{-2m(x-a)} & (m > 0) \end{cases}. \tag{S106}$$

According to [9], the non-positive-mean case with $m \leq 0$ is classified as *recurrent*, without explosion to ∞, because

$$\int_a^{\infty}\rho_{\mathrm{CE}}(y)dy = \infty \quad (m \leq 0). \tag{S107}$$

On the other hand, for positive mean mark $m > 0$, the system is classified as *explosive*. Indeed, we obtain

$$\int_a^{\infty}\rho_{\mathrm{CE}}(y)dy < \infty, \quad \int_a^{\infty}\frac{|s_{\mathrm{CE}}(x)|dx}{\rho_{\mathrm{CE}}(x)G(x)} = \frac{1}{2m}\int_a^{\infty}\frac{dx}{G(x)} < \infty \quad (m > 0), \tag{S108}$$

because $G^{-1}(x)$ decays faster than x^{-2} on the assumption of the fast-accelerating intensity.

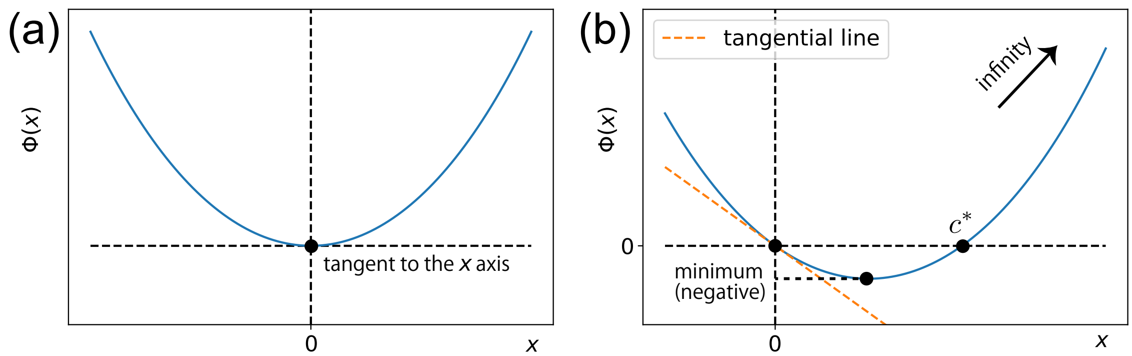


FIG. S4. Schematic figure of the moment-generating function $\Phi(x)$. $\Phi(x)$ is a strictly convex function with specific values $\Phi(0) = 0$, $\Phi(+\infty) = \infty$, and $d\Phi(0)/dt = m$. (a) Case with zero-mean mark $m = 0$, where the curve is tangent to the x axis at $x = 0$. (b) Case with negative-mean mark $m < 0$, where the tangential line at $x = 0$ has negative coefficient $d\Phi(0)/dt = m < 0$. Considering this geometrical shape, the minimum of $\Phi(x)$ occurs at a negative value of $\Phi(x)$ for $x > 0$ and the roots of $\Phi(c) = 0$ are given by $c = 0$ and $c = c^* > 0$.

D. **Derivation of solution** (S27)

Here we show that the solution of the integral equation

$$\int_{-\infty}^{\infty} dy\rho(y)\phi(\nu - y) - \phi(\nu) \simeq 0 \text{ for large } \nu. \tag{S109}$$

Let us first assume that the solution is given by an exponenital $\phi(\nu) = e^{-c\nu}$. By direct substitution, we obtain the self-consistent condition

$$\Phi(c) = 0, \tag{S110}$$

by defining

$$\Phi(x) := \int_{-\infty}^{\infty} dy\rho(y)(e^{xy} - 1). \tag{S111}$$

Remarkably, $\Phi(x)$ is a strictly convex function because

$$\frac{d^2\Phi(x)}{dx^2} = \int_{-\infty}^{\infty} y^2\rho(y)e^{xy}dy > 0. \tag{S112}$$

This means that $\Phi(x)$ has no more than one minimum. Here we assume that $\rho(y)$ decays sufficiently fast and $\Phi(x)$ exists. The general solution of Eq. (S109) is given by the superposition of exponentials (i.e., the two-sided Laplace representation),

$$\phi(\nu) \simeq \sum_i C_i e^{-c_i\nu} \tag{S113}$$

with the ith zero point c_i, satisfying $\Phi(c_i) = 0$ and $c_i < c_{i+1}$.

1. Assuming a zero-mean mark distribution

Let us assume that the mean mark is zero:

$$m := \int_{-\infty}^{\infty} y\rho(y)dy = 0. \tag{S114}$$

For the zero-mean mark distribution, the equation $\Phi(c) = 0$ has a single root at $c = 0$ (see Fig. S4a). Indeed, the minimum of $\Phi(x)$ is at $x = 0$ because

$$\left.\frac{d\Phi(x)}{dx}\right|_{x=0} = \int_{-\infty}^{\infty} y\rho(y)dy = m = 0. \tag{S115}$$

Because the minimum value of $\Phi(x)$ is given by $\Phi(0) = 0$, the only real solution of $\Phi(c) = 0$ is therefore given by $c = 0$.

Interestingly, $c = 0$ is the double root of $\Phi(c) = 0$ and thus a special treatment is necessary: One of the basic solutions of Eq. (S109) is given by a constant function

$$\phi(\nu) \simeq \sum_i C_i e^{-c_i \nu} = C_0. \tag{S116}$$

In addition, the affine function

$$\phi(\nu) \simeq C_0 + C_1 \nu \tag{S117}$$

with another constant C_1 is also a solution. Indeed, we obtain the consistent relation

$$\int_{-\infty}^{\infty} dy \rho(y) \left(C_0 + C_1(\nu - y)\right) - (C_0 + C_1 \nu) = 0, \tag{S118}$$

by considering $\int_{-\infty}^{\infty} dy \rho(y) = 1$ and $\int_{-\infty}^{\infty} y\rho(y) dy = 0$. Notably, the affine form of the solution (S117) can be systematically derived by considering the zero-mean limit of the negative-mean case, as discussed below.

