Theoretical Population Biology 93 (2014) 30–37

Contents lists available at ScienceDirect

Theoretical Population Biology

journal homepage: www.elsevier.com/locate/tpb

Stochastic population dynamics in a Markovian environment implies Taylor's power law of fluctuation scaling

Joel E. Cohen

The Rockefeller University and Columbia University, 1230 York Avenue, Box 20, New York, NY 10065, USA

ARTICLE INFO

Article history:
Received 27 October 2013
Available online 18 January 2014

Keywords:
Markov chain
Taylor's law
Power law
Fluctuation scaling
Random evolution
Variance function

ABSTRACT

Taylor's power law of fluctuation scaling (TL) states that for population density, population abundance, biomass density, biomass abundance, cell mass, protein copy number, or any other nonnegative-valued random variable in which the mean and the variance are positive, variance $= a(\text{mean})^b$, $a > 0$, or equivalently log variance $= \log a + b \times \log$ mean. Many empirical examples and practical applications of TL are known, but understanding of TL's origins and interpretations remains incomplete. We show here that, as time becomes large, TL arises from multiplicative population growth in which successive random factors are chosen by a Markov chain. We give exact formulas for a and b in terms of the Markov transition matrix and the values of the successive multiplicative factors. In this model, the mean and variance asymptotically increase exponentially if and only if $b > 2$ and asymptotically decrease exponentially if and only if $b < 2$.

1. Introduction

Fluctuation scaling is a name popular among physicists for a lawful relationship between the mean and variance of any random variable when the mean and variance are functions of some parameter. Among statisticians, such a relationship is often called a variance function. In population biology and ecology, Taylor's power law of fluctuation scaling (Taylor, 1961, 1984) states that when the mean and the variance exist and are positive functions of some parameter, they are related by a power law: variance $= a(\text{mean})^b$, $a > 0$, or equivalently log variance $= \log a + b \times \log$ mean.

Taylor's law (TL) began with empirical observations of insect population densities and was verified in hundreds of biological species (Eisler et al., 2008) including, recently, bacteria (Ramsayer et al., 2011; Kaltz et al., 2012), trees (Cohen et al., 2012, 2013a), and humans (Cohen et al., 2013b). TL is one of the most widely verified empirical relationships in ecology. TL has also been confirmed for cell populations within specific organs (Azevedo and Leroi, 2001), stem cell populations (Klein and Simons, 2011), counts of single nucleotide polymorphisms and genes (Kendal and Jørgensen, 2011), cases of measles and whooping cough (Keeling and Grenfell, 1999), the mass of single-celled organisms of different species (Giometto et al., 2013), and in diverse other fields (for additional references, see review by Eisler et al., 2008), including cancer metastases, single nucleotide polymorphisms and genes on chromosomes, and non-biological measurements such as precipitation, packet switching on the Internet, stock market trading, and number theory. TL has practical applications in the design of sampling plans for the control of insect pests (soybeans: Kogan et al., 1974, Bechinski and Pedigo, 1981; cotton: Wilson et al., 1989; glasshouse roses: Park and Cho, 2004).

There is little consensus about why TL is so widely observed and how its estimated parameters should be interpreted. The theoretical analysis of probability distributions in which the variance is a power-law function of the mean preceded TL (Tweedie, 1946, 1947) (in other words, Taylor did not invent Taylor's law) and TL has been much studied theoretically with or without recognition of its empirical roots in ecology (e.g., Anderson et al., 1982, Tweedie, 1984, Perry and Taylor, 1985, Gillis et al., 1986, Jørgensen, 1987, Kemp, 1987, Perry, 1988, Lepš, 1993, Jørgensen, 1997, Keeling, 2000, Azevedo and Leroi, 2001, Kilpatrick and Ives, 2003, Kendal, 2004, Ballantyne and Kerkhoff, 2007, Eisler et al., 2008, Engen et al., 2008, Kendal and Jørgensen, 2011, Cohen et al., 2013a). Davidian and Carroll (1987) and Wang and Zhao (2007) emphasized the importance of modeling correctly how the variance is related to the mean if one desires statistical efficiency in estimating the mean. They considered multiple variance functions including TL. But they did not identify a power-law variance function with TL or discuss models that might explain the origin of these variance functions.

