Analytical solution of a generalized Penna model

J. B. Coe and Y. Mao
Cavendish Laboratory, Madingley Road, Cambridge, CB3 0HE, United Kingdom
(Received 28 February 2003; published 23 June 2003)

In 1995 Penna introduced a simple model of biological aging. A modified Penna model has been demonstrated to exhibit behavior of real-life systems including catastrophic senescence in salmon and a mortality plateau at advanced ages. We present a general steady-state, analytic solution to the Penna model which is able to deal with arbitrary birth and survivability functions. This solution is employed to solve standard variant Penna models studied by simulation. Different Verhulst factors regulating both the birth rate and external death rate are considered.

DOI: 10.1103/PhysRevE.67.061909 PACS number(s): 87.23.−n, 87.10.+e

I. INTRODUCTION

While nothing in life is certain except death and taxes [1], only death is universal. Death and the preceding period of functional decline are a fate to be endured by all from medflies to men [2–5]. The steady decline in the functional ability of an organism over time is known as senescence or aging. The phenomenon of aging is of obvious interest and has attracted attention from biologists and physicists for some time [6].

The establishment and maintenance of harmful behavior by natural selection would appear to defy explanation. Medawar proposed [7,8] that the strength of selection on survival related genes is dependent on the age at which the genes exert their effects. For genes that express themselves late in the life of an organism, there is less impact on the dwindling population than for a gene whose effect is expressed earlier in life.

It has been suggested [9] that favorable mutations that act to increase survivability can be used to account for senescence. Such mutations will, under natural selection, increase the survivability at early ages and converting an initially constant mortality rate into an age-dependent one. It seems improbable that this mechanism can provide a full explanation of aging as positive mutations are very rare compared to harmful ones. Instead a better understanding can be gained by considering processes through which genes with harmful effects are introduced.

There are two such theories: antagonistic pleiotropy and mutation accumulation [10,11]. According to antagonistic pleiotropy, senescence occurs as a result of mutations that increase the functional ability of the young and decrease that of the old. Mutation accumulation proposes that aging occurs due to mutations that are initially harmless but take effect at later stages in the life of an organism. In either case, the force of natural selection is reduced once an organism ages beyond its point of reproductive maturity so the effects of either will be confined to older ages.

The Penna model is the most commonly used model of aging through mutation accumulation [12,13] and is ideally suited to computational implementation. Using Monte Carlo methods [14], the model predicts features found in real ecological systems, such as, the catastrophic senescence of Pacific salmon [15]. Analytical work by Almeida *et al.* [16] on a theoretical approach to biological aging can be adapted to apply to specific cases of the Penna model. This work presents a solution to a generalized Penna model, in particular, one that allows incorporation of arbitrary survivability and birth functions. Subtle modification to the survivability function has been found [18] to demonstrate a mortality plateau at older ages, a result that had so far eluded theories of mutation accumulation.

To ensure that the population is in a steady state, the total population is controlled through the use of a Verhulst factor [19,20]. Traditionally, this has been a genome-independent chance of death for every individual regardless of age. This model also considers a birth rate that decreases as the population grows and resources become more scarce [21]. In agreement with earlier work [16], we find a maximum permitted genetic lifespan and predict the existence of catastrophic senescence for organisms whose reproductive life is terminated by an upper age limit. The model is extended to a continuum case, which is explicitly solved.

II. THE PENNA MODEL

The Penna model as formulated by Penna represents a genome by a single string of 1's and 0's. Time is treated as a discrete variable. A 1 on a site i along the string means that the organism develops a disease at age i. Once an organism develops a number T of diseases, it dies. At each time step an organism reproduces with probability b. The offspring's genome is a copy of its parent's with a probability m of each bit mutating into a 1. Positive mutations are rare in nature and ignored in this model, there is no possibility of a 1 mutating into a 0. The bit string is traditionally 32 bits long for ease of computational implementation. The finite-length bit string is an artifact of simulation and is discounted in our analysis where there is no need for such a restriction. Along similar lines to Almeida and Thomas [17] we consider a fixed probability of mutation occurring on any site. The bit-string sites are labeled so that there is a zeroth site, which is read as soon as the organism is born.

