PHYSICAL REVIEW X **12,** 021056 (2022)

Nonparametric Power-Law Surrogates

Jack Murdoch Moore,[1,†] Gang Yan,[1,2,3,*] and Eduardo G. Altmann[4,‡]

[1]*MOE Key Laboratory of Advanced Micro-Structured Materials, and School of Physics Science and Engineering, Tongji University, Shanghai 200092, People's Republic of China*

[2]*Frontiers Science Center for Intelligent Autonomous Systems, Tongji University, Shanghai, 200092, People's Republic of China*

[3]*Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, Shanghai, 200031, People's Republic of China*

[4]*School of Mathematics and Statistics, University of Sydney, Sydney, New South Wales 2006, Australia*

(Received 25 June 2021; revised 12 March 2022; accepted 17 May 2022; published 10 June 2022)

Power-law distributions are widely used in computational and statistical investigations of extreme events and complex systems. The usual technique to generate power-law distributed data is to first infer the scale exponent α using the observed data of interest and then sample from the associated distribution. This approach has important limitations because it relies on a fixed α (e.g., it has limited applicability in testing the *family* of power-law distributions) and on the hypothesis of independent observations (e.g., it ignores temporal correlations and other constraints typically present in complex systems data). Here we propose a constrained surrogate method that overcomes these limitations by choosing uniformly at random from a set of sequences exactly as likely to be observed under a discrete power law as the original sequence (i.e., regardless of α) and by showing how additional constraints can be imposed in the sequence (e.g., the Markov transition probability between states). This nonparametric approach involves redistributing observed prime factors to randomize values in accordance with a power-law model but without restricting ourselves to independent observations or to a particular α. We test our results in simulated and real data, ranging from the intensity of earthquakes to the number of fatalities in disasters.

DOI: 10.1103/PhysRevX.12.021056

Subject Areas: Complex Systems,
Interdisciplinary Physics,
Statistical Physics

I. INTRODUCTION

Heavy-tailed distributions are one of the most pronounced characteristics from complex systems [1–3], appearing in paradigmatic models in statistical physics (e.g., critical phenomena, preferential attachment processes, self-organized criticality) [4–11] and in the analysis of a variety of datasets (e.g., city sizes, word frequencies, earthquake waiting times, and magnitude of disasters) [12–15]. A key computational tool to investigate all these systems is the generation of synthetic datasets (surrogates) that account for the heavy-tail characteristic [16] of the observation or system of interest. These surrogate sequences $\mathbf{x} = x_1, \ldots, x_N$ represent null models which allow tests of properties of the data—e.g., whether the data are compatible with a specific distribution—and to make estimations—e.g., the magnitude of extreme events.

The typical approach to generate surrogates of heavy-tailed data is to first fit [18–20] a power-law distribution,

$$p(x) = Cx^{-\alpha}, \tag{1}$$

to the data (or tail, $x \geq x_{\min}$) and then use the maximum-likelihood estimated parameter $\hat{\alpha}$ to generate the synthetic dataset [21–23]. The problems with this approach, which we overcome in this manuscript, are as follows.

(i) The synthetic dataset arises from a model for the fixed parameter $\hat{\alpha}$ and thus does not represent a null model for power-law distributions in general (arbitrary α).

(ii) Correlations in the data are not accounted for [complex systems data are not independent and identically distributed (IID)] [24–26].

Figure 1 illustrates how these issues affect data analysis: for small or correlated samples there are substantial differences between the fitted model and the underlying process

---

[*] Corresponding author.
gyan@tongji.edu.cn

[†] jackmoore@tongji.edu.cn

[‡] eduardo.altmann@sydney.edu.au

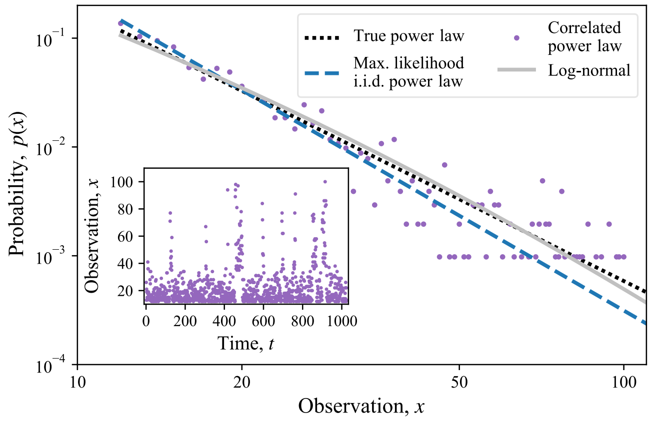

FIG. 1. Correlations and limited samples obscure power laws. A synthetic time series of length $N = 1024$ (see inset) is generated from a correlated (Markov order one) power law with exponent $\alpha = 2.5$ and lower cutoff $x_{\mathrm{min}} = 12$ (black dotted line; see Appendix A for details). The time series' empirical distribution is shown (purple filled circles), together with the maximum-likelihood fit to an IID power law (blue dashed line), and a log-normal distribution (gray solid line, location parameter $\mu = 0$ and scale parameter $\sigma = 1.5$, as defined in Table I).

(estimating scale exponents is hard) and it is difficult to distinguish between different distributions. Therefore, the two issues above directly affect the perennial debates about the ubiquity of power-law distributions: whether log-normals or power laws better describe city-size distribution [27–29], and whether power-law degree distributions are ubiquitous in complex networks [30–34] and more generally [35–38]. Progress on these fundamental debates, that lie at the foundation of complex systems research, requires methods that go beyond the typical approach.

Surrogates are numerically generated data sequences (time series) that share particular features of the original data (time series) of interest but that randomize other characteristics in accordance with a particular model (or fulfill a certain null hypothesis) [39,40]. Constrained surrogates fix particular features strictly, matching the exact value observed in the original data, thus allowing us to condition these out and draw principled conclusions about the remaining properties or features which are not being fixed and which we are interested in testing or analyzing [41,42]. Constrained surrogates can represent an entire family of processes, rather than the particular process with the parameters of best fit, and thus have a range of favorable properties explained and explored below. Our contribution in this paper is to develop and apply constrained surrogate techniques to sequences with discrete power-law distributions. We will account for different types of correlations without committing to particular values of the scale exponent α (i.e., nonparametrically) while allowing variation in the values of extreme events and other statistics of interest. We propose a constrained surrogate which redistributes observed prime factors to randomize values while fixing the likelihood of the sequence for all α and avoiding the challenging problem of estimating α. We also formulate variants which, in addition, preserve correlations (up to a given Markov order) or other constraints. We then apply these constrained surrogates to artificial and real data, showing that they lead to exact hypothesis tests, provide unbiased and lower-variance estimates of statistics, are more robust to model misspecification, and can accommodate and thus facilitate inference about correlations present in the time series. Finally, we show how the use of our constrained surrogates impacts the conclusions and estimates obtained from different heavy-tailed data. Our codes are available in Ref. [43].

II. SURROGATE METHOD

In this section, we introduce our constrained power-law surrogate methods. Let $\mathbf{x} = x_1, \ldots, x_N$ be an ordered sequence of integers, each of which is no less than a lower cutoff x_{min}. Given one such input sequence $\mathbf{x}^{\mathrm{input}}$, we are interested in generating an ensemble $\{\mathbf{x}_n\}$ of other surrogate sequences $\mathbf{x}$ constrained to the hypothesis of power-law distribution Eq. (1). A simple surrogate is the shuffle surrogate [44], which corresponds to a permutation, chosen uniformly at random, of the input sequence. Another is bootstrapping: sampling uniformly at random from the input sequence with replacement. Other important surrogates are designed to apply the null hypothesis that observations are a static transformation of a linear process. Surrogates for this hypothesis include the statically transformed autoregressive process [45], amplitude adjusted Fourier transform (AAFT) [46], and iterated AAFT [47], the last two of which we compare to our proposed surrogate methods in the Supplemental Material (SM) [48]. While these surrogates succeed in capturing features of the heavy-tailed distribution in $\mathbf{x}^{\mathrm{input}}$, a strong limitation is that they only offer values which have already been observed and therefore they cannot be used to explore unobserved cases (SM [48], Fig. S1), such as the extreme events which are particularly important in heavy-tailed processes. Constrained surrogates overcome this limitation, while showing other interesting properties not present in the usual approach based on maximum-likelihood fitting.

A. Constrained surrogates

Constrained surrogates incorporate hypotheses or constraints by fixing a set of properties $\{K\}$ to match those observed in $\mathbf{x}^{\mathrm{input}}$ such that

$$K(\mathbf{x}) = K(\mathbf{x}^{\mathrm{input}}), \qquad (2)$$

for all $\mathbf{x} \in \{\mathbf{x}_n\}$. A constrained property K could be, for example, the number of elements N of the sequence $\mathbf{x}$, the parity (odd or even) of the sum $\sum_t x_t$, or the truth value of the statement "$\max_t x_t \leq x_{\max}$," where $x_{\max}$ is known and fixed. Here we are particularly interested in the hypothesis that $\mathbf{x}^{\text{input}}$ is generated from a probability distribution with parameter(s) α and the corresponding property K to be constrained is the likelihood function $\mathcal{L}_{\mathbf{x}}$ which maps α to the probability $\mathcal{L}_{\mathbf{x}}(\alpha)$ of generating a sequence $\mathbf{x}$. The constraint Eq. (2) is then the condition

$$\mathcal{L}_{\mathbf{x}}(\alpha) = \mathcal{L}_{\mathbf{x}^{\text{input}}}(\alpha), \tag{3}$$

for all $\mathbf{x} \in \{\mathbf{x}_n\}$ and all α. Constrained surrogates $\{\mathbf{x}_n\}$ are obtained by sampling uniformly at random from a collection of (ideally all) sequences $\mathcal{U}(\mathbf{x}^{\text{input}})$ such that

(C1) $\mathbf{x}^{\text{input}} \in \mathcal{U}(\mathbf{x}^{\text{input}})$;
(C2) $\forall\, \mathbf{x} \in \mathcal{U}(\mathbf{x}^{\text{input}}), \mathcal{U}(\mathbf{x}) = \mathcal{U}(\mathbf{x}^{\text{input}})$; and
(C3) $\forall\, \mathbf{x} \in \mathcal{U}(\mathbf{x}^{\text{input}})$ and for each constrained property K, Eq. (2) is satisfied.

This procedure factors out the influence of constraints and (unknown) model parameters by conditioning on the output of the map $\mathbf{x}^{\text{input}} \mapsto \mathcal{U}(\mathbf{x}^{\text{input}})$. If we consider (i) $\mathbf{x}^{\text{input}}$ as a realization of a discrete random variable $\mathbf{X}$ drawn from a distribution with parameters α and (ii) the likelihood under the distribution to be one of the constrained properties K, then conditions C1–C3 above guarantee that the conditional probability of $\mathbf{X}$ given $\mathcal{U}(\mathbf{X})$ is independent of α—i.e., $P_\alpha(\mathbf{X}|\mathcal{U}(\mathbf{X})) \equiv P(\mathbf{X}|\mathcal{U}(\mathbf{X}))$—and uniform on $\mathcal{U}(\mathbf{X})$, in agreement with our prescription for constrained surrogates (for a proof, see SM [48], Sec. I). That is, as long as the likelihood under the null hypothesis is one of the fixed quantities, producing a constrained surrogate is equivalent to generating data under the hypothesized process subject to a condition—the particular value of $\mathcal{U}(\mathbf{X})$—which has been observed for the input sequence. In particular, by preserving the likelihood function we ensure that (maximum-likelihood) parameter estimations are the same for the input and surrogate sequences. Constrained surrogates are closely related to the idea of conditioning on a sufficient statistic [42,49–51].

