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

Bernat Corominas-Murtra\textsuperscript{1,2}, Rudolf Hanel\textsuperscript{1,2} and Stefan Thurner\textsuperscript{1,2,3,4,5,}\textsuperscript{*}

\textsuperscript{1} Section for the Science of Complex Systems, CeMSIIS, Medical University of Vienna, Spitalgasse 23, A-1090, Vienna, Austria
\textsuperscript{2} Complexity Science Hub Vienna, Josefstaetterstrasse 39, 1080 Vienna, Austria
\textsuperscript{3} Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA
\textsuperscript{4} IIASA, Schlossplatz 1, 2361 Laxenburg, Austria

(Dated: Version May 10, 2021)

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.

PACS numbers: 89.75.Da, 05.40.-a, 05.10.Ln, 05.40.Fb

I. INTRODUCTION

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)\textsuperscript{1–3}, (ii) multiplicative processes with constraints, see\textsuperscript{4–6}, (iii) criticality\textsuperscript{7–10}, (iv) self-organized criticality and cascading processes\textsuperscript{11–13}, (v) constraint optimization\textsuperscript{14,15}, and (vi) sample space reducing (SSR) processes\textsuperscript{16,17}. Most of these mechanisms are able to explain specific values of the exponent $\lambda$ 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 $\lambda \in [1/2, 5/2]$\textsuperscript{16}. Many exponents for avalanche processes are found within a range of $\lambda \in (0, 3)$\textsuperscript{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 dynamics\textsuperscript{18}. 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\textsuperscript{19–21}, earth quakes\textsuperscript{22,23}, precipitation\textsuperscript{24,25}, dynamics of combinatorial evolution\textsuperscript{26}, or failure in networks\textsuperscript{27,28}. 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 history-dependent processes and their relation to power laws was made in\textsuperscript{29,30}. Maybe
The simplest history-dependent processes are the sample space reducing (SSR) processes [14], which explain the origin of scaling in a very simple and intuitive way. They have been used in various applications in computational linguistics [30], fragmentation processes [14], and diffusion on directed networks and search processes [15].

II. SSR PROCESSES AND CASCADES

A SSR process is a stochastic process whose sample space reduces as it evolves in time. They can be depicted in a simple way, see figure (a). Imagine a set of $N$ states in a system, labelled by $i = 1, 2, \cdots, N$. 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 restarted. It was shown in [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 $p(i) = i^{-\lambda}$. This scaling law is extremely robust and occurs for a wide class of prior probabilities [15]. 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 52]. This is possible by imposing energy conservation on the SSR cascading processes. We discuss other potential cases where the theory of SSR cascades might apply.

A. 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, $q_1, \cdots, q_{N-1}$, being state $N$ the starting point of the cascade. The process starts at $t = 0$ with $\mu$ balls at state $N$ jumping to any state $i_1, \cdots, i_{\mu} < N$ with a probability proportional to $q_1, \cdots, q_{\mu}$, respectively. Suppose that the $\mu$ balls landed on states $i_1, \cdots, i_{\mu}$, respectively. At the next timestep, $t = 1$, each of these $\mu$ balls divide into $\mu$ 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 $\mu$, and the SSR process described above, see figure (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 figure (b).

The derivation of the visiting distribution of this cascading SSR process follows the arguments found in [14]. We first define the cumulative prior distribution function $g(k)$.

