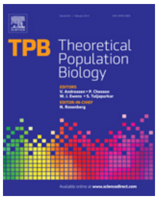
Theoretical Population Biology 93 (2014) 30–37

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 $E\left(N(t)\right)$ and the variance $\mathrm{Var}\left(N(t)\right)$ are finite and positive, so $\log \mathrm{Var}\left(N(t)\right)$ and $\log E\left(N(t)\right)$ 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 $t \in \Theta$, $\mathrm{Var}\left(N(t)\right) = a\left(E\left(N(t)\right)\right)^b$. 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

$$\log \mathrm{Var}\left(N(t)\right) - b \log E\left(N(t)\right) = \log a. \tag{1}$$

The mean $E\left(N(t)\right)$ and the variance $\mathrm{Var}\left(N(t)\right)$ 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

$$\lim_{t \to \theta}[\log \mathrm{Var}\left(N(t)\right) - b \log E\left(N(t)\right)] = \log a. \tag{2}$$

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

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

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 $A\left(t-1\right) = d_i$, then $N(t) = d_i N\left(t-1\right)$, 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 $P = \left(p_{ij}\right)$ with $p_{ij} \geq 0,\ \sum_{i=1}^{s} p_{ij} = 1,\ i, j = 1, \ldots, s$ and $\Pr\left\{A(t) = d_i | A\left(t-1\right) = d_j\right\} = p_{ij},\ i, j = 1, \ldots, s;\ t = 1, 2, \ldots$. (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 $D = \mathrm{diag}(d_i)$ 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, $r\left(Z^t\right) = (r(Z))^t \equiv r^t(Z),\ t = 0, 1, 2, \ldots$.

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 $A^T A$ 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 $1^T P = 1^T$. For any real number p, (3) implies

$$(N(t))^p = (A\left(t-1\right))^p \cdots (A(0))^p\,(N(0))^p\,, \quad t = 1, 2, \ldots. \tag{4}$$

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 $t = 1$,

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

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,

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

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).

By the Perron–Frobenius theorem (Gantmacher, 1960), D^pP has spectral radius $r(D^pP) > 0$, left (row) eigenvector $w(p)^T > 0$ and right (column) eigenvector $v(p) > 0$ such that $w(p)^T v(p) = 1$ and

$$\lim_{t\to\infty} \left(D^pP\right)^t / \left[r\left(D^pP\right)\right]^t = v(p)w(p)^T. \tag{7}$$

In particular ($p = 1$), the expected value of $N(t)$ grows geometrically in proportion to the tth power of $r(DP)$. So $E(N(t))$ neither grows nor declines in the limit of large t if and only if $r(DP) = 1$ or $\log r(DP) = 0$.

The main result is:

Theorem 1. *In the scalar stationary discrete-time Markovian model, assume that D is not scalar and P is two-fold irreducible and $\log[r(DP)] \neq 0$. Then $N(t)$ obeys TL (2), $\lim_{t\to\infty}[\log \mathrm{Var}(N(t)) - b \log E(N(t))] = \log a$, with*

$$\begin{aligned} b &= \log\left[r(D^2P)\right] / \log\left[r(DP)\right], \\ a &= \left(1^T v(2)w(2)^T \pi N(0)^2\right) / \left(1^T v(1)w(1)^T \pi N(0)\right)^b. \end{aligned} \tag{8}$$

A proof and further interpretation are provided in the Appendix.

Altenberg (personal communication, December 11, 2013) pointed out that if the leading right eigenvector $v(p)$ of D^pP is normalized so that $1^T v(p) = 1, p = 1, 2$ (which makes the leading left and right eigenvectors unique), in addition to the assumed $w(p)^T v(p) = 1$, then the formula for a simplifies to $a = N(0)^{2-b}\left(w(2)^T\pi\right) / \left(w(1)^T\pi\right)^b$.

3.1. When assumptions of Theorem 1 do not hold

If D is scalar, i.e., $D = dI$ for some $d > 0$, then $E(N(t)) = N(0)d^t = N(t)$ with probability 1, and $E((N(t))^2) = (N(0))^2 d^{2t} = (N(t))^2$ with probability 1, and $r(D^2P) = [r(DP)]^2 = d^2 r(P) = d^2$. So $\mathrm{Var}(N(t)) = 0$ while the second moment of $N(t)$ grows in proportion to d^{2t}. In this case, the formula for b in (8) reduces to $b = (2 \log d)/ \log d = 2$, but (1) and (2) no longer apply because the variance is 0. The behavior of the second moment in this example differs radically from the behavior of the variance.

Theorem 1 may also fail to hold for non-scalar D if, contrary to assumption, P is irreducible but not two-fold irreducible. The following example is taken from Cohen (in press) and is analyzed further here. Let $d > 1$ and $D = \begin{pmatrix} d & 0 \\ 0 & 1 \end{pmatrix}$, $P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. Here D is not scalar, P is column-stochastic and irreducible (but not primitive and not two-fold irreducible, because $P^TP = I$), and

$$\text{for } p \in (0, \infty), \quad D^p = \begin{pmatrix} d^p & 0 \\ 0 & 1 \end{pmatrix},$$

$$D^pP = \begin{pmatrix} 0 & d^p \\ 1 & 0 \end{pmatrix}, \qquad r\left(D^pP\right) = d^{p/2}.$$