Favorable properties of constrained surrogates include the following.

(i) They preserve the probability distribution of sequences and, as a consequence, provide unbiased estimates of the expectation of *any* statistic.
(ii) The expectation of any sample statistic estimated using the mean over many constrained surrogate datasets realized independently from the same observation has variance no larger than the variance of the original statistic and, as long as this variance is finite and the sample statistic is nonconstant over the collection $\mathcal{U}(\mathbf{x}^{\text{input}})$ for some $\mathbf{x}^{\text{input}}$, strictly smaller.
(iii) Constrained surrogates provide exact [52–54] (i.e., theoretically supported) hypothesis tests regardless of discriminating statistic or sample length [44]. In particular, this allows for hypothesis testing using composite [55] hypotheses and nonpivotal [60] test statistics, a requirement for testing power-law distributions (for all α) using test statistics of interest.

Property (ii) follows from a classical statistical result that gives the total variance of any sample statistic as a positively weighted sum of the variance within and between the level sets of a statistic [61], and which in our case implies, for constrained surrogates and any statistic s,

$$\mathbb{V}[s] = \mathbb{V}[\mathbb{E}_{\text{surr}}[s]] + \mathbb{E}[\mathbb{V}_{\text{surr}}[s]], \tag{4}$$

where $\mathbb{E}$ ($\mathbb{V}$) denotes the expectation (variance) over sequences from the original generative process and $\mathbb{E}_{\text{surr}}$ ($\mathbb{V}_{\text{surr}}$) denotes the expectation (variance) over surrogate sequences generated from a single input sequence $\mathbf{x}^{\text{input}}$.

B. Constrained power-law surrogates

The constrained power-law surrogate methods we propose here correspond to the null hypothesis that $\mathbf{x}^{\text{input}}$ follows a power-law distribution (1) with (an unknown) exponent α (in contrast, in the typical approach α is fixed to be equal to the maximum-likelihood estimation of α in $\mathbf{x}^{\text{input}}$). We are interested in constrained surrogates such that for any surrogate sequence $\mathbf{x} \in \{\mathbf{x}_n\}$ and for all α, the likelihood is the same, i.e., Eq. (3) is satisfied. For an IID power law governed by Eq. (1), the likelihood is

$$\mathcal{L}_{\mathbf{x}}(\alpha) = C^N \left(\prod_{t=1}^{N} x_t\right)^{-\alpha} = C^N \left(\prod_q q^{n_q}\right)^{-\alpha},$$

where n_q is the number of instances of the prime factor q in the product $\prod_{t=1}^{N} x_t$. It follows that maintaining the likelihood is equivalent to preserving the count of each prime factor which appears in the sequence. By the uniqueness of prime decompositions, when $x_{\min} = 1$, we can choose uniformly at random from among all sequences with the same likelihood by randomly assigning instances of prime factors to elements of the sequence such that, for each distinct prime factor, each possible sequence of counts is equally likely. This procedure satisfies conditions C1–C3 and so produces a constrained surrogate (Sec. II A). We illustrate the process in Fig. 2.

When $x_{\min} > 1$, it is necessary instead to choose uniformly at random from a set of sequences with the same product and each element of which is greater than or equal to the lower cutoff $x_{\min}$. This can be accomplished by (1) associating with each element of a dataset sufficiently many instances of prime factors that the element cannot dip below $x_{\min}$, (2) randomly allocating instances of prime factors among all elements not associated with any smaller prime factors, and (3) randomizing the order of the resulting dataset (see Appendix B for details of choosing uniformly at random from among all distinct distributions of prime factors by choosing a random weak integer composition into a fixed number of parts [62], generating

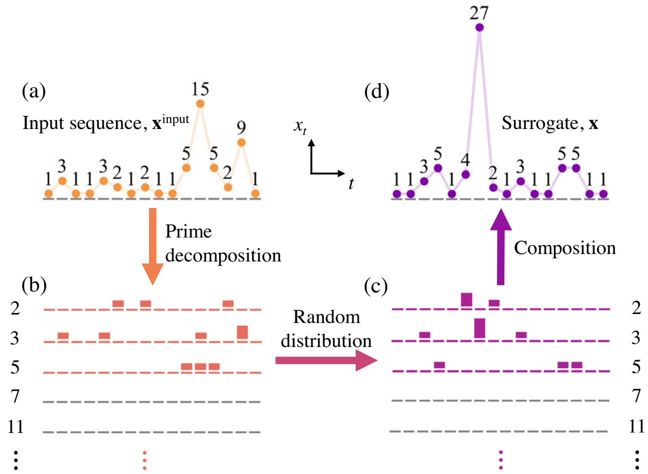

FIG. 2. Schematic illustration of the process to generate power-law surrogates. (a),(b) Each element x_t of the input sequence $\mathbf{x}^{\text{input}} = x_1, \ldots, x_N$ is separately decomposed into its prime factors. (b),(c) All observed instances of each specific prime factor are randomly redistributed in t such that each distinct distribution is equally likely. (c),(d) The redistributed prime factors are composed into a new surrogate sequence $\mathbf{x}$.

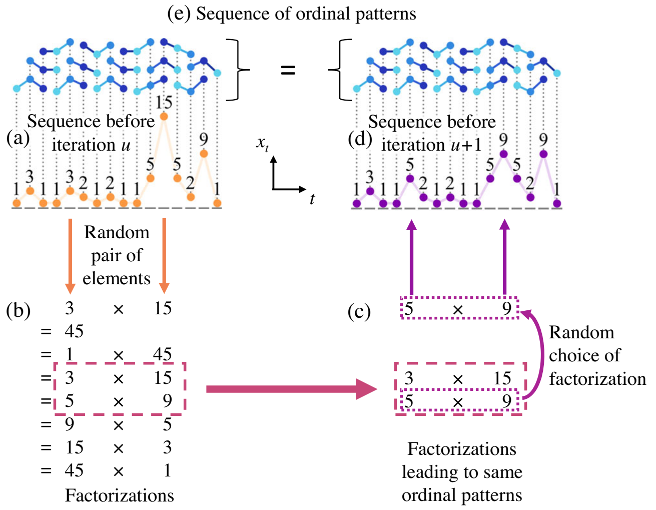

FIG. 3. Schematic illustration of the process to generate power-law surrogates with additional constraints (ordinal patterns of length $m + 1 = 3$). In each iteration u of the algorithm we (a) choose randomly a pair of elements x_i, x_j ($x_5 = 3$, $x_{12} = 15$) from the sequence $x_1, \ldots, x_N$, (b) compute all factorizations $x_i' \times x_j' = x_i \times x_j$ of the product $x_i \times x_j$ ($x_5 \times x_{12} = 45$), (c) choose a single factorization $x_i' \times x_j'$ ($x_5' \times x_{12}' = 5 \times 9$) uniformly at random from all factorizations $x_i' \times x_j'$ which preserve all ordinal patterns in the original sequence ($x_5' \times x_{12}' = 3 \times 15$, $x_5' \times x_{12}' = 5 \times 9$), and (d) update the sequence by setting $x_i = x_i'$ ($x_5 = x_5' = 5$) and $x_j = x_j'$ ($x_{12} = x_{12}' = 9$). The sequence (a)–(d) is repeated many times, with each iteration preserving (e) the same sequence of ordinal patterns.

constrained IID power-law surrogates with $x_{\min} > 1$, and fitting the lower cutoff).

C. Beyond IID

Here we show how to construct power-law surrogates which go beyond the IID hypothesis mentioned above and that consider temporal correlations of length $m \geq 1$. We build collections of surrogates $\{\mathbf{x}_n\}$ that, in addition to the power-law constraint in Eq. (3), are also constrained to have one of the following properties observed in $\mathbf{x}^{\text{input}}$:

(1) the same rank order of all size $m + 1$ subsequence $x_t, x_{t+1}, \ldots, x_{t+m}$ (length $m + 1$ ordinal patterns [63]),
(2) the same empirical transition probabilities $p(z_t | z_{t-1}, \ldots, z_{t-m})$ between states [69] constructed as nonoverlapping sets of integers (order m Markov).

We produce ordinal pattern power-law surrogates (case 1) using a Metropolis algorithm which, in the limit of a large number of transitions, samples uniformly at random from a set of sequences $\mathcal{U}(\mathbf{x}^{\text{input}}) \ni \mathbf{x}^{\text{input}}$, each of which satisfies Eq. (3) and also exhibits the same sequences of ordinal patterns as the input sequence $\mathbf{x}^{\text{input}}$. Because this prescription satisfies conditions C1–C3 (Sec. II A), it provides constrained surrogates. The algorithm begins with $\mathbf{x} = \mathbf{x}^{\text{input}}$ and for each iteration

(i) a pair of distinct observations x_i and x_j (in which $i \neq j$ but, possibly, $x_i = x_j$) is chosen uniformly at random from the sequence $x_1, \ldots, x_N$,
(ii) the observations x_i and x_j are replaced by x_i' and x_j', respectively, where $x_i' x_j' = x_i x_j$ is a factorization chosen uniformly at random from among all which would lead to a sequence which
  (A) has no element less than the lower cutoff $x_{\min}$ and
  (B) exhibits the same sequence (of length $N - m$) of ordinal patterns of length $m + 1$.

In Fig. 3 we illustrate an iteration of the Metropolis algorithm for generating constrained ordinal pattern power-law surrogates. The preceding Metropolis algorithm can be adapted to represent alternative assumptions about correlation structure. Markov power-law surrogates (case 2) are also obtained using the Metropolis algorithm above with condition (B) replaced by “exhibits the same sequence of Markov states.” Subsequently, the sequence is randomly reordered while preserving empirical transition probabilities between Markov states [70]; see Refs. [42,54,71,72]. In our studies we used 10^5 transitions of the Metropolis algorithm.

The correlated power-law surrogates defined above can be used to estimate the effective length m of correlations in a power-law sequence. To reach this, we emulate a popular framework for estimating Markov order for non-power-law discrete data [42,54,72]. We increase the hypothesized length m of correlations, starting from the IID case $m = 0$, until we cannot reject with our prespecified test statistic and level of significance. The lowest value of m for which this occurs provides a quantification of the correlation (or temporal dependencies) present in the sequence. If it appears that no value of m can prevent rejection, then a power-law model together with the type of correlations hypothesized is presumably not an appropriate explanation for the time series.

