

SCIENTIFIC REPORTS



OPEN

Sample space reducing cascading processes produce the full spectrum of scaling exponents

Bernat Corominas-Murtra^{1,2}, Rudolf Hanel^{1,2} & Stefan Thurner^{1,2,3,4}

Sample Space Reducing (SSR) processes are simple stochastic processes that offer a new route to understand scaling in path-dependent processes. Here we define a cascading process that generalises the recently defined SSR processes and is able to produce power laws with arbitrary exponents. We demonstrate analytically that the frequency distributions of states are power laws with exponents that coincide with the multiplication parameter of the cascading process. In addition, we show that imposing energy conservation in SSR cascades allows us to recover Fermi's classic result on the energy spectrum of cosmic rays, with the universal exponent -2 , which is independent of the multiplication parameter of the cascade. Applications of the proposed process include fragmentation processes or directed cascading diffusion on networks, such as rumour or epidemic spreading.

Practically all complex adaptive systems exhibit fat-tailed distribution functions in the statistics of their dynamical variables. Often these distribution functions are exact or almost exact asymptotic power-laws, $p(x) \sim x^{-\lambda}$. While Gaussian statistics can often be traced back to a single origin, the central limit theorem, for the origin of power laws there exist several routes. These include (i) Yule-Simon processes (preferential attachment)^{1–3}, (ii) multiplicative processes with constraints, see refs 4 and 5, (iii) criticality⁶, (iv) self-organized criticality and cascading processes^{7–10}, (v) constraint optimization^{11–13}, and (vi) sample space reducing (SSR) processes^{14–16}. Most of these mechanisms are able to explain specific values of the exponent λ or a range of exponents. None of them however explain the full range of exponents from zero to infinity, $\lambda \in [0, \infty)$ in a straightforward way. Depending on the generative process exponents belong to different ranges. For example, exponents near critical points are often fractional, and within the range $\lambda \in [1/2, 5/2]$ ⁶. Many exponents for avalanche processes are found within a range of $\lambda \in (0, 3)$ ¹⁷. Some processes, like the preferential attachment, can formally explain a wide range of exponents, although deviations from the standard values $\lambda \in (2, 3.5)$ are hard to map to realistic underlying stochastic dynamics¹⁸. Here we show that the combination of cascading processes with SSR processes is able to do exactly that, to provide a single one-parameter model that produces the full spectrum of all possible scaling exponents. The parameter is nothing but the multiplication ratio of the cascading processes. Finally, we show that generic disintegration processes can be mapped one-to-one to SSR cascades with an over imposed condition of conservation of whatever magnitude is represented by the states. This mapping allows us to derive a remarkable result: The histogram of visits to each state follows an exponent -2 , regardless the multiplication parameter. This result may have important consequences in order to understand generic properties of disintegration processes and the ubiquity of the exponent -2 in nature. This, for example, allows us to recover Fermi's classic result on the energy spectrum of cosmic rays, only appealing to combinatorial properties of the cascade. In addition, we provide a rigorous proof of that result in the appendix A, as a new contribution to the study of the random partition of the interval.

Cascading processes have played an important role in the understanding of power-law statistics in granular media^{7–10, 19}, earth quakes^{20–22}, precipitation^{23, 24}, dynamics of combinatorial evolution²⁵, or failure in networks^{26–28}. The scaling exponents of the probability distribution functions of quantities such as avalanche sizes, energy distributions, visiting times, event durations, etc., are found within a relatively narrow band. Cascading processes are often history-dependent processes in the sense that for a particular event taking place the temporal order of microscopic events is important. Recent progress in the understanding of the generic statistics of

¹Section for the Science of Complex Systems, CeMSIIS, Medical University of Vienna, Spitalgasse 23, A-1090, Vienna, Austria. ²Complexity Science Hub Vienna, Josefstädterstrasse 39, 1080, Vienna, Austria. ³Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM, 87501, USA. ⁴IIASA, Schlossplatz 1, 2361, Laxenburg, Austria. Correspondence and requests for materials should be addressed to S.T. (email: stefan.thurner@meduniwien.ac.at)