A. The solution of a simple Penna model

Consider a simple Penna model where an organism dies after a single disease ($T=1$) and can reproduce with equal

probability at any point during its life. An individual organism can be characterized uniquely by two variables, its age x and its string length l. The string length is the location of the 1 bit on the string and corresponds to the number of time steps for which the organism lives. To produce an organism with string-length l either a perfect copy of a length l organism or a mutated copy of an organism with longer string length must be born. To produce an organisms of string length l, an organism must give birth, with probability b, and the first l sites on the offspring must go unmutated, each with probability $(1-m)$. For the mutated offspring, the parent string must be longer, the parent must give birth, with probability b, the first l sites on the offspring must go unmutated and one site must be mutated with probability m. As all organisms are capable of reproduction, mutated and perfect copies of organisms of any age must be taken into account.

In our notation $n_j(x,l)$ is the number of organisms with age x and string length l at time step j. We define $e^{-\beta}$ to be $(1-m)$. New organisms are produced as mutated or unmutated copies of organisms in the previous time step:

$$n_{j+1}(0,l)=be^{-\beta l}\sum_{x=0}^{\infty} n_j(x,l)+mbe^{-\beta l}\sum_{l'>l}^{\infty}\sum_{x=0}^{\infty} n_j(x,l'). \tag{1}$$

For a steady state there is no difference between $n(x,l)$ at different time steps, so the time step indices can be dropped. The simple Penna model is constructed in such a way that an organism with string length l lives for l time steps. The probability of giving birth during each of these time steps is a constant b. Thus the sum over all ages is a sum from 0 to l of $n(x,l)$ which in the steady state can be written as $l \times n(0,l)$. Defining $n(l)=l\times n(0,l)$, the equation now reads

$$0=be^{-\beta l}n(l)-\frac{n(l)}{l}+mbe^{-\beta l}\sum_{l'>l}^{\infty} n(l'). \tag{2}$$

A similar equation can be written for $n(l+1)$. Manipulation of both will eliminate the sum over l' and give a recursion relation

$$\frac{n(l+1)}{n(l)}=\frac{l+1}{l}\,\frac{e^{\beta l}-bl}{e^{\beta(l+1)}-b(l+1)e^{-\beta}}. \tag{3}$$

If this expression is to be usefully employed, the steady state interdependence of b and β must be examined. In the steady state it is required that, in the statistical limit, the population size and distribution remain unchanged over time for constant values of b and β.

As the string length l gets longer, the probability of an unmutated copy being produced is reduced by an exponential factor $e^{-\beta l}$. For an organism to be able to produce a single perfect copy of itself during its lifespan, $lbe^{-\beta l}=1$. If any subpopulation is capable of maintaining itself, there must be no contribution to that subpopulation from mutations. As all subpopulations will have contributions from mutation from longer strings, it follows that a subpopulation capable of maintaining itself must have the longest string length in the population, $l_{\max}$. For all other subpopulations with $l<l_{\max}$ the probability of an organism producing a perfect copy of itself must be less than 1 to avoid population explosion when contributions from mutation are taken into account.

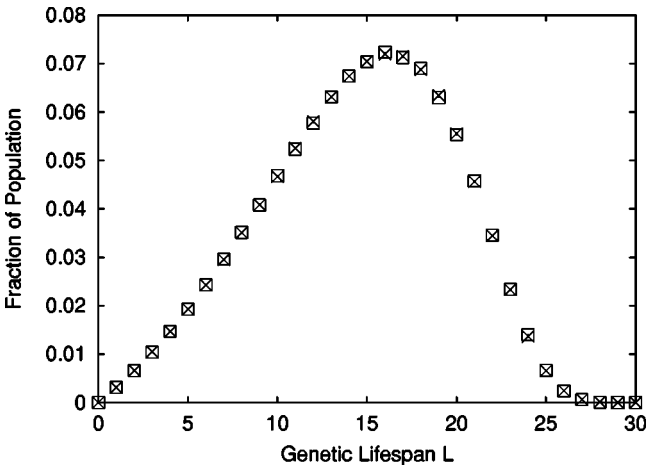

FIG. 1. Lifespan distribution for $l_{\max}=30$. Analytical results ($\times$) are compared with those from simulation ($\square$). Simulation size 10^7.