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

Without a multiplication factor $\mu$ 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

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

In a SSR cascade, if there is an element sitting at state $j$ at time $t$, there are now $\mu$ 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:

$$n(j \rightarrow i) = \mu p(i|j) \quad .$$  

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 $n_1, ..., n_N$ is identical to the histogram of visits. From equations (1) and (2) we get

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

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

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

or, when applied iteratively

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

Since $\mu \frac{q_i}{g(i)}$ is typically small the product term is well approximated by

$$\prod_{1<j\leq i} \left(1 + \mu \frac{q_j}{g(j-1)}\right)^{-1} \approx \exp \left[ \sum_{1<j\leq i} \log \left(1 + \mu \frac{q_j}{g(j-1)}\right) \right] \quad .$$

$$\approx \exp \left[ -\mu \sum_{1<j\leq i} \frac{q_j}{g(j-1)} \right] \quad .$$

$$\approx \exp \left[ -\mu \int^i \frac{dg}{dx} \frac{1}{g(x) dx} \right] \quad .$$

$$\approx \exp \left[ -\mu \log \frac{g(i)}{q_1} \right] \quad = \left( \frac{g(i)}{q_1} \right)^{-\mu} \quad ,$$

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

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

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

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

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

**FIG. 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 black 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 exponents [34].

We numerically compute the cascade size distribution as a function of the number of states $N$ and $\mu$. For a given realisation of a cascade $\psi$ with initial sample space $N$ and multiplicative parameter $\mu$, starting with a single element at $N$, we define the cascade size, $s_{\mu,N}(\psi)$, as the number of elements of the cascade $\psi$ that reach state $1$, $n_1(\psi)$. Numerical analysis suggests that the cascade size distribution $f(s_{\mu,N})$ can be well approximated by a $\Gamma$ distribution [33]. For the sake of simplicity, we drop the subscripts $\mu$ and $N$ for $s$. We thus find a purely phenomenological equation that reads

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

with $a = 0.82, b = 0.9, \alpha = \langle s^2 \rangle / \sigma^2, \lambda = \langle s \rangle / \sigma^2$. Numerical results and fits are shown in figure 3. The inset shows that the approximation for $\langle s \rangle$ is highly accurate. 
B. 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 $\mu$ random numbers between 0 and 1 and then renormalizing them such that their sum is 1 as in figure (4a,b) [53]. Our strategy can be seen as a particular choice of Dirichlet partitioning of the interval [53]. This strategy is different from the random selection of breaking points of the interval, described, e.g., in [53, 57]. 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 $1, 2, \ldots, N$ are associated with energy levels

$$E_1, E_2, \ldots, E_N.$$  

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

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

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 (4) 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 figure (4). Let us ignore energy conservation for the moment and define a continuous uniform random variable $u$ on the interval $[0,1]$. Let $u_1, u_2, \ldots, u_\mu$ be $\mu$ independent realisations of $u$, see figure (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 $\mu$ as

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

This is the continuous version of what we described in section II.A. Now we define a new random variable, $\phi_\mu$, which is the sum of $\mu$ realisations of the random variable $u$:

$$\phi_\mu = \sum_{k=1}^{\mu} u_k .$$

The sum of $\mu$ realisations of a random variable $u$ uniformly distributed on the interval $[0,1]$, $\phi_\mu$, follows the Irwin-Hall distribution, $f_\mu(\phi_\mu)$ [53]. This means that one can construct, for each $\mu$ realisations of the random variable $u$ a rescaled sequence, see figure (4b)

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

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 $E_{i_1}, E_{i_2}, \ldots, E_{i_\mu}$ for the emerging particles, where

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

This rescaling approach assumes that the $\mu$ new particles behave independently. The crucial issue is to map this process into a cascade, see figure (4d). To approach this problem, we first study the expected number of particles that jump to a given state $E$ if a given value of $\phi_\mu$ occurs. We then average over all potential values of $\phi_\mu$. We assume that the expected number of particles from $E \rightarrow E'$, $n(E \rightarrow E')$ goes as $\sim \mu p(E'|E)$, as we did

FIG. 3: Cascade size distribution after 3,000 realizations of individual SSR cascades with $\mu = 2.5$ and $N = 10^5$ states (circles), with a fit based on the $\Gamma$ distribution $f(s)$ (red dashed line) given in equation (1). Inset: dependence of $\langle s \rangle$ on $\mu$ (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).
above. Taking into account the rescaling imposed by energy conservation, $p(E'|E) \sim \frac{\phi}{E}$, see figure (4), one has that:

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

(6)

For a given value of $\phi_\mu$, 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 $\phi_\mu$, distributed as the Irwin-Hall distribution, $f_\mu$:

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

(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 $\alpha$:

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

(8)

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

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

(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 $\mu$ small. In figure (3) 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.

