Gompertz mortality law and scaling behavior of the Penna model

J. B. Coe[1,*] and Y. Mao[2]
[1]*Biomathematics & Statistics Scotland, James Clerk Maxwell Building, The King's Building, Edinburgh EH9 3JZ Scotland*
[2]*School of Physics and Astronomy, University of Nottingham, University Park, Nottingham NG7 2RD, United Kingdom*
(Received 10 May 2005; revised manuscript received 9 September 2005; published 28 November 2005)

The Penna model is a model of evolutionary ageing through mutation accumulation where traditionally time and the age of an organism are treated as discrete variables and an organism's genome is represented by a binary bit string. We reformulate the asexual Penna model and show that a universal scale invariance emerges as we increase the number of discrete genome bits to the limit of a continuum. The continuum model, introduced by Almeida and Thomas [Int. J. Mod. Phys. C **11**, 1209 (2000)] can be recovered from the discrete model in the limit of infinite bits coupled with a vanishing mutation rate per bit. Finally, we show that scale invariant properties may lead to the ubiquitous Gompertz law for mortality rates for early ages, which is generally regarded as being empirical.

DOI: 10.1103/PhysRevE.72.051925 PACS number(s): 87.23.−n, 87.10.+e

I. INTRODUCTION

The Penna model was devised in 1995 by Penna [1] to model the process of evolutionary ageing through mutation accumulation. The idea that natural selection would permit behavior such as ageing is initially baffling: it would seem that survival of the fittest would remove any such detrimental behavior. Medawar proposed [2] that certain genes may be age specific in their effects; if such genes are harmful and are activated later on in the reproductive life of an organism, natural selection against them will be much weaker than if they had become active earlier in the organism's life. Given the existence of such genes, it can be anticipated that harmful genetic conditions will become more common as an organism ages giving rise to increasing mortality rates with age. The Penna model is a means to model the evolution of an age-structured population under the influence of age-specific harmful mutations [3].

Traditionally, mortality rates are known to rise exponentially for early ages, giving rise to the Gompertz law [4] of mortality. More recent experiments using much larger sample populations have shown that the mortality rate for advanced ages is shown to slow substantially, giving rise to a mortality plateau or peaks [5–8]. It has been shown that a modified Penna model while continuing to show Gompertz growth in mortality rates at early ages can also exhibit a mortality plateau at advanced ages [9].

The original Penna model is discrete in nature, with time represented by an integer and an organism's genome by a bit string. Each 0 on the bit string represents a healthy site; each 1 is a harmful mutation which becomes active once the organism reaches age x where x is the index of the site on the bit string. Having activated T harmful mutations an organism dies. The bit string is taken to be finite in length (usually 32 bits) and each newborn organism has a number of mutations M introduced into the bit string. These mutations are taken to be harmful so they can only turn healthy sites into unhealthy ones—a mutation on an unhealthy site is ignored. This assumption is relaxed in Ref. [10] where a small rate of positive mutation is allowed: we confine ourselves here to the case of only harmful mutations.

Scaling behavior was considered by Malarz [11] who investigated the effects of different bit string lengths on the Penna model. Malarz inquired as to whether large bit strings were required or whether one could expect, after appropriate scaling of other parameters, one would get the same results for different genome lengths. Investigating the effects of string length through simulation, Malarz was unable to find scaling in the Penna model.

Almeida *et al.* [12] later considered a continuous Penna model and for certain mutation regimes were able to find simple scaling relations. To obtain such scaling the authors decoupled the string length and mutation rate so that the probability of finding a given number of mutations in a given string length was given by a Poisson distribution. They also observed that the Penna model is able to sustain a maximum possible lifespan in steady state, which we call $l_{\max}$. If the imposed string length is greater than $l_{\max}$ then it will have no effect on the properties of the population; if it is less than $l_{\max}$ then the imposed string length will impose a maximum lifespan on the population and the distribution will be accordingly altered. The authors suggested that the size of time steps in a discrete Penna model may have an effect on scaling behavior but did not investigate the size or nature of this effect.

Brigatti *et al.* [13] investigated scaling in a sexual Penna model through simulation and suggested that results from the continuum model of Almeida *et al.* [12] were not readily mapped onto the discrete model employed in simulation. Scaling effects in the sexual model were also investigated by Laszkiewicz *et al.* [14] through simulation.