It is sufficient to state that

$$(l_{\max}-1)be^{-\beta(l_{\max}-1)}<1. \tag{4}$$

The constraints on $l_{\max}$ lead to a maximum sustainable value and a corresponding steady-state birth rate,

$$l_{\max}<\frac{1}{1-e^{-\beta}}, \tag{5}$$

$$b=\frac{1}{l_{\max}}e^{\beta l_{\max}}. \tag{6}$$

In agreement with Ref. [16] we have predicted the existence of a maximum sustainable genetic lifespan, see Figs. 1–6.

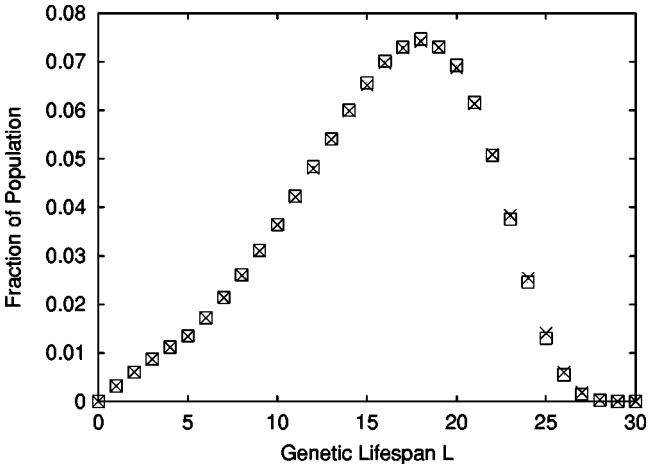

FIG. 2. Lifespan distribution for $l_{\max}=30$, $l_B=5$. Analytical results ($\times$) are compared with those from simulation ($\square$). Simulation size 10^7.

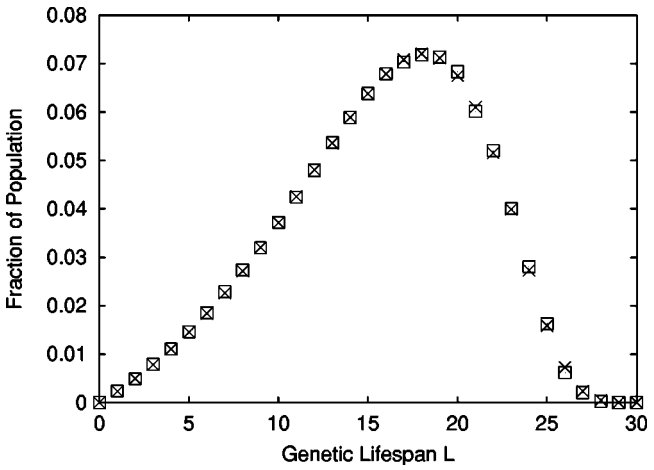

FIG. 3. Lifespan distribution for $l_{\max}=30$, $\gamma=0.02$. Analytical results ($\times$) are compared with those from simulation ($\square$). Simulation size 10^8.