III. 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
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 figure 1. An alternative view is given in figure 10 in this appendix.

Let \( u \) be a random variable whose probability density is uniform in the interval \([0, 1]\). Let \( u_1, ..., u_\mu \) be a sequence of independent drawings of the random variable \( u \) and \( \phi(\mu) \) a random variable defined over the interval \([0, \mu]\) as:

\[
\phi_\mu = \sum_{k \leq \mu} u_k .
\]  
(A1)

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

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

Differentiating,

\[
\frac{dn}{dE} = \frac{d}{dE} \int_0^\mu f_\mu(\phi_\mu) \left\{ \int_{E, E'}^{\infty} \frac{\phi_\mu}{E'} n(E')dE' \right\} d\phi_\mu
\]  
\[
= \int_0^\mu f_\mu(\phi_\mu) \frac{d}{dE} \left\{ \int_{E, E'}^{\infty} \frac{\phi_\mu}{E'} n(E')dE' \right\} d\phi_\mu
\]  
\[
= -\mu \int_0^\mu \phi_\mu f_\mu(\phi_\mu)n(\phi_\mu E)d\phi_\mu .
\]

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 \( \alpha \):

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

This is equation 8 and that is what we have to solve. The following theorem states that the solution in the limit of large \( \mu \)'s is \( \alpha = 2 \), independent of \( \mu \). To demonstrate that we need to prove 5 lemmas. After the demonstration, we approach the solution \( \alpha \to 2 \) 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 \( \alpha \) satisfying the following equation:

\[
\lim_{\mu \to \infty} \mu \int_0^\mu \phi_\mu^{1-\alpha} f_\mu(\phi_\mu)d\phi_\mu = \alpha \quad \text{(A2)}
\]

is \( \alpha = 2 \).

To prove this theorem, we first observe that a random variable following the Irwin-Hall distribution converges

![Figure 5: Histograms of SSR cascades with energy conservation](image-url)
to a random variable following a normal distribution with average $\frac{\mu}{2}$ and standard deviation $\sqrt{\frac{\mu}{12}}$.

**Lemma 1.** Let $\phi_\mu$ be 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 $E(u) = \frac{1}{2}$ and standard deviation $\sigma = \sqrt{\frac{\mu}{12}}$. By the central limit theorem, one has that, for an i.i.d. sequence of random variables $u_1, ..., u_n$:

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

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

$$\phi_\mu \rightarrow \frac{\mu}{2} + \sqrt{\frac{\mu}{12}} Y(0,1) \ ,$$
as we wanted to demonstrate. $\square$

This implies that the Irwin-Hall distribution $f_\mu$ can be fairly approached by a normal distribution with mean $\frac{\mu}{2}$ and standard deviation $\sqrt{\frac{\mu}{12}}$, $\Phi(x)$. However, one must be careful with this approach: It can lead the integral that we want to solve, $\int_0^\mu \phi_\mu^{-\alpha} f_\mu(\phi_\mu) d\phi_\mu$, 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 $\mu$’s.