2. *Assuming a negative-mean mark distribution*

Next, let us assume that the mean mark is negative and the probability of positive marks is nonzero

$$m := \int_{-\infty}^{\infty} y\rho(y) dy < 0, \quad m_+ := \int_{0}^{\infty} y\rho(y) dy > 0. \tag{S119}$$

Under this condition, the derivative of $\Phi(x)$ at $x = 0$ is negative,

$$\left.\frac{d\Phi(x)}{dx}\right|_{x=0} = \int_{-\infty}^{\infty} y\rho(y) dy = m < 0. \tag{S120}$$

In addition, $\lim_{x\to\infty} \Phi(x) = +\infty$ because

$$\Phi(x) = \int_0^{\infty} dy \rho(y)(e^{xy} - 1) + \int_{-\infty}^{0} dy \rho(y)(e^{xy} - 1) > \int_0^{\infty} dy \rho(y) xy + \int_{-\infty}^{0} dy \rho(y)(0 - 1) = m_+ x - p_- \to +\infty \tag{S121}$$

for $x \to +\infty$, where we have introduced $p_- := \int_{-\infty}^{0} \rho(y) dy$ and have used the following inequalities: $e^{xy} \geq xy + 1$ for $x \geq 0$ and $e^{xy} > 0$ for any x.

Considering these properties, the schematic picture of $\Phi(x)$ is given by Fig. S4b and all the roots of $\Phi(c) = 0$ are given by $c = 0$ and $c = c^* > 0$. We thus find that the solution of Eq. (S109) is given by

$$\phi(\nu) \simeq \sum_i C_i e^{-c_i \nu} = C_1 + C_0 e^{-c^* \nu}. \tag{S122}$$

For the zero-mean mark limit. For a reference, let us consider the zero-mean mark limit $m \uparrow 0$. Interestingly, for infinitesimal negative m, the positive root of $\Phi(c) = 0$ approaches zero, such that $c^* \downarrow 0$ for $m \uparrow 0$. Then, the solution (S122) can be expanded as

$$\phi(\nu) \simeq \sum_i C_i e^{-c_i \nu} = C_1 + C_0 - c^* C_0 \nu + O(C_0 c^{*2}) \tag{S123}$$

up to the second order by assuming infinitesimal positive c^*. Here we replace $C_0' := C_0 + C_1$ and $C_1' := -c^* C_0$ to obtain

$$\phi(\nu) \simeq \sum_i C_i e^{-c_i \nu} = C_0' + C_1' \nu + O(C_1' c^{*1}). \tag{S124}$$

By taking the limit $m \uparrow 0$ and thus $c^* \downarrow 0$, we obtain the affine form of the solution (S117) and the specific values of the constants C_0' and C_1'.

3. Example: Gaussian mark distribution

As an example, let us consider the case of the Gaussian mark distribution:

$$\rho(y)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-(y-m)^2/(2\sigma^2)} \tag{S125}$$

with mean m and variance σ^2. The moment-generating function is given by

$$\Phi(x)=\int_{-\infty}^{\infty}dy\rho(y)(e^{xy}-1)=e^{mx+\sigma^2x^2/2}-1. \tag{S126}$$

This means that the non-zero root of $\Phi(c^*)=0$ is given by

$$c^*:=-\frac{2m}{\sigma^2}. \tag{S127}$$

E. Derivation of Eq. (S29)

The integration in Eq. (S29) can be performed as follows. From the definition (S25) and the asymptotic solution (S28), the asymptotic steady state solution to the master equation (S21) is given by

$$P_{\rm ss}(\boldsymbol{z})=\{G(\boldsymbol{z})\}^{-1}\phi(\boldsymbol{z})=\{G(\boldsymbol{z})\}^{-1}\psi(W;\boldsymbol{Z}')\approx\{G(\boldsymbol{z})\}^{-1}C_0(\boldsymbol{Z}') \tag{S128}$$

with $G(\boldsymbol{z}):=g\left(\sum_{k=1}^K z_k\right)$, where the variable transformation $\boldsymbol{z}\to\boldsymbol{Z}:=(W,\boldsymbol{Z}')$ is defined by Eq. (S23). We then obtain

$$\begin{aligned}
&\int_{-\infty}^{\infty}d\boldsymbol{z}P_{\rm ss}(\boldsymbol{z})\delta\left(\nu-\sum_{k=1}^K z_k\right)\approx\int_{-\infty}^{\infty}d\boldsymbol{z}C_0(\boldsymbol{Z}')\{G(\boldsymbol{z})\}^{-1}\delta\left(\nu-\sum_{k=1}^K z_k\right)\\
&=\int_{-\infty}^{\infty}d\boldsymbol{z}C_0(\boldsymbol{Z}')\left\{g\left(\sum_{k=1}^K z_k\right)\right\}^{-1}\delta\left(\nu-\sum_{k=1}^K z_k\right)=\int_{-\infty}^{\infty}d\boldsymbol{z}C_0(\boldsymbol{Z}')\{g(\nu)\}^{-1}\delta\left(\nu-\sum_{k=1}^K z_k\right)\\
&=\{g(\nu)\}^{-1}\int_{-\infty}^{\infty}dz_1\int_{-\infty}^{\infty}\left(\prod_{j=2}^K dz_k\right)C_0\left(z_2-\frac{\tilde h_2}{\tilde h_1}z_1,\dots,z_K-\frac{\tilde h_K}{\tilde h_1}z_1\right)\delta\left(\nu-\sum_{k=1}^K z_k\right).
\end{aligned} \tag{S129}$$