The system has a range of possible values for $l_{\max}$ and will adopt one depending on the dynamics and the initial state of the system. In simulation, a population is often regulated through the use of a Verhulst factor [12,19,20]. This ensures that a population will find a steady state configuration. Such simulations also regulate the total population, which is realistic in any system with finite resources. It has been suggested [21] that by regulating the birth rate, the system will adopt a biologically realistic equilibrium. The simple Penna model solved here can be used to explain behavior of these simulations, which replace the constant birth rate with a population-dependent one. Where N is the total population, A a constant of the simulation, and $N_{\max}$ is the maximum population the simulation will allow, the birth rate at any time step i is given by

$$b_i = A\left(1-\frac{N_i}{N_{\max}}\right). \tag{7}$$

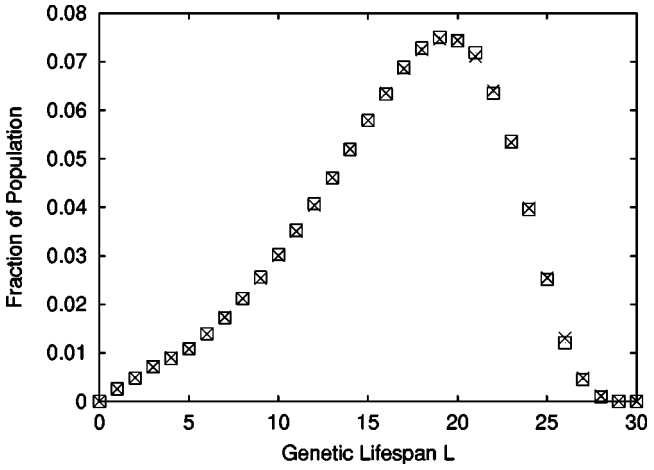

FIG. 4. Lifespan distribution for $l_{\max}=30$, $l_B=5$, $\gamma=0.02$. Analytical results ($\times$) are compared with those from simulation ($\square$). Simulation size 10^8.

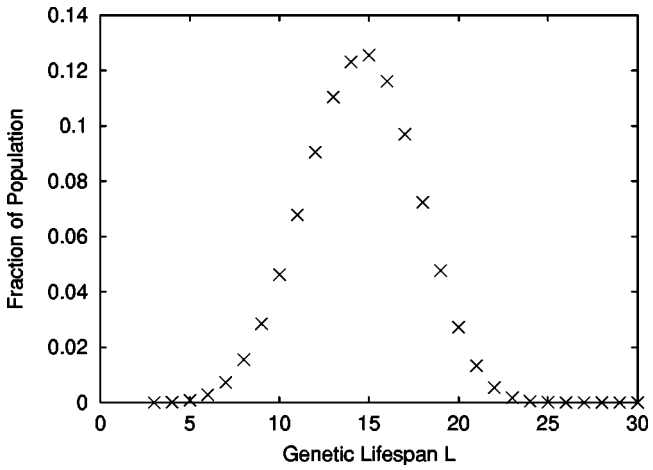

FIG. 5. Lifespan distribution for $l_{\max}=30$, $T=4$.

In the steady state the total population will be that required to set this dynamic Verhulst birth rate to the value required by the earlier equilibrium conditions:

$$N = N_{\max}\left(1-\frac{e^{\beta l_{\max}}}{A l_{\max}}\right). \tag{8}$$

B. The solution of a Penna model with reproductive threshold ages

Frequently the Penna model used [13] has birth cut-offs so that an organism begins giving birth at an age, l_B up to an age, l_S. When able to reproduce, an organisms does so at rate b as before. The equation for stepwise evolution of an age zero subpopulation remains the same except that the birth rate is now a function of the organism's age. An organism is no longer capable of reproducing in every time step, but only those for which its age is greater than or equal to l_B and less than l_S. The sum over ages must now sum over the age-dependent birth term as well:

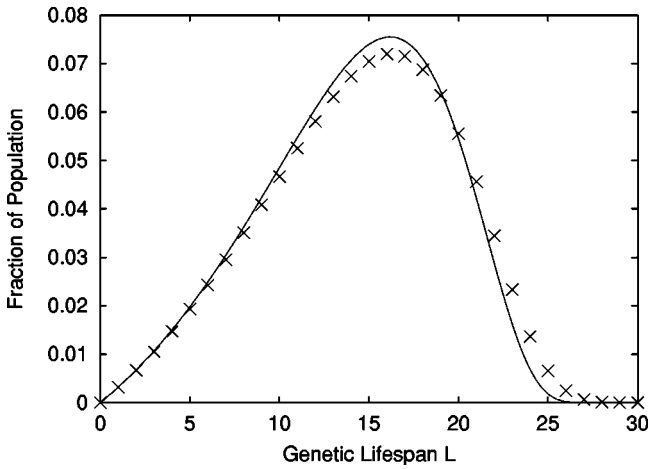

FIG. 6. A plot of genetic lifespan distribution for a discrete ($\times$) and continuum Penna model with $l_{\max}=30$.

$$n(0,l)=e^{-\beta l}\sum_{x=0}^{\infty} b_x n(x,l)+me^{-\beta l}\sum_{l'>l}^{\infty}\sum_{x=0}^{\infty} b_x n(x,l'). \tag{9}$$

As in the simple Penna model, $n(l)$ is defined to be $l \times n(0,l)$ as the survivability function is unaltered. Summing over reproductive ages leads to a simplified steady state equation

$$0=be^{-\beta l}\chi(l)n(l)-\frac{n(l)}{l}+mbe^{-\beta l}\sum_{l'>l}^{\infty}\chi(l')n(l'), \tag{10}$$

where $\chi(l)$ is defined as $[\Sigma_{x=0}^{\infty} b_x n(x,l)]/[bn(0,l)]$ and is given by

$$\chi(l)=0 \quad \text{for} \quad l\leqslant l_B$$
$$\chi(l)=l-l_B \quad \text{for} \quad l_B<l\leqslant l_S$$
$$\chi(l)=l_S-l_B \quad \text{for} \quad l>l_S. \tag{11}$$

Employing the same method as was used to solve the simple Penna model, a recursion relation can be obtained,

$$\frac{n(l+1)}{n(l)}=\frac{l+1}{l}\frac{e^{\beta l}-b\chi(l)}{e^{\beta(l+1)}-b\chi(l+1)e^{-\beta}}. \tag{12}$$

In the case of no birth thresholds, $\chi(l)=l$ and we retrieve Eq. (3).

The steady state conditions for a population will have been altered by the introduction of birth threshold ages. The conditions for $l_{\max}$ are, by the same reasoning as before,

$$l_{\max}be^{-\beta l_{\max}}\chi(l_{\max})=1, \tag{13}$$

$$(l_{\max}-1)be^{-\beta(l_{\max}-1)}\chi(l_{\max}-1)<1. \tag{14}$$

There are two distinct nontrivial cases for the position of $l_{\max}$, either it is below the upper reproductive boundary or it is contained within the reproductive window.

Consider $l_{\max}\leqslant l_S$. The boundary conditions can be written as

$$(l_{\max}-l_B)be^{-\beta l_{\max}}=1, \tag{15}$$

$$(l_{\max}-1-l_B)be^{-\beta(l_{\max}-1)}<1. \tag{16}$$

After rearrangement, this yields

$$l_{\max}<\frac{1+(e^{-\beta}-1)l_B}{1-e^{-\beta}} \tag{17}$$

with a corresponding birth rate of

$$b=\frac{e^{\beta l_{\max}}}{l_{\max}-l_B}. \tag{18}$$

Consider now the case of $l_{\max}>l_S$. The boundary conditions become

$$b(l_S-l_B)e^{-\beta l_{\max}}=1, \tag{19}$$

$$b(l_S-l_B)e^{-\beta(l_{\max}-1)}<1. \tag{20}$$

Requiring $e^{\beta}<1$. This cannot be satisfied as β is a mutation rate and consequently a positive number. As a result $l_{\max}$ is not permitted to be greater than the second birth threshold l_S for any value of mutation rate or birth rate. A steady state population will not maintain organisms that are no longer able to contribute to reproduction. This result does not apply to organisms who have some social structure and assist in rearing young once they themselves are no longer of a reproductive age. Such cooperative complications [22] are not taken into account here.

