Taylor's power law and fluctuation scaling explained by a central-limit-like convergence

Wayne S. Kendal* and Bent Jørgensen†

Division of Radiation Oncology, University of Ottawa, 501 Smyth Road, Ottawa, Ontario, Canada K1H 8L6

Department of Mathematics and Computer Science, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark

(Received 3 September 2010; revised manuscript received 3 May 2011; published 22 June 2011)

A power function relationship observed between the variance and the mean of many types of biological and physical systems has generated much debate as to its origins. This Taylor's law (or fluctuation scaling) has been recently hypothesized to result from the second law of thermodynamics and the behavior of the density of states. This hypothesis is predicated on physical quantities like free energy and an external field; the correspondence of these quantities with biological systems, though, remains unproven. Questions can be posed as to the applicability of this hypothesis to the diversity of observed phenomena as well as the range of spatial and temporal scales observed with Taylor's law. We note that the cumulant generating functions derived from this thermodynamic model correspond to those derived over a quarter century earlier for a class of probabilistic models known as the Tweedie exponential dispersion models. These latter models are characterized by variance-to-mean power functions; their phenomenological basis rests with a central-limit-theorem-like property that causes many statistical systems to converge mathematically toward a Tweedie form. We review evaluations of the Tweedie Poisson-gamma model for Taylor's law and provide three further cases to test: the clustering of single nucleotide polymorphisms (SNPs) within the horse chromosome 1, the clustering of genes within human chromosome 8, and the Mertens function. This latter case is a number theoretic function for which a thermodynamic model cannot explain Taylor's law, but where Tweedie convergence remains applicable. The Tweedie models are applicable to diverse biological, physical, and mathematical phenomena that express power variance functions over a wide range of measurement scales; they provide a probabilistic description for Taylor's law that allows mechanistic insight into complex systems without the assumption of a thermodynamic mechanism.

DOI: 10.1103/PhysRevE.83.066115 PACS number(s): 89.75.Da, 05.40.−a, 87.23.Cc, 89.75.Hc

I. INTRODUCTION

Taylor's power law was originally described as an empirical power function relationship between the variance and mean of the number of individuals of a species per unit area $\langle N\rangle$ as they are distributed over their habitat

$$\langle N^2\rangle - \langle N\rangle^2 = a\langle N\rangle^p \tag{1}$$

(a and p are constants) [1]. Since Taylor's description in 1961 [1], this scaling relationship has been repeatedly demonstrated within ecological systems and other biological processes ranging from chromosomal structure to population genetics to regional organ blood flow to cancer metastasis and epidemiology [2–8]. More recently, the physics community has shown interest in this relationship, in the context of the statistical mechanics of fluctuation scaling [9,10]. Taylor's law has its appeal to field ecologists, and other biologists, as a simple means to assess clustering: $p = 1$ indicates a random, or Poisson-distributed, pattern of individuals whereas $p > 1$ indicates clustering.

Fronczak and Fronczak have recently proposed that Taylor's law can be explained on the basis of the second law of thermodynamics through a maximum entropy principle [10]. By this means, they concluded, it is not necessary "to invoke any stochastic models to explain phenomena such as aggregation effects in different populations." Taylor's law is remarkable in that it is evident over the scale of a single chromosome [7,8] to the lungs of mice [2], a farmer's field [11], and upward to the breadth of the British Isles [12]. How the macroscopic fluctuations that characterize many of these observations might arise as a consequence of thermodynamic mechanisms is an intriguing question. On the basis of the fluctuation theorem [13] (assuming this theorem is applicable) one would expect a phase-space trajectory of duration t and with entropy production $\sum_t = \tilde{A}$ to be related to a trajectory of opposite magnitude and entropy loss such that

$$\frac{\mathrm{Prob}\left(\sum_t = \tilde{A}\right)}{\mathrm{Prob}\left(\sum_t = -\tilde{A}\right)} = \exp(\tilde{A}). \tag{2}$$

The probability of observing a transient or localized fluctuation associated with an entropy loss decreases exponentially as the time period, or size of the system, under observation increases. Thus in order for the larger fluctuations, evident with Taylor's law, to be explained by the second law of thermodynamics some additional mechanism(s) would seem necessary.

In this regard Fronczak and Fronczak have proposed to relate the macroscopic states to a microscopic system using a density of states function $g(N)$. They postulated a statistical distribution expressed in terms of the free energy of the system $F(\tilde{\mu})$, defined for an external field $\tilde{\mu}$ coupled to the ensemble of objects under study, and they related the variance in the mean density of objects $\langle N^2\rangle - \langle N\rangle^2$ to $F(\tilde{\mu})$ using a fluctuation dissipation relation. They then derived the probability generating function that corresponds to the frequency distribution of the objects yielding Taylor's law.

Over the past half-century since Taylor's initial description there have been many attempts to explain his law, yet none of these have found general acceptance [3,9,14–25].