Here we apply a variable transformation $z_j':=z_j-\frac{\tilde h_j}{\tilde h_1}z_1$ for $j=2,\dots,K$ to obtain the relation

$$\begin{aligned}
&\int_{-\infty}^{\infty}dz_1\int_{-\infty}^{\infty}\left(\prod_{j=2}^K dz_k\right)C_0\left(z_2-\frac{\tilde h_2}{\tilde h_1}z_1,\dots,z_K-\frac{\tilde h_K}{\tilde h_1}z_1\right)\delta\left(\nu-\sum_{k=1}^K z_k\right)\\
&=\int_{-\infty}^{\infty}dz_1\int_{-\infty}^{\infty}\left(\prod_{j=2}^K dz_k'\right)C_0\left(z_2',\dots,z_K'\right)\delta\left(\nu-\sum_{k=2}^K z_k'-rz_1\right)\\
&=\int_{-\infty}^{\infty}\left(\prod_{j=2}^K dz_k'\right)C_0\left(z_2',\dots,z_K'\right)\int_{-\infty}^{\infty}dz_1\delta\left(\nu-\sum_{k=2}^K z_k'-rz_1\right).
\end{aligned} \tag{S130}$$

Finally, by considering the identities for the δ functions

$$\delta(ax-b)=\frac{\delta(x-b/a)}{|a|},\quad \int_{-\infty}^{\infty}dz_1\delta(z_1-b)=1 \tag{S131}$$

for the constants $a\neq 0$ and b, we obtain

$$\int_{-\infty}^{\infty}dz_1\delta\left(\nu-\sum_{k=2}^K z_k'-rz_1\right)=\frac{1}{r}\int_{-\infty}^{\infty}dz_1\delta\left(z_1-\frac{\nu-\sum_{k=2}^K z_k'}{r}\right)=\frac{1}{r}, \tag{S132}$$

which allows us to deduce

$$\int_{-\infty}^{\infty} d\boldsymbol{z} P_{\rm ss}(\boldsymbol{z})\delta\left(\nu-\sum_{k=1}^{K} z_k\right) \approx \frac{\{g(\nu)\}^{-1}}{r}\int_{-\infty}^{\infty}\left(\prod_{j=2}^{K} dz'_k\right) C_0\left(z'_2,\ldots,z'_K\right) \tag{S133}$$

with a constant $r := (1/\tilde{h}_1)\sum_{k=1}^{K}\tilde{h}_k$. This implies Eq. (S29) by assuming $(1/r)\int_{-\infty}^{\infty} C_0(z'_2,\ldots,z'_K)\Pi_{j=2}^{K} dz'_j < \infty$.

F. Derivation of Eq. (S47)

The integration in Eq. (S47) can be performed as follows. From the definition (S25) and the asymptotic solution (S46), the steady state solution is given by

$$P_{\rm ss}(\boldsymbol{z}) = \{G(\boldsymbol{z})\}^{-1}\phi(\boldsymbol{z}) = \{G(\boldsymbol{z})\}^{-1}\psi(W;\boldsymbol{Z}') = \{G(\boldsymbol{z})\}^{-1} C_0(\boldsymbol{Z}')e^{-c^*W} \tag{S134}$$

with $G(\boldsymbol{z}) := g\left(\sum_{k=1}^{K} z_k\right)$, where the variable transformation $\boldsymbol{z}\to\boldsymbol{Z} := (W,\boldsymbol{Z}')$ is defined by Eq. (S23). We then obtain

$$\begin{aligned}
&\int_{-\infty}^{\infty} d\boldsymbol{z} P_{\rm ss}(\boldsymbol{z})\delta\left(\nu-\sum_{k=1}^{K} z_k\right) \approx \int_{-\infty}^{\infty} d\boldsymbol{z} C_0(\boldsymbol{Z}')e^{-c^*W}\{G(\boldsymbol{z})\}^{-1}\delta\left(\nu-\sum_{k=1}^{K} z_k\right)\\
&=\int_{-\infty}^{\infty} d\boldsymbol{z} C_0(\boldsymbol{Z}')e^{-c^*W}\left\{g\left(\sum_{k=1}^{K} z_k\right)\right\}^{-1}\delta\left(\nu-\sum_{k=1}^{K} z_k\right) = \int_{-\infty}^{\infty} d\boldsymbol{z} C_0(\boldsymbol{Z}')e^{-c^*W}\{g(\nu)\}^{-1}\delta\left(\nu-\sum_{k=1}^{K} z_k\right)\\
&=\{g(\nu)\}^{-1}\int_{-\infty}^{\infty} dz_1\int_{-\infty}^{\infty}\left(\prod_{j=2}^{K} dz_k\right) C_0\left(z_2-\frac{\tilde{h}_2}{\tilde{h}_1}z_1,\ldots,z_K-\frac{\tilde{h}_K}{\tilde{h}_1}z_1\right)\exp\left(-\frac{c^*}{\tilde{h}_1}z_1\right)\delta\left(\nu-\sum_{k=1}^{K} z_k\right).
\end{aligned} \tag{S135}$$

Here we apply a variable transformation $z'_j := z_j - \frac{\tilde{h}_j}{\tilde{h}_1}z_1$ for $j=2,\ldots,K$ to obtain the relation

$$\begin{aligned}
&\int_{-\infty}^{\infty} dz_1\int_{-\infty}^{\infty}\left(\prod_{j=2}^{K} dz_k\right) C_0\left(z_2-\frac{\tilde{h}_2}{\tilde{h}_1}z_1,\ldots,z_K-\frac{\tilde{h}_K}{\tilde{h}_1}z_1\right)\exp\left(-\frac{c^*}{\tilde{h}_1}z_1\right)\delta\left(\nu-\sum_{k=1}^{K} z_k\right)\\
&=\int_{-\infty}^{\infty} dz_1\int_{-\infty}^{\infty}\left(\prod_{j=2}^{K} dz'_k\right) C_0\left(z'_2,\ldots,z'_K\right)\exp\left(-\frac{c^*}{\tilde{h}_1}z_1\right)\delta\left(\nu-\sum_{k=2}^{K} z'_k - rz_1\right)\\
&=\int_{-\infty}^{\infty}\left(\prod_{j=2}^{K} dz'_k\right) C_0\left(z'_2,\ldots,z'_K\right)\int_{-\infty}^{\infty} dz_1\exp\left(-\frac{c^*}{\tilde{h}_1}z_1\right)\delta\left(\nu-\sum_{k=2}^{K} z'_k - rz_1\right)
\end{aligned} \tag{S136}$$