In this paper we extend our previous analytical solution of the asexual model [9,15] to examine the scaling behavior. We show that the scale invariance emerges as we increase the number of discrete genome bits, and that the scaling becomes exact in the continuum case, which can be regarded as

*J.B.C. is also affiliated to the Institute of Evolutionary Biology, University of Edinburgh and Department of Physics and Astronomy, University of Edinburgh.

 ©2005 The American Physical Society

the limit of infinite genome bits coupled with a vanishing mutation rate per bit. This establishes a clear relationship between the distribution, parameters, and scaling behavior of the continuum model of Almeida *et al.* and those of the traditional discrete model. Finally, we use scale invariance to analytically show that at early ages mortality rates grow exponentially in accordance with the Gompertz law [4] which, for the lack of a general proof, is still generally regarded as being empirical [16].

II. A CONTINUOUS PENNA MODEL

The asexual Penna model can be reformulated [12,15] so that rather than considering discrete time steps, time is treated as a continuous variable, t. The bit string of an organism is replaced by an axis representing the genome: position x on the genome is examined at age x. Harmful mutations are then represented by δ functions along the genome. After accumulating T δ functions an organism dies.

For our analytical solutions, we concern ourselves primarily with $T=1$ as generalizing a $T=1$ solution for a continuous model will be no more difficult than generalizing a discrete $T=1$ model, as done previously [9]. In the continuum Penna model an organism reproduces at a constant rate b and dies at age x where its genome has its first harmful mutation (δ function) at position x.

An organism can be characterized by its age x and its genetic lifespan l (the position on the genome of the first harmful mutation). Neither x nor l are constrained to be integers. $n(x,l)$ is now a density of organisms so that the number of organisms with age and genetic lifespan in the range $x \to x+dx$ and $l \to l+dl$ is given by $n(x,l)dxdl$. The probability of giving birth in time dt is given by bdt, the probability of a mutation being introduced in length dl is given by βdl. These definitions are consistent with the discrete Penna model where sites can be interpreted as infinitesimal lengths of genome and time steps as infinitesimal units of time. The probability of no mutations occurring in length dl is $1-\beta dl$ which is $e^{-\beta dl}$ for infinitesimal dl.

Newborn organisms may be produced as unmutated copies of organisms with equal genetic lifespan, or as mutated copies of naturally longer lived organisms. An equation can then be constructed for the production of new organisms within the population for the infinitesimal time period of t to $t'=t+dt$

$$n(0,l)_{t'}dtdl = bdtdle^{-\beta l}\int_0^\infty dx n(x,l)_t + bdt\beta dle^{-\beta l}\int_0^\infty dx\int_l^\infty dl' n(x,l')_t \qquad (1)$$

where subscripts t and t' denote time. At steady state, the subscripts may be dropped, the above equation can be simplified and an expression obtained [9,15] for the relative sizes of population densities (see Fig. 1)

$$\frac{n(l+x)}{n(l)} = \frac{l+x}{l}\frac{e^{\beta l}-bl}{e^{\beta(l+x)}-b(l+x)}\exp\left[\int_l^{l+x}\frac{\beta bl'}{bl'-e^{\beta l'}}dl'\right]. \qquad (2)$$

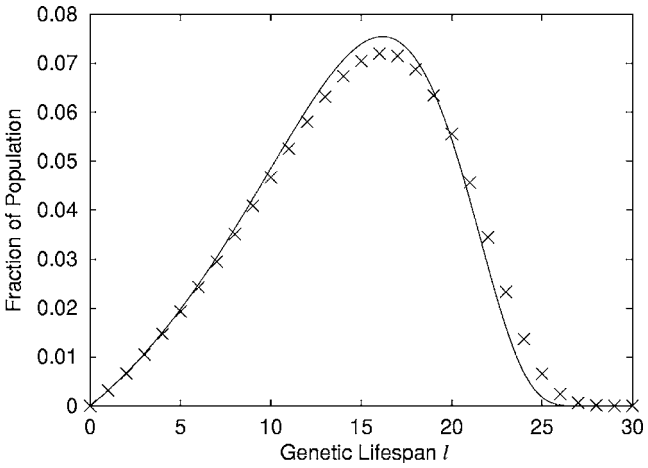

FIG. 1. A plot of genetic lifespan distribution for a discrete ($\times$) and continuous Penna model with $l_{\max}=30$.