Cohen et al. (2013a) showed that the Lewontin and Cohen (1969) (no relation to the present author) stochastic multiplicative population model (a geometric random walk with independently and identically distributed [i.i.d.] multiplicative increments) implies TL. Cohen et al. (2013a) calculated log a and b explicitly. Here

---

E-mail addresses: joel.cohen@rockefeller.edu, cohen@rockefeller.edu.

0040-5809/$ – see front matter © 2014 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.tpb.2014.01.001

we consider a more general model in which the factors that multiply the population density at each time step are history-dependent, not independent as in the Lewontin–Cohen model. We show that a multiplicative model of change in a Markovian environment leads to TL in the limit of large time, and we calculate $\log a$ and b explicitly.

2. Taylor's law

Let a family of nonnegative random variables $N(t)$ be parameterized by $t \in \Theta$, where Θ is an index set. Assume that, for all $t \in \Theta$, the mean $E(N(t))$ and the variance $\operatorname{Var}(N(t))$ are finite and positive, so $\log \operatorname{Var}(N(t))$ and $\log E(N(t))$ are well defined. We may think of $N(t)$ as population density at time t.

Definition. TL applies to $N(t)$ exactly for all $t \in \Theta$ if and only if there exist real constants $a > 0$ and b such that, for all $t \in \Theta$, $\operatorname{Var}(N(t)) = a\,(E(N(t)))^b$. Equivalently, TL applies to $N(t)$ exactly for all $t \in \Theta$ if and only if there exist constants $a > 0$ and b such that

$$\log \operatorname{Var}(N(t)) - b \log E(N(t)) = \log a. \qquad (1)$$

The mean $E(N(t))$ and the variance $\operatorname{Var}(N(t))$ refer to an ensemble mean and ensemble variance at t (the mean and the variance over independent realizations, e.g., in sufficiently separated regions in space), not to a mean and variance over t.

Definition. TL applies to $N(t)$ in the limit as t approaches some finite or infinite limit $\theta \in \Theta$ if and only if there exist real constants $a > 0$ and b such that

$$\lim_{t\to\theta}[\log \operatorname{Var}(N(t)) - b \log E(N(t))] = \log a. \qquad (2)$$

These definitions intentionally leave unspecified the base of the logarithms (e.g., e, 10, or 2) because TL is equally valid for logarithms to any base. For the following analysis, $\log = \log_e$.

3. Scalar discrete-time Markovian multiplicative growth

Assume $N(0)$ is a fixed positive number. Suppose that

$$N(t) = A(t-1)A(t-2)\cdots A(0)N(0), \quad t = 0, 1, 2, \ldots. \qquad (3)$$

Then $A(t-1) = N(t)/N(t-1)$, $t = 1, 2, \ldots$ represents the random factor of change from time $t-1$ to time t. Assume that each value of $A(t)$ is taken from a finite set of positive numbers $\{d_1, \ldots, d_s\}$, $s > 1$, at least two of which are distinct. Intuitively, s is the number of states of the environment. By assumption, each state of the environment determines a multiplicative factor of change: if $A(t-1) = d_i$, then $N(t) = d_i N(t-1)$, for $i = 1, \ldots, s$ and $t = 1, 2, \ldots$. Assume $\{A(t),\ t = 0, 1, 2, \ldots\}$ is a finite-state homogeneous Markov chain ("a Markovian environment") with an $s \times s$ column-to-row ($j \to i$) transition probability matrix $P = (p_{ij})$ with $p_{ij} \geq 0$, $\sum_{i=1}^{s} p_{ij} = 1$, $i, j = 1, \ldots, s$ and $\Pr\{A(t) = d_i | A(t-1) = d_j\} = p_{ij}$, $i, j = 1, \ldots, s$; $t = 1, 2, \ldots$. (Notational aside: in many works on Markov chains, it is customary for the transition matrix P to specify row-to-column ($i \to j$) transition probabilities. But in many works on stochastic population models in Markovian environments, the opposite convention has become usual because it conforms with usual matrix–vector multiplication. One must pick one convention or the other, and the latter is used here.) Assume P has equilibrium probability $s \times 1$ vector $\pi = (\pi_i)$ such that $\sum_{i=1}^{s} \pi_i = 1$ and $P\pi = \pi$, and that $\pi_i > 0$, $i = 1, \ldots, s$. Also assume that π is the initial distribution of the Markov chain, i.e., $\Pr\{A(0) = d_i\} = \pi_i$, $i = 1, \ldots, s$. Consequently, $\Pr\{A(t) = d_i\} = \pi_i$, $i = 1, \ldots, s$ for all $t = 0, 1, 2, \ldots$.