with a constant $r := (1/\tilde{h}_1)\sum_{k=1}^{K}\tilde{h}_k$. Considering the relation for the δ function

$$\begin{aligned}
\int_{-\infty}^{\infty} dz_1\exp\left(-\frac{c^*}{\tilde{h}_1}z_1\right)\delta\left(\nu-\sum_{k=2}^{K} z'_k - rz_1\right) &= \frac{1}{r}\int_{-\infty}^{\infty} dz_1\exp\left(-\frac{c^*}{\tilde{h}_1}z_1\right)\delta\left(z_1-\frac{\nu-\sum_{k=2}^{K} z'_k}{r}\right)\\
&=\frac{1}{r}\exp\left(-\frac{c^*}{\tilde{h}_1}\frac{\nu-\sum_{k=2}^{K} z'_k}{r}\right),
\end{aligned} \tag{S137}$$

we finally obtain

$$\int_{-\infty}^{\infty} d\boldsymbol{z} P_{\rm ss}(\boldsymbol{z})\delta\left(\nu-\sum_{k=1}^{K} z_k\right) \propto \{g(\nu)\}^{-1}\exp\left(-\frac{c^*}{\tilde{h}_{\rm tot}}\nu\right) \tag{S138}$$

and $\tilde{h}_{\rm tot} := r\tilde{h}_1 = \sum_{k=1}^{K}\tilde{h}_k$. This implies Eq. (S47) by assuming

$$\frac{1}{r}\int_{-\infty}^{\infty}\left(\prod_{j=2}^{K} dz'_k\right) C_0\left(z'_2,\ldots,z'_K\right)\exp\left(\frac{c^*}{\tilde{h}_{\rm tot}}\sum_{k=2}^{K} z'_k\right) < \infty. \tag{S139}$$

VII. AUTHOR CONTRIBUTIONS

KK conceived the technical framework and performed the analytical and numerical calculations. DS designed the research, contributed to and checked the analytical calculations and supervised this project. KK and DS discussed all of the results, developed their interpretation and wrote the manuscript.

VIII. CODE AVAILABILITY

- Numerical Code S1.pdf: the numerical source code to obtain Figure 1 in the main text. The code is written in Python and can be run on Jupyter Notebook.
- Numerical Code S2.cpp: the numerical source code to obtain Figure 2c in the main text. The code is written in C++ and its numerical computation was carried out at the Yukawa Institute Computer Facility.
- Numerical Code S3.pdf: the numerical source code to obtain Figure S1. The code is written in Python and can be run on Jupyter Notebook.

---

[1] J.-P. Bouchaud, J. Bonart, J. Donier and M. Gould, *Trades, Quotes and Prices* (Cambridge University Press, Cambridge, 2018).
[2] A. Dassios and H. Zhao, Advances in Applied Probability **43**, 814 (2011).
[3] K. Kanazawa and D. Sornette, Phys. Rev. Lett. **125**, 138301 (2020).
[4] K. Kanazawa and D. Sornette, Phys. Rev. Research **2**, 033442 (2020).
[5] C.W. Gardiner, *Stochastic Methods*, 4th ed. (Springer, Berlin, 2009).
[6] P. Blanc, J. Donier, and J.-P. Bouchaud, Quantitative Finance **17**, 171 (2017).
[7] H. Kesten, Acta Math. **131**, 207 (1973).
[8] N. G. Van Kampen, *Stochastic Processes in Physics and Chemistry* (Elsevier, New York, 1992).
[9] A.S. Cherny and H.-J. Engelbert, *Singular Stochastic Differential Equations.* (Springer-Verlag, Berlin, 2005).
[10] K. Kanazawa, T.G. Sano, T. Sagawa, and H. Hayakawa, Phys. Rev. Lett. **114**, 090601 (2015).
[11] K. Kanazawa, *Statistical Mechanics for Athermal Fluctuation: Non-Gaussian Noise in Physics* (Springer, Berlin, 2017).
[12] G.I. Barenblatt, *Scaling, self-similarity, and intermediate asymptotics* (Cambridge University Press, Cambridge, UK, 1996).

Numerical Code S1

Kiyoshi Kanazawa and Didier Sornette

August 9, 2021

[1]:
```
import numpy as np
import matplotlib.pyplot as plt
from numba import njit
from joblib import Parallel, delayed
```

1 Simulation

Trajectory of the non-linear Hawkes process with the exponential intensity map with truncation

$$\lambda = g(\mathbf{z}) = \max\left\{\lambda_0 e^{\beta\sum_{k=1}^K z_k}, \lambda_{\max}\right\}$$

under the following paramter set:

$$K = 3, \tau = (1.0, 0.5, 2.0), \tilde{h} = (0.5, 0.6, 0.1), \lambda_0 = 1, \lambda_{\max} = 10^7, \beta = 5, \rho(y) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{y^2}{2\sigma^2}}, \sigma = 0.1.$$

We employ an adaptive method for the time step, such that

$$dt_i := t_{i+1} - t_i = \min\left\{dt^{(1)}_{\min}, \frac{dt^{(2)}_{\min}}{g(\nu_i)}\right\}$$