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

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

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

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

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

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

Computing directly the integral, one has:

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

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

$$\int_1^1 \phi_\mu^{-\alpha} f_\mu(\phi_\mu) d\phi_\mu = \frac{1}{(\mu-1)!} \int_0^1 \phi_\mu^{\mu-\alpha} d\phi_\mu \ ,$$

$$= \frac{1}{(\mu-1)!} \frac{1}{(\mu-\alpha-1)} \ .$$

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

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

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

$$\left| \int_0^\mu \phi_\mu^{-\alpha} f_\mu(\phi_\mu) d\phi_\mu - \int_1^\mu \phi_\mu^{-\alpha} \Phi(\phi_\mu) d\phi_\mu \right| < \epsilon' + \epsilon(\mu,\delta) \ .$$
Finally, we observe that
\[
\lim_{\mu \to \infty} \epsilon(\mu, \delta) = 0 ,
\]
which demonstrates the lemma. \(\square\)

**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:
\[
\frac{d}{d\alpha} \int_1^\mu \phi_\mu^{1-\alpha} \Phi(\phi_\mu) d\phi_\mu = \int_1^\mu \left( \frac{d}{d\alpha} \phi_\mu^{1-\alpha} \right) \Phi(\phi_\mu) d\phi_\mu
\]
\[
= - \int_1^\mu \phi_\mu^{1-\alpha} \log \phi_\mu \Phi(\phi_\mu) d\phi_\mu < 0 ,
\]
since the term inside the integral, \(\phi_\mu^{1-\alpha} \log \phi_\mu \Phi(\phi_\mu)\), is strictly positive in the interval \((1, \mu)\). \(\square\)

Now take a monotonously increasing function that grows slower than the standard deviation \(\sigma_\mu = \sqrt{\frac{2}{\mu}},\) \(\varphi(\mu)\). For convenience, we define it as:
\[
\varphi(\mu) \equiv \left( \frac{\mu}{12} \right)^{\frac{1}{4}} .
\]
Clearly:
\[
\lim_{\mu \to \infty} \int_0^\mu f_\mu(\phi_\mu) d\phi_\mu = 1 .
\]

Now suppose that \(\alpha = 2\). Then, thanks to Lemma 2, one has that \((\forall \epsilon > 0) (\exists N) : (\forall \mu > N)\)
\[
\left| \int_0^\mu f_\mu(\phi_\mu) d\phi_\mu - \int_0^\mu \frac{f_\mu(\phi_\mu)}{\Phi(\phi_\mu)} d\phi_\mu \right| < \epsilon .
\]
From this we derive the third lemma of our demonstration:

**Lemma 4:** \((\forall \epsilon > 0) (\exists N) : (\forall \mu > N)\)
\[
\left| \int_1^\mu \frac{\Phi(\phi_\mu)}{\phi_\mu} d\phi_\mu - \int_1^\mu \frac{\Phi(\phi_\mu)}{\phi_\mu - \sigma_\mu \varphi(\mu)} d\phi_\mu \right| < \epsilon .
\]

**Proof:** We need to compute the parts that fall outside the integration limits and see that their contribution vanishes. First, we see that:
\[
\int_1^\mu \frac{\Phi(\phi_\mu)}{\phi_\mu} d\phi_\mu < \left( \frac{\mu}{2} + \sigma_\mu \varphi(\mu) \right) e^{-\varphi^2(\mu)} - O(\log \mu)
\]
\[
< \left( \frac{\mu}{2} + \sigma_\mu \varphi(\mu) \right) e^{-\sqrt{\pi}} .
\]
Analogously,
\[
\int_1^\mu \frac{\Phi(\phi_\mu)}{\phi_\mu} d\phi_\mu < \left( \frac{\mu}{2} + \sigma_\mu \varphi(\mu) \right) e^{-\sqrt{\pi}} .
\]
Now we define:
\[
\epsilon_1(\mu) = \left( \frac{\mu}{2} + \sigma_\mu \varphi(\mu) \right) e^{-\sqrt{\pi}},
\]
\[
\epsilon_2(\mu) = \left( \frac{\mu}{2} - \sigma_\mu \varphi(\mu) \right) e^{-\sqrt{\pi}} ,
\]
which leads to:
\[
\left| \int_1^\mu \frac{\Phi(\phi_\mu)}{\phi_\mu} d\phi_\mu - \int_1^\mu \frac{\Phi(\phi_\mu)}{\phi_\mu - \sigma_\mu \varphi(\mu)} d\phi_\mu \right| < \epsilon_1(\mu) + \epsilon_2(\mu) ,
\]
demonstrating the lemma. \(\square\)