Hence $r(D^2P) = d = [r(DP)]^2$. Now suppose the initial probability distribution of the Markov chain for $A(t)$ is the column vector $\pi = [\frac{1}{2}, \frac{1}{2}]^T$. Then $P\pi = \pi$, so the Markov chain is stationary with probability distribution π at every time. It is elementary to calculate that if t is odd, then

$$(E\,(N(t)))^2 = (N(0))^2 \left[\frac{d^{\frac{t+1}{2}} + d^{\frac{t-1}{2}}}{2}\right]^2,$$

$$E\left((N(t))^2\right) = (N(0))^2 \left[\frac{d^{t+1} + d^{t-1}}{2}\right],$$

$$\begin{aligned} \mathrm{Var}\,(N(t)) &= E\left((N(t))^2\right) - (E\,(N(t)))^2 \\ &= N(0)^2 d^{t-1}(d-1)^2/4 > 0, \quad \text{for } t \text{ odd}, \end{aligned}$$

and since we assumed $d > 1$, for increasing odd t we have

$$\lim_{\text{odd } t\to\infty} \mathrm{Var}\,(N(t)) = \infty.$$

But if t is even, then

$$(E\,(N(t)))^2 = (N(0))^2\, d^t, \qquad E\left((N(t))^2\right) = (N(0))^2\, d^t,$$

so for all even t we have $\mathrm{Var}\,(N(t)) = E\left((N(t))^2\right) - (E\,(N(t)))^2 = 0$. Hence, although $E(N(t))$ converges to an increasing exponential function for large t, $\mathrm{Var}(N(t))$ does not converge to an increasing exponential function for large t (the limit in (16) in the Appendix does not exist), so TL cannot hold as $t \to \infty$.

We do not know whether the conclusions of Theorem 1 always hold if D is not scalar and P is primitive but not two-fold irreducible. Our proof (Appendix) uses a strict eigenvalue inequality (27) to obtain the conclusions. That strict eigenvalue inequality holds if and only if P is two-fold irreducible (not merely primitive) and D is not scalar (Altenberg, 2013; Cohen, in press). Whether the conclusions of Theorem 1 can be reached by another route under weaker assumptions remains to be determined.

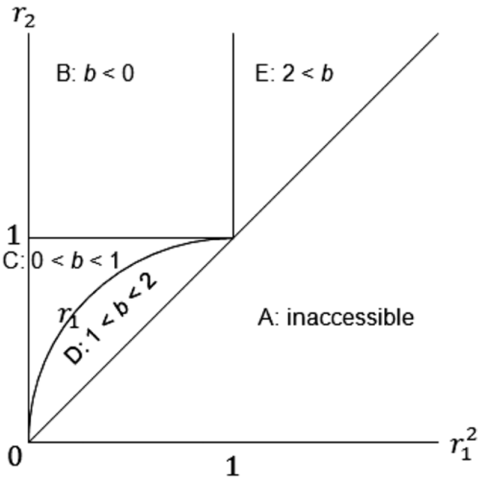


Fig. 1. Behavior of $b = \log[r(D^2P)]/\log[r(DP)]$ as a function of $r_1{}^2 = [r(DP)]^2$ and $r_2 = r(D^2P)$, assuming D has at least 2 distinct diagonal elements and P is column-stochastic and two-fold irreducible. The diagonal line represents $r_1{}^2 = r_2$. When $0 < r_1{}^2 < 1$, then $r_1 > r_1{}^2$.

3.2. Behavior of $b = \log[r(D^2P)]/\log[r(DP)]$

We shall describe the behavior of b over the nonnegative quadrant with abscissa $r_1{}^2 = [r(DP)]^2 = r(DPDP)$ and ordinate $r_2 = r(D^2P)$ (Fig. 1). The denominator of b is r_1 but the horizontal axis of Fig. 1 is $r_1{}^2$ (the exponent is 2). The most interesting (and unexpected) finding is that when $r_2 > 1$, as $r_1{}^2$ increases smoothly from left to right across the vertical half line in Fig. 1 at $r_1{}^2 = 1$, b jumps discontinuously from negative values (approaching $-\infty$) to positive values (starting from $+\infty$).

Because of (27) in the Appendix, (r_1^2, r_2) never falls on or below the diagonal line in Fig. 1 and the zone labeled A (including the diagonal line $r_1^2 = r_2$, where $b = 2$) is inaccessible under the conditions of Theorem 1. The value of b in the remaining zones is calculated in Table 1. Across the semi-infinite vertical line $r_1^2 = 1$, $r_2 \geq 1$, the value of b jumps discontinuously from $-\infty$ just left of the line to $+\infty$ just right of the line. Cohen (2013b) gave a concrete example of this discontinuity and a discussion of its ecological implications.

As t increases, $E(N(t))$ asymptotically increases exponentially if $r_1 > 1$ and $b > 2$ (zone E, Table 1). As t increases, $E(N(t))$ asymptotically decreases exponentially if $r_1 < 1$ and $b < 1$ (zones B, C, D, Table 1).

$N(t)$ decreases as t increases, or at least does not grow, in every environment if $\max d_{ii} \leq 1$. If $\max d_{ii} \leq 1$, then $1 < b < 2$ (zone D), because then at least one diagonal element of D is strictly less than 1 and none is greater than 1, so every non-zero element of DP is not greater than the corresponding element of P and at least one element of DP is less than the corresponding element of P. Hence $r(DP) < r(P) = 1$. Moreover, by the same reasoning, $D^2 \leq D$, $D^2 \neq D$, so $r_2 = r\left(D^2P\right) < r(DP) = r_1$. While $\max d_{ii} \leq 1$ is sufficient to assure $1 < b < 2$, it is not necessary for $1 < b < 2$.

As t increases, $N(t)$ increases (or at least does not decrease) in every environment if $\min d_{ii} \geq 1$. If $\min d_{ii} \geq 1$, then $2 < b$ (zone E), because then at least one diagonal element of D is strictly greater than 1 and none is less than 1, so every non-zero element of DP is at least as large as the corresponding element of P and at least one element of DP is larger than the corresponding element of P. Hence $r(DP) > r(P) = 1$. While $\min d_{ii} \geq 1$ is sufficient to assure $2 < b$, it is not necessary for $2 < b$.