III. APPLICATIONS

In this section, we show how the methods introduced above perform in different applications. This is done by generating surrogates from the following different input sequences $\mathbf{x}^{\mathrm{input}}$:

(i) IID power laws,
(ii) IID sequences from three other distributions which can be mistaken for power laws at small sample lengths: a power law truncated at its $1/1024$ upper quantile, a power law with exponential cutoff and a discretized log-normal distribution truncated at x_{min}; see Table I,
(iii) Markov sequences with power-law limiting distributions,
(iv) real data in the form of records of fatalities from historical epidemics and terrorist attacks, as well as intensities of solar flares, energy release by earthquakes, customers affected by blackouts, and frequencies of words in the novel *Moby Dick* by Herman Melville; see Table II.

We compare constrained power-law surrogates of different Markov order with the conventional method and, where appropriate, with shuffling and bootstrapping on different hypothesis testing and estimation tasks.

A. Hypothesis testing

In hypothesis testing, a null hypothesis is rejected when the p value—i.e., the probability (under this hypothesis) of the value of a discriminating statistic (computed from the input sequence $\mathbf{x}^{\mathrm{input}}$)—is smaller than a predetermined threshold (or nominal size parameter, typically set to 0.05 or 0.1). Surrogates provide a computationally efficient procedure to perform hypothesis testing because the probability of different discriminating statistics can be estimated by computing their value in the surrogate ensemble $\{\mathbf{x}_n\}$. Here we use different discriminating statistics—the mean (an average), variance (an indicator of spread), maximum (the most extreme event observed), conditional entropies of order one and two (the conditional entropy of order $m+1$ quantifies Markov properties of order $m+1$ and is appropriate for assessing the null hypothesis that data are Markov of order m; see Appendix C), and the Kolmogorov-Smirnov (KS) distance, relative to its maximum-likelihood parameter $\hat{\alpha}$, of the part of the dataset no less than the lower cutoff (a popular way to assess goodness of fit to a power law [21])—and one-sided hypothesis tests (see Appendix C for details on the implementation of hypothesis tests). To quantify the efficiency of different surrogate methods in each such hypothesis test, we will compute two key quantities:

(i) The *size* of a test is the rate of rejection of the null hypothesis when it is indeed true (incorrect rejection). A test is *exact* when its size equals the predetermined nominal size.
(ii) The *power* of a test as the rate of rejection of the null hypothesis when this hypothesis is incorrect, which depends also on the process underlying the input sequence.

First we test the IID power-law hypothesis and, using shuffle surrogates, a more general IID hypothesis, for input sequences generated from an IID power law. In this case,

TABLE I. The probability distributions considered. In each case the constant C is defined such that the total probability mass is unity.

Name	Probability	Support	Parameters
Power law	$p(x) = Cx^{-\alpha}$	$\{x_{\mathrm{min}}, x_{\mathrm{min}}+1, \ldots\}$	$\alpha = 2.5,\ x_{\mathrm{min}} = 1,\ 12$
Truncated power law[a]	$p(x) = Cx^{-\alpha}$	$\{x_{\mathrm{min}}, x_{\mathrm{min}}+1, \ldots, x_{\mathrm{max}}\}$	$\alpha = 2.5,\ x_{\mathrm{min}} = 1,\ x_{\mathrm{max}} = 64$
Power law with cutoff (discretized[b])	$p(x) = Cx^{-\alpha}\exp(-\lambda x)$	$\{x_{\mathrm{min}}, x_{\mathrm{min}}+1, \ldots\}$	$\alpha = 2.5,\ \lambda = 0.01,\ x_{\mathrm{min}} = 1$
Log-normal (Truncated and discretized[b])	$p(x) = C[1/(x\sigma\sqrt{2\pi})]\exp[-(\log x - \mu/\sqrt{2}\sigma)^2]$	$\{x_{\mathrm{min}}, x_{\mathrm{min}}+1, \ldots\}$	$\mu = 0,\ \sigma = 1.5,\ x_{\mathrm{min}} = 1,\ 12$
Power law of Markov order m[c]	$p(x) = Cx^{-\alpha}$	$\{x_{\mathrm{min}}, x_{\mathrm{min}}+1, \ldots\}$	$\alpha = 2.5,\ x_{\mathrm{min}} = 1$
Power law with correlation (Lyapunov) time τ[c]	$p(x) = Cx^{-\alpha}$	$\{x_{\mathrm{min}}, x_{\mathrm{min}}+1, \ldots\}$	$\alpha = 2.5,\ x_{\mathrm{min}} = 1$

[a] The truncated power law is a power law truncated at its $1/1024$ upper quantile.

[b] Initially continuous random variables were rounded to the nearest integer, so that the final probability of observing an integer x is proportional to $\int_{x-0.5}^{x+0.5} p(y)\mathrm{d}y$, where p is the probability density listed above.

[c] The generation of correlated power-law sequences is detailed in Appendix A.

TABLE II. Constrained and typical power-law surrogates can lead to different conclusions about empirical data. The p values identified via tests using the KS distance, mean, variance, and maximum, and 999 surrogates of the hypothesis that the N_G elements of a dataset which exceed the fitted lower cutoff $\hat{x}_{\min}$ (Appendix B) arose IID under a power law [21]. Values which, ignoring issues associated with multiple tests, would allow rejection at the 10% level of significance [21] are shown in bold font. The lower cutoff $x_{\min}$ is fitted to the value which minimizes the KS distance between the empirical probability distribution function and the maximum-likelihood discrete power law on $x \geq x_{\min}$ [19,21,30], and surrogates are conditioned on fitted lower cutoff $\hat{x}_{\min}$ (Appendix B). When continuous datasets are discretized or converted to units of 10^3 (in the cases indicated in the footnotes to this table), we are considering power-law models for the coarsened sequences rather than the data to their original precision.

				Typical				Constrained			
Name	N	$\hat{x}_{\min}$	N_G	p_{KS}	p_{mean}	p_{Var}	$p_{\max}$	p_{KS}	p_{Mean}	p_{Var}	$p_{\max}$
Diseases[a]	72	2317	27	0.834	0.363	0.343	0.318	0.282	0.388	0.396	0.300
Blackouts[b]	211	235	57	0.883	0.435	0.390	0.476	0.557	0.369	0.466	0.581
Terrorism[c]	9,101	12	547	0.679	0.680	0.743	0.763	0.939	**0.027**	**0.023**	0.293
Flares[d]	1,711	1	1,711	**0.008**	**0.058**	**0.012**	**0.003**	**0.006**	**0.065**	**0.011**	**0.001**
Words[e]	18,855	7	2,958	0.712	0.191	**0.097**	0.141	0.346	0.161	0.122	0.169
Earthquakes[f]	59,555	1	59,555	**0.001**	0.798	0.752	0.714	**0.003**	0.913	0.866	0.840

[a] "Diseases" corresponds to average estimates in units of 10^3 of fatalities due to historical epidemics, rescaled as outlined in Ref. [37].

[b] "Blackouts" comprises the numbers of customers, in 10^3 and rounded to the nearest integer, affected by electrical blackouts in the United States between 1984 and 2002 [1,21]: Ref. [73].

[c] "Terrorism" lists the number of deaths as a direct result of terrorist attacks which took place between February 1968 and June 2006 [74]: Ref. [75].

[d] "Flares" lists the peak gamma-ray intensity of solar flares, in counts per second, made from a particular satellite between 1980 and 1989 [1]: Ref. [76]. Intensities less than 323 counts per second were discarded [21], then the time series was divided by twice this lower cut-off, to obtain a sequence of real numbers bounded below by 0.5. Finally, each element was rounded to the nearest integer, rounding up when two integers were equally close.

[e] "Words" comprises the count of unique words in the novel *Moby Dick* by Herman Melville [1,21]: Ref. [77].

[f] "Earthquakes" records the approximate energies released by the 59 555 earthquakes of magnitude at least 2.0 [78] detected in southern California between 1981 and 2000 [79]: Ref. [80]. Each earthquake magnitude M was converted to an approximate energy E, in Joules, using the formula [81]

$$\log_{10} E = 5.24 + 1.44M.$$

These energies were divided by twice the energy required for an earthquake of magnitude 2.0, to obtain a sequence of real numbers bounded below by 0.5. Finally, each element of this sequence of energies was rounded to the nearest integer.

the distribution of p values is expected to be flat [82,83]. Figure 4 shows this is obtained for constrained power-law surrogates, regardless of discriminating statistic or sample length, but not for typical power-law surrogates. Even when the KS distance is used as a discriminating statistic, as recommended in Ref. [21], typical surrogates lead to small but statistically significant deviations from uniformity which are particularly relevant for small sample sizes N. Figure 5 confirms these results by showing how the size of different tests scale with N. The constrained power-law surrogates have an exact size for all discriminating statistics, while the typical approach shows pronounced deviations from the desired nominal value. When the conditional entropy of order one is used as a discriminating statistic, the traditional approach shows a deviation from uniformity that even increases with N. Differences between true and nominal size arise for typical surrogates because the method allows large variation in the parameter of best fit which, in turn, lead to excessive variability in discriminating statistics [44].

As a next test we consider sequences which actually arose from an IID log-normal distribution, and compare the rate at which hypothesis tests based on typical and constrained surrogates can rule out the hypothesis of an IID power-law origin. Figure 6 shows that when the KS distance is used as a statistic, typical surrogates provide slightly greater power, but that for other (nonpivotal) statistics constrained surrogates often lead to similar or higher power, as reported more generally in Ref. [44] for nonpivotal statistics. Most importantly, constrained surrogates seem to perform systematically better in the crucial case of small sample size N. In the limit of small sample length N, the power of tests based on typical surrogates can approach zero, but the power arising from constrained surrogates approaches the nominal size.

Finally, we apply hypothesis tests to heavy-tailed empirical data recording deaths from historical epidemics, the numbers of customers affected by blackouts, word frequencies in the novel *Moby Dick*, the number of deaths as a direct result of terrorist attacks, intensities of solar flares, and energies released by earthquakes. In Table II we show that the p values associated with the hypothesis that real data are IID power law above a lower cutoff can vary considerably with statistic and surrogate method. The

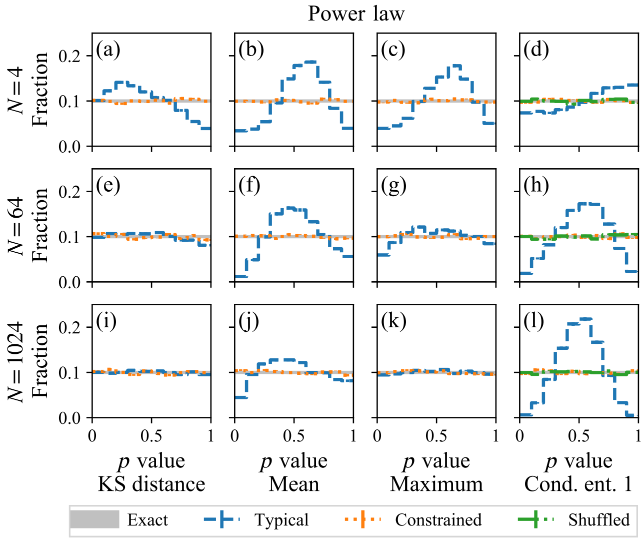

FIG. 4. Constrained surrogates lead to flat distributions of p values. Distributions of p values estimated from 10 000 hypothesis tests each using an independent sample of length $N = 4$, 64, or 1024 from an IID power law with scale exponent $\alpha = 2.5$ and lower cutoff $x_{\min} = 1$. Each test uses nine typical, constrained, or shuffled surrogates, with nominal size 10%, and uses as test statistic the KS distance, mean, maximum, or conditional entropy of order one (from left to right). The gray line spans one standard error above and below the expected fraction of p values in each interval, and error bars of other lines correspond to standard error but are smaller than the linewidth.