C. The solution of a Penna model with an external death rate

Consider the simple Penna model. An external death rate is introduced into this model so that in any time step an organisms has a chance γ of dying independent of its bitstring composition. We define the survival rate σ as $1-\gamma$. Counting contributions from unmutated and mutated reproduction, the steady-state equation is

$$n(0,l)=be^{-\beta l}\sum_{x=0}^{\infty} n(x,l)+mbe^{-\beta l}\sum_{l'>l}^{\infty}\sum_{x=0}^{\infty} n(x,l'). \tag{21}$$

The sums over all ages can be evaluated by taking into account the external source of death:

$$\sum_{x=0}^{\infty} n(x,l)=\sum_{x=0}^{l-1}\sigma^x n(0,l)=\frac{1-\sigma^l}{\gamma}n(0,l). \tag{22}$$

Defining $n(l)=[(1-\sigma^l)/\gamma]n(0,l)$, the steady state equation can be rearranged to give

$$0=be^{-\beta l}n(l)-\frac{\gamma}{1-\sigma^l}n(l)+mbe^{-\beta l}\sum_{l'>l}^{\infty} n(l'). \tag{23}$$

The same approach is used as before to give an recursive relationship between $n(l)$ and $n(l+1)$:

$$\frac{n(l+1)}{n(l)}=\frac{1-\sigma^{l+1}}{1-\sigma^l}\frac{\gamma e^{\beta l}-1+\sigma^l}{\gamma e^{\beta(l+1)}-(1-\sigma^{l+1})e^{-\beta}}. \tag{24}$$

In the limit of a vanishing external death rate, (small γ) power-series expansion of these expressions will, to leading order, give the conditions from the simple Penna model (3).

To find the steady state relationship between b, β, and γ, the same conditions are imposed as before. $n(l_{\max})$ must be self-sustaining and all lower string lengths are partly reliant on mutation:

$$\frac{1-\sigma^{l_{\max}}}{\gamma}be^{-\beta l_{\max}}=1,$$
$$\frac{1-\sigma^{l_{\max}-1}}{\gamma}be^{-\beta(l_{\max}-1)}<1. \tag{25}$$

These conditions give the birth rate and limit the value of $l_{\max}$:

$$l_{\max} < \frac{\ln\left(\frac{e^{\beta}-1}{e^{\beta}\sigma^{-1}-1}\right)}{\ln\sigma}, \tag{26}$$

$$b = \frac{\gamma e^{\beta l_{\max}}}{1-\sigma^{l_{\max}}}. \tag{27}$$

In the limit of a small external death rate these expressions become those from the simple Penna model case, Eqs. (5) and (6).

An external death rate is commonly used in simulations [12,13] as a Verhulst factor to regulate the population. In any time step an organism has probability of death given by $N_i/N_{\max}$. The preceding analysis can be used to explain how systems using this Verhulst factor behave. In the steady state the external death rate provided by this Verhulst factor must satisfy the above conditions for stability at and around $l_{\max}$. For specified b, β, and $l_{\max}$, γ can be found numerically and the total population can be found from $N=\gamma N_{\max}$.

D. The solution of a Penna model with an external death rate and reproductive threshold ages

It is a relatively trivial matter to extend the analysis of a Penna model with an external death rate to incorporate birth threshold ages. This type of model has been extensively studied in simulation, so is of sufficient interest to consider separately. Only the lower birth threshold needs to be considered. An upper threshold, as demonstrated, will only serve to artificially lower the maximum allowable string length $l_{\max}$. As ever, in the steady state

$$n(0,l) = e^{-\beta l}\sum_{x=0}^{\infty} b_x n(x,l) + m e^{-\beta l}\sum_{l'>l}^{\infty}\sum_{x=0}^{\infty} b_x n(x,l'). \tag{28}$$