To investigate numerically how often values of b fall in each zone in Fig. 1, we set the number of states of the environment to be $s = 2$ in two simple probabilistic models, both chosen for mathematical simplicity rather than realism. In model 1, we assigned each diagonal element of D an independent exponential distribution with parameter 1 (calculated as $d_{ii} = -\log U$, $i = 1, 2$, where U is uniformly distributed on [0,1] and independent for each i). Each element of the first row of P took the value of an independent uniform random variable on [0, 1], and $p_{2j} = 1 - p_{1j}, j = 1, 2$. For each pair (D, P) randomly generated in this way, we calculated b and the zone in Fig. 1. Approximately 49% of one million simulations yielded $1 < b < 2$ and approximately 40% yielded $2 < b$. In model 2, we assigned each diagonal element of D an independent lognormal distribution $\exp(\mathcal{N}(0, 1))$, each element of the first row of P an independent Beta(1/2, 1/2) distribution on [0,1] (probability density was concentrated at low and high extremes compared to the uniform distribution), and $p_{2j} = 1 - p_{1j}$, $j = 1, 2$. Approximately 35% of one million simulations yielded $1 < b < 2$ and approximately 58% yielded $2 < b$. The quantitative proportions in each zone varied between models, but both models assigned most values of b to $1 < b < 2$ or $2 < b$.

The estimates of the frequencies of each zone were subject to multinomial sampling error. To estimate the sampling uncertainty of a single proportion, it is valid to ignore the multinomial dependence among proportions and use a binomial distribution for a single proportion. For example, the standard deviation of the estimate of 5.4% for the frequency of occurrence of zone B in model 1 was approximately $\sqrt{0.054 \times (1 - 0.054)/10^6} = 0.02\%$. The standard deviation of the estimate of 49.1% for zone D in model 1 was approximately 0.05%.

The value of b in these simulations has no sampling error, because the simulation does not generate a trajectory of the dynamic process but generates only the values of D and P. Values of b are calculated numerically from D and P by an exact formula (8) without sampling error. Numerical error in the calculation of b is negligible here because no knife-edge equalities are reported, only inequalities involving b.

Both models yielded a few percent of simulations with $0 < b < 1$ (zone C). The existence of values of D and P that imply $0 < b < 1$ demonstrates that exponential dispersion models and Tweedie distributions do not include all distributions that obey TL in the limit of large t, because $0 < b < 1$ is impossible for exponential dispersion models and Tweedie distributions (Jørgensen, 1987, p. 133, Theorem 2; Jørgensen, 1997, p. 130, his Table 4.1).

The high proportion of simulated values of b such that $1 < b < 2$ and the low proportion of values such that $b < 1$ agreed qualitatively with the histograms of empirical values of b assembled by Anderson et al. (1982, their Fig. 2) and Kendal (2004, p. 196, his Fig. 1). This qualitative agreement supported the plausibility of the Markovian model analyzed here but did not verify any details of the mechanisms it assumed.

Table 1

Relations of $r_1^2 = [r(DP)]^2$, $r_2 = r(D^2P)$, and $b = \log[r(D^2P)]/\log[r(DP)]$ in the zones of Fig. 1 and in 10^6 simulations of each of two stochastic models described in the text.

Zone	Relations of r_1, r_2	$b = \log r_2/\log r_1$	Frequency (%)	
			Model 1	Model 2
A	$0 < r_2 \leq r_1^2$	Inaccessible	0	0
B	$0 < r_1^2 < r_1 < 1 < r_2$	$-\infty < b < 0$	5.4%	3.7%
C	$0 < r_1^2 < r_1 < r_2 < 1$	$0 < b < 1$	5.6%	3.2%
D	$0 < r_1^2 < r_2 < r_1 < 1$	$1 < b < 2$	49.1%	34.9%
E	$1 < r_1 < r_1^2 < r_2$	$2 < b$	39.9%	58.1%

4. Branching processes and birth-and-death processes

Taylor's power law of fluctuation applies asymptotically to other Markovian population processes in addition to the example just studied.

First, the discrete-generation Galton–Watson branching process (Bartlett, 1955, 1966, Section 2.3, Eq. (3), p. 40 in 1955, p. 42 in 1966) assumes that each individual of the tth generation independently has a stochastically distributed number of offspring in the next generation. If the initial number of individuals is $m_0 > 0$, and the mean and variance of the number of offspring per individual are $m \neq 1$ and v, respectively, then

$$m(t) = m^t m_0, \qquad v(t) = m^{2t} v_0 + m^{t-1}(1 - m^t) m_0 v/(1 - m).$$

Hence $\lim_{t\to\infty} t^{-1} \log m(t) = \log m$. The variance $v(t)$ of the number of individuals in generation t is a sum of two terms. The first term contains a factor v_0, here assumed positive, representing the variance in the 0th generation. The second term contains a factor v. If $v = 0$, then $v(t) = m^{2t} v_0 = v_0(m(t)/m_0)^2 = (v_0/m_0^2)(m(t))^2$ so TL holds exactly for all t with $b = 2$. Henceforth assume $v > 0$. Then

$$\begin{aligned} t^{-1} \log v(t) &= t^{-1} \log[m^{2t} v_0 + m^{t-1}(1 - m^t) m_0 v/(1 - m)] \\ &= t^{-1} \log[m^t \{m^t v_0 + m^{-1}(1 - m^t) m_0 v/(1 - m)\}] \\ &= \log m + t^{-1} \log[m^t v_0 \\ &\quad + m^{-1}(1 - m^t) m_0 v/(1 - m)]. \end{aligned}$$

If $m > 1$, then $t^{-1} \log v(t) \to 2 \log m$ as $t \to \infty$ and from (23), $b = \lim_{t\to\infty} t^{-1} \log v(t)/\lim_{t\to\infty} t^{-1} \log m(t) = 2$. If $0 < m < 1$, then as $t \to \infty$, $m^t v_0 + m^{-1}(1 - m^t) m_0 v/(1 - m) \to 0 + m_0 v/[m(1 - m)]$. Hence

$$\begin{aligned} &\lim_{t\to\infty} t^{-1} \log v(t) \\ &\quad = \log m + \lim_{t\to\infty} t^{-1} \log[m^t v_0 + m^{-1}(1 - m^t) m_0 v/(1 - m)] \\ &\quad = \log m + \lim_{t\to\infty} t^{-1} \log\left[\frac{m_0 v}{m(1 - m)}\right] = \log m. \end{aligned}$$

Hence $b = 1$. Thus if $0 < m < 1$, TL holds with $b = 1$, whereas if $m > 1$ or $v = 0$, then TL holds with $b = 2$. If $m = 1$, $v \neq 0$, then $v(t) = v_0 + t m_0 v$. Hence $\lim_{t\to\infty} t^{-1} \log v(t) = 0$ and $\lim_{t\to\infty} t^{-1} \log m(t) = 0$, so b is not defined.