For a steady state to exist there must be a longest lived subpopulation which is self-sustaining, i.e., not reliant on mutated births. No other subpopulation can be self-sustaining if the population is to remain bounded, as shorter-lived organisms can always be created by mutated copies of longer-lived ones. For the longest-lived subpopulation to be self-sustaining, each organism must produce one perfect copy of itself during its lifetime

$$l_{\max}be^{-\beta l_{\max}} = 1. \qquad (3)$$

All other populations, with $l<l_{\max}$, gain from mutated births of the longest lived, so unmutated birth per individual must, on average, be less than unity

$$lbe^{-\beta l} < 1 \quad \forall \; l < l_{\max}. \qquad (4)$$

These conditions can be combined to give [15]

$$l_{\max} \leqslant \frac{1}{\beta}, \qquad (5)$$

$$b = \frac{1}{l_{\max}}e^{\beta l_{\max}}. \qquad (6)$$

In the discrete Penna model the probability of no mutation for 1 site or bit is $1-m$ where m is the mutation rate per site. The probability for l sites without mutations would be $(1-m)^l$. In the limit of $m \to 0$, $(1-m)^l \approx e^{-ml}$ and therefore m play the same role as β in the continuous case, where the probability of no mutations in genome length l is $e^{-\beta l}$. Thus we can identify the continuum Penna model as the limit of the discrete model as the mutation rate per site tends to zero. For vanishingly small units of discretization, a discrete model becomes a continuous one. A measure of the extent of discretization is the size of one of the discrete units divided by the total size of the system; for the Penna model this is $1/l_{\max}$. As the extent of discretization gets smaller, $l_{\max}$ tends to infinity which implies a vanishing mutation rate. Thus, the two limits of mutation rate tending to zero, and of increasingly fine grained discretization, are identical.

III. SCALING PROPERTIES

We examine the discrete and continuous Penna models in turn to examine how they behave under rescaling. Informed by this behavior we interpret the continuous model as the limit of a discrete model with a vanishingly small mutation rate.

A. The discrete Penna model

The traditional asexual Penna model has one unit of discretization for each unit of time. It is possible to rescale the discrete Penna model so that each unit of time is broken up into several time steps. This can be done by taking a Penna model with a maximum lifespan of $al_{\max}$ and rescaling l so that, in the rescaled time units, the model has maximum lifespan $l_{\max}$ and a distinct time steps in one unit of time. For example, an $l_{\max}=30$ model could be rescaled to give an $l_{\max}=15$ model with two time steps per unit of time.

When discussing rescaled Penna models we require that the steady-state conditions are invariant under rescaling. For a population with $l_{\max}=60$, the steady-state conditions should be the same regardless of how many time steps one unit of time has been broken up into. For steady-state conditions to be invariant under rescaling the population with genetic lifespan $l_{\max}$ must be self-sustaining and all other populations partly dependent on mutation. The first condition can be written as

$$bl_{\max}e^{-\beta l_{\max}}=1. \tag{7}$$

A model rescaled by a factor a will allow $n(l_{\max}-1/a)$ to exist, where a gives the number of units of discretization per time interval. The same conditions, Eqs. (4)–(6), apply as before. A population is then identified by the largest (unscaled) value of $l_{\max}$ it can sustain. When steady state is required to be robust under rescaling of the model a population can be uniquely identified by the maximum genetic lifespan it can maintain.

For rescaled models to be the same they should give the same population sizes at comparable points up to an arbitrary scaling factor. If the discrete Penna model is scale invariant, it should be possible to rescale a model to obtain an unscaled model with shorter $l_{\max}$. For instance: an $l_{\max}=30$ model scaled by a factor of 2 will have a rescaled maximum lifespan of 15; if the Penna model is scale invariant, this rescaled model will, at comparable points give identical results to an unscaled $l_{\max}=15$ model (up to a constant normalization factor for finite size scaling). Where n_{30} denotes a model with unscaled maximum lifespan 30, we require that $n_{30}(2l)/n_{15}(l)$ is constant. In a general case for models to be identical after scaling we require that

$$n_{al_{\max}}(al)\propto n_{l_{\max}}(l). \tag{8}$$

This can be satisfied, eliminating the constant of proportionality by

$$\frac{n_{al_{\max}}(al+a)}{n_{al_{\max}}(al)}=\frac{n_{l_{\max}}(l+1)}{n_{l_{\max}}(l)}. \tag{9}$$