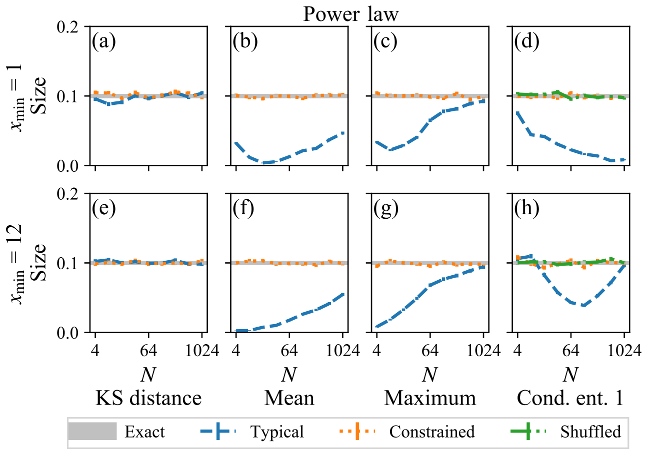

FIG. 5. Constrained surrogates lead to exact tests (true size equals nominal size), regardless of sample length or discriminating statistic. Sizes estimated from 10 000 hypothesis tests, each using an independent sample of length N from an IID power law with scale exponent $\alpha = 2.5$ and lower cutoff $x_{\min} = 1$ or 12. Tests use nine typical, constrained, or shuffled surrogates, with nominal size 10%, and use as test statistics the KS distance, mean, maximum, or conditional entropy of order one (from left to right). The gray line spans one standard error above and below the nominal size, and the error bars of other lines correspond to standard error but are at most only slightly larger than the linewidth.

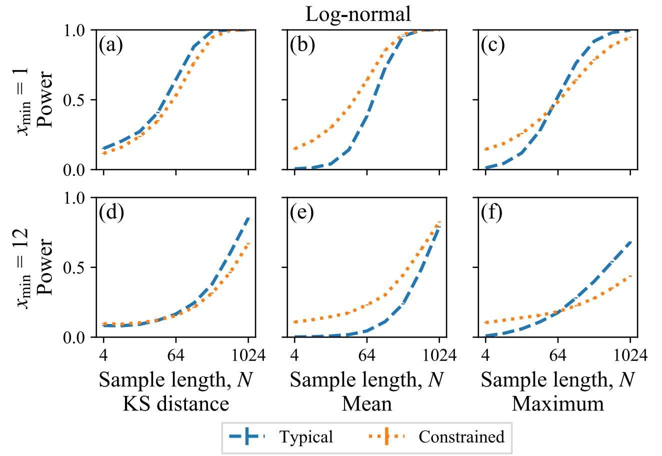

FIG. 6. Constrained surrogates can increase power for small sample lengths. Powers estimated from 10 000 hypothesis tests (power-law null hypothesis) each using an independent sample of length N from an IID log-normal distribution with location parameter $\mu = 0$, shape parameter $\sigma = 1.5$, and lower cutoff $x_{\min} = 1$ or 12. Tests use nine typical or constrained surrogates, with nominal size 10%, and use as discriminating statistics the KS distance, mean, or maximum (from left to right). Error bars show standard error and are at most only slightly larger than the linewidth.

difference between surrogate methods is especially noticeable for the "Terrorism" data when the mean or variance is used as a discriminating statistic.

B. Estimation

Now we show the advantages of using constrained surrogates to estimate quantities of interest. We are particularly interested in extreme events (e.g., expected sample maxima) because of their significance in processes with heavy-tailed distributions such as power laws and, throughout this section, employ statistics sensitive to the tail of the distribution. In contrast to shuffling and bootstrapping, constrained surrogates and the typical approach allow for the estimation of the probability of unobserved (extreme) events. The benefit of constrained surrogates over the typical approach is that they avoid biasing estimations of extreme events which, as we will see below, is particularly relevant for small N.

First, we consider independently generating power-law sequences, and for each surrogate sequence estimating the sample maximum, ratio of the two largest values, and index of dispersion (the ratio of variance to mean [84]). For each statistic s we also calculate the relative bias $(\tilde{\mathbb{E}}[s] - \mathbb{E}[s])/\mathbb{E}[s]$ in the expectation $\tilde{\mathbb{E}}[s]$ computed using the considered surrogate method. Figure 7 shows that constrained surrogates provide an unbiased estimate in all cases, in contrast to bootstrapping and the typical approach. This advantage is particularly important in the relevant case of small sample size N (e.g., the expectation of sample maximum

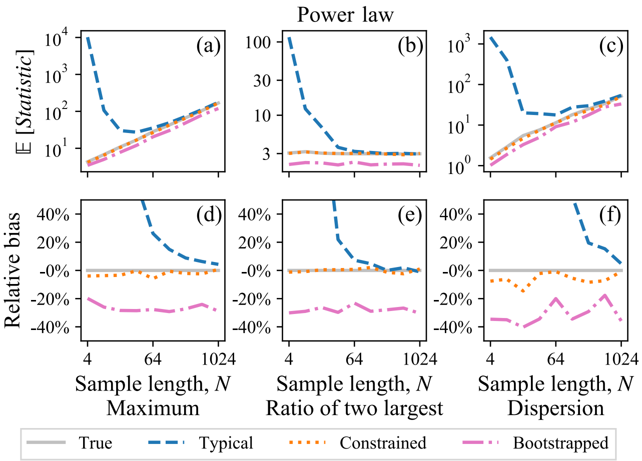

FIG. 7. Constrained surrogates do not bias sample statistics. (a)–(c) Expectation $\mathbb{E}$ of sample statistics and (d)–(f) relative bias of the calculated expectation. The expectation $\mathbb{E}$ is calculated using 10 000 independent input sequences $\mathbf{x}^{\mathrm{input}}$, each comprising N values drawn IID from a power-law distribution with $\alpha = 2.5$, using each input sequence to generate 100 surrogates. The surrogates considered are either generated independently from the true underlying process (solid gray), typical, constrained, or bootstrapped. Surrogates have the same length N as the original sequence (horizontal axis). Results for the sample maximum, ratio of the two largest observations, and index of dispersion are shown in each column and were computed in sequences of length N.

is much larger when based on typical surrogates, and smaller when based on bootstrapping). Constrained power-law surrogates also reduce finite variance, as predicted by Eq. (4), without introducing bias (see SM [48], Fig. S3, where, in addition to statistics investigated herein, we consider a measure of inequality or heterogeneity called the coefficient of variation [85]).

Now we consider time series which are not power-law distributed, but which could reasonably be misidentified as such at small sample length (log-normal distribution, truncated power law, or power law with exponential cutoff; see Table I). Although the model will not be optimal for these time series, in the absence of knowledge of the true underlying process decision makers may decide to apply power-law surrogates, especially when trying to allow for the possibility of extreme events. Figure 8 shows that constrained surrogates also improve estimates of the expectation of statistics in this context. In particular, the overestimation of the sample maxima is not as extreme as the one obtained using the typical approach. For these non-power-law distributions, bootstrapping, which does not apply a power-law model, provides estimates of maxima closer to the true expected maximum than estimates arising from either typical or constrained power-law surrogates. However, bootstrapping cannot produce new values, with the consequence that resulting maxima are always less than or equal to the maximum of the original observation.

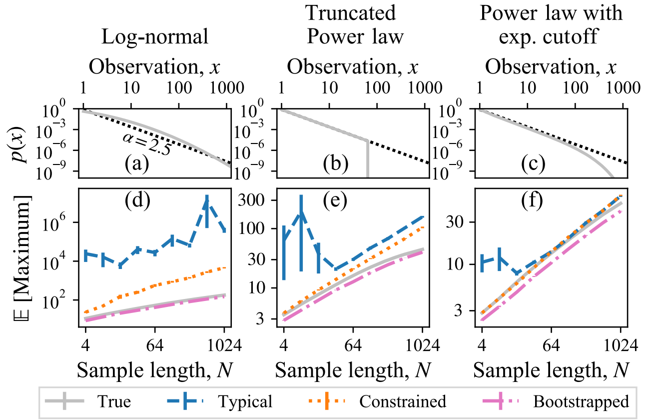

FIG. 8. Constrained surrogates ameliorate overestimation of the expected maximum arising from misapplication of a power-law model. The expected maximum in N samples was estimated as the mean over 10 000 independent realizations of input sequences $\mathbf{x}^{\mathrm{input}}$ drawn IID from three distributions (left to right): log-normal, power law truncated at its $1/1024$ upper quantile, and a power law with an exponential cutoff. In the upper panels (a)–(c), the probability mass functions $p(x)$ of these distributions (solid gray line) is compared with that of a power law with scale exponent $\alpha = 2.5$ and lower cutoff $x_{\min} = 1$ (densely dotted black line). In the lower panels (d)–(f), the true values of the expected maxima are compared to the ones estimated using three surrogate methods: typical, constrained, and bootstrap. Error bars show standard error.

Finally, we consider again heavy-tailed empirical data of varied origin. In Fig. 9, we compare the predictions made by typical and constrained power-law surrogate methods of the expected maximum of random subsamples of varied length N. This includes estimates under an IID power-law model of the expected maximum number of people killed in a single event among the next N epidemics or terrorist attacks. Typical surrogates often produce estimates of expected sample maxima which are alarmingly large. Conversely, bootstrapping, because it cannot provide unobserved values, systematically underestimates the expectation of sample maximum. Constrained surrogates provide a compromise which avoids systematic underestimation but leads to sample maxima which, in expectation, are usually closer to the expected maximum of samples drawn IID from the original data than are the sample maxima of typical surrogates (though not as close as the systematic underestimates available from bootstrapping).

C. Correlated data

In this section, we explore power-law and empirical datasets that are not IID We show that correlations impact sample statistics and hypothesis tests based on traditional surrogates, but can be accommodated by constrained surrogates. We focus consistently on the possibility that empirical data are correlated, rather than the alternative

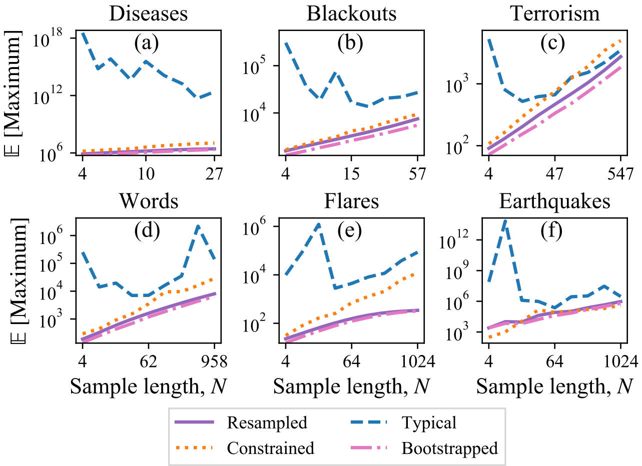

FIG. 9. Constrained surrogates avoid extreme values of expected maximum when applying a power-law model to empirical observations (Table II). Expectations of sample maximum estimated as the mean over 10 000 independent samples of length N drawn IID from the original data (solid purple) or a typical, constrained, or bootstrap surrogate produced from this sample. Only values of the original data which equal or exceed the fitted lower cutoff $\hat{x}_{\min}$ are considered.

explanation of deviations from IID which could be provided by nonstationarity.