For the Bellman–Harris and other processes, an asymptotic TL can be proved by very similar arguments from formulas for their mean and variance (Haccou et al., 2005, pp. 74–75).

Second, the continuous-time birth and death process assumes a birth rate λ per individual and a death rate μ per individual. Assuming one initial individual, independence among that individual and all offspring, and $\lambda > 0$, $\mu > 0$, $\lambda \neq \mu$, the mean and variance of the number of individuals at time t are (Bartlett, Section 3.4, Eq. (9), p. 70 in 1955, p. 74 in 1966)

$$m(t) = e^{(\lambda-\mu)t}, \qquad v(t) = \left(\frac{\lambda + \mu}{\lambda - \mu}\right) e^{(\lambda-\mu)t}[e^{(\lambda-\mu)t} - 1].$$

Hence $\lim_{t\to\infty} t^{-1}\log m(t) = \lambda - \mu$ and $t^{-1}\log v(t) = \lambda - \mu + t^{-1}\log\{(\frac{\lambda+\mu}{\lambda-\mu})[e^{(\lambda-\mu)t} - 1]\}$.

If $\lambda > \mu$, then as $t \to \infty$, $t^{-1}\log v(t) \to 2(\lambda - \mu)$. Hence $b = 2$. If $\mu > \lambda$, then as $t \to \infty$, $(\frac{\lambda+\mu}{\lambda-\mu})[e^{(\lambda-\mu)t} - 1] \to (\frac{\lambda+\mu}{\mu-\lambda})$ and $t^{-1}\log v(t) \to \lambda - \mu$. Hence $b = 1$. Thus if $\mu > \lambda$, TL holds with $b = 1$, whereas if $\lambda > \mu$, TL holds with $b = 2$. If $\lambda = \mu$, $v(t) = 2\lambda t$. Hence $\lim_{t\to\infty} t^{-1}\log v(t) = \lim_{t\to\infty} t^{-1}\log m(t) = 0$ and b is not defined.

Bartlett (1955, 1966) did not comment on the asymptotic relation between the mean and the variance in these models.

5. Multiple models lead to Taylor's law

A wide range of models can yield TL exactly or in the limit of large time. For example, a deterministic model of exponential clonal growth (Cohen, 2013a; Cohen et al., 2013b), the Galton–Watson branching process (with $v = 0$ or $m > 1$), and the birth and death process (with $\lambda > \mu$) all converge (for large time) to TL with $b = 2$. The gamma distribution satisfies TL exactly with $b = 2$. Hence $b = 2$ in TL need not indicate deterministic population growth. Both the model of Lewontin and Cohen (1969) with i.i.d. multiplicative factors (Cohen et al., 2013a) and the model of Markovian scalar multiplicative increments analyzed here predict TL asymptotically with b that may range from $-\infty$ to $+\infty$, depending on the parameters of the model. Anderson et al. (1982) gave further examples of Markovian processes consistent with TL. The Tweedie distributions analyzed by Jørgensen (1997) lead to TL with any real value of b excluding the interval (0, 1).

Consequently the power-law form and parameter values of TL can yield at most limited information about the underlying process. Interpreting the parameters of TL in terms of a specific model requires investigating whether the model's assumptions are verified. It is impossible to infer the underlying stochastic process solely from the observation that a random variable satisfies TL.

It is not surprising that diverse processes satisfy TL. If the mean changes as a non-constant exponential function of time (exactly or asymptotically) and the variance also behaves as any exponential function (constant or non-constant) of time (exactly or asymptotically), then the variance is a power-law function (constant or non-constant) of the mean (exactly or asymptotically) (Cohen, 2013a, Eqs. (13)–(15)). In the models just mentioned, the mean and variance are asymptotically exponential functions of time, so TL follows.

5.1. *Historical context*

In various stochastic ergodic theorems about population age structure and long-run population growth rate, Cohen (1976, 1977a,b) assumed Markovian transition matrices that were aperiodic and irreducible (i.e., primitive). Corollary 2 of Cohen (1977a, p. 25) also assumed that all the row-stochastic transition matrices of the Markov chain had a positive column. This assumption made them all two-fold irreducible (as Lee Altenberg pointed out privately). Cohen made no statements about the long-run growth rate of the variance (the central second moment) of population size.

Tuljapurkar (1982, p. 118) quoted formulas for asymptotic rates of growth of all moments about zero of population size, assuming that "the temporal sequence of environments is an ergodic aperiodic Markov process" (Tuljapurkar, 1982, p. 117). When the number of environmental states is finite, the transition matrix of an ergodic aperiodic Markov process is primitive, but not necessarily two-fold irreducible. Using β to denote the asymptotic long-run growth rate of the second moment about zero (his β equals our $r_2 = r(D^2P)$), he asserted (p. 118, after his Eq. (5)), but did not prove, that "Asymptotically then the variance must grow geometrically at rate β".