*wkendal@ottawahospital.on.ca

†bentj@stat.sdu.dk

 ©2011 American Physical Society

Here we examine the Fronczak hypothesis with regard to its applicability in biological and numerical systems. We will show that this model is mathematically equivalent to one predicated on a mathematical, central-limit-theorem-like effect that does not require the assumption of a thermodynamic mechanism.

II. DIFFICULTIES WITH THE THERMODYNAMIC MODEL

Schrödinger, in his book *What is life?* [26], was one of the first physicists to apply thermodynamic theory to living systems. Since his seminal description there has been only slow progress with such applications, given the difficulties with the nonequilibrium nature of biological systems and their complexity. The validity of physical descriptions of this nature requires that the physical quantities being proposed have a direct and verifiable correspondence to the system(s) being described. For the systems known to exhibit Taylor's law, we found it difficult (if not impossible) to establish a biological correspondence with the postulated free energy and external physical field. To the best of our knowledge, within ecological systems, there has been no demonstration of an external field that relates to the aggregation of animals and plants.

With regard to the clustering of gene structures and single nucleotide polymorphisms (SNPs) that manifest Taylor's law, there exists an experimental and theoretical basis to explain the segregation of polymorphisms at the level of the cistron [27], and there exists evidence for evolutionary rearrangements, duplications, and deletions of protochromosomal segments that presumably now comprise present day chromosomes [28]. However, it is not clear how the statistical mechanical mechanism proposed by Fronczak and Fronczak would apply to the clustering of SNPs and gene structures that have evolved over many eons within different organisms.

For the heterogeneity of regional organ blood flow, which also obeys Taylor's law, there exists data to support gamma-distributed blood flow at the level of capillaries [29], and data to show how such flow at the microscopic level could relate to macroscopic heterogeneities [6]. However, it is not clear how the postulated thermodynamic parameters could correspond to measurable biological variables.

Similarly, the placement of houses over the Tonami Plain in Japan has been shown to obey Taylor's law [3]. It is not clear here, either, what the postulated external field $\tilde{\mu}$ would correspond to or how the free energy $F(\tilde{\mu})$ would be assessed. Fronczak and Fronczak, in the Appendix of their paper, acknowledged a lack of understanding of the meaning of their external field parameter $\tilde{\mu}$, yet they went on to speculate that $1/\tilde{\mu}$ be considered analogous to temperature in thermodynamics [10]. The question we have then is: What would be the physical nature of this temperature analogue that could act over large geographic regions to affect the clustering of plants and animals, that could act across species barriers and over millions of years of evolution to affect the clustering of chromosomal structures, that could act upon different germlines separated by expanses of geography and over hundreds of thousands of years to affect SNP maps, that could act within organs to affect regional blood flow heterogeneity, and that could act over an entire coastal plain to affect where homes are built?

In addition, we found the constraints on the exponent p implied by the Fronczak formulas for $\langle N\rangle$ [their Eq. (11)] and the related free energy $F(\tilde{\mu})$ [their Eq. (12)] of interest. Do these constraints imply that values of the exponent $p<1$ do not occur and that values of $p>2$ could be expected in ecological systems? A comprehensive review of ecological data has shown that, within the limits of measurement error, the exponent p ranged within $1<p<2$ [25]. In addition, several other biological processes have appeared similarly constrained [2–8]. Fronczak and Fronczak's paper did not appear to provide any explanation for such empirical constraints, although we will provide one below.

III. TWEEDIE EXPONENTIAL DISPERSION MODELS

To further investigate this issue we used Fronczak and Fronczak's equations (11),(12), and (B9) [10] to derive the cumulant generating function (CGF) $K(s)$ corresponding to each of three different cases defined by the exponent p

$$K(s)=\begin{cases}(Xe^{-a\tilde{\mu}}/a)(e^{as}-1) & \text{for} \quad p=1,\\ \frac{[(p-1)a\tilde{\mu}+X]^{\alpha}}{a(2-p)}\left\{\left(1-\frac{s(p-1)a}{(p-1)a\tilde{\mu}+X}\right)^{\alpha}-1\right\} & \text{for} \quad p\neq 1,2,\\ \frac{-1}{a}\log\left(1-\frac{as}{a\tilde{\mu}+X}\right) & \text{for} \quad p=2.\end{cases} \tag{3}$$

Here s is the generating function variable, $\tilde{\mu}$ describes the external field, X is a constant, and $\alpha=(p-2)/(p-1)$. The first case from Eq. (1) ($p=1$) describes a Poisson distribution, the second (with$1<p<2$) a compound Poisson-gamma distribution, and the third ($p=2$) a gamma distribution. We will discuss these CGFs in more detail below, but first it would be useful to review another application of the maximum entropy principle.