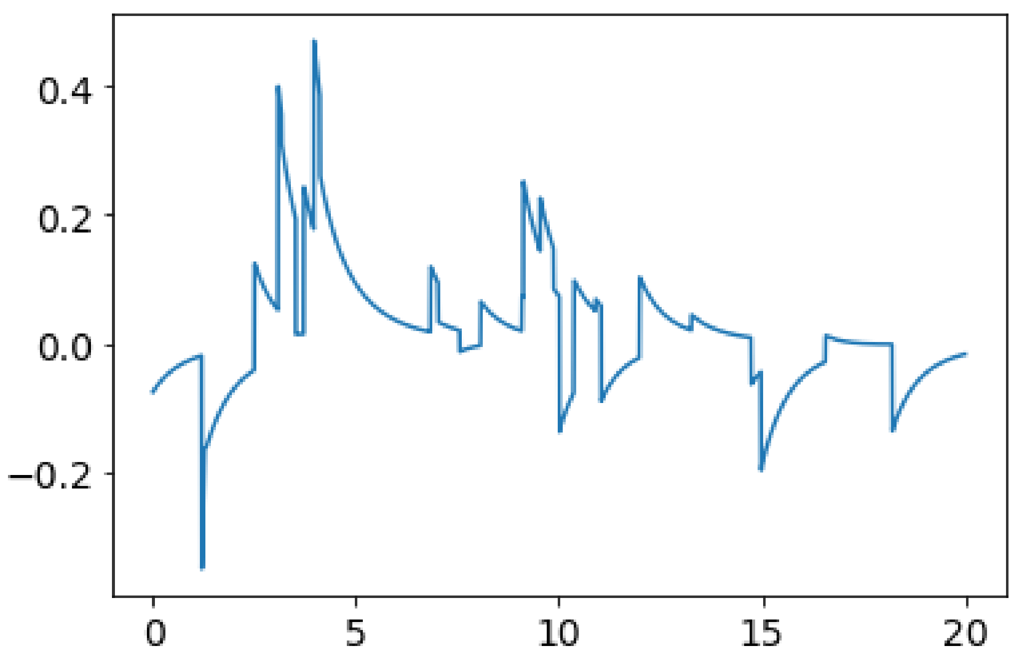
with $dt^{(1)}_{\min} = 0.1$ and $dt^{(2)}_{\min} = 0.01$.

2 Figure 1(a)

A sample tension trajectory $\{\nu(t)\}_t$.

[2]:
```
@njit
def g(x):
    return min(1*np.exp(5*x.sum()), 10**7)

@njit
def gen_trj(seed=0):
    np.random.seed(seed)
    t = 0.0
    T = 10000
    x = np.array([0.0,0.0,0.0])
```

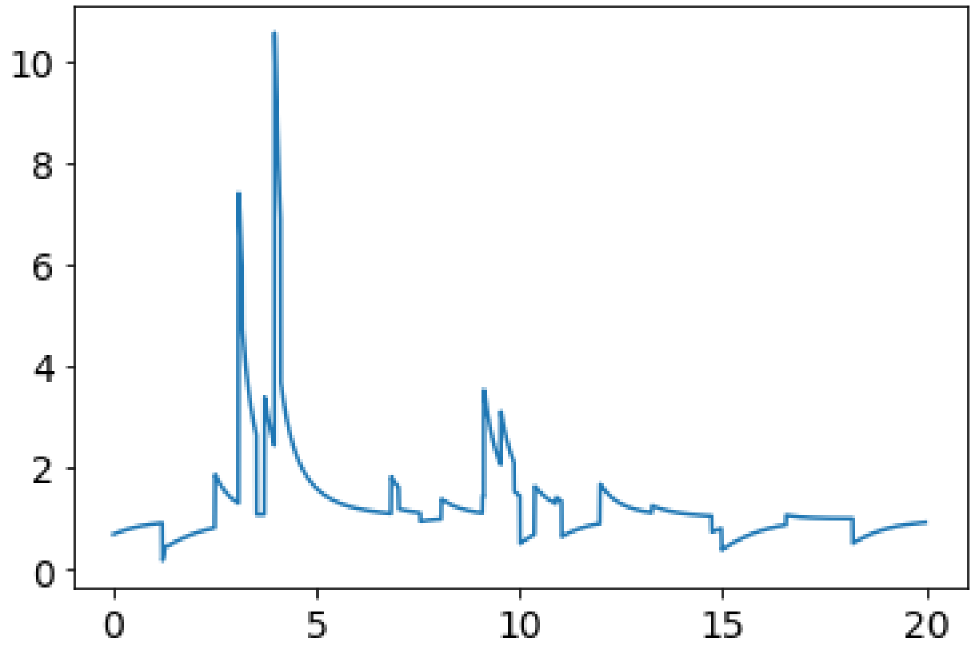
```
    tau = np.array([1.0,0.5,2.0])
    h = np.array([0.5,0.6,0.1])

    ts, zs, ls = [], [], []
    while t < T:
        dt = min(0.1,(1.0/g(x))/100)
        x = x*np.exp(-dt/tau)
        r = np.random.normal() * 0.1
        if np.random.random() <= g(x)*dt:
            x += h * r
        ts.append(t)
        zs.append(x.sum())
        ls.append(g(x))
        t += dt
    return np.array(ts), np.array(zs), np.array(ls)
ts, zs, ls = gen_trj(10)
plt.rcParams["font.size"] = 14
t_ini = 1120; t_window = 20
loc=np.logical_and(t_ini < ts, ts < t_ini+t_window)
plt.plot(ts[loc]-t_ini,zs[loc])
plt.savefig('trajectory_tension.pdf')
```



3 Figure 1(b)

A sample intensity trajectory $\{\lambda(t)\}_t$, corresponding to the tension trajectory in the figure 1(a).