First, in Fig. 10, we consider data with a stochastic component: these are indeed power law, but also Markov of order one or two (Appendix A). Whether the KS distance or a conditional entropy is used as a discriminating statistic, typical and constrained power-law surrogates—which enforce an IID power-law hypothesis—both lead to high rates of rejection of the power-law hypothesis (at least, for long samples). This rejection occurs not because the sequences are not power law, but because they are not IID. The same is true of shuffle surrogates, designed to apply an IID hypothesis without utilizing a power-law model, when the conditional entropy is used as a discriminating statistic. Constrained Markov order power-law surrogates, designed to enforce the null hypothesis that an observed sequence is power law and Markov of order one with a given set of Markov states, lead to different behavior: constrained Markov order power-law surrogates avoid inappropriate (and provide appropriate) rejections. When the original power-law data are Markov of order one, a constrained power-law Markov order one surrogate (with correctly chosen Markov states) leads to a rate of rejection which closely matches the nominal size of the test. When the original power-law sequence is Markov of order two, and the constrained power-law Markov order one surrogate is used, rates of rejection are, once again, close to the nominal size when either the KS distance or conditional entropy of order one is used as a discriminating statistic. This result is reasonable, because the KS distance and conditional entropy of order one are not designed to be sensitive to Markov properties of order greater than one. When the conditional entropy of order two, which is sensitive to order two Markov properties, is used as a discriminating statistic, the rate of rejection correctly approaches unity as the sample length increases. Constrained Markov order power-law surrogates maintain similar advantages across a wider range of discriminating statistics and competing surrogate methods than shown here (see SM [48], Fig. S4).

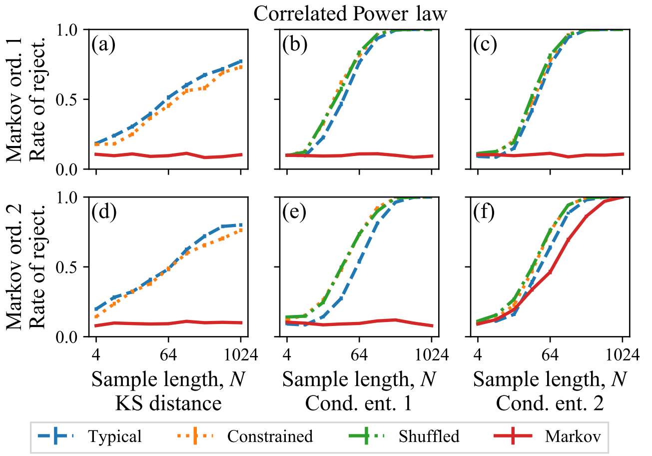

FIG. 10. Constrained Markov order power-law surrogates can reduce correlation-based rejection of power-law hypotheses. Rates of rejection estimated from 1000 hypothesis tests, each using a sample of length N from a power law which has scale exponent $\alpha = 2.5$, lower cutoff $x_{\min} = 1$, and Markov order (a)–(c) one or (d)–(f) two (Appendix A). Tests use nine typical, constrained, shuffled, or constrained Markov order one power-law surrogates, have a nominal size 10%, and use as test statistics the KS distance or conditional entropy of order one or two. Note that the conditional entropy of order $m + 1$ is an appropriate statistic for testing for Markov order m (Appendix C). Markov power-law surrogates use the same Markov states as the original time series. Error bars show standard error and are at most only slightly larger than the linewidth.

Next, in Fig. 11, we use IID and constrained ordinal pattern power-law surrogates to investigate power-law data of deterministic chaotic origin, having correlation time (Lyapunov time) $\tau = 5.0$, 10.2, and 17.9 seconds per natural unit of information (sec/nat) (Appendix A). We apply hypothesis tests which employ the discriminating statistic most widely used in tests of a power-law null hypothesis: the KS distance. As correlation time τ increases, the considered power-law sequences deviate more from IID, leading to increases in the rejection rates arising from both typical and constrained IID power-law surrogates. In contrast, constrained ordinal pattern power-law surrogates can accommodate correlations. The length $m + 1$ of ordinal patterns which must be preserved to avoid rates of rejection above the nominal size increases as the correlation time τ grows, showing how ordinal pattern power-law surrogates can be used to resolve correlation structure in observed data.

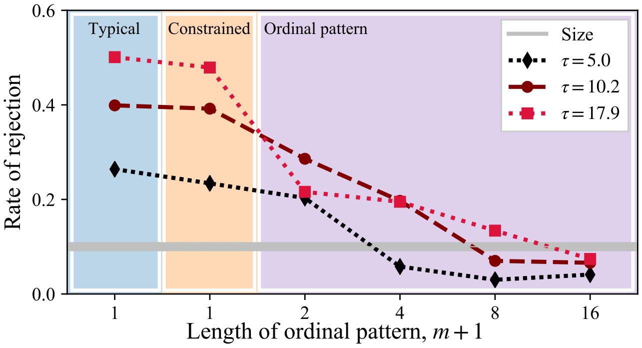

FIG. 11. By constraining longer ordinal patterns we can accommodate longer correlations in power-law data. Rates of rejection estimated from 1000 hypothesis tests with KS distance as discriminating statistic. Each test uses a sample of length $N = 1024$ from a power law which has scale exponent $\alpha = 2.5$ and Lyapunov time τ equal to 5.0 (black diamonds with densely dotted lines), 10.2 (maroon circles with dashed lines), and 17.9 sec/nat (red squares with dotted lines) (see Appendix A). Tests use nine typical, constrained, or constrained length $m + 1$ ordinal pattern power-law surrogates, have a nominal size 10%, and use the KS distance as test statistic. The gray line spans one standard error above and below the nominal size. Error bars show standard error, but are smaller than the marker size.

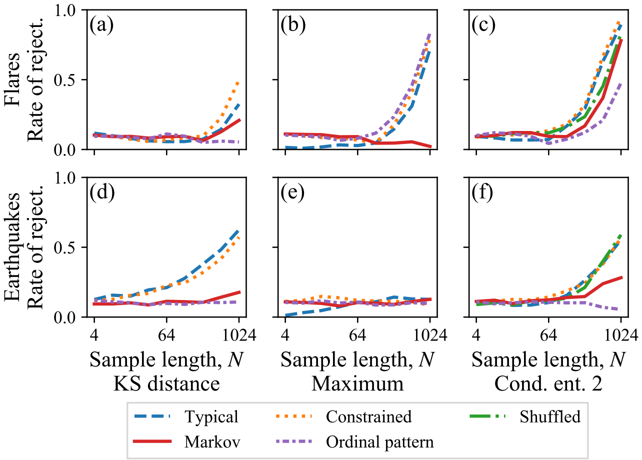

FIG. 12. Incorporating correlation constraints impacts the rate of rejection of power-law hypotheses for empirical observations. Rates of rejection estimated from 1000 hypothesis tests of sequences of length N of (a)–(c) intensities of solar flares and (d)–(f) energy released by earthquakes (Table II). Each input sequence $\mathbf{x}^{\mathrm{input}}$ considered begins at an independently and randomly chosen point in the full empirical sequence. Tests use nine typical IID, constrained IID, shuffled, constrained Markov order one, or constrained length 16 ordinal pattern surrogates, have nominal size 10%, and employ as discriminating statistics the KS distance, maximum, or conditional entropy of order two. Note that the conditional entropy of order $m + 1$ is an appropriate statistic for testing for Markov order m (Appendix C).

Finally, in Fig. 12 we illustrate how the choice of surrogate method impacts the conclusions we make about the correlated empirical datasets comprising sequences of solar flare intensities and earthquake energy release. Employing a constrained Markov order one power-law surrogate instead of a typical or constrained IID power-law surrogate consistently leads to low rates of rejection of a power-law hypothesis on the basis of sample maximum or the KS distance. However, the type of order one Markov constraints which we consider do not explain the observed order two Markov properties (captured by the conditional entropy of order two) of sequences of solar flare intensities. In contrast, for sequences of earthquake energies these Markov constraints substantially decrease rates of rejection when using the conditional entropy of order two and so appear to be able to at least partly explain the observed memory properties. Constrained length $m + 1 = 16$ ordinal pattern power-law surrogates, which enforce the null hypothesis that sequences are power law but have correlations which can be captured by ordinal patterns of length $m + 1 = 16$, also lead to low rates of rejection based on the KS distance (for ordinal patterns of other lengths, other discriminating statistics, and results for another empirical dataset, see SM [48], Figs. S5 and S6). However, for solar flare intensities, using ordinal pattern surrogates only slightly reduces the rate of rejection when the conditional entropy of order two is used as a discriminating statistic, and actually increases the rate of rejection when using the maximum. When correlations in earthquake intensities are accommodated by constraining ordinal patterns, the discriminating statistics considered consistently lead to low rates of rejection of a power-law model.

IV. CONCLUSIONS

Generating sequences with heavy-tailed distributions is a critical step in quantitative investigations of complex systems. The traditional parametric approach, based on fitting a power law to the data, has the important limitations that it leads to sequences from a single power-law exponent α and ignores correlations in the data. In this paper, we proposed nonparametric methods to obtain constrained power-law surrogates which overcome these limitations by not restricting α and by accommodating correlations in the data up to a predefined length m. We explored the benefit of our surrogates over alternative approaches (shuffling, bootstrapping, and the typical power-law fitting approach) in a variety of settings and datasets. Our approach leads to uniformly distributed p values, exact hypothesis tests, and unbiased estimates of expectation regardless of the sample statistics. The benefits are particularly important for small sample lengths N, which is the most important regime because this is when the determination of the validity of power-law distributions is challenging. This regime is also relevant in large datasets because of the reduction of the effective sample size N

that occurs when a power law is fit only to the tail of a distribution [21,30] and when sequences are downsampled to reduce correlations [26].

The significance of our methods and findings is that they provide an improved methodology to address problems recently raised in the long-standing debates on the ubiquity of power-law distributions in complex systems. First, it has recently been shown that ignoring correlations, which are ubiquitous in complex systems and affect statistical properties of time series, leads to wrong conclusions about the validity of power-law distributions [26]. The usual methods do not account for correlations, but our constrained surrogates do. Second, it is becoming increasingly accepted that whether a distribution is a power law is less important than whether it has a heavy or fat [86] tail, and that it is important to move beyond goodness-of-fit statistics [35,36]. Constrained power-law surrogates align with this view because (i) they provide theoretical support for arbitrary discriminating statistics and, hence, a license to investigate whichever types of deviation from power-law behavior are most likely to affect the best course of action, and (ii) they provide estimates substantially more accurate than the typical approach when the underlying distribution is lognormal rather than power law, and we would expect similar advantages to hold for other heavy-tailed distributions. Third, the application of our constrained surrogates modifies conclusions obtained in data analysis, as shown in Fig. 12 for the case of the analysis of earthquake datasets; the application of the usual approaches leads to a rejection of the power-law hypothesis (Gutenberg-Richter law), but the incorporation of correlations within our constrained surrogate approach shows that this hypothesis cannot be rejected at a rate significantly higher than the nominal size of the test.