Consider the derivation of the Maxwell-Boltzmann distribution: In the context of physical laws one can maximize Boltzmann's H function, or entropy, with the constraint that the system's kinetic energy should be proportional to the absolute temperature, to provide the required result [30]. Alternatively, one might consider an ideal gas as a statistical system, in which a particle's velocity is derived from the momentum transfer from n collisions with other particles. A velocity distribution thus results from the convolution of n momentum exchange distributions. Provided that the momentum distributions are statistically independent and they have finite means and variances, the central limit theorem (CLT) would imply a Gaussian distribution for large n, yielding the Maxwell-Boltzmann distribution [30]. This is an equivalent derivation, grounded in the abstraction of mathematics rather than physical principle. In the context of the Maxwell-Boltzmann distribution both derivations are equally valid, yet the CLT has an applicability that goes beyond the *ad hoc* physical model.

Wigner recognized "the sovereign role of mathematics" in physical theory and the richness of insight that could be so derived [31]. We propose an alternative explanation for Taylor's law, based on the mathematical theory of errors and exponential dispersion models. These models were developed to analyze error distributions arising from generalized linear models [32]. Consider the class

of exponential dispersion models that are invariant under scale transformation. Consequent to this requirement, this class is characterized by Taylor's law $\langle N^2\rangle - \langle N\rangle^2 = a\langle N\rangle^p$ [32]. These particular models have come to be known as the Tweedie exponential dispersion models, being named after the man who first described them in 1984 [33].

There are several different Tweedie models, each determined by the values expressed by the exponent p: For $p < 0$ we have the extreme stable distributions; for $p = 0$, the Gaussian distribution; $p = 1$, the Poisson distribution; $1 < p < 2$, the compound Poisson-gamma (PG) distribution; $p = 2$, the gamma distribution; $2 < p < 3$, the positive stable distributions; $p = 3$, the inverse Gaussian distribution; $p > 3$, the positive stable distributions; and $p = \infty$, the extreme stable distributions. For the range $0 < p < 1$ no Tweedie model exists.

For comparison to Eq. (3) we provide here the CGFs for the additive forms of the Tweedie models [32]

$$K_p^*(s;\theta,\lambda) = \begin{cases} \lambda e^{\theta}\,(e^s - 1) & \text{for} \quad p = 1, \\ \lambda\kappa_p(\theta)\left\{\left(1 + \frac{s}{\theta}\right)^{\alpha} - 1\right\} & \text{for} \quad p \neq 1,2, \\ -\lambda \log\left(1 + \frac{s}{\theta}\right) & \text{for} \quad p = 2. \end{cases} \tag{4}$$

Here λ is the index parameter, θ the canonical parameter, and the cumulant function $\kappa_p(\theta)$ is given by [32]

$$\kappa_p(\theta) = \begin{cases} e^{\theta} & \text{for} \quad p = 1, \\ \frac{\alpha-1}{\alpha}\left(\frac{\theta}{\alpha-1}\right)^{\alpha} & \text{for} \quad p \neq 1,2, \\ -\log(-\theta) & \text{for} \quad p = 2. \end{cases} \tag{5}$$

Indeed, a comparison of Eqs. (3) and (4) reveals similarities in form. If we choose the index parameter such that

$$\lambda = \begin{cases} X & \text{for} \quad p = 1, \\ a^{\alpha-1} & \text{for} \quad p \neq 1,2, \\ 1/a & \text{for} \quad p = 2, \end{cases} \tag{6}$$

and the canonical parameter

$$\theta = \begin{cases} -\tilde{\mu} & \text{for} \quad p = 1, \\ -\tilde{\mu} + (\alpha - 1)X/a & \text{for} \quad p \neq 1,2, \\ -\tilde{\mu} - X/a & \text{for} \quad p = 2, \end{cases} \tag{7}$$

then the Fronczak CGFs [Eq. (3)] and the Tweedie CGFs [Eq. (4)] are seen to be equivalent, and the problematic external field $\tilde{\mu}$ can be directly related to the canonical parameter θ.

The Tweedie PG distribution, for the case $1 < p < 2$, corresponds to the majority of biological observations of Taylor's law. Its probability density function is not known in closed form but can be expressed as [32]

$$p^*(z;\theta,\lambda,\alpha) = c_p^*(z;\lambda)\exp[\theta z - \lambda\kappa_p(\theta)], \tag{8}$$

where

$$c_p^*(z;\lambda) = \begin{cases} \frac{1}{z}\sum_{n=1}^{\infty} \lambda^n \kappa_p^n(-1/z)\big/\Gamma(-\alpha n)n! & \text{for} \quad z > 0, \\ 1 & \text{for} \quad z = 0. \end{cases} \tag{9}$$

These Tweedie distributions are specified by three independent adjustable parameters: α, λ, and θ. The parameter α relates to Taylor's power law exponent p, $\alpha = (p-2)/(p-1)$; λ and θ are analogous to the shape and scale parameters conventionally used with statistical distributions. One can fit the Tweedie PG cumulative distribution function (CDF) to an empirical CDF, derived from data that exhibit Taylor's law, to test the model. One may also compare the value for p derived from the CDF with that estimated from Taylor's law to further test the Tweedie distribution. If the comparison of the CDFs is discordant, or if the values for p are significantly different, then the hypothesis that the PG distribution can be used to describe Taylor's law would be falsified.