[3]:
```
plt.plot(ts[loc]-t_ini,ls[loc])
plt.savefig('trajectory_intensity.pdf')
```



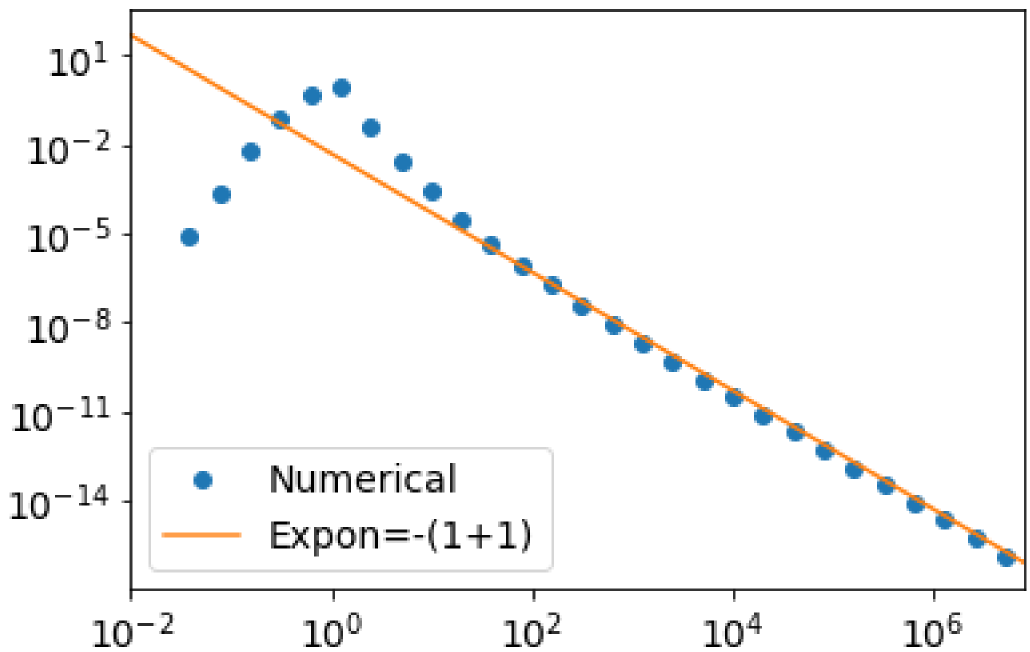
4 Figure 1(c)

This model theoretically exhibits the Zipf law in the intensity distribution:

$$P_{\mathrm{ss}}(\lambda) \simeq \lambda^{-2}.$$

[4]:
```
@njit
def g(x):
    return min(1*np.exp(5*x.sum()), 10**7)

@njit
def positionInBin(x,bins):
    for i in range(len(bins)-1):
        if bins[i] <= x < bins[i+1]:
            return i
    return -1

@njit
def gen(seed=0):
    np.random.seed(seed)
    N_bin = 60
```

```
    hist = np.zeros(N_bin-1)
    bins_hist = 0.0001*2.0**np.arange(0,N_bin,1) # Logarithmic bins

    t = 0.0
    T = 50000
    x = np.array([0.0,0.0,0.0])
    tau = np.array([1.0,0.5,2.0])
    h = np.array([0.5,0.6,0.1])

    while t < T:
        dt = min(0.1,(1.0/g(x))/100)
        x = x*np.exp(-dt/tau)
        r = np.random.normal() * 0.1
        if np.random.random() <= g(x)*dt:
            x += h * r

        pos = positionInBin(g(x),bins_hist)
        if pos >= 0:
            hist[pos] += dt
        t += dt

    return bins_hist, hist

# Parallel computation and reduction
n_jobs=8
result = Parallel(n_jobs=n_jobs)([delayed(gen)(n+11) for n in range(n_jobs)])
hist = np.array(result)[:,1].sum(axis=0)
bins_hist = result[0][0]
hist /= hist.sum()

bins_width = np.array([bins_hist[i+1]-bins_hist[i] for i in␣
 ↪range(len(bins_hist)-1)])
x = np.array([(bins_hist[i+1]+bins_hist[i])/2 for i in range(len(bins_hist)-1)])
PDF = hist/bins_width

plt.rcParams["font.size"] = 14
plt.xscale('log');plt.yscale('log')
x_min = 0.01; x_max = 0.8*10**7
plt.xlim(x_min,x_max)
plt.plot(x,PDF,'o',label='Numerical')
x = np.linspace(x_min,x_max,10)
plt.plot(x,x**(-2.00)/200,label='Expon=-(1+1)')
plt.legend(loc='lower left')
plt.savefig('IntensityPDF.pdf')
```



10^{1}
10^{-2}
10^{-5}
10^{-8}
10^{-11}
10^{-14}
Numerical
Expon=-(1+1)
10^{-2}
10^{0}
10^{2}
10^{4}
10^{6}

Numerical Code S3

Kiyoshi Kanazawa and Didier Sornette

August 9, 2021

[1]:
```
import numpy as np
import matplotlib.pyplot as plt
from numba import njit
from joblib import Parallel, delayed
```

1 Model

Trajectory of the non-linear Hawkes process with the exponential intensity map with truncation

$$\lambda = g(\mathbf{z}) = \max\left\{\lambda_0 e^{\beta\sum_{k=1}^K z_k}, \lambda_{\max}\right\}$$

under the following paramter set:

$$K=3, \tau=(1.0, 0.5, 2.0), \tilde{h}=(0.5, 0.6, 0.1), \lambda_0 = 1, \lambda_{\max}=10^7, \rho(y) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-(y-m)^2/(2\sigma^2)}, \beta = 1.$$

Here m characterizes the asymmetry of the marks. We employ an adaptive method for the time step, such that

$$dt_i := t_{i+1} - t_i = \min\left\{dt^{(1)}_{\min}, \frac{dt^{(2)}_{\min}}{g(\nu_i)}\right\}$$

with $dt^{(1)}_{\min} = 0.1$ and $dt^{(2)}_{\min} = 0.01$.