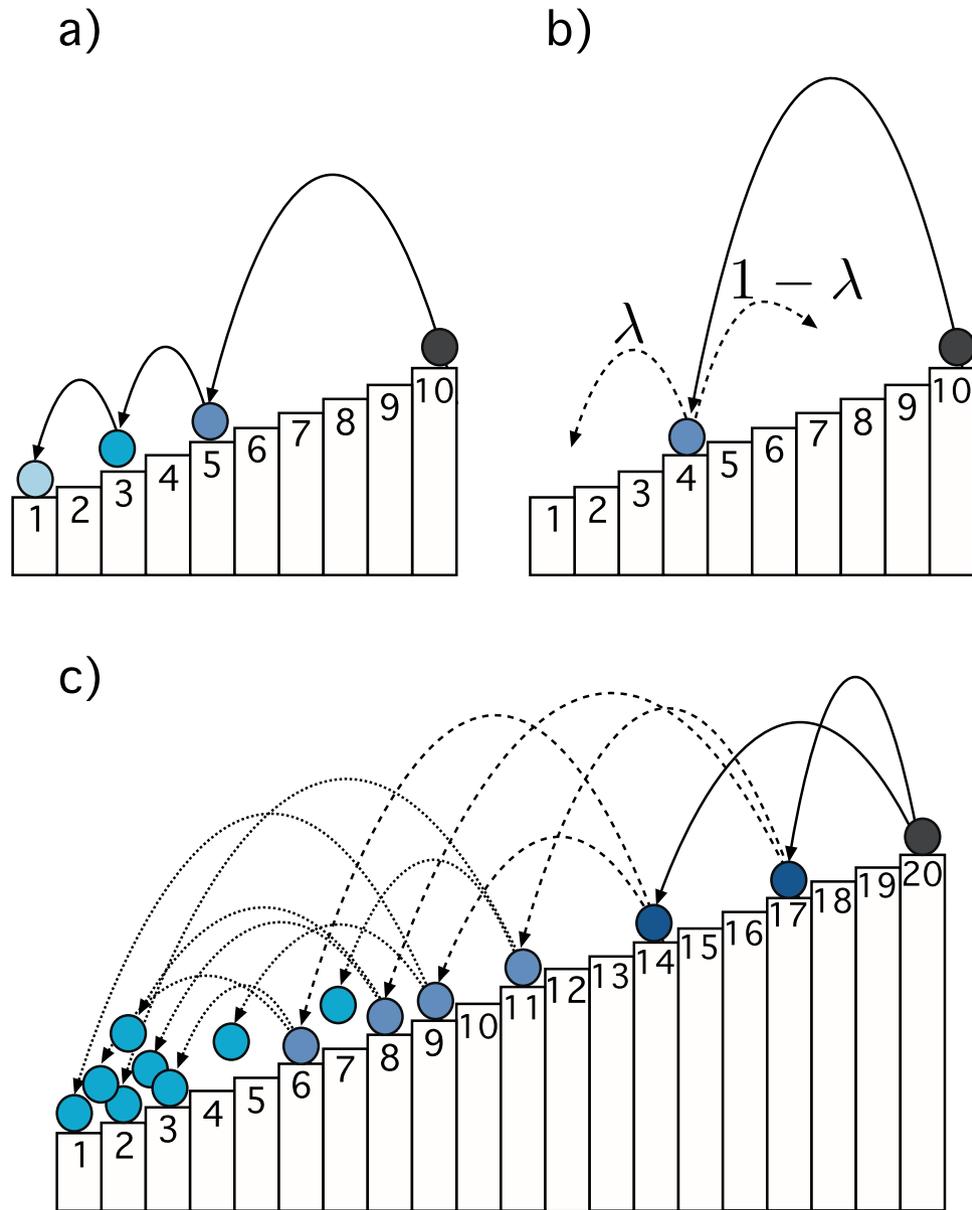


Figure 1. (a) SSR process. A ball starts at the highest state $i = N = 10$, and may sequentially jump randomly toward any lower state. Once the ball hits the lowest state $i = 1$ the process starts again at state N . If repeated many times, the probability of visiting states is an exact Zipf's law, $p(i) \propto i^{-1}$. (b) If with probability $1 - \lambda$, $\lambda \in [0, 1]$ the ball is allowed to jump to any state, the visiting probability becomes $p(i) \propto i^{-\lambda}$. $1 - \lambda$ can be seen as the noise strength in a noisy SSR process, see ref. 14 (c) SSR cascading process with a multiplicative parameter $\mu = 2$. A ball starts at the highest state, $i = N = 20$, and splits into μ new balls which independently jump to lower states as before. Whenever a ball hits a state it creates μ balls which continue their random jumps. It becomes a cascading or an avalanche process, and the visiting probability scales as $p(i) \propto i^{-\mu}$. We observe that (b) is automatically recovered if the multiplication parameter is $\mu < 1$.

history-dependent processes and their relation to power laws was made in refs 14–16 and 29. Maybe the simplest history-dependent processes are the sample space reducing (SSR) processes¹⁴, which explain the origin of scaling in a very simple and intuitive way. They have been used in various applications in computational linguistics³⁰, fragmentation processes¹⁴, and diffusion on directed networks and search processes¹⁵.

SSR Processes and Cascades

A SSR process is a stochastic processes whose sample space reduces as it evolves in time. They can be depicted in a simple way, see Fig. 1(a). Imagine a set of N states in a system, labelled by $i = 1, 2, \dots, N$. The states are ordered by the label. The only rule that defines the SSR process is that transitions between states may only occur from higher to lower labels. This means that transition from state $j \rightarrow i$ is possible only if label $j > i$. When the lowest state $i = 1$ is reached, the process stops or is re-started. It was shown in ref. 14 that this dynamics leads to a Zipf's law in the

frequency of state visits, i.e. the probability to visit state i is given by $p(i) = i^{-1}$. This scaling law is extremely robust and occurs for a wide class of prior probabilities¹⁵. We now show that the combination of SSR processes with the simplest cascading process allows us to obtain a mechanism that can produce power laws with any scaling exponent. We will comment on how the model can be used to recover the Fermi's classic result on the cosmic ray spectrum^{31, 32}. This is possible by imposing energy conservation on SSR cascading processes. We discuss other potential cases where the theory of SSR cascades might apply.

SSR cascades. To define SSR cascades, imagine a system with a set of N ordered states, each of which has a prior probability of appearing, q_1, \dots, q_{N-1} , being state N the starting point of the cascade. The process starts at $t = 0$ with μ balls at state N jumping to any state $i_1, \dots, i_\mu < N$ with a probability proportional to $q_{i_1}, \dots, q_{i_\mu}$, respectively. Suppose that the μ balls landed on states i_1, \dots, i_μ , respectively. At the next timestep, $t = 1$, each of these μ balls divide into μ new balls which all jump to any state *below* their original state. The multiplicative process continues downwards. Whenever a ball hits the lowest state, it is eliminated from the system. Effectively we superimpose a multiplicative process that is characterized by the multiplicative parameter μ , and the SSR process described above, see Fig. 1(c). The case $\mu = 1$ is exactly the standard SSR process, where no new elements are created, and the case $\mu < 1$ corresponds to the *noisy* SSR, where there is the possibility that the process gets cut at some step, see Fig. 1(b).

The derivation of the visiting distribution of this cascading SSR process follows the arguments found in ref. 15. We first define the cumulative prior distribution function $g(k)$,

$$g(k) = \sum_{i \leq k} q_i.$$

Without a multiplication factor μ the transition probabilities $p(i|j)$ determine the probability to reach state i at timestep $t + 1$, given that the system is in state j at time t , and are given by

$$p(i|j) = \begin{cases} \frac{q_i}{g(j-1)} & \text{for } i < j \\ 0 & \text{for } i \geq j. \end{cases} \tag{1}$$

In a SSR cascade, if there is an element sitting at state j at time t , there are now μ trials to reach any state $i < j$ at $t + 1$. Since the number of particles is not conserved throughout the process, we talk about the *expected* number of jumps from j to i . Since the jumps from j to i of each ball is independent, the expected number of jumps from j to i we denoted by $n(j \rightarrow i)$ can be approximated as follows:

$$n(j \rightarrow i) = \mu p(i|j). \tag{2}$$

We denote the expected number of elements that will hit state i in a given SSR cascade by n_i . Up to a factor the sequence n_1, \dots, n_N is identical to the histogram of visits. From equations (1) and (2) we get

$$n_i = \sum_{j>i} n(j \rightarrow i) n_j = \mu q_i \sum_{j>i} \frac{n_j}{g(j-1)}.$$

By subtracting $n_{i+1} - n_i$ and after re-arranging terms we find

$$\frac{n_{i+1}}{q_{i+1}} \left(1 + \mu \frac{q_{i+1}}{g(i)} \right) = \frac{n_i}{q_i},$$

or, when applied iteratively

$$n_i = n_1 \frac{q_i}{q_1} \prod_{1 < j \leq i} \left(1 + \mu \frac{q_j}{g(j-1)} \right)^{-1}.$$

Since $\mu(q_j)/(g(j-1))$ is typically small the product term is well approximated by

$$\begin{aligned} \prod_{1 < j \leq i} (\dots)^{-1} &= \exp \left[- \sum_{1 < j \leq i} \log \left(1 + \mu \frac{q_j}{g(j-1)} \right) \right] \\ &\approx \exp \left(- \mu \sum_{1 < j \leq i} \frac{q_j}{g(j-1)} \right) \\ &\approx \exp \left(- \mu \int_1^i \frac{dq}{dx} \frac{1}{g(x)} dx \right) \\ &\approx \exp \left(- \mu \log \frac{g(i)}{q_1} \right) \\ &= \left(\frac{g(i)}{q_1} \right)^{-\mu}, \end{aligned}$$

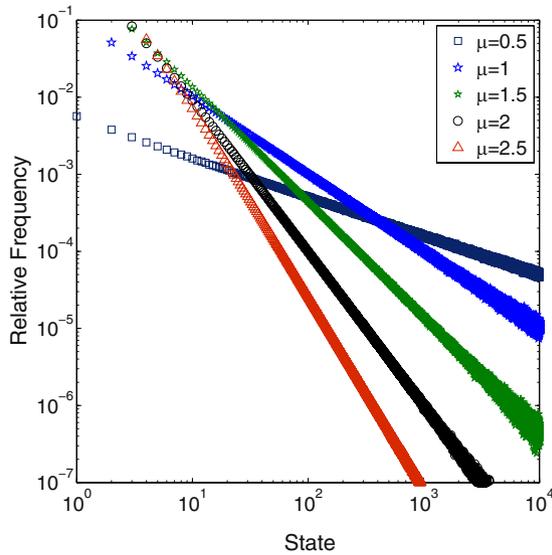


Figure 2. Normalized histograms of state visits of $N = 10,000$ states, obtained from numerical simulations of SSR cascades with a multiplicative parameters $\mu = 0.5$ (dark blue squares), $\mu = 1$ (blue circles), $\mu = 1.5$ (green circles), $\mu = 2$ (black diamonds) and $\mu = 2.5$ (red triangles). Histograms are clearly power laws, $\sim i^{-\mu}$. Curves were fitted using the matlab rplfit package using likelihood estimations for all exponents³⁷.