Constrained power-law surrogates can preserve arbitrary additional properties present in the original sequence, but precisely which characteristics should be preserved depends on cases of interest. Further research is needed to overcome limitations of our work and expand the results to new settings of interest. In the setting of networks, analysis of whether the degree distribution is power-law distributed (scale-free hypothesis) needs to account for constraints related to the formation of networks (e.g., impose that the degree sequence of a simple undirected graph must be graphical [87,88] and account for the process of network creation [33,89]). It would be interesting to investigate what effect these additional constraints would have on conclusions made about the ubiquity of scale-free networks [30]. Another limitation of our results is that they apply to discrete power laws and originally continuous datasets have to be rounded. It would be valuable to formally extend results to the continuous setting and to systematically study the impact of discretization (rounding)—common steps in data collection, interpretation, and analysis [90]—on conclusions and estimates about heavy-tailed datasets. For instance, relationships which are theoretically assured for discrete data might hold only approximately for continuous power laws. For this reason, and also because the additional analytic properties of real continuous data could provide new opportunities [41,45,91–93], it would be valuable separately to develop constrained continuous power-law surrogates which can be used to explore unobserved events. Finally, in order to support the use of constrained surrogates in estimating confidence intervals, it would be important to account systematically for our finding of lower variance of constrained power-law surrogates (when compared with the IID case).

ACKNOWLEDGMENTS

J. M. M. and G. Y. are supported by National Natural Science Foundation of China (Grants No. 11875043, No. 12150410309, and No. 12161141016), Shanghai Municipal Science and Technology Major Project (Grant No. 2021SHZDZX0100), Shanghai Municipal Commission of Science and Technology Project (Grant No. 19511132101), and the Fundamental Research Funds for the Central Universities (Grant No. 22120190251). We thank Michael Small, Thomas Jüngling, Débora Cristina Corrêa, Yinqi Xuan, and anonymous reviewers for valuable input.

APPENDIX A: CORRELATED POWER-LAW SEQUENCES

Here we describe how we generate input sequences $\mathbf{x}^{\text{input}}$ which are power-law distributed, but not IID. We consider two cases: (1) a stochastic generative process of Markov order $m > 0$ and (2) a deterministic chaotic system with tunable correlation time τ.

1. Stochastic process

Sequences from power laws of Markov order $m > 0$ are obtained using as Markov states the partition $\mathcal{A} = \{a_i | i \in \mathbb{Z}^+\}$ of the set of integers greater than or equal to the lower cutoff $x_{\min}$ which comprises intervals of equal logarithmic width defined by $a_i = \{k \in \mathbb{Z} | 3^{i-1}(x_{\min} - 0.5) \leq k < 3^i (x_{\min} - 0.5)\}$. The first m elements $x_1, \ldots, x_m$ of the sequence are generated IID from the limiting power-law distribution. For each integer time $t > m$, with probability ν the element x_t is also generated IID from the limiting power-law distribution. Otherwise, the element x_t is instead generated by first choosing one of the previous m values of x (say, x_*), identifying the interval a_{i_*} such that $x_* \in a_{i_*}$, and choosing the value x_t randomly within a_{i_*} according to the corresponding power-law distribution:

$$p(x|x \in a_{i_*}) = \begin{cases} x^{-\alpha}/\sum_{y\in a_{i_*}} y^{-\alpha} & x \in a_{i_*} \\ 0 & x \notin a_{i_*}. \end{cases}$$

We now check that this algorithm leads to the desired limiting distribution in the case of Markov of order one (the same reasoning applies for higher order). The probability $p_{t+1}(a)$ of Markov state $a \in \mathcal{A}$ at time $t+1$ given the probability $p_t(a)$ at time t is

$$p_{t+1}(a) = \nu p_{\text{limiting}}(a) + (1-\nu)p_t(a), \qquad \text{(A1)}$$

where p_{limiting} is the desired limiting distribution of Markov states and ν is a parameter governing the probability of choosing the next value according to p_{limiting}. Iterating Eq. (A1) we find that, for $a \in \mathcal{A}$ and non-negative integer t,

$$p_t(a) = \nu\left(\sum_{j=0}^{t}(1-\nu)^j\right)p_{\text{limiting}}(a) + (1-\nu)^t p_0(a),$$

and so $\lim_{t\to\infty} p_t(a) = p_{\text{limiting}}(a)$ for any $0 < \nu \leq 1$. In our case, p_{limiting} is a power law with scale exponent $\alpha = 2.5$ and lower cutoff $x_{\min} = 1$, and $\nu = 0.1$.

2. Deterministic chaotic system

We generate power-law sequences with tunable correlation time τ using a one-dimensional deterministic chaotic system. This is achieved by first generating a sequence from a map with a tunable correlation (Lyapunov time) and then applying a transformation of variables that maps the invariant density to a power-law distribution. We use the asymmetric tent map $f\colon [0,1] \to [0,1]$ given by

$$w_{t+1} = f(w_t) = \begin{cases} r^{-1}w_t & 0 \leq w_t \leq r \\ (1-r)^{-1}(1-w_t) & r < w_t \leq 1, \end{cases}$$

where $r \in [0,1]$ is a parameter, and iterate starting from a point $w_1 \in [0,1]$ chosen uniformly at random to produce a sequence $\mathbf{w} = w_1, \ldots, w_N$ of length N. This map is uniformly hyperbolic with a Lyapunov exponent given by [94]

$$\lambda = |r\log r + (1-r)\log(1-r)|.$$

The inverse of the Lyapunov exponent $\tau = 1/\lambda$ provides a measure of the correlation (memory) in the sequence. We consider parameters $r = 0.95$, 0.98, and 0.99, corresponding to $\tau = 5.0$, 10.2, and 17.9 sec/nat, respectively (the Lyapunov exponent has units of nat/sec because we use natural logarithms). The invariant density of the map is uniform in [0, 1] [95], which allows us to obtain a correlated power-law sequence $\mathbf{x} = x_1, \ldots, x_N$. This is done mapping each value $w_t \in [0,1]$ to an integer x_t no smaller than the desired lower cutoff $x_{\min}$ via the inverse of the cumulative distribution function of a discrete power law with the desired scale exponent α.

APPENDIX B: SURROGATE ALGORITHM

In this appendix, we detail our method for producing constrained power-law surrogates with arbitrary lower cutoff $x_{\min} \in \mathbb{Z}^+$ (the code is available in Ref. [43]). Given an input sequence $\mathbf{x}^{\text{input}} = x_1, x_2, \ldots, x_N$, let $x_t = \prod_{j=1}^{r_t} q_{t,j}$ be the prime decomposition of element x_t, where $q_{t,1} \geq q_{t,2} \geq \cdots \geq q_{t,r_t}$ and r_t is the total number of instances of prime factors in x_t. As mentioned in the main text, our strategy involves three steps.

1. Associate with each element $x_t = \prod_{j=1}^{r_t} q_{t,j}$ the k_t prime numbers $q_{t,1}, q_{t,2}, \ldots, q_{t,k_t}$, where k_t is chosen such that $\prod_{j=1}^{k_t} q_{t,j} \geq x_{\min}$ but, for $l = 1, 2, \ldots, k_t - 1$, $\prod_{j=1}^{l} q_{t,j} < x_{\min}$. Our strategy involves keeping these k_t instances of prime factors with element x_t, and thus ensuring that x_t does not decrease beneath the lower cutoff $x_{\min}$ or, indeed, beneath its personal lower cutoff $x_{t,\min} = \prod_{j=1}^{k_t} q_{t,j} \geq x_{\min}$. For each element x_t, note the maximum admissible prime factor q_{t,k_t}.
2. Randomly allocate all instances of prime factors which are not associated with an element of the sequence. For each distinct prime q, all instances of this prime factor which are not associated with a particular element of the input sequence $\mathbf{x}^{\text{input}}$ are distributed among all elements of the sequence for which the maximum admissible factor is no smaller than q (i.e., all elements x_t for which $q_{t,k_t} \geq q$). So that surrogates are drawn uniformly at random from a set of sequences which have the same product and the same likelihood, the instances of the prime factor q are distributed such that all distinct redistributions are equally likely. Because an element x_t will receive no additional instances of prime factors $q > q_{t,k_t}$, its personal lower cutoff $x_{t,\min} = \prod_{j=1}^{k_t} q_{t,j}$ will not change. Furthermore, each distinct time series with the same likelihood and the same sequence of personal lower cutoffs $x_{1,\min}, x_{2,\min}, \ldots, x_{N,\min}$ is equally likely.
3. Randomly permute the time series so that more distinct sequences are accessible. Whether this permutation is performed in the first or final stage of the surrogate algorithm does not affect the probability distribution of sequences.

The algorithm just described involves ordering prime factors from largest to smallest. Any other ordering of the prime numbers would be valid, but would lead to a distinct surrogate algorithm. Our choice preferentially fixes large prime factors and allows smaller factors (which we expect to tend to be more numerous) to be redistributed.

Choosing uniformly at random from all distributions of k instances of a prime factor among N elements is equivalent to choosing a random weak composition of k into N parts, for which we use the algorithm RANCOM [62]. We follow Ref. [96] in choosing a random $(N-1)$ subset using the $O(N)$ algorithm RKS2 instead of the $O(N \log N)$ algorithm RKSB originally employed as a subroutine of algorithm RANCOM [62]. The computational times required for constrained and typical IID power-law surrogates both scale almost linearly in sample length N (see SM [48], Fig. S7).

When either constructing constrained IID power-law surrogates with $x_{\min} \geq 2$ or generating constrained correlated power-law surrogates, we would expect our methods to benefit from the presence of more instances of prime factors, because this should allow more randomization.