IV. TWEEDIE CONVERGENCE THEOREM

The CLT provides insight into the origin of Maxwell-Boltzmann statistics; the Tweedie models are founded on a related convergence property that provides insight into the origin of Taylor's law and fluctuation scaling. The Tweedie convergence theorem [34] shows that for exponential dispersion models $\mathrm{E_D}(\mu,\sigma^2)$, with mean μ and variance $\sigma^2 V(\mu)$, and unit variance functions that approximate the form $V(\mu) \propto \mu^p$ as either $\mu \to 0$ or $\mu \to \infty$ then $\mathrm{c}^{-1}\mathrm{E_D}(c\mu,\sigma^2 c^{2-p})$ will converge to the form of a Tweedie model as the constant $c \to 0$ or $c \to \infty$. Since the variance functions for many probability distributions will approximate $V(\mu) \propto \mu^p$ for very small or very large values of μ, the variance-to-mean power function will appear as a focus of convergence for a wide variety of distributions. This convergence theorem appears to relate to stable generalizations of the CLT [32], and for these reasons many types of non-Gaussian data will manifest with Taylor's law; it is thus more general than the alternative convergence properties that have been proposed [9]. Summarized here, the Tweedie convergence theorem tells us that any statistical model or simulation designed to produce Taylor's law must, on mathematical grounds alone, converge to the form of one of the Tweedie models.

V. EVIDENCE TO SUPPORT THE TWEEDIE POISSON-GAMMA MODEL FOR TAYLOR'S LAW

At this point it could be surmised that we have two different hypotheses to explain Taylor's law that are both based upon the same CGFs [Eq. (4)]. Like with the Maxwell-Boltzmann distribution one might surmise that both hypotheses appear equally justifiable. However, we have not yet reviewed in detail systems where the Tweedie hypothesis has been tested.

In cancer biology the experimental metastasis assay has been extensively used to assess the metastatic potential of tumor cell lines. This assay involves the intravenous injection of suspensions of isolated tumor cells into groups of age-matched syngeneic mice with the subsequent enumeration of the numbers of lung metastases from each animal. When data from identically treated mice are compared, the numbers of lung metastases per mouse varies more than would be expected on the basis of a Poisson distribution. If different cancer cell clones, each with distinct metastatic potentials, are evaluated this way one can plot the variance of the numbers of metastases per animal for each clone against the corresponding

mean to find Taylor's law [35]. Taylor's law can also be demonstrated with human metastases; these patterns appear to parallel physiological heterogeneities in regional organ blood flow [36,37]. Indeed, physiologists have long recognized an empirical power function relationship between the relative dispersion of blood flow measured over tissue blocks of different sizes [38]. This blood flow relationship exactly corresponds to Taylor's law; it can be attributed to a gamma-distributed capillary blood flow through Poisson-distributed restrictive sites in the microcirculation, consistent with a PG distribution [39]. If the number of organ metastases represents a random function directly proportional to regional blood flow, Taylor's law would then manifest through the numbers of hematogenous metastasis [40]. Tests of the PG distribution for metastases can be conducted by comparing the empirical and theoretical CDFs for the numbers of experimental metastases sustained within by groups of syngeneic mice that that been administered tumor cells from different clones, as well as by comparing the estimates of p derived from Taylor's law and from the CDF [40].

The PG model has been tested with field data from the spatial distribution of the Colorado potato beetle [41]. In this case Taylor's law was evident and the raw data could be used to construct an empirical CDF to compare with the PG CDF. This theoretical CDF agreed well with the data, moreover, the value of Taylor's exponent p, determined from the CDF, agreed with Taylor's law.

The SNP Consortium and the public Human Genome Project have compiled a dense map of human genome sequence variation over the 22 human autosomes as well as the X and Y chromosomes [42]. The density of SNPs can be assessed as equal-sized nonoverlapping enumerative bins that span each chromosome. If these assessments are repeated for sets of bins of different sizes, a range of values for the mean SNP density and the corresponding variances can be obtained. The variance-to-mean plot obtained closely agreed with Taylor's law; plots of the empirical and Tweedie PG CDFs agreed well for each chromosome; and the values of p derived from the CDFs similarly corresponded to those from the plots of Taylor's law [7].

A further example of Taylor's law with DNA sequence variations is provided here from the SNP map of the domestic horse. Figure 1(a) details the variations in the density of SNPs along chromosome 1. Figure 1(b) provides the fit of Taylor's law, determined by the method of expanding enumerative bins $p = 1.58$. A close agreement was found between the empirical and PG CDFs [Fig. 1(c)]. Furthermore, the PG CDF yielded $p = 1.57$, further validating the Tweedie hypothesis.