where we used $q_i \sim dg/dx|_i$ and $\log(1 + x) \sim x$. Finally, we have

$$n_i \sim \frac{n_1}{q_1^{1-\mu}} \left(\frac{q_i}{g(i)^\mu} \right). \tag{3}$$

For equal prior probabilities, $q_i = \frac{1}{N-1}$ for all states, we get the expected visiting probability to be

$$n_i \sim i^{-\mu}.$$

The multiplication factor μ becomes the scaling exponent, for $\mu = 1$ the standard SSR processes is recovered¹⁴. Figure (2) shows numerical results which are in perfect agreement with the theoretical predictions. Note that the argument also holds for non-integer μ , where, on average, μ balls are created at every step. In the numerical implementation, a non-integer μ is introduced as follows: Let $\mu = \lfloor \mu \rfloor + \delta$, with $\delta < 1$. Then, with probability δ , $\lfloor \mu \rfloor + 1$ balls are created and, with probability $1 - \delta$, $\lfloor \mu \rfloor$ balls are created. Also the case of multiplication factors $\mu < 1$ are possible, reproducing the previously defined noisy SSR case, see Fig. (1b) and ref. 14. In this situation at each step the process can be restarted with the probability $1 - \mu$.

We numerically compute the cascade size distribution as a function of the number of states N and μ . For a given realisation of a cascade ψ with initial sample space N and multiplicative parameter μ , starting with a single element at N , we define the *cascade size*, $s_{\mu,N}(\psi)$, as the *number of elements of the cascade ψ that reach state 1*, $n_1^{(\psi)}$. Numerical analysis suggests that the cascade size distribution $f(s_{m,\mu})$ can be well approximated by a Γ distribution³³. For the sake of simplicity, we drop the subscripts μ and N for s . We thus find a purely phenomenological equation that reads.

$$f(s) \propto s^{\alpha-1} e^{-\lambda s}, \quad \langle s \rangle \propto \frac{N^\mu}{e^{a\mu}}, \quad \sigma^2 \propto \frac{N^{b\mu}}{e^{(\frac{1}{2}+a)\mu}} \langle s \rangle, \tag{4}$$

with $a = 0.82$, $b = 0.9$, $\alpha = \langle s \rangle^2 / \sigma^2$, $\lambda = \langle s \rangle / \sigma^2$. Numerical results and fits are shown in Fig. (3). The inset shows that the approximation for $\langle s \rangle$ is highly accurate.

Energy conservation and the prevalence of -2 exponent. In the sequel we derive the statistics of visits to the states of our system when our cascade observes an *energy conservation* constraint. Remarkably, we see that the histogram of visits to each state along the whole cascading process follows a power-law of exponent -2 , regardless the multiplication parameter of the avalanche. The strategy followed here is based on the *partition through renormalization* and works basically as follows: out of an interval $[0, 1]$, a partition is performed by throwing μ random numbers between 0 and 1 and then renormalizing them such that their sum is 1 as in Fig. (4a,b)³⁴. Our strategy can be seen as a particular choice of Dirichlet partitioning of the interval³⁵. This strategy is different from the random selection of breaking points of the interval, described, e.g., in refs 33 and 36. In the appendix A we provide a complete proof of our result.

To study SSR cascades with a superimposed conservation law, let us assume, with any loss of generality, that the states $1, 2, \dots, N$ are associated with energy levels

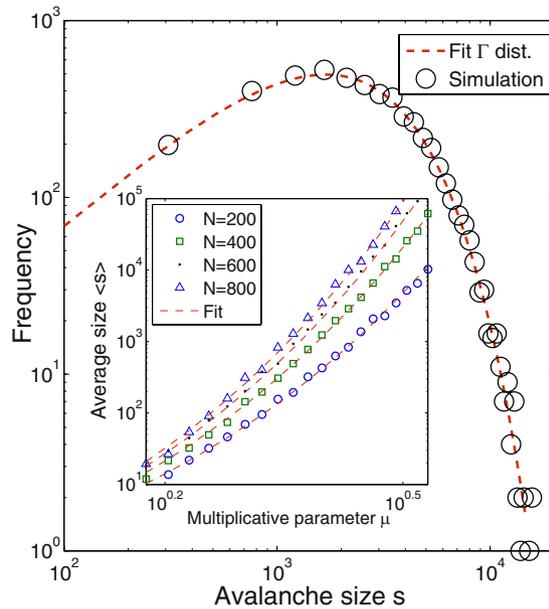


Figure 3. Cascade size distribution after 3,000 realisations of individual SSR cascades with $\mu = 2.5$ and $N = 10^4$ states (circles), with a fit based on the Γ distribution $f(s)$ (red dashed line) given in equation (4). Inset: dependence of $\langle s \rangle$ on μ (from 1.5 to 3.5) for four different system sizes, $N = 200$ (blue circles), $N = 400$ (green circles), $N = 600$ (black diamonds) and $N = 800$ (blue triangles). Dashed curves are fits of the average size $\langle s \rangle$ given in equation (4).

$$E_1, E_2, \dots, E_N.$$

Energy conservation imposes the following constraint: If a particle with energy E and splits into μ particles i_1, i_2, \dots, i_μ , with respective energies $E_{i_1}, E_{i_2}, \dots, E_{i_\mu}$, then:

$$\sum_{k=1}^{\mu} E_{i_k} = E. \tag{5}$$

We are interested in the energy spectrum, i.e. number of observations of particles at a particular energy level at any point of the cascading process, $n(E)$.

The first task is to impose the energy conservation constraint given in equation (5) in the schema of transition probabilities of the SSR cascade. To compute it we use a rescaling technique and we will assume that the energy spectrum is continuous. The rescaling technique is outlined in Fig. (4). Let us ignore energy conservation for the moment and define a continuous uniform random variable u on the interval $[0, 1]$. Let u_1, u_2, \dots, u_μ be μ independent realisations of u , see Fig. (4a). Let us suppose that we are at level E . From this sequence of random variables one can derive the target sites of the newly created particles in a SSR avalanche with multiplicative parameter μ as

$$u_1 \cdot E, u_2 \cdot E, \dots, u_\mu \cdot E.$$

This is the continuous version of what we described in section 2.1. Now we define a new random variable, ϕ_μ , which is the sum of μ realisations of the random variable u :

$$\phi_\mu = \sum_{k \leq \mu} u_k.$$

The sum of μ realisations of a random variable u uniformly distributed on the interval $[0, 1]$, ϕ_μ , follows the Irwin-Hall distribution, $f_\mu(\phi_\mu)$ ³³. This means that one can construct, for each μ realisations of the random variable u a rescaled sequence, see Fig. (4b)

$$\phi_\mu^{-1} u_1 \cdot E, \dots, \phi_\mu^{-1} u_\mu \cdot E,$$

such that sum up to the total energy E ,

$$\phi_\mu^{-1} \sum_{i \leq \mu} u_i \cdot E = E.$$

Thus, by imposing energy conservation we actually expect the following sequence of rescaled energies $E_{i_1}, E_{i_2}, \dots, E_{i_\mu}$ for the emerging particles, where