In the case of $a=2$ we require that

$$\frac{n_{[l_{\max},2]}(l+1)}{n_{[l_{\max},2]}\left(l+\frac{1}{2}\right)}\frac{n_{[l_{\max},2]}\left(l+\frac{1}{2}\right)}{n_{[l_{\max},2]}(l)}=\frac{n_{[l_{\max},1]}(l+1)}{n_{[l_{\max},1]}(l)}. \tag{10}$$

For a Penna model to have $l_{\max}$ a factor of a greater, the mutation rate and birth rate must be a factor of a smaller. If the parameters of the model which is rescaled are labeled as l', β', m', and b', then scaled and unscaled parameters are related by

$$l'=al, \tag{11}$$

$$l'_{\max}=al_{\max}, \tag{12}$$

$$\beta'=\frac{\beta}{a}, \tag{13}$$

$$b'=\frac{b}{a}. \tag{14}$$

Application of these scaling rules, the recursion relation between successive subpopulations at steady state, and our condition for scale invariance of the model gives a relation, in terms of birth and mutation rate, which must be satisfied for the discrete model to be scale invariant.

For a rescaling by a factor of 2 we require that

$$\frac{e^{\beta l}-bl}{e^{\beta(l+1/2)}-b\left(l+\frac{1}{2}\right)e^{-\beta/2}}\frac{e^{\beta l+1/2}-b\left(l+\frac{1}{2}\right)}{e^{\beta(l+1)}-b(l+1)e^{-\beta/2}}=\frac{e^{\beta l}-bl}{e^{\beta(l+1)}-b(l+1)e^{-\beta}}. \tag{15}$$

This equality cannot be satisfied due to the factor of $e^{-\beta/2}$ on the bottom of the recursion relation. As such, the discrete Penna model does not exhibit scale invariance. In the limit of a vanishing mutation rate: $e^{-\beta}$ approaches unity, the differences between scaled and unscaled models vanish and the discrete model will become scale invariant. Figure 2 confirms that the scaled results of the discrete model do approach a limiting "master curve." This limit is the same as that which gives the continuous model, so we expect to find the continuous model to be scale invariant. Note, only comparable points have been plotted and distributions have been normalized so $\Sigma_l a^{-1}n(l)=1$ where a is the scale factor.

B. The continuous Penna model

As in the discrete Penna model, we identify a population by the largest value of $l_{\max}$ it can sustain. This value is no longer constrained to be an integer and can be simply expressed as $l_{\max}=1/\beta$. Rescaling of a continuous Penna model is carried out in much the same way as in the discrete case: $l_{\max}$ is divided by a scale factor a and the new model has a correspondingly reduced maximum lifespan. In the continuous model time is not broken into distinct time steps, but is treated as a continuum: as a result rescaling will not alter the number of time steps in one unit of time. If a continuous model is to be invariant under rescaling by a factor a, through similar reasoning as in the discrete case,

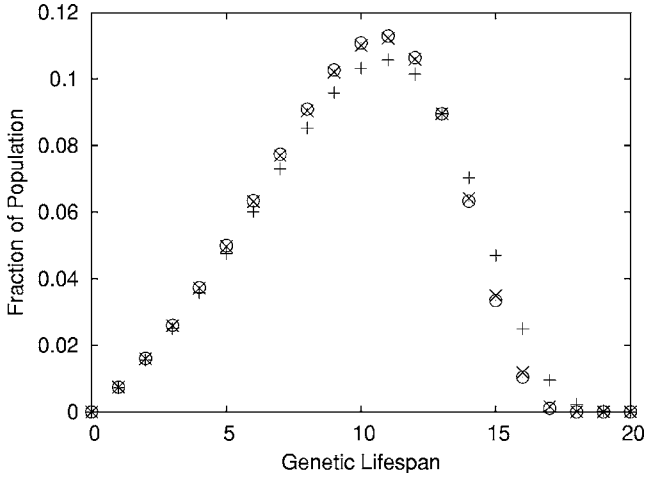

FIG. 2. A plot of genetic lifespan distribution for an unscaled Penna model with $l_{\max}=20$ (+), a model with $l_{\max}=200$ scaled down by a factor of 10 (×), and a model with $l_{\max}=2000$ scaled down by a factor of 100 (○). Only comparable points have been shown.