The coalescent model of population genetics describes the history of sequence variations derived from gene genealogies of homologous sampled sequences [43]. Watterson provided a theoretical basis for this model where he showed that the number of segregating sites within corresponding cistrons sampled from a population should approximate a geometric distribution [27]. If within each enumerative bin there was a Poisson-distributed number of genomic blocks, themselves resulting from recombination and containing, on average, a gamma-distributed number of polymorphic loci, the PG distribution would seem potentially applicable to these data.

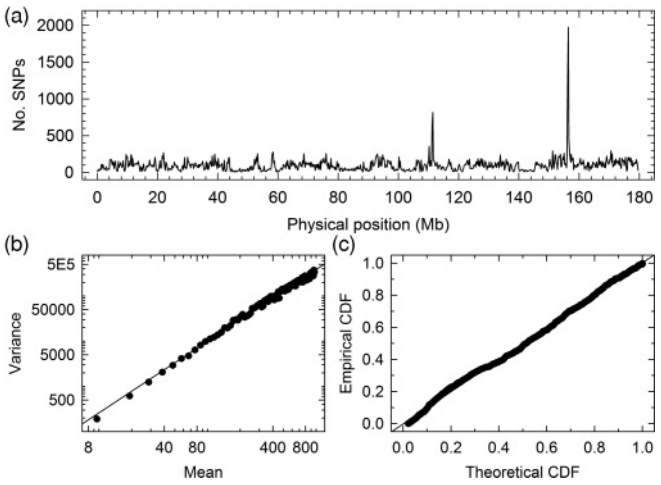

FIG. 1. (a) SNP density along domestic horse chromosome 1. The Horse Genome Project provided a map of 86033 SNPs from chromosome 1 (EquCab2, released September 2007, http://www.broadinstitute.org/mammals/horse). Chromosome 1 was divided into nonoverlapping enumerative bins 200 kb in length and the number of SNPs contained within each bin were counted. (b) The variance function for the density of SNPs. The enumeration of SNPs was repeated for covers of equal-sized bins ranging from 200 kb to 2 Mb in size. The variance and mean number of SNPs per bin were evaluated for each cover and plotted. The log-log plot revealed a straight-line relationship indicative of Taylor's law with exponent $p = 1.58$ and constant $a = 1.92$. (c) Probability-probability plot. A frequency histogram was constructed from the 200 kb bin data. A theoretical Tweedie PG CDF was fitted to these data yielding the parameters $\theta = -0.067$, $\lambda = 0.128$, $\alpha = -0.764$, and $p = 1.57$. The empirical CDF was plotted versus the theoretical CDF to indicate a straight-line relationship and thus agreement of the theoretical model with observation.

The Human Genome Project has provided us with detailed genetic maps of all human chromosomes. Similar to the approach taken with SNP maps one may assess the number of gene structures contained within equal-sized adjacent enumerative bins that span a given chromosome. This analysis has been done for human chromosomes 1, 2, and 7 [8,25,44]. In all cases the chromosomal distribution of gene structures, when assessed by the method of expanding bins, yielded Taylor's law and obeyed the PG distribution. In addition, the values derived for Taylor's exponent from the fit of Taylor's law and the PG CDF agreed [8,25,44].

A further example of the clustering of gene structures is provided here. Figure 2(a) documents the fluctuations in density of gene structures along human chromosome 8. A close agreement of Taylor's law was found with these data using the method of expanding bins $p = 1.59$ [Fig. 2(b)]. In addition, the empirical and PG CDFs fitted closely and yielded $p = 1.59$.

Chromosomal structure is thought to reflect the accumulation of multiple rearrangements, deletions, insertions, and duplications that have occurred through evolution [45]. An evolutionary expansion and modification of chromosomal segments has been proposed to further explain long-range correlations evident within genomic sequences [46]. Within the enumerative bins one could expect a Poisson-distributed