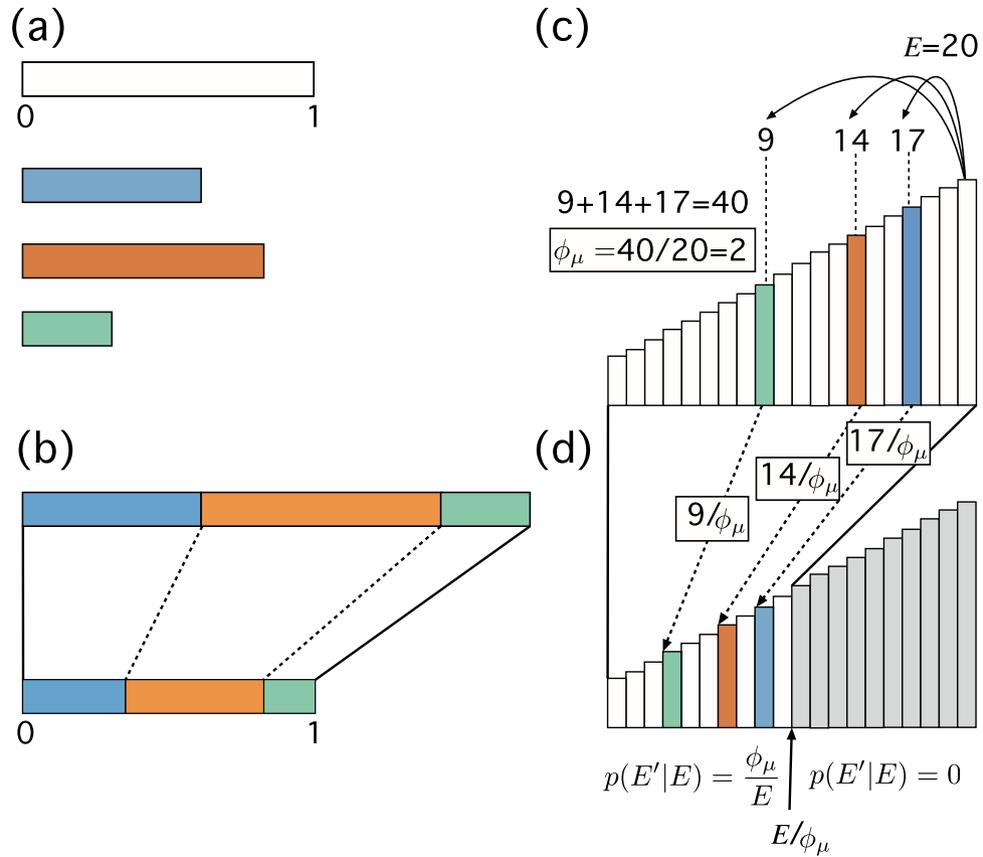


Figure 4. SSR cascades with energy conservation and the random partition of the interval. **(a)** Consider a stick of length 1 and chose three sticks (blue, orange and green) of random sizes between 0 and 1. **(b)** Put them together sequentially and glue them. This will create another stick made of the blue, orange and green sticks of size larger than 1. Divide the stick by its size and one gets a stick of length 1 partitioned in different random segments proportional to the sizes of the sticks chosen in **(a)**. This process of random partition of the interval is exactly analogous to a SSR cascade with energy conservation. **(c)** At a given energy level (20) three balls jump downwards to a randomly chosen sites (17, 14, 9). Energy conservation imposes that the sum of the energies of the landing sites is equal to 20, although the sum of the obtained energies is 40. The rescaling parameter ϕ_μ will be thus $40/20 = 2$. **(d)** We rescale all the outcomes by the factor ϕ_μ and project to a new staircase. The effect of the rescaling is that, in this particular realisation of there cascading process, any value E' such that $\phi_\mu E' > 20$ will be forbidden –grey region in the picture. A cascade would iterate the described process until all created balls reach state 1.

$$E_{i_k} = \phi_\mu^{-1}(u_k \cdot E).$$

This rescaling approach assumes that the μ new particles behave independently. The crucial issue is to map this process into a cascade, see Fig. (4c,d). To approach this problem, we first study the expected number of particles that jump to a given state E if a given value of ϕ_μ occurs. We then average over all potential values of ϕ_μ . We assume that the expected number of particles from $E \rightarrow E'$, $n(E \rightarrow E')$ goes as $\sim \mu p(E'|E)$. Taking into account the rescaling imposed by energy conservation, $p(E'|E) \sim \frac{\phi_\mu}{E}$, see Fig. (4)–on has that:

$$n(E \rightarrow E', \phi_\mu) = \begin{cases} \mu \frac{\phi_\mu}{E} & \text{for } \phi_\mu E' < E \\ 0 & \text{for } \phi_\mu E' \geq E. \end{cases} \quad (6)$$

Assuming a continuous spectrum of energies, one has, for a given value of ϕ_μ , that the expected number of particles that will visit state E at some point of the cascade, $n(E, \phi_\mu)$ is:

$$n(E, \phi_\mu) = \int_{\phi_\mu E}^{\infty} \mu \frac{\phi_\mu}{E'} n(E') dE',$$

where $n(E')$ is the total number of particles that are expected to visit state E' during the cascade. $n(E)$ will be obtained by averaging $n(E, \phi_\mu)$ over all potential values of ϕ_μ , distributed as the Irwin-Hall distribution, f_μ :

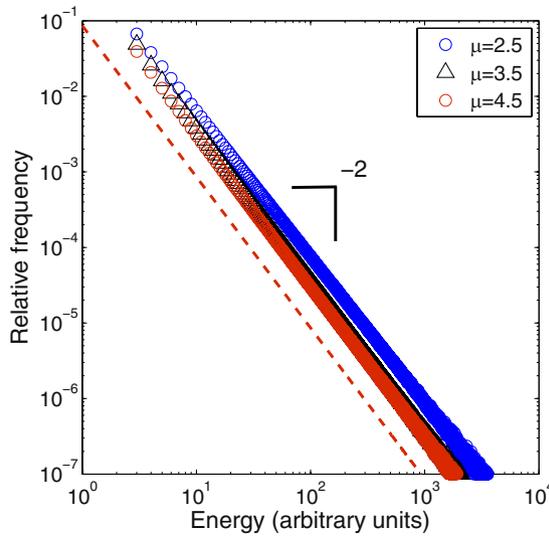


Figure 5. Histograms of SSR cascades with energy conservation at each multiplication event, for $\mu = 2.5$ (blue dots), $\mu = 3.5$ (red triangles) and $\mu = 4.5$ (black diamonds). As predicted only exponents with a value of 2 occur. The dashed line shows the slope a perfect power-law with exponent $\alpha = 2$ for comparison. Curves were fitted using the matlab rplfit package using likelihood estimations for all exponents³⁷.

$$n(E) = \int_0^\mu f_\mu(\phi_\mu) \left\{ \int_{\phi_\mu E}^\infty \mu \frac{\phi_\mu}{E'} n(E') dE' \right\} d\phi_\mu. \tag{7}$$

Differentiating $n(E)$, one arrives at the following equation with displacement:

$$\frac{dn}{dE} = -\mu \int_0^\mu \phi_\mu f_\mu(\phi_\mu) n(\phi_\mu E) d\phi_\mu.$$

Assuming that $n(E) \propto E^{-\alpha}$, we arrive at the following self-consistent equation for α :

$$\alpha = \mu \int_0^\mu \phi_\mu^{1-\alpha} f_\mu(\phi_\mu) d\phi_\mu. \tag{8}$$

whose only solution, for large μ 's converges to $\alpha = 2$, leading to the general result.

$$n(E) \propto E^{-2}. \tag{9}$$

In the appendix A we provide a rigorous derivation of this result. In spite of the asymptotic nature of the proof given in the appendix A, numerical simulations show an excellent agreement with this theoretical prediction, even for α small. In Fig. (5) we show the frequency plots for 10^5 avalanches with $\mu = 2.5, 3.5, 4.5$, and the convergence of the histograms to E^{-2} can be perfectly appreciated. This result is the same that is obtained from Fermi's particle acceleration model to explain the spectrum of cosmic rays^{31,32}. Here we derived it on the basis of simple combinatorial reasonings of SSR processes with a superimposed constraint.