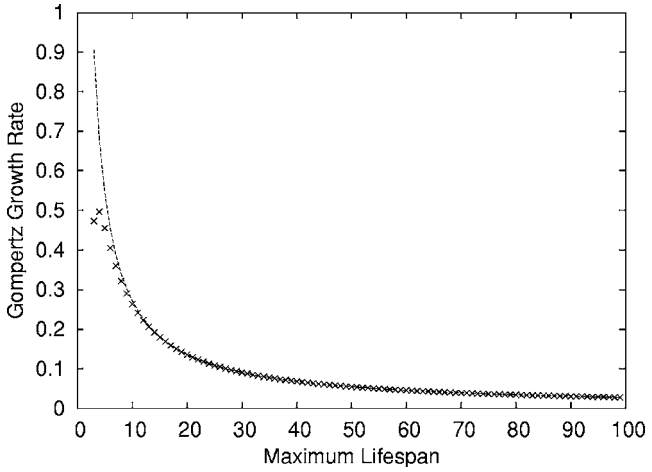

FIG. 3. The exponential coefficient of Gompertz growth in mortality rate estimated from early age mortality rates (×) is plotted against the maximum lifespan the population can sustain ($l_{\max}$). The dashed line gives the birth rate at each value of $l_{\max}$.

$$\frac{n_{al_{\max}}(al+ax)}{n_{al_{\max}}(al)}=\frac{n_{l_{\max}}(l+x)}{n_{l_{\max}}(l)}. \tag{16}$$

Upon substitution of the steady-state relation for continuous Penna model populations, this is satisfied by

$$\frac{l+x}{l}\frac{e^{\beta l}-bl}{e^{\beta(l+x)}-b(l+x)}\exp\left[\int_l^{l+x}\frac{\beta bl'}{bl'-e^{\beta l'}}dl'\right]$$

$$=\frac{al+ax}{al}\frac{e^{\beta' al}-b'al}{e^{\beta'(al+ax)}-b'(al+ax)}$$

$$\times\exp\left[\int_{al}^{al+ax}\frac{\beta' b'al'}{b'al'-e^{\beta' al'}}dal'\right]. \tag{17}$$

The mutation rate and birth rate in the rescaled model are labeled β' and b'. If, by rescaling the model by a factor a, the maximum lifespans are to be the same then the mutation and birth rates must be related by: $\beta'=\beta/a$, $b'=b/a$. After this substitution the continuous model is clearly scale invaraint as both sides of the equation give

$$\frac{l+x}{l}\frac{e^{\beta l}-bl}{e^{\beta(l+x)}-b(l+x)}\exp\left[\int_l^{l+x}\frac{\beta bl'}{bl'-e^{\beta l'}}dl'\right]. \tag{18}$$

It has been shown that in the limit of a vanishing mutation rate coupled with an infinite maximum lifespan, the discrete model becomes a continuous one. In other words, for a vanishing mutation rate, the discrete model becomes scale invariant. As the limits of a vanishing mutation rate and maximum genetic lifespan tending to infinity are equivalent, approximate scale invariance becomes more realistic for discrete Penna models of increasingly large $l_{\max}$.

IV. MORTALITY RATES

Early age Penna mortality rates display the exponential growth predicted by the Gompertz law. Using our analytical solution to the simple Penna model we evaluate the growth exponent γ where the mortality rate at age x is proportional to $e^{\gamma x}$. Evaluation of the Gompertz growth rate in terms of the Penna model parameters will facilitate the fitting of Penna parameters to real world data. Throughout, we assume that any model has adopted the maximum genetic lifespan allowed by its mutation rate.

