Zipf’s and Taylor’s laws

Charlotte James
Department of Engineering Mathematics, University of Bristol, Woodland Road, BS8 1UB, Bristol, UK

Sandro Azaele
School of Mathematics, University of Leeds, Leeds, UK.

Amos Maritan
Dipartimento di Fisica “G. Galilei”, Università di Padova, INFN, via Marzolo 8, 35131 Padova, Italy.

Filippo Simini
Department of Engineering Mathematics, University of Bristol,
Merchant Venturers Building, Woodland Road, BS8 1UB, Bristol, UK
(Dated: August 7, 2018)

Zipf’s law states that the frequency of an observation with a given value is inversely proportional to the square of that value; Taylor’s law, instead, describes the scaling between fluctuations in the size of a population and its mean. Empirical evidence of the validity of these laws has been found in many and diverse domains. Despite the numerous models proposed to explain the presence of Zipf’s law, there is no consensus on how it originates from a microscopic process of individuals dynamics without fine tuning. Here we show that Zipf’s law and Taylor’s law can emerge from a general class of stochastic processes at the individual level, which incorporate one of two features: environmental variability, i.e. fluctuations of parameters, or correlations, i.e. dependence between individuals. Under these assumptions, we show numerically and with theoretical arguments that the conditional variance of the population increments scales as the square of the population and that the corresponding stationary distribution of the processes follows Zipf’s law.

Formally, a random variable follows Zipf’s law if its probability density function is a power law with exponent $-2$. Evidence for Zipf’s law has been found in many and diverse empirical domains, including systems where observations correspond to groups of individuals and the variable of interest is the group size, such as the number of employees in firms [1] or the distribution of family names [2]. In particular, one of the most documented empirical findings in human geography is Zipf’s law for the distribution of city sizes: the probability to find a city with a given number of dwellers, $n$, is inversely proportional to the square of that number: $P(n) \sim n^{-1-\gamma}$, where $\gamma \approx 1$. The exponent $\gamma \approx 1$ has been shown to apply to city sizes both globally and historically with surprisingly small deviations. Zipf’s law also applies to the distribution of population of larger regions, for example countries in Europe [3] as well as cities and counties in the United States (see Fig.1).

A number of general mechanisms exist to account for the emergence of Zipf’s law in various systems [2, 6, 7]. A series of recent works [8, 9] has demonstrated that models with latent variables can lead to Zipf’s law without fine tuning by mixing together narrow distributions with very different means. However, these works focus on static systems, without an explicit time dependence. Several models have been proposed to explain how Zipf’s law can emerge as the stationary distribution of dynamical processes for the sizes of groups of individuals. These models can be divided into two classes based on the scale considered: mesoscopic models at the scale of the groups (e.g., cities) and microscopic models at the scale of the individuals (e.g., dwellers). The mesoscopic models are stochastic processes describing the evolution of the group’s population as a whole. Examples of mesoscopic models are the random multiplicative process [10], also called Gibrat’s law [11] or proportionate random growth in the economics literature [12, 13], and those based on the interplay between intermittency and diffusion [14]. Mesoscopic models are able to explain the emergence of Zipf’s law without the need to fine-tune their parameters to specific values [15], however they are coarse-grained descriptions of population dynamics and lack an explicit link to the underlying microscopic processes. Microscopic models provide a more fundamental description because they are stochastic processes describing the events experienced by an individual, namely births, deaths and migrations, that ultimately determine the change in the size of a population. Examples of microscopic models include Yule’s and Simon’s models based on the rich-get-richer mechanism [16, 17], cluster growth and aggregation [18, 19], preferential migration to large aggregates [20], and network growth with redirection [21]. However, microscopic models are only able to produce the power law exponent $\gamma = 1$ for specific values of their parameters.

Here we present a class of microscopic stochastic processes that are able to reproduce Zipf’s law with exponent $\gamma = 1$ without fine tuning. These processes are charac-
characterized by an anomalous scaling of the fluctuations of the population increments, commonly known as Taylor’s law in ecology \cite{22,24}. In particular, we describe two general mechanisms to construct microscopic stochastic processes where the conditional variance of the population increments scale as the square of the population, and we demonstrate that the stationary distribution of such processes follows Zipf’s law. Our processes are applicable to dynamical systems with an explicit time dependence where the stationary distribution of group sizes can be described by Zipf’s law. In this respect, our derivation of Zipf’s law differs from static models without an explicit time dependence and from models where groups can only grow \cite{6,8,17,25,26}.