Discussion

We have shown, both analytically and numerically, that SSR processes represent a general route to scaling, that is able to generate *any* scaling exponent. To obtain exponents larger than 1, which was the main target of this paper, we introduced the concept of SSR cascades, a simple multiplicative process which combines SSR- and cascading processes. Our results also add a new way of the interpretation of scaling laws observed in multiplicative processes, and avalanche- or cascading processes. The quality of the SSR view rests in its extreme simplicity, intuitive nature and its generality. In addition, SSR cascades can be mapped to physical process of successive disintegration. By applying the approach to a physical cascading process of particle cascades from cosmic rays it is sufficient to impose energy conservation in the cascading process. In doing so, we recover the classic result of Fermi for the energy spectrum of cosmic rays. More generally, we have shown that energy conservation (or any other similar constraint) leads to a universal scaling exponent of 2, regardless of the details of the microscopic cascading process. In the case of no constraint and no cascading present the ubiquitous Zipf's law is obtained, with its exponent 1. Our findings imply that the presence of simple history-dependence imposed by SSR processes deforms statistics of in a highly non-intuitive way and leads naturally to power laws, and at the same time explains why exponents 1 and 2 are special. We believe that our results might be useful to understand scaling in problems of statistical inference whenever observation biases exist, for problems of fragmentation and cascading, including



Figure 6. SSR cascades with energy conservation can be seen as follows: Consider a polygon with a given area (e.g., 1), $t=0$. At $t=1$ throw three random numbers between 0 and 1, x_1, x_2, x_3 and then rescale them $x'_i = x_i / (x_1 + x_2 + x_3)$. Now embed three polygons of arbitrary shape with area x'_1, x'_2, x'_3 , respectively, and put them inside the first polygon. At $t=2$ we repeat the operation for each of the subpolygons created and we iterate the process for $t \gg 1$. In this appendix we prove that the probability that area x has been found at some point of the partitioning process scales as $p(x) \sim x^{-2}$. Notice that the 2D polygon is just a representation and that our result does not depend on the dimension of the partitioned object.

the meteorite energy spectrum, particle cascades, and for multiplicative directed diffusion on networks such as e.g. rumour spreading, or the spreading of viral loads in populations.

Appendix A: Derivation of exponent -2 for cascades with energy conservation. We derive the main result of section IIB. The strategy followed is summarised in Fig. (4). An alternative view is given in Fig. (6) in this appendix.

We start with the definition of the Irwin-Hall distribution: Let u be a random variable whose probability density is uniform in the interval $[0, 1]$. Let u_1, \dots, u_μ be a sequence of independent drawings of the random variable u and $\phi(\mu)$ a random variable defined over the interval $[0, \mu]$ as:

$$\phi_\mu = \sum_{k \leq \mu} u_k. \quad (\text{A1})$$

The probability density that governs the random variable ϕ_μ is the Irwin-Hall distribution. Now we go to equation (7),

$$n(E) = \int_0^\mu f_\mu(\phi_\mu) \left\{ \int_{\phi_\mu E}^\infty \mu \frac{\phi_\mu}{E'} n(E') dE' \right\} d\phi_\mu.$$

Differentiating,

$$\begin{aligned} \frac{dn}{dE} &= \frac{d}{dE} \int_0^\mu f_\mu(\phi_\mu) \left\{ \int_{\phi_\mu E}^\infty \mu \frac{\phi_\mu}{E'} n(E') dE' \right\} d\phi_\mu \\ &= \int_0^\mu f_\mu(\phi_\mu) \frac{d}{dE} \left\{ \int_{\phi_\mu E}^\infty \mu \frac{\phi_\mu}{E'} n(E') dE' \right\} d\phi_\mu \\ &= -\mu \int_0^\mu \phi_\mu f_\mu(\phi_\mu) n(\phi_\mu E) d\phi_\mu, \end{aligned}$$

one arrives at the following equation with displacement:

$$\frac{dn}{dE} = -\mu \int_0^\mu \phi_\mu f_\mu(\phi_\mu) n(\phi_\mu E) d\phi_\mu.$$

Assuming that $n(E) \propto E^{-\alpha}$, we arrive at the following self-consistent equation for α :

$$\alpha = \mu \int_0^\mu \phi_\mu^{1-\alpha} f_\mu(\phi_\mu) d\phi_\mu.$$

This is equation (8) and that is what we have to solve. The following theorem states that the solution in the limit of large μ 's is $\alpha=2$, independent of μ . To demonstrate that we need to proof 5 lemmas. After the demonstration, we approach the solution $\alpha \rightarrow 2$ using a mean field approach. Finally, we report a side observation concerning the behaviour of the average value of a random variable following the Irwin-hall distribution.

Theorem: *The only α satisfying the following equation:*

$$\lim_{\mu \rightarrow \infty} \mu \int_0^\mu \phi_\mu^{1-\alpha} f_\mu(\phi_\mu) d\phi_\mu = 2. \tag{A2}$$

is $\alpha=2$.

To prove this theorem, we observe first observe that a random variable following the Irwin-Hall distribution converges to a random variable following a normal distribution with average $\frac{\mu}{2}$ and standard deviation $\sqrt{\frac{\mu}{12}}$.

Lemma 1: *Let ϕ_μ a random variable following the Irwin-Hall distribution. Let Y be a random variable following a normal distribution centred at 0 and with standard deviation 1. Then, the following limit holds:*

$$\phi_\mu \rightarrow \frac{\mu}{2} + \sqrt{\frac{\mu}{12}} Y(0, 1).$$

in probability.

Proof: The average value of a uniformly distributed random variable u is $\mathbb{E}(u) = \frac{1}{2}$ and standard deviation $\sigma = \sqrt{\frac{1}{12}}$. By of the central limit theorem, one has that, for an i.i.d. sequence of random variables u_1, \dots, u_μ :

$$\frac{\sum_{i \leq \mu} u_i - \frac{\mu}{2}}{\sqrt{\frac{\mu}{12}}} \rightarrow Y,$$

being Y a random variable following a normal distribution centred at 0 and with standard deviation 1. Therefore, by realising that ϕ_μ is actually a sum of μ i.i.d random variables u , one has:

$$\phi_\mu \rightarrow \frac{\mu}{2} + \sqrt{\frac{\mu}{12}} Y(0, 1),$$

as we wanted to demonstrate. □

This implies that the Irwin-Hall distribution f_μ can be fairly approached by a normal distribution with mean $\frac{\mu}{2}$ and standard deviation $\sqrt{\frac{\mu}{12}}$, $\Phi_\mu(x)$. However, one must be careful with this approach: It can lead the integral that we want to solve, $\int_0^\mu \phi_\mu^{1-\alpha} f_\mu(\phi_\mu) d\phi_\mu$, to a singularity at 0 which is inexistent in the Irwin-Hall distribution. Therefore, for the interval $[0, 1)$ we will maintain the original form of the distribution. In the following lemma we demonstrate that this has no impact in the limit of large μ 's.