**Corollary of Lemma 4:** \((\forall \epsilon > 0) (\exists N) : (\forall \mu > N)\)
\[
\left| \int_0^\mu f_\mu(\phi_\mu) d\phi_\mu - \int_0^\mu \frac{\Phi(\phi_\mu)}{\phi_\mu - \sigma_\mu \varphi(\mu)} d\phi_\mu \right| < \epsilon .
\]

**Proof:** By direct application of lemmas 2 and 4. \(\square\)

Now we define the following functions of the limits of the integral \(\int_\frac{\mu}{2}^{\mu + \sigma_\mu \varphi(\mu)} \): \(r_1(\mu)\) and \(r_2(\mu)\):
\[
r_1(\mu) \equiv \left( \frac{\mu}{2} + \sigma_\mu \varphi(\mu) \right)^{-1} = \frac{2}{\mu} - \frac{2\sigma_\mu \varphi(\mu)}{\mu \left( \frac{\mu}{2} + \sigma_\mu \varphi(\mu) \right)} ,
\]
\[
r_2(\mu) \equiv \left( \frac{\mu}{2} - \sigma_\mu \varphi(\mu) \right)^{-1} = \frac{2}{\mu} + \frac{2\sigma_\mu \varphi(\mu)}{\mu \left( \frac{\mu}{2} - \sigma_\mu \varphi(\mu) \right)} .
\]
Clearly,
\[
r_{1,2}(\mu) \sim \frac{r_{1,2}(\mu)}{\mu} + O \left( \mu^{-\frac{3}{4}} \right) ,
\]
where the subscript 1,2 means that both functions satisfy the property.

**Lemma 5:** \((\forall \epsilon > 0) (\exists N) : (\forall \mu > N)\)
\[
\left| r_{1,2}(\mu) \cdot \mu \int_\frac{\mu}{2}^{\mu + \sigma_\mu \varphi(\mu)} \frac{\Phi(\phi_\mu)}{\phi_\mu} d\phi_\mu - \mu \int_\frac{\mu}{2}^{\mu + \sigma_\mu \varphi(\mu)} \frac{\Phi(\phi_\mu)}{\phi_\mu - \sigma_\mu \varphi(\mu)} d\phi_\mu \right| < \epsilon .
\]

**Proof:** We first observe that, by substituting \(\phi_\mu^{-1}\) by the integration limits, we have the following chain of inequalities, in terms of the above defined functions \(r_{1,2}(\mu)\):
\[
r_2(\mu) \cdot \int_\frac{\mu}{2}^{\mu + \sigma_\mu \varphi(\mu)} \frac{\Phi(\phi_\mu)}{\phi_\mu} d\phi_\mu < \int_\frac{\mu}{2}^{\mu + \sigma_\mu \varphi(\mu)} \frac{\Phi(\phi_\mu)}{\phi_\mu} d\phi_\mu < r_1(\mu) \cdot \int_\frac{\mu}{2}^{\mu + \sigma_\mu \varphi(\mu)} \frac{\Phi(\phi_\mu)}{\phi_\mu} d\phi_\mu .
\]
Therefore, is enough to demonstrate that \((\forall \epsilon > 0)(\exists N)\) : 
\[(\forall \mu > N) \quad |\mu r_1(\mu) - \mu r_2(\mu)| < \epsilon \quad .\]
This can be proven directly from equation (A6), leading to:
\[|\mu r_1(\mu) - \mu r_2(\mu)| \sim O \left(\mu^{1/2}\right) \quad ,\]
which demonstrates the lemma.\(\square\)