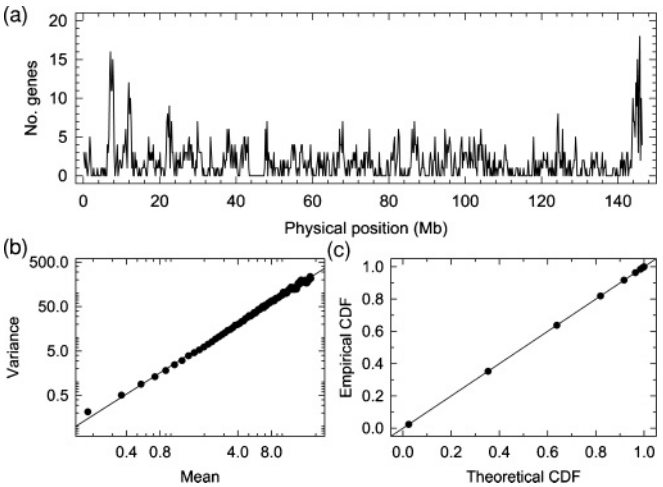

FIG. 2. (a) Gene density along human chromosome 8. The National Center for Biotechnology Information (NCBI) provided a map of 1317 gene structures from chromosome 8 (Build 37.2 http://www.ncbi.nlm.nih.gov/mapview/). Chromosome 8 was divided into nonoverlapping enumerative bins 200 kb in length and the number of gene structures contained within each bin were counted. (b) The variance function for the density of genes. The enumeration of gene structures was repeated for covers of equal-sized bins ranging from 200 kb to 2 Mb in size. The variance and mean number of gene structures per bin were evaluated for each cover and plotted. The log-log plot revealed a straight-line relationship indicative of Taylor's law with exponent $p = 1.59$ and constant $a = 0.84$. (c) Probability-probability plot. A frequency histogram was constructed from the 200 kb bin data. A theoretical Tweedie PG CDF was fitted to these data yielding the parameters $\theta = -1.428, \lambda = 1.354, \alpha = -0.688$ and $p = 1.59$. The empirical CDF was plotted versus the theoretical CDF to indicate a straight-line relationship and thus agreement of the theoretical model with observation.

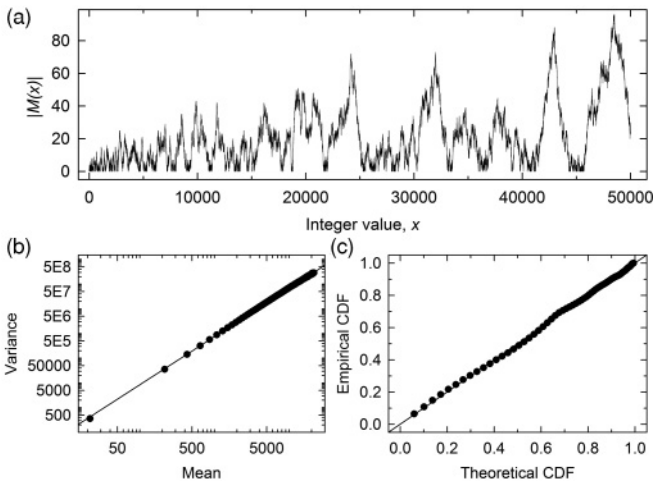

FIG. 3. (a) The absolute value $|M(n)|$ of the Mertens function for the first 50,000 integers. (b) The variance function for the Mertens function. The sequence $|M(n)|$ was used to estimate the variance function and plotted on logarithmic axes. Linear regression yielded the power function with exponent $p = 1.95$ (95% confidence interval by bootstrap method 1.936–1.970), constant $a = 0.066$, and correlation coefficient squared of 1.00. (c) Probability-probability plot. A frequency histogram was constructed from the values $|M(n)|$. The empirical CDF function was fitted to a PG CDF, $\theta = -0.072, \lambda = 0.492, \alpha = -0.302$, and $p = 1.77$. The probability-probability plot revealed agreement between the theoretical and empirical models.

number of chromosomal segments. The mean number of genes per chromosomal segment would be expected to obey a gamma distribution, consequent to the age distribution of different segments and a presumed constant rate of appearance of new genes. In this context a PG distribution would plausibly reflect the local variations in gene density [44].

Genetic rearrangement, deletion, insertion, and duplication are well understood processes for which there is no evidence to indicate the involvement of external physical fields. Granted these genetic processes are chemically driven, but it remains unclear how a thermodynamic process would otherwise have a causative role in the aggregation of gene structures.

In all of the examples so far presented these tests would also seem to support the Fronczak hypothesis (ignoring the lack of a direct and verifiable correspondence of the Fronczak hypothesis to these biological systems) since it too yields the Tweedie distributions. However, there remains a domain of examples that are inconsistent with Fronczak's thermodynamic hypothesis that shall be discussed next.

VI. EXAMPLE OF TAYLOR'S LAW FROM NUMBER THEORY

One further example of Taylor's law and test of the Tweedie hypothesis will be provided from number theory. Since this example is purely numerical, a thermodynamic explanation would be inappropriate. We will examine the summatory Mertens function that is defined by the equation

$$M(n) = \sum_{i=1}^{n} \mu(i), \tag{10}$$

expressed in terms of the Möbius function on the integer values n

$$\mu(n) = \begin{cases} 1 & \text{if } n = 1, \\ 0 & \text{if } n \text{ has one or more repeated prime factors}, \\ (-1)^k & \text{if } n \text{ is a product of } k \text{ distinct primes}. \end{cases} \tag{11}$$

The Mertens function exhibits an aperiodic behavior. Figure 3(a) shows how its absolute value $|M(n)|$ fluctuates for the first 50,000 integers. By the method of expanding enumerative bins the variance function taken from this data segment yielded Taylor's law with $p = 1.95$ [Fig. 3(b)]. A PG CDF was fitted to these data and was found in close agreement with the empirical CDF [Fig. 3(c)]. The PG CDF yielded a value of $p = 1.77$, in acceptable agreement with Taylor's exponent. Although the Mertens function was defined on the integers, the sampling of 50,000 values appeared sufficiently dense that the use of the PG distribution was appropriate. In this application, 1072 of the first 50,000 values of $|M(n)|$ were exactly zero, in agreement with the nonzero probability mass that the PG distribution has at zero.