Lemma 2: *Let Φ_μ be a normal distribution with mean at $\frac{\mu}{2}$ and standard deviation $\sigma_\mu = \sqrt{\frac{\mu}{12}}$. Then, $(\forall \epsilon > 0) (\exists N): (\forall \mu > N)$*

$$\left| \int_0^\mu \phi_\mu^{1-\alpha} f_\mu(\phi_\mu) d\phi_\mu - \int_1^\mu \phi_\mu^{1-\alpha} \Phi_\mu(\phi_\mu) d\phi_\mu \right| < \epsilon.$$

Proof: From lemma 1 we know that $(\forall \epsilon' > 0) (\exists N): (\forall \mu > N)$

$$\left| \int_1^\mu \phi_\mu^{1-\alpha} f_\mu(\phi_\mu) d\phi_\mu - \int_1^\mu \phi_\mu^{1-\alpha} \Phi_\mu(\phi_\mu) d\phi_\mu \right| < \epsilon'.$$

Now we observe that the Irwin-Hall distribution can be defined per intervals using different polynomials. In the case of the interval $[0, 1]$, the polynomial reads:

$$f_\mu(\phi_\mu) = \frac{1}{(\mu - 1)!} \phi_\mu^{\mu-1}; \quad \phi_\mu \in [0, 1]. \tag{A3}$$

Computing directly the integral, one has:

$$\begin{aligned} \int_0^\mu \phi_\mu^{1-\alpha} f_\mu(\phi_\mu) d\phi_\mu &= \int_0^1 \phi_\mu^{1-\alpha} f_\mu(\phi_\mu) d\phi_\mu \\ &+ \int_1^\mu \phi_\mu^{1-\alpha} f_\mu(\phi_\mu) d\phi_\mu, \end{aligned}$$

where the first integral, according to equation (A3) leads to:

$$\begin{aligned} \int_0^1 \phi_\mu^{1-\alpha} f_\mu(\phi_\mu) d\phi_\mu &= \frac{1}{(\mu - 1)!} \int_0^1 \phi_\mu^{\mu-\alpha} d\phi_\mu \\ &= \frac{1}{(\mu - 1)! (\mu - \alpha - 1)}. \end{aligned}$$

Now take $\delta \in (0, 1)$ and define $\epsilon(\mu, \delta)$ as:

$$\epsilon(\mu) \equiv \frac{1 + \delta}{(\mu - 1)! (\mu - \alpha - 1)}$$

From lemma 1, $(\forall \epsilon' > 0) (\exists N): (\forall \mu > N)$ we can define the following bound:

$$\left| \int_0^\mu \phi_\mu^{1-\alpha} f_\mu(\phi_\mu) d\phi_\mu - \int_1^\mu \phi_\mu^{1-\alpha} \Phi_\mu(\phi_\mu) d\phi_\mu \right| < \epsilon' + \epsilon(\mu, \delta).$$

Finally, we observe that

$$\lim_{\mu \rightarrow \infty} \epsilon(\mu, \delta) = 0,$$

which demonstrates the lemma. □

Lemma 3: *The function of $G(\alpha)$ defined by the integral*

$$G(\alpha) = \int_1^\mu \phi_\mu^{1-\alpha} \Phi(\phi_\mu) d\phi_\mu,$$

is strictly decreasing.

Proof: It is enough to compute the derivative:

$$\begin{aligned} \frac{d}{d\alpha} \int_1^\mu \phi_\mu^{1-\alpha} \Phi(\phi_\mu) d\phi_\mu &= \int_1^\mu \left(\frac{d}{d\alpha} \phi_\mu^{1-\alpha} \right) \Phi(\phi_\mu) d\phi_\mu \\ &= - \int_1^\mu \phi_\mu^{1-\alpha} \log \phi_\mu \Phi(\phi_\mu) d\phi_\mu \\ &< 0, \end{aligned}$$

since the term inside the integral, $\phi_\mu^{1-\alpha} \log \phi_\mu \Phi(\phi_\mu)$, is strictly positive in the interval $(1, \mu)$. □

Now take a monotonously increasing function that grows slower than the standard deviation $\sigma_\mu = \sqrt{\frac{\mu}{12}}$, $\varphi(\mu)$. For convenience, we define define it as:

$$\varphi(\mu) \equiv \left(\frac{\mu}{12} \right)^{\frac{1}{4}}. \tag{A4}$$

Clearly:

$$\lim_{\mu \rightarrow \infty} \int_{\frac{\mu}{2} - \sigma_\mu \varphi(\mu)}^{\frac{\mu}{2} + \sigma_\mu \varphi(\mu)} \Phi(\phi_\mu) d\phi_\mu = 1. \tag{A5}$$

Now suppose, that $\alpha = 2$. Then, thanks to Lemma 2, one has that $(\forall \epsilon > 0) (\exists N): (\forall \mu > N)$

$$\left| \int_0^\mu \frac{f_\mu(\phi_\mu)}{\phi_\mu} d\phi_\mu - \int_1^\mu \frac{\Phi_\mu(\phi_\mu)}{\phi_\mu} d\phi_\mu \right| < \epsilon.$$

From this we derive the third lemma of our demonstration:

Lemma 4: $(\forall \epsilon > 0) (\exists N): (\forall \mu > N),$

$$\left| \int_1^\mu \frac{\Phi(\phi_\mu)}{\phi_\mu} d\phi_\mu - \int_{\frac{\mu}{2} - \sigma_\mu \varphi(\mu)}^{\frac{\mu}{2} + \sigma_\mu \varphi(\mu)} \frac{\Phi(\phi_\mu)}{\phi_\mu} d\phi_\mu \right| < \epsilon.$$

Proof: We need to compute the parts that fall outside the integration limits and see that their contribution vanishes. First, we see that:

$$\begin{aligned} \int_{\frac{\mu}{2} + \sigma_\mu \varphi(\mu)}^\mu \frac{\Phi(\phi_\mu)}{\phi_\mu} d\phi_\mu &< \left(\frac{\mu}{2} + \sigma_\mu \varphi(\mu) \right) e^{-\varphi^2(\mu) - O(\log \mu)} \\ &< \left(\frac{\mu}{2} + \sigma_\mu \varphi(\mu) \right) e^{-\sqrt{\mu}}. \end{aligned}$$

Analogously,

$$\int_1^{\frac{\mu}{2} - \sigma_\mu \varphi(\mu)} \frac{\Phi(\phi_\mu)}{\phi_\mu} d\phi_\mu < \left(\frac{\mu}{2} - \sigma_\mu \varphi(\mu) \right) e^{-\sqrt{\mu}}.$$