1.1 Power-law exponent for asymmetric $\rho(y)$

This model theoretically exhibits a power-law in the intensity distribution:

$$P_{\rm ss}(\lambda) \simeq \lambda^{-2-\beta^{-1}a}$$

with a positive constant a. In particular, let us consider the moment-generating function

$$\Phi(x) = \int_{-\infty}^{\infty}\rho(y)(e^{xy}-1)dy = \exp\left(mx + \sigma^2x^2/2\right) - 1.$$

The non-zero root of $\Phi(c^*)=0$ is given by

$$c^* = -\frac{2m}{\sigma^2}$$

by assuming negative m. The deviation constant a is given by

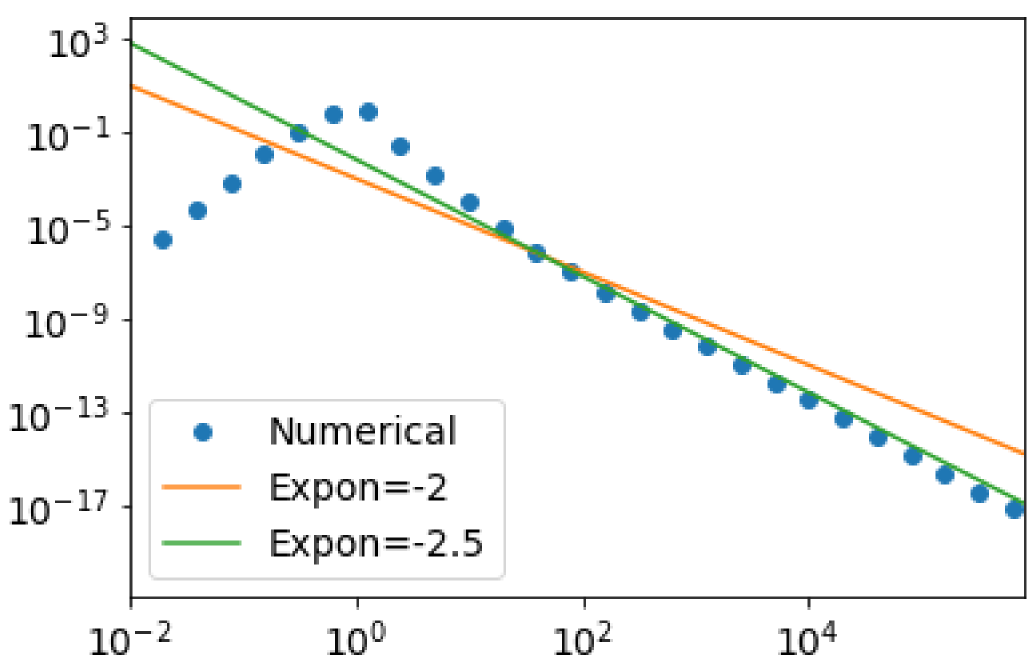
$$a := \frac{c^*}{h(0)} \quad \Longleftrightarrow \quad m = -\frac{\sigma^2 a h(0)}{2}.$$

1.1.1 Case 1

$$a = 0.5, \quad \sigma = 0.5, \quad \beta = 1$$

[2]:
```
@njit
def g(x):
    return min(np.exp(x.sum()), 10**7)

@njit
def positionInBin(x,bins):
    for i in range(len(bins)-1):
        if bins[i] <= x < bins[i+1]:
            return i
    return -1

@njit
def gen(seed=0):
    np.random.seed(seed)
    N_bin = 60
    hist = np.zeros(N_bin-1)
    bins_hist = 0.0001*2.0**np.arange(0,N_bin,1) # Logarithmic bins

    t = 0.0
    T = 5000000
    x = np.array([0.0,0.0,0.0])
    tau = np.array([1.0,0.5,2.0])
    h = np.array([0.5, 0.6, 0.1])
    sigma = 0.5
    a = 0.5
    h_0 = np.sum(h)
    m = -sigma**2 * a * h_0 / 2.0

    while t < T:
        dt = min(0.1,(1.0/g(x))/100)
        x = x*np.exp(-dt/tau)
        r = np.random.normal(loc=m,scale=sigma)
        if np.random.random() <= g(x)*dt:
            x += h * r

        pos = positionInBin(g(x),bins_hist)
        if pos >= 0:
            hist[pos] += dt
```

```
        t += dt

    return bins_hist, hist

# Parallel computation and reduction
n_jobs=8
result = Parallel(n_jobs=n_jobs)([delayed(gen)(n+111) for n in range(n_jobs)])
hist = np.array(result,dtype=object)[:,1].sum(axis=0)
bins_hist = result[0][0]
hist /= hist.sum()

bins_width = np.array([bins_hist[i+1]-bins_hist[i] for i in␣
 ↪range(len(bins_hist)-1)])
x = np.array([(bins_hist[i+1]+bins_hist[i])/2 for i in range(len(bins_hist)-1)])
PDF = hist/bins_width

plt.rcParams["font.size"] = 14
plt.xscale('log');plt.yscale('log')
x_min = 0.01; x_max = 0.8*10**6
plt.xlim(x_min,x_max)
plt.plot(x,PDF,'o',label='Numerical')
y = np.linspace(x_min,x_max,10)
plt.plot(y,y**(-2.0)/1000,label='Expon=-2')
plt.plot(y,y**(-2.50)/150,label='Expon=-2.5')
plt.legend(loc='lower left')
plt.savefig('PDF_asym_2.5.pdf')
```