This last example of Taylor's law was derived from a purely numerical process; other examples of Taylor's law can be demonstrated from number theory that are consistent with the Tweedie PG model.

VII. DISCUSSION

The explanation for Taylor's law presented here was justified by the transformational and convergence properties of error distributions. It was not necessary to invoke the second law of thermodynamics, for which large-scale fluctuations might be considered improbable. And since the Tweedie convergence theorem obviated the need to postulate a thermodynamic mechanism, the problematic issue of the nature of the putative external field in the various biological and numerical systems where Taylor's law has been observed was avoided. An explanation for Taylor's law and fluctuation scaling based on the Tweedie convergence theorem would thus appear more general and parsimonious than one that relies on the existence of external fields or physical processes, and it would be applicable to purely numerical processes. In addition, the restriction $1 < p < 2$, apparent to many ecological and biological examples of Taylor's law, can be explained in mechanistic terms through the application of the Tweedie compound Poisson distribution [6–8,11,44,47].

An explanation for Taylor's law and fluctuation scaling based on the Tweedie hypothesis is premised on a mathematically proven convergence theorem; further tests of this hypothesis come from the many biological and numerical observations that demonstrate Taylor's law through the agreement derived from the PG CDF with both the empirical CDFs and the Taylor's law fits. In contrast, the Fronczak hypothesis is based on thermodynamic theory; a specific confirmation of it would mandate the demonstration of the putative thermodynamic mechanism(s) in each manifestation of Taylor's law. Such demonstrations for the Fronczak hypothesis become particularly problematic given the range of manifestations of Taylor's law. It would seem unlikely that the same external physical field, temperature analogue, or other physical quantity proposed by Fronczak and Fronczak could be responsible for all such manifestations (if any). In contrast, proof of the Tweedie convergence theorem comes from mathematics that shows how Taylor's law can manifest through convergence effects on non-Gaussian processes.

If the physical quantities upon which the Fronczak hypothesis is premised (the entropy, the free energy, the external field $\tilde{\mu}$, and the temperature analogue $1/\tilde{\mu}$) cannot be measured or demonstrated, then the Fronczak hypothesis cannot be tested. If it is not possible to falsify (and thus to test) the Fronczak hypothesis then, by Popper's criteria [48], this hypothesis should not be considered scientific. Popper's criteria are equally applicable to the Tweedie hypothesis: If one were to demonstrate Taylor's law from data that did not converge toward one of the Tweedie distributions this would falsify the Tweedie hypothesis. Or, if the fitted values for p derived from Taylor's law and the PG CDF were significantly discordant, the Tweedie hypothesis would be falsified.

Demonstrations of Taylor's law from purely numerical systems constitute counterexamples where the Fronczak hypothesis is inapplicable. Physical mechanisms do not influence number theory, although number theory can be applied to help understand physics. Eventually, perhaps, physical or even biological examples might be found where the thermodynamic mechanism postulated by Fronczak and Fronczak can be demonstrated. Nonetheless, as with the CLT in Maxwell-Boltzmann statistics, the Tweedie convergence theorem would make the Tweedie hypothesis more general in scope than Fronczak and Fronczak's *ad hoc* hypothesis.

We find it nonetheless remarkable that the thermodynamic approach taken by Fronczak and Fronczak should yield CGFs identical in form to the mathematically based Tweedie models. Possibly, there might be a means to reconcile these two approaches, without the need to postulate the external physical field $\tilde{\mu}$. In the absence of observable evidence for the postulated external field and other physical quantities, as well as specific mechanistic explanations for how their thermodynamic hypothesis would apply to the diverse manifestations of Taylor's law, we would suggest that Fronczak and Fronczak have provided a maximum entropy derivation for the Tweedie exponential dispersion models. Shannon has demonstrated how this maximum entropy principle can be applied to derive probability distributions for mathematical processes [49]; the challenge would be to frame the Fronczak derivation in this context.

Wigner recognized the profound insights that mathematics can provide to physics [31]. We believe that, the Tweedie convergence theorem provides a mathematically demonstrable, and generally applicable, explanation for the genesis of Taylor's law and fluctuation scaling, evident within non-Gaussian biological, physical, and numerical systems. Potentially the Tweedie convergence theorem could provide further insight into other aspects of statistical physics.

ACKNOWLEDGMENTS

The authors acknowledge support from Patricia Rinaldo and the Danish Natural Science Research Council.