Tuljapurkar (1982, p. 117) did not explicitly assume that the matrix of vital rates (his multidimensional version of our scalar growth rate d_i) differed in at least two different environmental states, but it is reasonable to suppose that he thought so. If all $d_i = d > 0$, we showed (in an example above and in the Appendix Theorem 4) that the variance will not grow at the same long-run rate as the second raw moment because the variance over realizations at any time t is zero while the second moment about zero grows like d^{2t}. Assuming at least two different d_i (or matrices of vital rates, in Tuljapurkar's multidimensional case) in different environmental states, it appears to be unknown at present whether having a primitive transition matrix P guarantees that the variance has a long-run growth rate equal to the long-run growth rate of the second raw moment. A heuristic argument in favor of Tuljapurkar's (1982) claim is that, if P is primitive, then for a finite time T^*, $P^{T^*} > 0$ (every element is positive). So if we consider time in blocks of length T^*, then the blockwise transition matrix P^{T^*} is positive and therefore two-fold irreducible. Then our Theorem 1 would apply to the multiplicative process in blocks of time. If Tuljapurkar's (1982) claim that "the variance must grow geometrically at (his) rate β" is true, a detailed proof remains to be worked out.

Acknowledgments

I thank Lee Altenberg (as colleague and self-identified reviewer), Roy Malka, Michael Plank, Shripad Tuljapurkar, Meng Xu, and an anonymous reviewer for very helpful comments, Priscilla K. Rogerson for assistance, Michael Plank and the family of William T. Golden for hospitality during this work, and the US National Science Foundation grants EF-1038337 and DMS-1225529 and the Marsden Fund of the Royal Society of New Zealand (08-UOC-034) for partial support.

Appendix. Proofs of mathematical claims

The following proof of Theorem 1 assumes the two-fold irreducibility of P (the transition matrix of the Markov chain of environmental states) as a sufficient condition to prove that the second term of the variance, $[E(N(t))]^2$, can be neglected asymptotically compared to the first term, $E([N(t)]^2)$, and hence that the asymptotic rate of growth of $\mathrm{Var}(N(t))$ is given by $\log r(D^2P)$. Whether it is necessary that P be two-fold irreducible to attain this conclusion remains unknown.

The mean M and the variance V of every $A(t)$, for all $t = 0, 1, 2, \ldots$, are the mean and variance of $A(0)$ because, by construction, the sequence $\{A(t), t = 0, 1, 2, \ldots\}$ is stationary. So

$$M = E(A(0)) = \sum_{i=1}^{s} \pi_i d_i, \tag{9}$$

$$\begin{aligned} V &= \mathrm{Var}(A(0)) = E(A^2(0)) - (E(A(0)))^2 \\ &= \sum_{i=1}^{s} \pi_i d_i^2 - M^2 \geqslant 0. \end{aligned} \tag{10}$$

We assumed at least two distinct values in $\{d_1, \ldots, d_s\}$. Since $\pi > 0$, we have $V > 0$. Combining (6) and (7) gives

$$\begin{aligned} \lim_{t\to\infty} E[(N(t))^p]/[r(D^pP)]^t &= \lim_{t\to\infty} 1^T(D^pP)^t \pi N(0)^p/[r(D^pP)]^t \\ &= 1^T v(p) w(p)^T \pi N(0)^p. \end{aligned} \tag{11}$$

Taking $t^{-1}\log$ of both sides of (11) yields

$$\lim_{t\to\infty} t^{-1}\log E[(N(t))^p] - \log[r(D^pP)] = 0. \tag{12}$$

If $p = 1$, (12) becomes

$$\lim_{t\to\infty} t^{-1}\log E[N(t)] = \log[r(DP)]. \tag{13}$$

If $p = 2$, (12) becomes

$$\lim_{t\to\infty} t^{-1} \log E[(N(t))^2] = \log r(D^2P). \tag{14}$$

Now, assuming temporarily that the limits in (15) and (16) exist, define

$$\log \mu := \lim_{t\to\infty} t^{-1} \log E(N(t)), \tag{15}$$

$$\begin{aligned}\log \beta &:= \lim_{t\to\infty} t^{-1} \log \mathrm{Var}(N(t)) \\ &= \lim_{t\to\infty} t^{-1} \log(E(N^2(t)) - [E(N(t))]^2).\end{aligned} \tag{16}$$

(We use β for the long-run growth rate of the variance, not of the second raw moment, and *prove* that β equals the growth rate of the second raw moment under certain conditions.) By (13), the limit on the right side of (15) exists because $r(DP) > 0$, and

$$\log \mu = \log r(DP). \tag{17}$$

Moreover, since the variance is always nonnegative, the rate of growth of the second moment must be at least as large as the rate of growth of the square of the mean (Cauchy's inequality). Hence (Cohen, in press)

$$r(D^2P) \geq [r(DP)]^2. \tag{18}$$

These results are known (Bharucha, 1960; Furstenberg and Kesten, 1960; Bharucha, 1961; Cohen, 1976, 1977a,b; Tuljapurkar and Orzack, 1980; Tuljapurkar, 1982, 1986, 1990). But to prove that the long-run rate of growth of the variance (16) exists and equals the rate of growth of the second moment (14), we now invoke an additional assumption, two-fold irreducibility of P. We do not know whether this assumption can be weakened.

Theorem 2. *If P is two-fold irreducible and D is not scalar, then the limit in (16) exists and is given by (14), i.e.,*

$$\log \beta = \log r(D^2P). \tag{19}$$

Proof. Assuming temporarily that the limit in (16) exists, we have by definition (16) and result (14),

$$\begin{aligned}\log \beta &= \lim_{t\to\infty} t^{-1} \log \mathrm{Var}(N(t)) \\ &= \lim_{t\to\infty} t^{-1} \log \left\{ \frac{\mathrm{Var}(N(t))}{E[(N(t))^2]} E[(N(t))^2] \right\} \\ &= \lim_{t\to\infty} t^{-1} \log \left(\frac{E[(N(t))^2] - [E(N(t))]^2}{E[(N(t))^2]} \right) \\ &\quad + \lim_{t\to\infty} t^{-1} \log E[(N(t))^2] \\ &= \lim_{t\to\infty} t^{-1} \log(1 - [E(N(t))]^2/E[(N(t))^2]) \\ &\quad + \log r(D^2P).\end{aligned} \tag{20}$$