This model is a scalar case of much studied Markovian multiplicative models for age- and stage-structured populations (Bharucha, 1960, Furstenberg and Kesten, 1960, Bharucha, 1961, Cohen, 1976, 1977a,b, Tuljapurkar and Orzack, 1980, Tuljapurkar, 1982, 1986, 1990; review by Caswell, 2001; Tuljapurkar et al., 2009). The increments $\{A(t)\}$ are Markovian (by assumption) and therefore the pair $(A(t), N(t))$ is Markovian, but $N(t)$ by itself is not Markovian unless $\{A(t)\}$ are independent. (In the Lewontin and Cohen (1969) model, $N(t)$ *is* Markovian because $\{A(t)\}$ are independent.)

Our main result is that, under certain conditions, this model predicts TL in the limit of large time, and the parameters a and b of TL can be expressed as functions of the parameters of the Markovian model of $A(t)$. To state this theorem precisely, we give some definitions and notation.

Define the $s \times s$ diagonal matrix $D = \operatorname{diag}(d_i)$ to be zero everywhere except on the diagonal. The diagonal element $d_{ii} = d_i > 0$ is the value of $A(t)$ in state i of the environment. We say that a diagonal matrix is scalar if it is some scalar multiple of the identity matrix I, i.e., if all its diagonal elements are equal. Our assumption that at least two values in $\{d_1, \ldots, d_s\}$ are distinct means that D is not scalar.

For any $s \times s$ matrix Z, let $r(Z)$ be the spectral radius of Z. The spectral radius is the maximum of the magnitude of any eigenvalue of Z. For any square matrix Z, $r(Z^t) = (r(Z))^t \equiv r^t(Z)$, $t = 0, 1, 2, \ldots$.

By definition, a nonnegative $s \times s$ matrix A with $s > 1$ is irreducible if and only if for each row i and each column j with $1 \leq i, j \leq n$, there exists an integer p such that $(A^p)_{ij} > 0$. The transpose of A is written A^T. A nonnegative $s \times s$ matrix A with $s > 1$ is, by definition, two-fold irreducible if and only if A is irreducible and A^TA is irreducible (O'Cinneide, 2000; Altenberg, 2013).

Define 1^T to be the row s-vector with each element 1. The assumption that P is column-stochastic (each column sum is 1) is equivalent to $1^TP = 1^T$. For any real number p, (3) implies

$$(N(t))^p = (A(t-1))^p \cdots (A(0))^p\,(N(0))^p, \quad t = 1, 2, \ldots. \qquad (4)$$

Henceforth assume $p > 0$. Then because $0 < (N(t))^p \leq [(\max_{i=1,\ldots,s} d_i)^t N(0)]^p < \infty$ with probability 1, $E\left[(N(t))^p\right] > 0$ is well defined for $p > 0$. For $t = 1$,

$$\begin{aligned} E\left[(N(1))^p\right] &= E\left(A(0)^p\right) N(0)^p = \left(\sum_{i=1}^{s} (d_i)^p \pi_i\right) N(0)^p \\ &= 1^T(D^p\pi)N(0)^p = 1^T(D^pP\pi)N(0)^p. \end{aligned} \qquad (5)$$

The last equality in (5) follows from $\pi = P\pi$. In (5) and (6), π appears because the Markov chain was assumed stationary, i.e., starting at its equilibrium distribution. Summing (4) over all trajectories,

$$\begin{aligned} E\left[(N(t))^p\right] &= 1^T\left(D^pP\right)^{t-1}\left(D^pP\pi\right)N(0)^p \\ &= 1^T\left(D^pP\right)^{t}\pi N(0)^p, \quad t = 1, 2, \ldots. \end{aligned} \qquad (6)$$

We assumed the diagonal of D is positive. We assume further that P is two-fold irreducible. In an empirical application, P is likely to be positive. Since every positive matrix is two-fold irreducible, the assumption that P is two-fold irreducible is likely to be easily satisfied in an empirical application. The assumption that P is two-fold irreducible is the minimal condition necessary and sufficient to prove the eigenvalue inequalities which we use to prove our main result (Cohen, in press). It remains unknown whether our main result could be proved under the weaker assumption of Tuljapurkar (1982) that P is primitive (irreducible and aperiodic or ergodic).