[1] L. R. Taylor, Nature (London) **189**, 732 (1961).
[2] W. S. Kendal and P. Frost, J. Natl. Cancer Inst. **79**, 1113 (1987).
[3] W. S. Kendal, Ecol. Model. **80**, 293 (1995).
[4] M. Keeling and B. Grenfell, Philos. Trans. R. Soc. London Ser. B Biol. Sci. **354**, 769 (1999).
[5] P. Philippe, J. Theor. Biol. **199**, 371 (1999).
[6] W. S. Kendal, Proc. Nat. Acad. Sci. USA **98**, 837 (2001).
[7] W. S. Kendal, Mol. Biol. Evol. **20**, 579 (2003).
[8] W. S. Kendal, BMC Evol. Biol. **4**, 3 (2004).
[9] Z. Eisler, I. Bartos, and J. Kertesz, Adv. Phys. **57**, 89 (2008).
[10] A. Fronczak and P. Fronczak, Phys. Rev. E **81**, 066112 (2010).
[11] W. S. Kendal, Ecol. Model. **151**, 261 (2002).
[12] L. R. Taylor and R. A. Taylor, Nature (London) **265**, 415 (1977).
[13] G. M. Wang, E. M. Sevick, E. Mittag, D. J. Searles, and D. J. Evans, Phys. Rev. Lett. **89**, 050601 (2002).
[14] L. Taylor and R. Taylor, Nature (London) **265**, 415 (1977).
[15] I. Hanski, Oikos **34**, 293 (1980).
[16] R. Anderson *et al.*, Nature (London) **296**, 245 (1982).

[17] L. R. Taylor *et al*., Nature (London) **303**, 801 (1983).
[18] J. N. Perry and L. R. Taylor, Biometrics **44**, 881 (1988).
[19] I. Hanski and I. P. Woiwod, Oikos **67**, 29 (1993).
[20] M. R. Binns, Oikos **47**, 315 (1986).
[21] A. W. Kemp, Biometrics **43**, 693 (1987).
[22] J. N. Perry, Proc. R. Soc. London Ser. B Biol. Sci. **257**, 221 (1994).
[23] M. Keeling, Theor. J. Pop. Biol. **58**, 21 (2000).
[24] A. M. Kilpatrick and A. R. Ives, Nature (London) **422**, 65 (2003).
[25] W. S. Kendal, Ecol. Compl. **1**, 193 (2004).
[26] E. Schrödinger, *What is Life?* (Cambridge University Press, Cambridge, England, 1967).
[27] G. A. Watterson, Theor. Pop. Biol. **7**, 256 (1975).
[28] G. Blanc *et al*., Plant Cell **12**, 1093 (2000).
[29] C. R. Honig, M. L. Feldstein, and J. L. Frierson, Am. J. Physiol. Heart. Circ. Physiol. **233**, H122 (1977).
[30] S. Thurner and R. Hanel, e-print arXiv:0804.3477
[31] E. P. Wigner, Commun. Pure Appl. Math. **13**, 1 (1960).
[32] B. Jørgensen, *The Theory of Exponential Dispersion Models* (Chapman & Hall, London, 1997).
[33] M. C. K. Tweedie, in *Statistics: Applications and New Directions, Proceedings of the Indian Statistical Institute Golden Jubilee International Conference, Indian Statistical Institute*, edited by J.K. Ghosh and J. Roy, (Indian Statistical Institute, Calcutta, India, 1984), pp. 579–604.
[34] B. Jørgensen, J. R. Martínez, and M. Tsao, Scand. J. Statist. **213**, 223 (1994).
[35] W. S. Kendal and P. Frost, J. Natl. Cancer Inst. **79**, 1113 (1987).
[36] W. S. Kendal, Invasion and Metastasis **18**, 285 (1999).
[37] W. S. Kendal, J. Surg. Oncol. **74**, 116 (2000).
[38] J. B. Bassingthwaighte, R. B. King, and S. A. Roger, Circ. Res. **65**, 578 (1989).
[39] W. S. Kendal, Proc. Natl. Acad. Sci. USA **98**, 837 (2001).
[40] W. S. Kendal, J. Theor. Biol. **217**, 203 (2002).
[41] W. S. Kendal, Ecol. Model **151**, 261 (2002).
[42] The International SNP Map Working Group, Nature (London) **409**, 928 (2001).
[43] J. F. C. Kingman, Stochast. Proc. Appl. **13**, 235 (1982).
[44] W. S. Kendal, J. Theor. Biol. **245**, 329 (2007).
[45] E. E. Eichler and D. Sankoff, Science **301**, 793 (2003).
[46] W. Li and D. Holste, Phys. Rev. E. Stat. Nonlin. Soft. Matter. Phys. **71**, (2005).
[47] W. S. Kendal, J. Theor. Biol. **217**, 203 (2002).
[48] K. Popper, *The Logic of Scientific Discovery* (Harper Collins, New York, 1965).
[49] C. E. Shannon, in *The Mathematical Theory of Communication*, edited by C. E. Shannon and W. Weaver, (University of Illinois Press, Urbana, IL, 1949), p. 1.