As stated in the main text, the lower cutoff $x_{\min}$ describing an empirical dataset of length N is chosen as the value $\hat{x}_{\min}$ which minimizes the KS distance between the empirical distribution for the N_G values above the lower cutoff and the corresponding maximum-likelihood discrete power law on $x \geq x_{\min}$ [19,21,30]. This fitted lower cutoff is actually treated in three slightly different ways depending on the surrogate methods considered. (1) We fit the lower cutoff $x_{\min}$ only once for each empirical dataset, using all available data. Thereafter we work with (subsamples or subsequences of) the sequence of elements no less than this fitted lower cutoff $\hat{x}_{\min}$, and treat the previously fitted value $\hat{x}_{\min}$ as the known value of the lower cutoff $x_{\min}$. In particular, the KS distance is calculated using this previously fitted $\hat{x}_{\min}$. This approach has been applied in all cases, except in Table II. (2) Following Ref. [21], each typical surrogate considered in Table II comprises a sequence of N values chosen IID, as follows. With probability N_G/N, the value is drawn IID from the maximum-likelihood power law for the N_G values no less than the fitted lower cutoff $\hat{x}_{\min}$. Otherwise [i.e., with probability $(N - N_G)/N$], the value is drawn IID from the $(N - N_G)$ elements which fall below the fitted lower cutoff $x_{\min}$. The KS distance is calculated only after refitting the lower cutoff to the surrogate sequence [21]. (3) Each constrained surrogate of Table II is generated while constraining the fitted lower cutoff $\hat{x}_{\min}$ as well as the maximum likelihood value $\hat{\alpha}$ of the scale exponent α. Each such surrogate comprises the values in the original observation which fell below the fitted lower cutoff $\hat{x}_{\min}$, together with a constrained power-law surrogate generated from the observations with values above the fitted lower cutoff. To constrain the fitted lower cutoff $\hat{x}_{\min}$, we accept a surrogate only when it leads to the same fitted value of $\hat{x}_{\min}$ as does the original observation, but in other cases discard the result and repeat the surrogate generation process. Conditioning on the fitted lower cutoff $\hat{x}_{\min}$ increases computational cost, which is why we consider this approach only in Table II.

APPENDIX C: IMPLEMENTING HYPOTHESIS TESTS

In this appendix, we detail the implementation of hypothesis tests for an observed real sequence $\mathbf{x}^{\text{input}} = x_1, \ldots, x_N$ and a given null hypothesis. Before beginning a test, we choose (1) a discriminating statistic $s : \mathbb{R}^N \to \mathbb{R}$ (in the main text we use the KS distance, the mean, the variance, the maximum, and conditional entropies of order one and two), (2) a size $\alpha_{\text{size}} \in (0, 1)$, which corresponds to the maximum allowable probability of incorrectly rejecting the null hypothesis even when it is true, also called the rate of false positives or type I error (in the main text we use $\alpha_{\text{size}} = 0.1$), (3) a number N_s of surrogates $\mathbf{x}$ independently to generate (in the main text we use $N_s = 9$, except in Table II, where we use $N_s = 999$), and (4) whether the test should be left tailed (appropriate when lower values of the discriminating statistic are more extreme or less consistent with the null hypothesis), right tailed (appropriate when higher values of the discriminating statistic are more extreme or less consistent with the null hypothesis), or two tailed (appropriate when both unexpectedly low and unexpectedly high values of the discriminating statistic correspond to noteworthy evidence against the null hypothesis). A quantile q is estimated as

$$q = \frac{r - 0.5}{N_s + 1} \in (0, 1),$$

where the rank $r \in \{1, 2, \ldots, N_s + 1\}$ is the position of the value $s^{\text{input}} = s(\mathbf{x}^{\text{input}})$ of the discriminating statistic observed for the input sequence $\mathbf{x}^{\text{input}}$ when it is combined with the N_s values $\{s_n = s(\mathbf{x}_n)\}$ observed for the N_s surrogate sequences $\mathbf{x}_n$, and these $N_s + 1$ values are ranked from smallest to largest, with ties broken by adding small random perturbations. The p value $p \in (0, 1)$ can then be computed from the quantile q according to

$$p = \begin{cases} q & \text{if the test is left tailed} \\ 1 - q & \text{if the test is right tailed} \\ 2 \ \min\{q, 1 - q\} & \text{if the test is two tailed.} \end{cases}$$

Finally, if $p < \alpha_{\text{size}}$ ($p \geq \alpha_{\text{size}}$), then we reject (fail to reject) the null hypothesis at the level of significance α_{size}.

The most popular method to assess a power-law hypothesis involves a right-tailed test using the KS distance as the discriminating statistic [21,30]. The best combination of statistic and tail depends on the specific type of deviations of most interest or practical importance. We have used left-tailed tests together with the mean, maximum, and variance because power-law distributions are often used to investigate extreme events or to represent heavy-tailed processes. We would intuitively expect lower values for these discriminating statistics to correspond to lower risk from extreme events, less support for the hypothesis of a heavy-tailed process, and less motivation to model these processes using power-law distributions.

A test of the null hypothesis that a sequence arose from a Markov chain of order m should use a statistic which is sensitive to Markov properties of order greater than m, because this allows detection of differences between surrogates of Markov order m and original data of higher Markov order. Such a statistic is an estimate of the conditional entropy h_{m+1} of order $m+1$, which for a stationary sequence $z_1, z_2, \ldots, z_N$ is given by

$$h_{m+1} = H_{m+1} - H_m,$$

where H_m is the entropy of order m:

$$H_m = -\sum_{z_t, z_{t+1}, \ldots, z_{t+m}} p(z_t, z_{t+1}, \ldots, z_{t+m}) \times \log p(z_t, z_{t+1}, \ldots, z_{t+m}),$$

and $p(z_t, z_{t+1}, \ldots, z_{t+m})$ is the probability that a subsequence of length $m+1$ will be $z_t, z_{t+1}, \ldots, z_{t+m}$. By stationarity, this probability is independent of t. Because we are interested in both the power-law and Markov properties of sequences, we consider conditional entropies defined in terms of the original time series $x_1, x_2, \ldots, x_N$ rather than the sequence of Markov states $z_1, z_2, \ldots, z_N$. Joint probabilities are estimated from the observed sequence using the maximum-likelihood method, for which probabilities are proportional to the frequency. We use natural logarithms, so that entropies have units of nats. Following, e.g., Refs. [42,54,72], tests involving conditional entropies are also left tailed.

[1] M. E. J. Newman, *Power Laws, Pareto Distributions and Zipf's Law*, Contemp. Phys. **46**, 323 (2005).

[2] M. Mitzenmacher, *A Brief History of Generative Models for Power Law and Lognormal Distributions*, Internet Math. **1**, 226 (2004).

[3] G. K. Zipf, *The Psycho-Biology of Language: An Introduction to Dynamic Philology* (Routledge, London, 1936), Vol. 21.

[4] J. Sethna, *Statistical Mechanics: Entropy, Order Parameters, and Complexity* (Clarendon Press, Oxford, 2021), Vol. 14.

[5] K. G. Wilson and J. Kogut, *The Renormalization Group and the ϵ Expansion*, Phys. Rep. **12**, 75 (1974).

[6] J. C. Willis and G. U. Yule, *Some Statistics of Evolution and Geographical Distribution in Plants and Animals, and Their Significance*, Nature (London) **109**, 177 (1922).

[7] D. J. d. S. Price, *Networks of Scientific Papers*, Science **149**, 510 (1965).

[8] D. J. d. S. Price, *A General Theory of Bibliometric and Other Cumulative Advantage Processes*, J. Am. Soc. Inf. Sci. **27**, 292 (1976).

[9] Z. Olami, Hans Jacob S. Feder, and K. Christensen, *Self-Organized Criticality in a Continuous, Nonconservative Cellular Automaton Modeling Earthquakes*, Phys. Rev. Lett. **68**, 1244 (1992).

[10] A.-L. Barabási and R. Albert, *Emergence of Scaling in Random Networks*, Science **286**, 509 (1999).

[11] P. Bak, C. Tang, and K. Wiesenfeld, *Self-Organized Criticality: An Explanation of the $1/f$ Noise*, Phys. Rev. Lett. **59**, 381 (1987).

[12] H. A. Simon, *On a Class of Skew Distribution Functions*, Biometrika **42**, 425 (1955).

[13] G. K. Zipf, *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology* (Addison-Wesley Press, Oxford, 1949).

[14] B. Gutenberg and C. F. Richter, *Frequency of Earthquakes in California*, Bull. Seismol. Soc. Am. **34**, 185 (1944).

[15] P. Bak, K. Christensen, L. Danon, and T. Scanlon, *Unified Scaling Law for Earthquakes*, Phys. Rev. Lett. **88**, 178501 (2002).

[16] While various definitions appear in the literature [2], we consider a distribution to be heavy tailed when its tail decays more slowly than an exponential [17].

[17] S. Foss, D. Korshunov, and S. Zachary, *An Introduction to Heavy-Tailed and Subexponential Distributions* (Springer, New York, 2011), Vol. 6.

[18] M. L. Goldstein, S. A. Morris, and G. G. Yen, *Problems with Fitting to the Power-Law Distribution*, Eur. Phys. J. B **41**, 255 (2004).

[19] H. Bauke, *Parameter Estimation for Power-Law Distributions by Maximum Likelihood Methods*, Eur. Phys. J. B **58**, 167 (2007).

[20] R. Hanel, B. Corominas-Murtra, B. Liu, and S. Thurner, *Fitting Power-Laws in Empirical Data with Estimators that Work for All Exponents*, PLoS One **12**, e0170920 (2017).

[21] A. Clauset, C. R. Shalizi, and M. E. J. Newman, *Power-Law Distributions in Empirical Data*, SIAM Rev. **51**, 661 (2009).

[22] A. Deluca and Á. Corral, *Fitting and Goodness-of-Fit Test of Non-Truncated and Truncated Power-Law Distributions*, Acta Geophys. **61**, 1351 (2013).

[23] Á. Corral and Á. González, *Power Law Size Distributions in Geoscience Revisited*, Earth Space Sci. **6**, 673 (2019).

[24] S. T. Bramwell, K. Christensen, J.-Y. Fortin, P. C. W. Holdsworth, H. J. Jensen, S. Lise, J. M. López, M. Nicodemi, J.-F. Pinton, and M. Sellitto, *Universal Fluctuations in Correlated Systems*, Phys. Rev. Lett. **84**, 3744 (2000).

[25] Á. Corral, L. Telesca, and R. Lasaponara, *Scaling and Correlations in the Dynamics of Forest-Fire Occurrence*, Phys. Rev. E **77**, 016101 (2008).

[26] M. Gerlach and E. G. Altmann, *Testing Statistical Laws in Complex Systems*, Phys. Rev. Lett. **122**, 168301 (2019).

[27] J. Eeckhout, *Gibrat's Law for (All) Cities*, Am. Econ. Rev. **94**, 1429 (2004).

[28] M. Levy, *Gibrat's Law for (All) Cities: Comment*, Am. Econ. Rev. **99**, 1672 (2009).

[29] J. Eeckhout, *Gibrat's Law for (All) Cities: Reply*, Am. Econ. Rev. **99**, 1676 (2009).

[30] A. D. Broido and A. Clauset, *Scale-Free Networks Are Rare*, Nat. Commun. **10**, 1017 (2019).

[31] I. Voitalov, P. van der Hoorn, R. van der Hofstad, and D. Krioukov, *Scale-Free Networks Well Done*, Phys. Rev. Research **1**, 033034 (2019).

[32] B. Zhou, X. Meng, and H. E. Stanley, *Power-Law Distribution of Degree–Degree Distance: A Better Representation of the Scale-Free Property of Complex Networks*, Proc. Natl. Acad. Sci. U.S.A. **117**, 14812 (2020).

[33] M. Falkenberg, J.-H. Lee, S.-i. Amano, K.-i. Ogawa, K. Yano, Y. Miyake, T. S. Evans, and K. Christensen, *Identifying Time Dependence in Network Growth*, Phys. Rev. Research **2**, 023352 (2020).