Now we define:

$$\begin{aligned} \epsilon_1(\mu) &= \left(\frac{\mu}{2} + \sigma_\mu \varphi(\mu) \right) e^{-\sqrt{\mu}}, \\ \epsilon_2(\mu) &= \left(\frac{\mu}{2} - \sigma_\mu \varphi(\mu) \right) e^{-\sqrt{\mu}}, \end{aligned}$$

which leads to:

$$\left| \int_1^\mu \frac{\Phi(\phi_\mu)}{\phi_\mu} d\phi_\mu - \int_{\frac{\mu}{2} - \sigma_\mu \varphi(\mu)}^{\frac{\mu}{2} + \sigma_\mu \varphi(\mu)} \frac{\Phi(\phi_\mu)}{\phi_\mu} d\phi_\mu \right| < \epsilon_1(\mu) + \epsilon_2(\mu),$$

demonstrating the lemma. □

Corollary of Lemma 4: $(\forall \epsilon > 0) (\exists N): (\forall \mu > N),$

$$\left| \int_0^\mu \phi_\mu^{1-\alpha} f_\mu(\phi_\mu) d\phi_\mu - \int_{\frac{\mu}{2} - \sigma_\mu \varphi(\mu)}^{\frac{\mu}{2} + \sigma_\mu \varphi(\mu)} \frac{\Phi(\phi_\mu)}{\phi_\mu} d\phi_\mu \right| < \epsilon.$$

Proof: By direct application of lemmas 2 and 4. □

Now we define the following functions of the limits of the integral $\int_{\frac{\mu}{2} - \sigma_\mu \varphi(\mu)}^{\frac{\mu}{2} + \sigma_\mu \varphi(\mu)} \dots$:

$$\begin{aligned} r_1(\mu) &\equiv \left(\frac{\mu}{2} + \sigma_\mu \varphi(\mu) \right)^{-1} = \frac{2}{\mu} - \frac{2\sigma_\mu \varphi_\mu}{\mu \left(\frac{\mu}{2} + \sigma_\mu \varphi(\mu) \right)}, \\ r_2(\mu) &\equiv \left(\frac{\mu}{2} - \sigma_\mu \varphi(\mu) \right)^{-1} = \frac{2}{\mu} + \frac{2\sigma_\mu \varphi_\mu}{\mu \left(\frac{\mu}{2} - \sigma_\mu \varphi(\mu) \right)}. \end{aligned}$$

Clearly,

$$r_{1,2}(\mu) \sim \frac{2}{\mu} + \mathcal{O}\left(\mu^{-\frac{5}{4}}\right), \tag{A6}$$

where the subscript $_{1,2}$ means that both functions satisfy the property.

Lemma 5: $(\forall \epsilon > 0) (\exists N): (\forall \mu > N)$

$$\left| r_{1,2}(\mu) \cdot \mu \int_{\frac{\mu}{2} - \sigma_\mu \varphi(\mu)}^{\frac{\mu}{2} + \sigma_\mu \varphi(\mu)} \Phi(\phi_\mu) d\phi_\mu - \mu \int_{\frac{\mu}{2} - \sigma_\mu \varphi(\mu)}^{\frac{\mu}{2} + \sigma_\mu \varphi(\mu)} \frac{\Phi(\phi_\mu)}{\phi_\mu} d\phi_\mu \right| < \epsilon.$$

Proof: We first observe that, by substituting ϕ_μ^{-1} by the integration limits, we have the following chain of inequalities, in terms of the above defined functions $r_{1,2}(\mu)$:

$$\begin{aligned} r_2(\mu) \cdot \int_{\frac{\mu}{2}-\sigma_\mu\varphi(\mu)}^{\frac{\mu}{2}+\sigma_\mu\varphi(\mu)} \Phi(\phi_\mu) d\phi_\mu &< \int_{\frac{\mu}{2}-\sigma_\mu\varphi(\mu)}^{\frac{\mu}{2}+\sigma_\mu\varphi(\mu)} \frac{\Phi(\phi_\mu)}{\phi_\mu} d\phi_\mu \\ &< r_1(\mu) \cdot \int_{\frac{\mu}{2}-\sigma_\mu\varphi(\mu)}^{\frac{\mu}{2}+\sigma_\mu\varphi(\mu)} \Phi(\phi_\mu) d\phi_\mu. \end{aligned}$$

Therefore, is enough to demonstrate that $(\forall \epsilon' > 0) (\exists N): (\forall \mu > N)$

$$|\mu r_1(\mu) - \mu r_2(\mu)| < \epsilon'.$$

This can be proven directly from equation (A6), leading to:

$$|\mu r_1(\mu) - \mu r_2(\mu)| \sim \mathcal{O}\left(\mu^{-\frac{1}{4}}\right),$$

which demonstrates the lemma. □

Collecting lemmas 1, 2, 4, and 5, we have demonstrated that, under the assumption that $\alpha = 2$,

$$\mu r_1(\mu) \rightarrow \mu \int_0^\mu \phi_\mu^{1-\alpha} f_\mu(\phi_\mu) d\phi_\mu.$$

From equation (8), the only remaining issue is to demonstrate the consistency of our hypothesis is that indeed $\lim_{\mu \rightarrow \infty} \mu r_1(\mu) = 2$. It is not difficult to check, from equation (A6), that:

$$\lim_{\mu \rightarrow \infty} \mu r_1(\mu) = \lim_{\mu \rightarrow \infty} \left[2 + \mathcal{O}\left(\mu^{-\frac{1}{4}}\right) \right] = 2.$$

So far we have demonstrated that the solution $\alpha = 2$ is consistent with the statement of the theorem. Now it remains to demonstrate that this is the only solution. To see that, we observe that thanks to lemma 3, we know that the function $\mu G(\alpha)$ is decreasing. In addition, we have proven that the statement of the theorem is consistent for $\alpha = 2$. Therefore, if $\alpha = 2 + \beta$, with $\beta > 0$, then:

$$\mu \int_0^\mu \phi_\mu^{1-\alpha} f_\mu(\phi_\mu) d\phi_\mu < 2 + \beta,$$

which contradicts the statement of the theorem. The same happens if one imposes if $\alpha = 2 - \beta$, with $\beta > 0$, since one gets:

$$\mu \int_0^\mu \phi_\mu^{1-\alpha} f_\mu(\phi_\mu) d\phi_\mu > 2 - \beta,$$

which is, again inconsistent, thereby proving the lemma. □

By direct application of equation (8), Lemma 5 puts the last piece to demonstrate the theorem. □

Mean-field approach. In a less rigorous way, we observe that we can approach the solution as follows: We know that the expected value of a random variable ϕ_μ following the Irwin-Hall distribution f_μ is:

$$\mathbb{E}(\phi_\mu) = \frac{\mu}{2}.$$

Now assume that $\phi_\mu \approx \frac{\mu}{2}$. This implies that, in the integral of the statement of the theorem, equation (A2), we replace $f_\mu(\phi_\mu)$ by $\delta\left(\phi_\mu - \frac{\mu}{2}\right)$, where δ is the Dirac δ function:

$$\mu \int_0^\mu \phi_\mu^{1-\alpha} f_\mu(\phi_\mu) d\phi_\mu \approx \mu \int_0^\mu \phi_\mu^{1-\alpha} \delta\left(\phi_\mu - \frac{\mu}{2}\right) d\phi_\mu.$$

Solving the integral, and thanks to equation (8), we obtain the following relation:

$$\frac{\alpha}{\mu} = \left(\frac{\mu}{2}\right)^{1-\alpha},$$

whose only solution is $\alpha = 2$.