Recall that for the simple discrete Penna model the mortality rate is given by

$$M(x)=\frac{n(x)/x}{\sum_{l=0}^{\infty}n(l)/l}. \tag{19}$$

Using the steady state recursion relation from the simple Penna model, the ratio between successive mortality rates can be evaluated analytically

$$\frac{M(x+1)}{M(x)}=\frac{e^{\beta x}-bx}{e^{\beta(x+1)}-b(x+1)e^{\beta}}\left[\frac{n(x)/x}{\sum_{l=x+1}^{\infty}n(l)/l}+1\right]. \tag{20}$$

To usefully exploit this expression, we consider the limit of small x, and small β where the Penna model becomes scale invariant; numerical evaluation of the summation term and predicted scaling behavior can be used to simplify Eq. (20). Numerically, we find for $x\ll l_{\max}$

$$\frac{n(x)/x}{\sum_{l=x+1}^{\infty}n(l)/l}\simeq\frac{1}{l_{\max}}. \tag{21}$$

Crucially, if the Penna model exhibits universality as discussed earlier, this result remains valid for *all* values of $l_{\max}$. Therefore, noting the continuous Penna model result $l_{\max}=1/\beta$ and $b=\beta e$, in the regime of small $x\ll l_{\max}$, a first order

expansion of Eq. (20) leads to

$$\frac{M(x+1)}{M(x)} \approx e^{b}, \tag{22}$$

which then implies

$$M(x) \propto e^{bx}, \tag{23}$$

namely the Gompertz law, which states that the mortality rate increases exponentially at early ages. Furthermore, it predicts that the exponential coefficient of the Gompertz growth rate is given by b, the birth rate. In Fig. 3, we compare this birth rate with the exponential Gompertz coefficients, extracted by taking the difference between the logs of mortality rates at ages $x=2$ and $x=1$ for each population.

Our approximation depends on $x \ll l_{\max}$, therefore, deviation from Gompertz behavior at later ages (large x) is expected as the numerical approximation, Eq. (21), breaks down as x increases. Similarly, as shown in Fig. 3, for small values of $l_{\max}$ this approximation works less well, but for larger values of $l_{\max}$ it becomes increasingly accurate.

V. CONCLUSION

We have shown by means of exact analytic solution that, in the asexual Penna model, a universal scale invariance emerges as we increase the number of genome bits/sites, with the invariance becoming exact in the limit of the continuum model. In addition, we have built on this result and shown that scale invariance may be employed to derive an analytical expression for the Gompertz law of mortality, which has been generally regarded as empirical.

ACKNOWLEDGMENTS

The authors would like to thank Krzysztof Malarz for helpful discussion. This work has been financially supported by the Schiff foundation and the EPSRC under grant number GR/TR11777/01.

[1] T. J. P. Penna, J. Stat. Phys. **78**, 1629 (1995).
[2] P. B. Medawar, *An Unsolved Problem of Biology* (H. K. Lewis, London, 1952).
[3] D. Stauffer, *Biological Evolution and Statistical Physics* (Springer, Berlin, 2002).
[4] B. Gompertz, Philos. Trans. R. Soc. London **123**, 513 (1825).
[5] J. W. Vaupel *et al.*, Science **280**, 855 (1998).
[6] J. W. Curtsinger, H. H. Fukui, D. R. Townsend, and J. W. Vaupel, Science **258**, 461 (1992).
[7] L. Mueller and M. R. Rose, Proc. Natl. Acad. Sci. U.S.A. **93**, 15294 (1996).
[8] K. W. Watcher, Proc. Natl. Acad. Sci. U.S.A. **96**, 10544 (1999).
[9] J. B. Coe, Y. Mao, and M. E. Cates, Phys. Rev. Lett. **89**, 288103 (2002).
[10] J. B. Coe, Y. Mao, and M. E. Cates, Phys. Rev. E **70**, 021907 (2004).
[11] K. Malarz, Int. J. Mod. Phys. C **11**, 309 (2000).
[12] R. M. C. Almeida and G. L. Thomas, Int. J. Mod. Phys. C **11**, 1209 (2000).
[13] E. Brigatti, J. S. Sa Martins, and I. Roditi, Eur. Phys. J. B **42**, 431 (2004).
[14] A. Laszkiewicz, S. Cebrat, and D. Stauffer, e-print q-bio/0411051.
[15] J. B. Coe and Y. Mao, Phys. Rev. E **67**, 061909 (2003).
[16] M. Rose, *Evolutionary Biology of Aging* (Oxford University Press, New York, 1991).