So $\log \beta = \log r(D^2P)$ if

$$\lim_{t\to\infty} t^{-1} \log(1 - [E(N(t))]^2/E[(N(t))^2]) = 0. \tag{21}$$

To prove (21), it suffices to show that

$$\lim_{t\to\infty} [E(N(t))]^2/E[(N(t))^2] = 0. \tag{22}$$

Using (11) with $p = 1$, squaring both sides, and dividing the result by (11) with $p = 2$ gives

$$\begin{aligned}&\lim_{t\to\infty} \left\{ \left(\frac{E[(N(t))]}{[r(DP)]^t} \right)^2 \left(\frac{[r(D^2P)]^t}{E[(N(t))^2]} \right) \right\} \\ &= \lim_{t\to\infty} \left\{ \left(\frac{(E[N(t)])^2}{E[(N(t))^2]} \right) \left(\frac{r(D^2P)}{[r(DP)]^2} \right)^t \right\} \\ &= \frac{(1^T v(1) w(1)^T \pi)^2}{1^T v(2) w(2)^T \pi} \equiv K, \quad 0 < K < \infty.\end{aligned}$$

If A is two-fold irreducible and D is not scalar, then $r(D^2P) > r^2(DP)$ (Cohen in press) so as $t \to \infty$, $(r(D^2P)/[r(DP)]^2)^t \to \infty$ geometrically fast and therefore necessarily $(E[N(t)])^2/E[(N(t))^2] \to 0$ geometrically fast. □

If every column of P equals the positive stationary probability distribution π, then $\{A(t),\ t = 0, 1, 2, \ldots\}$ is an i.i.d. sequence. In this case (Cohen et al., 2013a), $\log \mu = \log E(A(0)) = \log M$ and $\log \beta = \log E[(A(0))^2] = \log(V + M^2)$, where M and V are given explicitly by (9) and (10).

In the scalar non-stationary discrete Markovian model, the initial (equilibrium) probability vector π may be replaced by an arbitrary initial probability vector π_0. In this extension, $\{A(t),\ t = 0, 1, \ldots\}$ is no longer stationary. Then (5) would be replaced by $E[(N(1))^p] = 1^T(D^p\pi_0)N(0)^p$ and (6) by $E[(N(t))^p] = 1^T(D^pP)^{t-1}(D^p\pi_0)N(0)^p$. But a two-fold irreducible transition matrix is primitive (Altenberg, 2013), and a Markov chain specified by a primitive transition matrix and any initial probability vector is ergodic, so we speculate that the same limiting arguments should remain valid and at most minor modifications of the limiting formulas (8) might be expected.

Theorem 3 (*Cohen et al., 2013a*)**.** *If TL applies to $N(t)$ exactly for all $t \in \Theta$ or in the limit as $t \to \infty$, then $\log \beta = b \log \mu$. If $\log \mu \neq 0$, then*

$$b = \log \beta / \log \mu. \tag{23}$$

The proof is so easy that we repeat it. Divide (2) by t and let $t \to \infty$. Then, from the definitions of $\log \mu$ (15) and $\log \beta$ (16), $\log \beta = 0 + b \log \mu$. If $\log \mu \neq 0$, then (23) is well defined. □

We now prove the main result.

Theorem 1. *In the scalar stationary discrete Markovian model, assume that D is not scalar, P is two-fold irreducible, and $\log \mu = \log[r(DP)] \neq 0$. Then $N(t)$ obeys TL (2) in the limit as $t \to \infty$ with*

$$\begin{aligned}b &= \log[r(D^2P)]/\log[r(DP)], \\ a &= (1^T v(2) w(2)^T \pi N(0)^2)/(1^T v(1) w(1)^T \pi N(0))^b.\end{aligned} \tag{24}$$

Proof. The assumptions imply that the denominators in both expressions in (24) are non-zero, so both expressions are well defined. Now taking log of the left-most member and the right-most member of (11) yields, for any real p,

$$\begin{aligned}&\lim_{t\to\infty} \{\log E[(N(t))^p] - t \log[r(D^pP)]\} \\ &= \log(1^T v(p) w(p)^T \pi N(0)^p).\end{aligned} \tag{25}$$

Let γ be any real number. In (25), put $p = 1$, multiply the resulting equation by γ, and subtract that result from (25) in which $p = 2$. Then

$$\begin{aligned}&\lim_{t\to\infty} \{\log E[(N(t))^2] - t \log[r(D^2P)] \\ &\quad - \gamma \log E[(N(t))] + \gamma t \log[r(DP)]\} \\ &= \log(1^T v(2) w(2)^T \pi N(0)^2) - \gamma \log(1^T v(1) w(1)^T \pi N(0)).\end{aligned}$$

Then if we set $\gamma = b$ and use a taken from (24), we find

$$\lim_{t\to\infty} \{\log E[(N(t))^2] - b \log E[N(t)]\} = \log a. \tag{26}$$

Then (14), (19) and (26) imply TL (2):

$$\lim_{t\to\infty}\{\log \mathrm{Var}[N(t)] - b\log E[N(t)]\} = \log a. \quad \square$$

Theorem 4. *Assume that D is any diagonal matrix with positive diagonal elements, P is any column-stochastic matrix, and $\log[r(DP)] \neq 0$. Define $r_h := r(D^hP) > 0,\ h = 1, 2$. Then $b := \log r_2/\log r_1$ is well defined and finite since (by assumption) $\log r_1 \neq 0$. Also assume that $N(t)$ is defined by the multiplicative process (3). If the growth rate is the same in every environment, i.e., $D = dI,\ d > 0, d \neq 1$, or if the sequence of environments is a deterministically repeated cycle, i.e., P is a blockwise permutation matrix with constant values of $A(t)$ within each block, or both, then $b = 2$. But if at least two diagonal elements of D are distinct, all diagonal elements of D are positive, and P is two-fold irreducible, then $b \neq 2$.*

Proof. If $D = dI,\ d > 0, d \neq 1$, then $b = 2\log d/\log d = 2$, for any column-stochastic P. (The assumption $d \neq 1$ excludes $r_1 = r(DP) = 1$ and guarantees that b is well defined and finite.)