We present our microscopic processes as models in which individuals belong to different local areas (groups or patches). To describe how the population in each group evolves in time, we use a modified version of the Galton-Watson process. The original Galton-Watson process \cite{27} is a discrete-time branching stochastic process describing the evolution of a population of $n_i$ individuals at time $t$ according to the equation

$$n_{t+1} = \sum_{i=1}^{n_t} x_i$$  \hspace{1cm} (1)$$

where $x_i$ are independent and identically distributed random variables over the integers, with finite mean and variance and with probability mass function $P(x|\lambda)$, where $\lambda$ are parameters. If $x_i = 0$, individual $i$ dies, if $x_i = 1$, they do nothing and if $x_i = 2, 3, \ldots$ then they have $1, 2, 3, \ldots$ children. Extinction will occur with probability 1 if the average number of offspring per individual is less than or equal to one, $\mathbb{E}(x) \leq 1$. To avoid extinction and ensure the process has a stationary state we include
the Galton-Watson process, namely processes with envi-
ronmental variability and processes with correlated indi-
viduals, and show that for large populations they have exponents \( \gamma = 1 \) and \( \alpha = 1 \).

We first present the case of processes with envi-
ronmental variability. To this end, we consider a modified
Galton-Watson process where the parameters \( \lambda \) of the distribution of the individuals are not constant values but random variables drawn from a distribution \( G(\lambda) \) at each time step. To be specific, we assume that the \( x_i \) are Poisson random variables with distribution

\[
P(x|\lambda) = \text{Pois}(\lambda) = e^{-\lambda} \frac{\lambda^x}{x!}.
\]

In order to determine the fluctuations of the population
increments, we compute the variance of the popula-
tion change, \( \Delta n = n_{t+1} - n_t \), conditioned to the
population at time \( t \) by applying the law of total vari-
ance \([30]\),

\[
Var(\Delta n|n_t = n) = E_\lambda[Var_{\Delta n|\lambda}(\Delta n)] + Var_{\lambda}(E_{\Delta n|\lambda}(\Delta n)).
\]

Here \( E_{\Delta n|\lambda} \) and \( Var_{\Delta n|\lambda} \) denote the mean and variance of \( \Delta n \) for a fixed \( \lambda \), and \( E_{\lambda} \) and \( Var_{\lambda} \) denote the mean and variance with respect to \( G(\lambda) \). We obtain,

\[
Var(\Delta n|n_t = n) = n \ E_\lambda(\lambda) + n^2 Var_\lambda(\lambda).
\] (3)

Note that the fluctuations in the size of a population are propor-
tional to \( n \) and follow Eq. [2] with exponent \( \alpha = 1 \) for small populations, \( n \ll n_c = \frac{E_\lambda(\lambda)}{Var_\lambda(\lambda)} \), whereas they scale as \( n^2 \) for large populations \( n \gg n_c \).

The crossover population \( n_c \) marks the transition be-
tween these two scaling regimes. Empirical evidence of the
presence of this crossover can be found analysing the fluctua-
tions of the populations of cities and coun-
ties in the United States, where data shows a transition
between Taylor exponents \( \alpha = 1/2 \) and \( \alpha = 1 \) around

\( n_c \sim 10^4 \) (Fig. 1b).

Within our framework, this re-
result means that on average the ratio between variance
and mean of growth rates is very small, around

\( n_c^{-1} = \frac{Var_\lambda(\lambda)}{E_\lambda(\lambda)} \sim 10^{-4} \). For example, if the distribution of \( \lambda \) is a Gamma distribution, \( G(\lambda) = \text{Gamma}(\lambda|\kappa, \nu) \), then Eq. [3] becomes

\[
Var(\Delta n|n_t = n) = n \ k \nu + n^2 \ k^2 \nu^2 \ 	ext{and} \ n_c = \nu^{-2}.
\]

We verify the validity of this prediction with
numerical simulations of Eq. [4] shown in Fig 2a,
where the \( x_i \) are independent and identical Poisson random vari-
bles with distribution Pois(\( \lambda \)) and a new parameter \( \lambda \)

is drawn at each discrete time step from a Gamma dis-
tribution with fixed parameters \( \kappa = 10^4 \) and \( \nu = 10^{-4} \).