We end observing that the theorem that we demonstrated has a curious consequence: Let ϕ_μ be a random variable following the Irwin-Hall distribution. We observe that, if $\alpha = 2$, then:

$$\begin{aligned}
\frac{\alpha}{\mu} &= \int_0^\mu \phi_\mu^{1-\alpha} f_\mu(\phi_\mu) d\phi_\mu \\
&= \int_0^\mu \frac{f_\mu(\phi_\mu)}{\phi_\mu} d\phi_\mu \\
&= \mathbb{E}\left(\frac{1}{\phi_\mu}\right) \\
&\rightarrow \frac{2}{\mu}.
\end{aligned}$$

We know that $\mathbb{E}(\phi_\mu) = \frac{\mu}{2}$. Therefore, a direct consequence of the theorem is that:

$$\mathbb{E}\left(\frac{1}{\phi_\mu}\right) \rightarrow \frac{1}{\mathbb{E}(\phi_\mu)}.$$

References

1. Yule, G. U. A Mathematical Theory of Evolution, Based on the Conclusions of Dr. J. C. Willis. *Philos. Trans. R. Soc. London B* **213**, 21 (1925).
2. Simon, H. A. On a class of skew distribution functions. *Biometrika* **42**, 425 (1955).
3. Barabási, A.-L. & Albert, R. Emergence of scaling in random networks. *Science* **286**, 509 (1999).
4. Mitzenmacher, M. A Brief History of Generative Models for Power Law and Lognormal Distributions. *Internet Mathematics* **1**(2), 226 (2003).
5. Newman, M. E. J. Power laws, Pareto distributions and Zipf's law. *Contemp. Phys.* **46**(5), 323 (2005).
6. Stanley, H. E. *Introduction to Phase Transitions and Critical Phenomena* (Oxford University Press: Oxford) (1987).
7. Bak, P., Tang, C. & Wiesenfeld, K. Self-organized criticality: An explanation of the 1/f noise. *Phys. Rev. Lett.* **59**, 381 (1987).
8. Kadanoff, L. P., Nagel, S. R., Wu, L. & Zhou, S.-M. Scaling and universality in avalanches. *Phys. Rev. A* **39**, 6524 (1989).
9. Jensen, H. J. *Self-Organized Criticality* (Cambridge University Press, Cambridge) (1996).
10. Christensen, K. & Moloney, N. R. *Complexity and Criticality* (Imperial College Press, London, UK) (2005).
11. Mandelbrot, B. An Informational Theory of the Statistical Structure of Languages. In *Communication Theory*, Jackson, W. editor 486502 (Woburn, MA: Butterworth) (1953).
12. Harremoës, P. & Topsøe, P. Maximum Entropy Fundamentals. *Entropy* **3**, 191 (2001).
13. Corominas-Murtra, B., Fortuny, J. & Solé, R. V. Emergence of Zipf's law in the evolution of communication. *Phys. Rev. E* **83**, 036115 (2011).
14. Corominas-Murtra, B., Hanel, R. & Thurner, S. Understanding scaling through history-dependent processes with collapsing sample space. *Proc. Natl. Acad. Sci. USA* **112**(17), 5348 (2015).
15. Corominas-Murtra, B., Hanel, R. & Thurner, S. Extreme robustness of scaling in sample space reducing processes explains Zipf-law in diffusion on directed networks. *New Journal of Physics* **18**(9), 093010 (2016).
16. Hanel, R., Thurner, S. & Gell-Mann, M. How multiplicity determines entropy and the derivation of the maximum entropy principle for complex systems. *Proc. of the Natl. Acad. of Sci. USA* **111**, 6905 (2014).
17. Paczuski, M., Maslov, S. & Bak, P. Avalanche dynamics in evolution, growth, and depinning models. *Phys. Rev. E* **53**, 414 (1996).
18. Jackson M. O. *Social and Economic Networks* (Princeton University Press, Princeton, NJ) (2010).
19. Frette, V., Christensen, K., Malthé-Sørensen, A., Feder, J., Jossang, T. & Meakin, P. Avalanche dynamics in a pile of rice. *Nature* **379**, 49 (1996).
20. Sornette, A. & Sornette, D. Self-organized criticality and earthquakes. *EPL (Europhysics Letters)* **9**(3), 197 (1989).
21. Turcotte, D. L. *Fractals and Chaos in Geology and Geophysics* (Cambridge University Press, Cambridge), 2nd ed (1997).
22. Corral, Á. Long-term clustering, scaling, and universality in the temporal occurrence of earthquakes. *Phys. Rev. Lett.* **92**(10), 108501 (2004).
23. Peters, O. & Neelin, D. Critical phenomena in atmospheric precipitation. *Nature Physics* **2**, 393 (2006).
24. Corral, Á., Osso, A. & Llebot, J. E. Scaling of tropical-cyclone dissipation. *Nature Phys* **6**, 693 (2010).
25. Thurner, S., Klimek, P. & Hanel, R. Schumpeterian economic dynamics as a quantifiable minimum model of evolution. *New Journal of Physics* **12**, 075029 (2010).
26. Boss, M., Summer, M. & Thurner, S. Contagion flow through banking networks. *Lecture Notes in Computer Science* **3038**, 1070 (2004).
27. Buldyrev, S. V., Parshani, R., Paul, G., Stanley, H. E. & Havlin, S. Catastrophic cascade of failures in interdependent networks. *Nature* **464**(7291), 1025 (2010).
28. Thurner, S., Farmer, J. D. & Geanakoplos, J. Leverage causes fat tails and clustered volatility. *Quantitative Finance* **12**, 695 (2012).
29. Hanel, R. & Thurner, S. Generalized (c,d)-Entropy and Aging Random Walks. *Entropy* **15**, 5324–5337 (2013).
30. Thurner, S., Hanel, R., Liu, B. & Corominas-Murtra, B. Understanding Zipf's law of word frequencies through sample-space collapse in sentence formation. *Journal of the Royal Society Interface* **12**, 20150330 (2016).
31. Fermi, E. On the Origin of the Cosmic Radiation. *Phys. Rev.* **75**, 1169 (1949).
32. Longair, M. S. *High Energy Astrophysics, Vol. 2: Stars, the Galaxy and the Interstellar Medium* (Cambridge University Press, Cambridge, MA) (2008).
33. Feller, W. *An Introduction to Probability Theory and its Applications*, Vols I and II, third edition (John Wiley and Sons, New York, NY) (1968).
34. Kingman, J. F. C. *Poisson processes, Oxford Studies in Probability*, vol. 3 (The Clarendon Press, Oxford University Press, New York) (1993).
35. Huillet, T. Sampling formulae arising from random Dirichlet populations. *Communications in Statistics - Theory and Methods* **34**(5), 1019–1040 (2005).
36. Krapivsky, P. L. & Ben-Naim, E. Scaling and multiscaling in models of fragmentation. *Phys. Rev. E* **50**, 3502–3507 (1994).
37. Hanel, R., Corominas-Murtra, B., Liu, B. & Thurner, S. Fitting Power-laws in empirical data with estimators that work for all exponents. *PLoS One* **12**(2), e0170920 (2016).

Acknowledgements

This work was supported by the Austrian Science Fund FWF under the P29032 and P 29252 projects. We acknowledge an anonymous reviewer for the constructive comments on our first version of the manuscript.

Author Contributions

B.C.-M., R.H. and S.T. designed, performed the research and wrote the manuscript.

Additional Information

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017