If $D \neq dI$ for any $d > 0$ and if $P = I$, then $r_h = r(D^h) = (\max d_{ii})^h,\ h = 1, 2$, hence if $\max d_{ii} \neq 1$, then $b = 2$. If $P \neq I$ is a blockwise permutation matrix with constant values of $A(t)$ within each block, again $r_2 = r_1^2$. Hence if $r_1 \neq 1$ then $b = 2$. (The long but still elementary proof reduces this case to the previous case by raising P to the power needed to produce a block diagonal matrix and replacing the diagonal elements of D by appropriate corresponding products of the d_{ii}.)

Now resume the assumptions of Theorem 1: at least two diagonal elements of D are distinct, all diagonal elements of D are positive, and P is two-fold irreducible. (If $A(t)$ is independently and identically distributed for all τ, then $P = \pi 1^T$ is positive and therefore two-fold irreducible.) Now strict inequality holds in (18) (Cohen, in press), or equivalently

$$r_2 > r_1^2, \tag{27}$$

which implies that b cannot equal 2. $\square$

References

Altenberg, Lee, 2013. A sharpened condition for strict log-convexity of the spectral radius via the bipartite graph. Linear Algebra Appl. 438 (9), 3702–3718. http://dx.doi.org/10.1016/j.laa.2013.01.008.

Anderson, R.M., Gordon, D.M., Crawley, M.J., Hassell, M.P., 1982. Variability in the abundance of animal and plant species. Nature 296, 245–248.

Azevedo, R.B.R., Leroi, A.M., 2001. A power law for cells. Proc. Natl. Acad. Sci. 98, 5699–5704.

Ballantyne IV, F., Kerkhoff, A.J., 2007. The observed range for temporal mean–variance scaling exponents can be explained by reproductive correlation. Oikos 116, 174–180.

Bartlett, M.S., 1955,1966. Introduction to Stochastic Processes with Special Reference to Methods and Applications. Cambridge University Press, Cambridge, first ed., 2nd ed..

Bechinski, E.J., Pedigo, L.P., 1981. Population dispersion and development of sampling plans for Orius insidiosus and Nabis spp. in soybeans. Environ. Entomol. 10 (6), 956–959.

Bharucha, B.H., 1960. Stability of Random Linear Systems with Markov Parameters, Vol. 60. University of California, Institute of Engineering Research, issue no. 289, June.

Bharucha, B.H., 1961. On the stability of randomly varying systems. In: Institute of Engineering Research Series No. 60. Issue No. 381 July 5. Contract No. AF 19(604)-5466, Project No. 5632, Task No. 56320, Scientific Report No. 7. Armed Services Technical Information Agency (ASTIA). University of California, Arlington, p. 83.

Caswell, H., 2001. Matrix Population Models: Construction, Analysis and Interpretation, 2nd ed. Sinauer Associates, Inc., Sunderland, MA.

Cohen, J.E., 1976. Ergodicity of age structure in populations with Markovian vital rates, I: countable states. J. Amer. Statist. Assoc. 71, 335–339.

Cohen, J.E., 1977a. Ergodicity of age structure in populations with Markovian vital rates, II: general states. Adv. Appl. Probab. 9, 18–37.

Cohen, J.E., 1977b. Ergodicity of age structure in populations with Markovian vital rates, III: finite-state moments and growth rates; an illustration. Adv. Appl. Probab. 9, 462–475.

Cohen, J.E., 2013a. Taylor's power law of fluctuation scaling and the growth-rate theorem. Theor. Popul. Biol. 88, 94–100. http://www.sciencedirect.com/science/article/pii/S004058091300035X.

Cohen, J.E., 2013b. Taylor's law and abrupt biotic change in a smoothly changing environment. Theor. Ecol. http://dx.doi.org/10.1007/s12080-013-0199-z. Published Online 12 September 2013.

Cohen, J.E., 2014. Cauchy inequalities for the spectral radius of products of diagonal and nonnegative matrices. Proc. Amer. Math. Soc. in press.

Cohen, J.E., Xu, M., Schuster, W.S.F., 2012. Allometric scaling of population variance with mean body size is predicted from Taylor's law and density-mass allometry. Proc. Natl. Acad. Sci. 109 (39), 15829–15834. http://dx.doi.org/10.1073/pnas.1212883109.

Cohen, J.E., Xu, M., Brunborg, Helge, 2013b. Taylor's law applies to spatial variation in a human population. Genus 69 (1), 25–60. http://scistat.cilea.it/index.php/genus/article/view/491.

Cohen, J.E., Xu, M., Schuster, W.S.F., 2013a. Stochastic multiplicative population growth predicts and interprets Taylor's power law of fluctuation scaling. Proc. R. Soc. Ser. B 280, 20122955. http://dx.doi.org/10.1098/rspb.2012.2955.

Davidian, M., Carroll, R.J., 1987. Variance function estimation. J. Amer. Statist. Assoc. 82 (400), 1079–1091. Stable URL: http://www.jstor.org/stable/2289384.

Eisler, Z., Bartos, I., Kertész, J., 2008. Fluctuation scaling in complex systems: Taylor's law and beyond. Adv. Phys. 57, 89–142.

Engen, S., Lande, R., Saether, B.-E., 2008. A general model for analyzing Taylor's spatial scaling laws. Ecology 89, 2612–2622.

Furstenberg, H., Kesten, H., 1960. Products of random matrices. Ann. Math. Statist. 31 (2), 457–469.

Gantmacher, F.R., 1960. Theory of Matrices. Chelsea Publishing Company, New York.

Gillis, D.M., Kramer, D.L., Bell, G., 1986. Taylor's power law as a consequence of Fretwell's ideal free distribution. J. Theoret. Biol. 123, 281–287.