Next we demonstrate that the stationary distribution
of group sizes follows Zipf’s law for large populations,
as shown in Fig 2a and summarised in Table I. When
the total population is sufficiently large, we expect that
Eq. [1] can be approximated as (see martingale Central
Limit Theorem [31])

\[
\Delta n \equiv n_{t+1} - n_t \approx \hat{\mu}_n + \hat{\sigma}_n \xi(t),
\] (4)

where \( \Delta n(t) \) is a continuous random variable,
\( \hat{\mu}_n = E(\Delta n|n_t = n) \), \( \hat{\sigma}_n = \sqrt{Var(\Delta n|n_t = n)} \), and \( \xi(t) \) is a zero-mean Gaussian white noise with autocorrelation

\[
\langle \xi(t)\xi(t') \rangle = 2\delta(t - t').
\]

Figures 3a-b demonstrate that numerical simulations support the validity of the
ansatz of Eq. [4]. Using the law of total expectation
and Eq. [3] we obtain

\[
\hat{\mu}_n = n \ E_\lambda(\lambda - 1) \equiv n \hat{\mu} \ 	ext{and} \ \hat{\sigma}_n = \sqrt{n E_\lambda(\lambda) + n^2 Var_\lambda(\lambda)} \equiv \sqrt{n \mu \sigma_1^2 + n^2 \sigma_2^2}.
\]

Using the formal substitutions \( t + 1 \rightarrow t + dt \), \( \hat{\mu} \rightarrow \hat{\mu} dt \), \( \hat{\sigma}_n \rightarrow \hat{\sigma}_n dt \), and \( \hat{\sigma}_n \rightarrow \hat{\sigma}_n dt \) to first order in \( dt \), we ob-
tain the following stochastic differential equation:

\[
d\hat{\xi}(t) = \hat{\mu}(t) + \sqrt{\hat{\sigma}_n^2(t)} d\hat{\xi}(t),
\]

with a reflecting boundary at \( n = 1 \). Hence, the proposed birth-death micro-
scopic process can be well approximated by a meso-
scopic proportionate random growth dynamics, when
populations are large. In fact, for large populations,
\( n \gg n_c = \sigma_1^2 / \sigma_2^2 \), the above equation becomes a ran-
dom multiplicative process with growth rate of mean \( \mu \) and variance \( \sigma_2^2 \).

In this limit, we obtain a stationary

<table>
<thead>
<tr>
<th>Distribution</th>
<th>( \gamma )</th>
<th>( D )</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>US Counties</td>
<td>0.86</td>
<td>0.04</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>US Cities</td>
<td>0.96</td>
<td>0.07</td>
<td>0.01</td>
</tr>
<tr>
<td>Environmental variability</td>
<td>0.91</td>
<td>0.02</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>Correlated individuals</td>
<td>1.00</td>
<td>0.01</td>
<td>&lt; 0.01</td>
</tr>
</tbody>
</table>

TABLE I. Power law fit: exponent \( \gamma \) for the distribution of group sizes obtained using a Maximum likelihood estimate, Kolmogorov-Smirnov statistic \( D \) and p-value for the PDF distributions in Figs 1a and 2a. All numbers are rounded to two decimal places.
distribution with a power-law tail $P(n) \sim n^{-2+\frac{2}{\sigma_2^2}}$.

Notice that we get the exponent of Zipf’s law, $\gamma = 1$, for $|\mu| < \sigma_2^2$. The parameter values used in the simulations of Fig. 2 are chosen to fit the models to city data, capturing both the Zipf exponent of the tail and the large value of $n_c$. Because of the specific choice of a Poisson distribution for $P(x|\lambda)$, this requires to use a set of parameter values that are very close to criticality (i.e., $\lambda = 1$). However, it is important to emphasize that it is possible to find various distributions $P(x|\lambda)$ and $G(\lambda)$ that produce non-critical systems with stationary distributions with exponent $\gamma = -2 + \frac{\mu}{\sigma_2^2}$ close to Zipf’s law for any value of $n_c = \sigma_1^2/\sigma_2^2$. In general, for generic distributions $P(x)$ and $G(\lambda)$, we have:

\[
\begin{align*}
\sigma_1^2 &= \mathbb{E}_\lambda (\text{Var}_{\Delta n|\lambda}(\Delta n)) \\
\sigma_2^2 &= \mathbb{V}ar_\lambda \left( \mathbb{E}_{\Delta n|\lambda}(\Delta n) \right) \\
\mu &= \mathbb{E}_\lambda (\mathbb{E}_{\Delta n|\lambda}(\Delta n)) - 1
\end{align*}
\]

so it is possible to find many combinations of functions and parameter values such that $|\mu| < \sigma_2^2 < \sigma_1^2$, corresponding to Zipf’s law and large $n_c$. For example, it might be possible to satisfy these conditions using a Negative Binomial distribution for $P(x)$, whose mean and variance can be independently set, and choosing an appropriate distribution over its parameters, $G(\lambda)$. The second case we present is the class of processes where individuals are correlated. To this end, we consider a modified Galton-Watson process with correlated individuals, where the joint probability $P(x_1, \ldots, x_n|\lambda)$ does not factorize into the product of the individual probabilities $P(x_i|\lambda)$. To be specific, we assume that the $x$ random variables have a Poisson distribution with fixed parameter $\lambda$ (the same for all individuals) and correlation matrix $\rho_{ij} = \text{Cov}(x_i, x_j)/\lambda$, where $\text{Cov}(x_i, x_j) = \mathbb{E}[(x_i - \lambda)(x_j - \lambda)]$. When individuals are correlated the off-diagonal terms of the covariance matrix are non-zero, and in the simplest case that we consider, they are all equal: $\text{Cov}(x_i, x_j) = \delta_{ij}\lambda + (1 - \delta_{ij})\rho\lambda$. The fluctuations of the population increments for the process with correlated individuals have the same scaling form of the fluctuations for the process with environmental variability. The conditional variance of $\Delta n = n_{t+1} - n_t$ for a given population $n_t = n$ is:

\[
\text{Var}(\Delta n|n_t = n) = \sum_{i,j} \text{Cov}(x_i, x_j) = n\lambda(1 - \rho) + n^2\rho\lambda
\]

Here the cross-over population between the regimes with Taylor exponents $\alpha = 1/2$ and $\alpha = 1$ is $n_c = 1/\rho - 1$. Within this framework, the cross-over population $n_c \sim 10^4$ of United States cities and counties (Fig. 1b) corresponds to a very small value of correlation between people, $\rho \sim 1/n_c \sim 10^{-4}$. The physical meaning of this correlation can be interpreted considering that the inverse of the correlation matrix corresponds to the interaction between individuals [33, 34]. Since the off-diagonal elements of the inverse of the covariance matrix vanish in the small $\rho$ limit we get that only a small amount of interaction is needed for large group sizes to follow Taylor’s law with exponent $\alpha = 1$, yet in the absence of correlations this will not be present.

Using the ansatz of Eq. 4 and following the same steps taken for the case with environmental variability, it is possible to show that the process with correlated individuals has a stationary distribution of group sizes which for large populations, $n > n_c$, is a power-law, $P(n) \sim n^{-2+\frac{2}{\sigma_2^2} - \frac{1}{\rho^2}}$, that follows Zipf’s law when $|\lambda - 1/\lambda| \ll \rho$. Numerical simulations support the validity of the approximation of Eq. 4 on the scaling of the population increments (Figs. 3c,d and Fig. 2c) and the presence of Zipf’s law in the distribution tail (Fig. 2a, Table I).

In the numerical simulations, we used two methods to generate correlated Poisson variables. In the first method, used in the simulations of Fig. 2 we consider
the modified Galton-Watson process \( n_{t+1} = \sum_{i=1}^{n_t} (x_i + z) \) where the \( x_i \) are independent and identical random variables with Poisson distribution \( \text{Pois} (\lambda (1 - \rho)) \) and we introduce the random variable \( z \), independent of the \( x_i \), with Poisson distribution \( \text{Pois} (\rho \lambda) \). One can show that for this process the conditional variance of \( \Delta n \) is identical to Eq. [5]. In Fig. 2 we use \( \lambda = 1 \) and \( \rho = 10^{-4} \). In the second method, used in the simulations of Fig. 3a-c, we use a Gaussian copula model to link the Poisson marginals [35]. We verified with numerical simulations that for appropriate values of \( \lambda \) and covariance matrix \( \text{Cov}(x_i, x_j) = C_{ij} \); the conditional variance of \( \Delta n \) is again identical to Eq. [5].

Both methods give results compatible with the theoretical predictions.

To summarise our results, we have shown that environmental variability and correlations between individuals are able to produce Zipf’s law and Taylor’s law [36]. The exponent of Taylor’s law shifts from \( \alpha = 1/2 \) for small populations to \( \alpha = 1 \) for large populations. Also, when population is large Zipf’s law emerges naturally without fine tuning whenever the growth rate’s mean is much larger than the variance. The proposed framework can also naturally account for exponent values close but not identical to Zipf’s law. This reveals a general connection between Zipf’s law and Taylor’s law in microscopic stochastic processes of population dynamics under realistic assumptions.

ACKNOWLEDGMENTS

FS is supported by EPSRC First Grant EP/P012906/1.

[26] Modelling immigration using a boundary condition is a solution adopted in several models of population dynamics [37]. We verified that the specific implementation of the boundary (as a hard reflecting boundary or as a constant influx of individuals) and its value do not affect our results.
[27] Taylor’s law originates from ecology, where it was originally defined as the scaling relationship between the variance and mean of populations in different areas [24][38]. Since then, the name “Taylor’s law” has been often used to denote the general scaling relationship between the variance and the mean of random variables in complex systems [22][23]. In this case, the random variable of interest is the population increment in one time step, \( \Delta n \), and we measure its mean and variance in patches of different sizes, \( \eta \).
[28] Indeed, numerical simulations show that the same results can be obtained when both mechanisms are present.