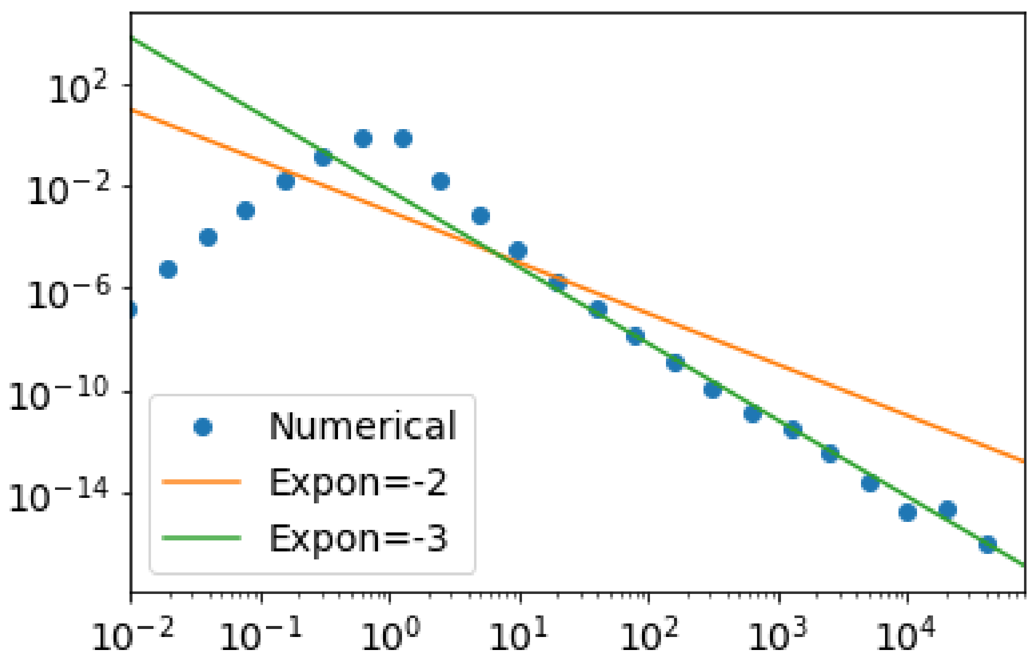


1.1.2 Case 2

$$a = 1.0, \quad \sigma = 0.5, \quad \beta = 1$$

[3]:
```
@njit
def g(x):
    return min(np.exp(x.sum()), 10**7)

@njit
def positionInBin(x,bins):
    for i in range(len(bins)-1):
        if bins[i] <= x < bins[i+1]:
            return i
    return -1

@njit
def gen(seed=0):
    np.random.seed(seed)
    N_bin = 60
    hist = np.zeros(N_bin-1)
    bins_hist = 0.0001*2.0**np.arange(0,N_bin,1) # Logarithmic bins

    t = 0.0
    T = 5000000
    x = np.array([0.0,0.0,0.0])
    tau = np.array([1.0,0.5,2.0])
    h = np.array([0.5, 0.6, 0.1])
    sigma = 0.5
    a = 1.0
    h_0 = np.sum(h)
    m = -sigma**2 * a * h_0 / 2.0

    while t < T:
        dt = min(0.1,(1.0/g(x))/100)
        x = x*np.exp(-dt/tau)
        r = np.random.normal(loc=m,scale=sigma)
        if np.random.random() <= g(x)*dt:
            x += h * r

        pos = positionInBin(g(x),bins_hist)
        if pos >= 0:
            hist[pos] += dt
        t += dt

    return bins_hist, hist
```

```
# Parallel computation and reduction
n_jobs=8
result = Parallel(n_jobs=n_jobs)([delayed(gen)(n+111) for n in range(n_jobs)])
hist = np.array(result,dtype=object)[:,1].sum(axis=0)
bins_hist = result[0][0]
hist /= hist.sum()

bins_width = np.array([bins_hist[i+1]-bins_hist[i] for i in␣
 ↪range(len(bins_hist)-1)])
x = np.array([(bins_hist[i+1]+bins_hist[i])/2 for i in range(len(bins_hist)-1)])
PDF = hist/bins_width

plt.rcParams["font.size"] = 14
plt.xscale('log');plt.yscale('log')
x_min = 0.01; x_max = 0.8*10**5
plt.xlim(x_min,x_max)
plt.plot(x,PDF,'o',label='Numerical')
y = np.linspace(x_min,x_max,10)
plt.plot(y,y**(-2.0)/1000,label='Expon=-2')
plt.plot(y,y**(-3.0)/150,label='Expon=-3')
plt.legend(loc='lower left')
plt.savefig('PDF_asym_3.0.pdf')
```



1.1.3 Case 3

$$a = 1.5, \quad \sigma = 0.5, \quad \beta = 1$$

[4]:
```python
@njit
def g(x):
    return min(np.exp(x.sum()), 10**7)

@njit
def positionInBin(x,bins):
    for i in range(len(bins)-1):
        if bins[i] <= x < bins[i+1]:
            return i
    return -1

@njit
def gen(seed=0):
    np.random.seed(seed)
    N_bin = 60
    hist = np.zeros(N_bin-1)
    bins_hist = 0.0001*2.0**np.arange(0,N_bin,1) # Logarithmic bins

    t = 0.0
    T = 5000000
    x = np.array([0.0,0.0,0.0])
    tau = np.array([1.0,0.5,2.0])
    h = np.array([0.5, 0.6, 0.1])
    sigma = 0.5
    a = 1.5
    h_0 = np.sum(h)
    m = -sigma**2 * a * h_0 / 2.0

    while t < T:
        dt = min(0.1,(1.0/g(x))/100)
        x = x*np.exp(-dt/tau)
        r = np.random.normal(loc=m,scale=sigma)
        if np.random.random() <= g(x)*dt:
            x += h * r

        pos = positionInBin(g(x),bins_hist)
        if pos >= 0:
            hist[pos] += dt
        t += dt

    return bins_hist, hist

# Parallel computation and reduction
n_jobs=8
```

```
result = Parallel(n_jobs=n_jobs)([delayed(gen)(n+111) for n in range(n_jobs)])
hist = np.array(result,dtype=object)[:,1].sum(axis=0)
bins_hist = result[0][0]
hist /= hist.sum()

bins_width = np.array([bins_hist[i+1]-bins_hist[i] for i in␣
 ↪range(len(bins_hist)-1)])
x = np.array([(bins_hist[i+1]+bins_hist[i])/2 for i in range(len(bins_hist)-1)])
PDF = hist/bins_width

plt.rcParams["font.size"] = 14
plt.xscale('log');plt.yscale('log')
x_min = 0.01; x_max = 0.8*10**4
plt.xlim(x_min,x_max)
y_min = 1e-15; y_max = 1e+2
plt.ylim(y_min,y_max)
plt.plot(x,PDF,'o',label='Numerical')
y = np.linspace(x_min,x_max,10)
plt.plot(y,y**(-2.0)/500,label='Expon=-2')
plt.plot(y,y**(-3.5)/50,label='Expon=-3.5')
plt.legend(loc='lower left')
plt.savefig('PDF_asym_3.5.pdf')
```