Giometto, Andrea, Altermatt, Florian, Carrara, Francesco, Maritan, Amos, Rinaldo, Andrea, 2013. Scaling body size fluctuations. Proc. Natl. Acad. Sci. USA 110 (12), 4646–4650. www.pnas.org/cgi/doi/10.1073/pnas.1301552110.

Haccou, P., Jagers, P., Vatutin, V.A., 2005. Branching Processes: Variation, Growth, and Extinction of Populations. Cambridge University Press, Cambridge, New York.

Jørgensen, B., 1987. Exponential dispersion models. J. R. Stat. Soc. Ser. B 49, 127–162.

Jørgensen, B., 1997. The Theory of Dispersion Models. Chapman & Hall, London.

Kaltz, O., Escobar-Páramo, P., Hochberg, M.E., Cohen, J.E., 2012. Bacterial microcosms obey Taylor's law: effects of abiotic and biotic stress and genetics on mean and variance of population density. Ecol. Process. 1, 5.

Keeling, M.J., 2000. Simple stochastic models and their power-law type behavior. Theor. Popul. Biol. 58, 21–31.

Keeling, M., Grenfell, B., 1999. Stochastic dynamics and a power law for measles variability. Philos. Trans. R. Soc. Lond. Ser. B 354, 769–776.

Kemp, A.W., 1987. Families of discrete distributions satisfying Taylor's power law. Biometrics 43, 693–699.

Kendal, W.S., 2004. Taylor's ecological power law as a consequence of scale invariant exponential dispersion models. Ecol. Complex. 1, 193–209.

Kendal, W.S., Jørgensen, Bent, 2011. Taylor's power law and fluctuation scaling explained by a central-limit-like convergence. Phys. Rev. E 83, 066115:1–066115:7. http://dx.doi.org/10.1103/PhysRevE.83.066115.

Kilpatrick, A.M., Ives, A.R., 2003. Species interactions can explain Taylor's power law for ecological time series. Nature 422, 65–68.

Klein, A.M., Simons, B.D., 2011. Universal patterns of stem cell fate in cycling adult tissues. Development 138, 3103–3111.

Kogan, Marcos, Ruesink, William G., McDowell, Kenneth, 1974. Spatial and temporal distribution patterns of the bean leaf beetle, Cerotoma trifurcata (Forster), on soybeans in Illinois. Environ. Entomol. 3 (4), 607–617.

Lepš, J., 1993. Taylor's power law and the measurement of variation in the size of populations in space and time. Oikos 68 (2), 349–356.

Lewontin, R.C., Cohen, D., 1969. On population growth in a randomly varying environment. Proc. Natl. Acad. Sci. USA 62, 1056–1060.

O'Cinneide, Colm, 2000. Markov additive processes and Perron–Frobenius eigenvalue inequalities. Ann. Probab. 28 (3), 1230–1258. Stable URL: http://www.jstor.org/stable/2652986.

Park, Heungsun, Cho, Kijong, 2004. Use of covariates in Taylor's power law for sequential sampling in pest management. J. Agric. Biol. Environ. Stat. 9 (4), 462–478. http://dx.doi.org/10.1198/108571104X15746.

Perry, J.N., 1988. Some models for spatial variability of animal species. Oikos 51, 124–130.

Perry, J.N., Taylor, L.R., 1985. Ades: new ecological families of species-specific frequency distributions that describe repeated spatial samples with an intrinsic power-law variance–mean property. J. Anim. Ecol. 54, 931–953.

Ramsayer, J., Fellous, S., Cohen, J.E., Hochberg, M.E., 2011. Taylor's law holds in experimental bacterial populations but competition does not influence the slope. Biol. Lett. 1–4. http://dx.doi.org/10.1098/rsbl.2011.0895.

Taylor, L.R., 1961. Aggregation, variance and the mean. Nature 189, 732–735.

Taylor, L.R., 1984. Assessing and interpreting the spatial distributions of insect populations. Annu. Rev. Entomol. 29, 321–357.

Tuljapurkar, S.D., 1982. Population dynamics in variable environments. II. Correlated environments, sensitivity analysis and dynamics. Theor. Popul. Biol. 21, 114–140.

Tuljapurkar, S.D., 1986. Demographic applications of random matrix products. In: Cohen, J.E., Kesten, H., Newman, C.M. (Eds.), Random Matrices and Their Applications. In: Contemporary Mathematics, vol. 50. American Mathematical Society, Providence, RI.

Tuljapurkar, S.D., 1990. Population Dynamics in Variable Environments. In: Lecture Notes in Biomathematics, vol. 85. Springer-Verlag, Berlin, New York.

Tuljapurkar, Shripad, Gaillard, Jean-Michel, Coulson, T., 2009. From stochastic demography to life histories and back. Phil. Trans. R. Soc. B 364, 1499–1509.

Tuljapurkar, S.D., Orzack, S.H., 1980. Population dynamics in variable environments. I. Long run growth rates and extinction. Theor. Popul. Biol. 18, 314–342.

Tweedie, M.C.K., 1946. The regression of the sample variance on the sample mean. J. Lond. Math. Soc. 21, 22–28.

Tweedie, M.C.K., 1947. Functions of a statistical variate with given means, with special reference to Laplacian distributions. Proc. Cambridge Philos. Soc. 43, 41–49.

Tweedie, M.C.K., 1984. An index which distinguishes between some important exponential families. In: Ghosh, J.K., Roy, J. (Eds.), Statistics: Applications and New Directions. Indian Statistical Institute, Calcutta, pp. 579–604.

Wang, You-Gan, Zhao, Yuning, 2007. A modified pseudolikelihood approach for analysis of longitudinal data. Biometrics 63 (3), 681–689. Stable URL: http://www.jstor.org/stable/4541399.

Wilson, L.T., Sterling, W.L., Rummel, D.R., DeVay, J.E., 1989. Quantitative sampling principles in cotton. In: Frisbie, Raymond E., El-Zik, Kamal M., Ted Wilson, L. (Eds.), Integrated Pest Management Systems and Cotton Production. John Wiley & Sons, Inc, pp. 85–119 (Chapter 5).