[34] M. Serafino, G. Cimini, A. Maritan, A. Rinaldo, S. Suweis, J. R. Banavar, and G. Caldarelli, *True Scale-Free Networks Hidden by Finite Size Effects*, Proc. Natl. Acad. Sci. U.S.A. **118**, e2013825118 (2021).

[35] M. P. H. Stumpf and M. A. Porter, *Critical Truths about Power Laws*, Science **335**, 665 (2012).

[36] P. Holme, *Rare and Everywhere: Perspectives on Scale-Free Networks*, Nat. Commun. **10**, 1016 (2019).

[37] P. Cirillo and N. N. Taleb, *Tail Risk of Contagious Diseases*, Nat. Phys. **16**, 606 (2020).

[38] Á. Corral, *Tail of the Distribution of Fatalities in Epidemics*, Phys. Rev. E **103**, 022315 (2021).

[39] E. Bradley and H. Kantz, *Nonlinear Time-Series Analysis Revisited*, Chaos **25**, 097610 (2015).

[40] M. Paluš, *From Nonlinearity to Causality: Statistical Testing and Inference of Physical Mechanisms Underlying Complex Fynamics*, Contemp. Phys. **48**, 307 (2007).

[41] G. Lancaster, D. Iatsenko, A. Pidde, V. Ticcinelli, and A. Stefanovska, *Surrogate Data for Hypothesis Testing of Physical Systems*, Phys. Rep. **748**, 1 (2018).

[42] M. J. Van der Heyden, C. G. C. Diks, B. P. T. Hoekstra, and J. DeGoede, *Testing the Order of Discrete Markov Chains Using Surrogate Data*, Physica (Amsterdam) **117D**, 299 (1998).

[43] J. M. Moore, Power-Law, https://github.com/JackMurdochMoore/power-law.git.

[44] J. Theiler and D. Prichard, *Constrained-Realization Monte-Carlo Method for Hypothesis Testing*, Physica (Amsterdam) **94D**, 221 (1996).

[45] D. Kugiumtzis, *Statically Transformed Autoregressive Process and Surrogate Data Test for Nonlinearity*, Phys. Rev. E **66**, 025201(R) (2002).

[46] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. D. Farmer, *Testing for Nonlinearity in Time Series: The Method of Surrogate Data*, Physica (Amsterdam) **58D**, 77 (1992).

[47] T. Schreiber and A. Schmitz, *Improved Surrogate Data for Nonlinearity Tests*, Phys. Rev. Lett. **77**, 635 (1996).

[48] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevX.12.021056 for additional text and figures supporting our results.

[49] R. A. Fisher, *On the Mathematical Foundations of Theoretical Statistics*, Phil. Trans. R. Soc. A **222**, 309 (1922).

[50] N. Reid, *The Roles of Conditioning in Inference*, Stat. Sci. **10**, 138 (1995).

[51] E. L. Lehmann, J. P. Romano, and G. Casella, *Testing Statistical Hypotheses* (Springer, New York, 2005), Vol. 3.

[52] J. Theiler and D. Prichard, *Using "Surrogate Surrogate Data" to Calibrate the Actual Rate of False Positives in Tests for Nonlinearity in Time Series*, in *Nonlinear Dynamics and Time Series: Building a Bridge Between the Natural and Statistical Sciences* edited by C. D. Cutler and D. T. Kaplan (American Mathematical Society, Providence, 1997), Vol. 11, p. 99.

[53] J. Besag and D. Mondal, *Exact Goodness-of-Fit Tests for Markov Chains*, Biometrics **69**, 488 (2013).

[54] S. D. Pethel and D. W. Hahs, *Exact Significance Test for Markov Order*, Physica (Amsterdam) **269D**, 42 (2014).

[55] A hypothesis is called simple when it is consistent with precisely one process [40,44,56,57]; more general hypotheses (e.g., hypotheses which do not specify parameter values) are called composite [40,42,44,56–59].

[56] M. Small and K. Judd, *Correlation Dimension: A Pivotal Statistic for Non-Constrained Realizations of Composite Hypotheses in Surrogate Data Analysis*, Physica (Amsterdam) **120D**, 386 (1998).

[57] M. Small, *Applied Nonlinear Time Series Analysis: Applications in Physics, Physiology and Finance* (World Scientific, Singapore, 2005), Vol. 52.

[58] T. Schreiber, *Interdisciplinary Application of Nonlinear Time Series Methods*, Phys. Rep. **308**, 1 (1999).

[59] T. Schreiber and A. Schmitz, *Surrogate Time Series*, Physica (Amsterdam) **142D**, 346 (2000).

[60] A statistic is called pivotal when its distribution is the same for all processes consistent with the tested hypothesis; other statistics are nonpivotal [40,44,56,57].

[61] R. A. Fisher, *Statistical Methods for Research Workers* (Oliver and Boyd, London, 1938).

[62] A. Nijenhuis and H. S. Wilf, *Combinatorial Algorithms: For Computers and Calculators* (Academic Press, Inc., New York, 1978).

[63] Given a real time series $x_1, x_2, \ldots, x_N \in \mathbb{R}$, the corresponding sequence of ordinal patterns of length $m+1$ is the sequence of real $(m+1)$ vectors $\boldsymbol{y}_t, \boldsymbol{y}_{t+1}, \ldots, \boldsymbol{y}_{N-m}$, in which $\boldsymbol{y}_t \in \mathbb{R}^{m+1}$ comprises the rank order of the subsequence $x_t, x_{t+1}, \ldots, x_{t+m}$, with tied values replaced by their mean rank. Ordinal patterns are discussed in, e.g., Refs. [64–66], and Refs. [67,68] discuss (non-power-law) ordinal pattern surrogates (in these references, the definition of ordinal patterns involves a different treatment of tied values).

[64] C. Bandt and B. Pompe, *Permutation Entropy: A Natural Complexity Measure for Time Series*, Phys. Rev. Lett. **88**, 174102 (2002).

[65] J. M. Amigó, S. Zambrano, and M. A. Sanjuán, *True and False Forbidden Patterns in Deterministic and Random Dynamics*, Europhys. Lett. **79**, 50001 (2007).

[66] J. M. Amigó, *Permutation Complexity in Dynamical Systems: Ordinal Patterns, Permutation Entropy and All That* (Springer Science & Business Media, Heidelberg, 2010).

[67] M. McCullough, K. Sakellariou, T. Stemler, and M. Small, *Regenerating Time Series from Ordinal Networks*, Chaos **27**, 035814 (2017).

[68] Y. Hirata, M. Shiro, and J. M. Amigó, *Surrogate Data Preserving All the Properties of Ordinal Patterns Up to a Certain Length*, Entropy **21**, 713 (2019).

[69] The set of all Markov states is a partition $\mathcal{A}$ of the integers greater than or equal to the lower cutoff $x_{\min}$. The Markov state z_t at time t is the unique element of the partition $\mathcal{A}$ which contains the value x_t at time t; $x_t \in z_t \in \mathcal{A}$.

[70] As a result of preserving empirical transition probabilities, the Markov order one surrogate precisely maintains the conditional entropy of order m with respect to the Markov states (see SM [48], Fig. S2).

[71] D. Kandel, Y. Matias, R. Unger, and P. Winkler, *Shuffling Biological Sequences*, Discrete Appl. Math. **71**, 171 (1996).

[72] D. C. Corrêa, J. M. Moore, T. Jüngling, and M. Small, *Constrained Markov Order Surrogates*, Physica (Amsterdam) **406D**, 132437 (2020).

[73] https://aaronclauset.github.io/powerlaws/data/blackouts.txt.

[74] A. Clauset, M. Young, and K. S. Gleditsch, *On the Frequency of Severe Terrorist Events*, J. Conflict Resolution **51**, 58 (2007).

[75] https://aaronclauset.github.io/powerlaws/data/terrorism.txt.

[76] https://aaronclauset.github.io/powerlaws/data/flares.txt.

[77] https://aaronclauset.github.io/powerlaws/data/words.txt.

[78] Á. Corral, *Long-Term Clustering, Scaling, and Universality in the Temporal Occurrence of Earthquakes*, Phys. Rev. Lett. **92**, 108501 (2004).

[79] W. Yang, E. Hauksson, and P. M. Shearer, *Computing a Large Refined Catalog of Focal Mechanisms for Southern California (1981–2010): Temporal Stability of the Style of Faulting*, Bull. Seismol. Soc. Am. **102**, 1179 (2012).

[80] https://scedc.caltech.edu/data/alt-2011-yang-hauksson-shearer.html.

[81] M. Båth, *Earthquake Energy and Magnitude*, Phys. Chem. Earth **7**, 115 (1966).

[82] H. M. J. Hung, R. T. O'Neill, P. Bauer, and K. Kohne, *The Behavior of the p-Value When the Alternative Hypothesis Is True*, Biometrics **53**, 11 (1997).

[83] B. Bhattacharya and D. Habtzghi, *Median of the p Value Under the Alternative Hypothesis*, Am. Stat. **56**, 202 (2002).

[84] J. N. Perry and R. Mead, *On the Power of the Index of Dispersion Test to Detect Spatial Pattern*, Biometrics **35**, 613 (1979).

[85] D. G. Champernowne and F. A. Cowell, *Economic Inequality and Income Distribution* (Cambridge University Press, Cambridge, England, 1998).

[86] J. Nair, A. Wierman, and B. Zwart, *The Fundamentals of Heavy Tails* (Cambridge University Press, Cambridge, England, 2022).

[87] P. Erdős and T. Gallai, *Graphs with Given Degrees of Vertices*, Mat. Lapok **11**, 264 (1960).

[88] R. Arratia and T. M. Liggett, *How Likely Is an I.I.D. Degree Sequence to Be Graphical?*, Ann. Appl. Probab. **15**, 652 (2005).

[89] D. Chorozoglou, E. Papadimitriou, and D. Kugiumtzis, *Investigating Small-World and Scale-Free Structure of Earthquake Networks in Greece*, Chaos Solitons Fractals **122**, 143 (2019).

[90] F. Serinaldi and F. Lombardo, *General Simulation Algorithm for Autocorrelated Binary Processes*, Phys. Rev. E **95**, 023312 (2017).

[91] C. Diks, J. C. Van Houwelingen, F. Takens, and J. DeGoede, *Reversibility as a Criterion for Discriminating Time Series*, Phys. Lett. A **201**, 221 (1995).

[92] D. Kugiumtzis, *Test Your Surrogate Data Before You Test for Nonlinearity*, Phys. Rev. E **60**, 2808 (1999).

[93] P. Dimitriadis and D. Koutsoyiannis, *Stochastic Synthesis Approximating Any Process Dependence and Distribution*, Stoch. Environ. Res. Risk Assess. **32**, 1493 (2018).

[94] M. A. van Wyk and W.-H. Steeb, *Chaos in Electronics* (Springer, Dordrecht, 1997).

[95] P. Gaspard, *Chaos, Scattering and Statistical Mechanics*, Cambridge Nonlinear Science Series (Cambridge University Press, Cambridge, England, 2005).

[96] S. Stumpf, Combalg-py https://pythonhosted.org/combalg-py/ (2017), accessed October 2019.