Collecting lemmas 1,2,4, and 5, we have demonstrated that, under the assumption that \(\alpha = 2\),
\[\mu r_1(\mu) \rightarrow \mu \int_0^\mu \phi^{1-\alpha}_\mu f_\mu(\phi_\mu) d\phi_\mu \quad .\]
From equation (3), the only remaining issue to demonstrate the consistency of our hypothesis is that indeed
\[\lim_{\mu \rightarrow \infty} \mu r_1(\mu) = 2 \quad .\]
It is not difficult to check, from equation (A6), that:
\[\lim_{\mu \rightarrow \infty} \mu r_1(\mu) = \lim_{\mu \rightarrow \infty} \left[2 + O \left(\mu^{1/2}\right)\right] = 2 \quad .\]
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 on \(\alpha\). 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:
\[\mu \int_0^\mu \phi^{1-\alpha}_\mu f_\mu(\phi_\mu) d\phi_\mu < 2 + \beta \quad ,\]
which contradicts the statement of the theorem. The same happens if one imposes if \(\alpha = 2 - \beta\), with \(\beta > 0\), since one gets:
\[\mu \int_0^\mu \phi^{1-\alpha}_\mu f_\mu(\phi_\mu) d\phi_\mu > 2 - \beta \quad ,\]
which is, again inconsistent, thereby proving the lemma.\(\square\)

By direct application of equation (3), Lemma 5 puts the last piece to demonstrate the theorem.\(\blacksquare\)

---

[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, and 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 Zipfs 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., and 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., and 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. and 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, W. Jackson, editor, (Woburn, MA: Butterworth) (1953).
[12] Harremoës, P., and Topsøe P. Maximum Entropy Fundamentals. Entropy 3, 191 (2001).
[13] Corominas-Murtra, B., Fortuny, J., and 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., and 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., and 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., and 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(5), 6905 (2014).
[17] Paczuski, M., Maslov, S., and 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., Malthe-Sørensen, A., Feder, J., Jossang, T., and Meakin, P. Avalanche dynamics in a pile of rice, Nature, 379, 49 (1996).
[20] Sornette, A., and 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, A. Long-term clustering, scaling, and universality in the temporal occurrence of earthquakes. Phys. Rev. Lett. 92 (10), 108501 (2004).
[23] Peters, O., and Neelin, D. Critical phenomena in atmospheric precipitation. Nature Physics 2, 393 (2006).
[24] Corral, A., Osso, A., and Llebot, J. E. Scaling of tropical-cyclone dissipation. Nature Phys. 6, 693 (2010).
[25] Thurner, S., Klimek, P., and Hanel, R. Schumpeterian economic dynamics as a quantifiable minimum model of evolution New Journal of Physics 12, 075029, (2010).
[26] Boss, M., Sumner, M., and 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., and Geanakoplos, J. Lever age causes fat tails and clustered volatility. Quantitative Finance 12, 695, (2012).
[29] Hanel, R., and Thurner, S. Generalized (c,d)-Entropy and Aging Random Walks. Entropy 15, 5324-5337, (2013).
[30] Thurner, S., Hanel, R., Liu, B., and 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. II, third edition (John Wiley and Sons, New York, NY) (1968).
[34] Hanel, R., Corominas-Murtra, B. Liu, B., and Thurner, S. Fitting Power-laws in empirical data with estimators that work for all exponents PLoS ONE 12(2): e0170920, (2016).
[35] Kingman, J. F. C. Poisson processes, Oxford Studies in Probability, vol. 3 (The Clarendon Press, Oxford University Press, New York) (1993).
[36] Huillet, T. Sampling formulæ arising from random Dirichlet populations. Communications in Statistics - Theory and Methods 34 5, 1019–1040, (2005).
[37] Krapivsky, P. L., and Ben-Naim, E. Scaling and multiscale models of fragmentation. Phys. Rev. E 50, 3502–3507 (1994).

Additional information

Acknowledgments
This work was supported by the Austrian Science Fund FWF under the P29232 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.

Competing interests
The authors declare no competing financial interest.