Taking external deaths into account when summing over reproductive ages, we define $\chi_\gamma(l)$ as

$$\chi_\gamma(l) = 0 \quad \text{for} \quad l < l_B$$

$$\chi_\gamma(l) = \frac{\sigma^{l_B}-\sigma^{l}}{\gamma} \quad \text{for} \quad l_B \leqslant l. \tag{29}$$

With $n(l)$ given by $[(1-\sigma^l)/\gamma]n(0,l)$, a recursion relation for $n(l)$ can be given

$$\frac{n(l+1)}{n(l)} = \frac{1-\sigma^{l+1}}{1-\sigma^{l}}\,\frac{e^{\beta l}-b\chi_\gamma(l)}{e^{\beta(l+1)}-b\chi_\gamma(l+1)e^{-\beta}}. \tag{30}$$

E. The solution of a multiple disease Penna model

An organism in the simple Penna model dies upon encountering a single 1 on its bitstring. Models are frequently set up so that an organism must develop T diseases before death. Each site on the string is occupied by one or no deleterious mutations, multiple mutations on a single site are not allowed.

The relevant part of an individual string is that containing the first T 1s, as bits after this point are irrelevant. Rather than uniquely classifying an organism by its genetic lifespan l, as was done in the single mutation case, we must now specify the position of each of the first T mutations on the organism's bit string. The only other property an organism has is its age x. Any individual can thus be uniquely classified by its type $x;l_1,l_2,\ldots,l_T$, where $n(x;l_1,l_2,\ldots,l_T)$ is the number of the specified organisms. The position of the final disease on the string determines an organism's age of death, the position of the other bits play no part in this, nor do they determine birth rate. This inspires the ansatz that $n(x;\{l\})$ has no dependency on the positions of the nonterminal diseases.

Contributions to a child subpopulation from a single mutation come from all organisms with $T-1$ bits in common with the child subpopulation, and the final bit at a position $l'>l_T$. For $T=4$, where l_1, l_2, l_3, and l_4 are the deleterious bit positions in the child subpopulation, there will be contributions from mutation from $n(x;l_1,l_2,l_3,l')$, $n(x;l_1,l_2,l_4,l')$, $n(x;l_1,l_3,l_4,l')$, and $n(x;l_2,l_3,l_4,l')$. Our ansatz means that all these terms are the same as they are dependent on only age x and the position of the terminal bit l'. Where $n_T(x,l)$ is the number of organisms with age x and terminal mutation at site l, the number of organisms capable of contributing through a single mutation is $Tn_T(x,l')$. In a similar vein to the single disease Penna model, we can now account for all birth terms and generate a steady state equation:

$$n_T(0,l) = b e^{\beta(l-T+1)}\sum_{x=0}^{\infty} n_T(x,l) + Tmb e^{\beta(l-T+1)}\sum_{l'>l}^{\infty}\sum_{x=0}^{\infty} n_T(x,l'). \tag{31}$$

This can be solved in the same manner as that for the simple Penna model, where $n_T(l)=l\times n_T(0,l)$,

$$\frac{n_T(l+1)}{n_T(l)} = \frac{l+1}{l}\times\frac{e^{-\beta(l-T+1)}-bl}{e^{-\beta(l+1-T+1)}-b(l+1)(e^{\beta}+T-1)}. \tag{32}$$

$n_T(l)$ is not of direct interest in determining age distributions or mortality rates as it is only one of several configurations of nonterminal mutations. This is amended by summing over all possible configurations so that $n(l)=C_{T-1}^{l}n_T(l)$.

The steady state correspondence between b, β, and $l_{\max}$ for arbitrary T is

$$l_{\max} < \frac{1}{1-e^{-\beta}}, \tag{33}$$

$$b = \frac{1}{l_{\max}} e^{\beta(l_{\max}-T+1)}. \tag{34}$$

Confirming the solution of a multiple disease Penna model through simulation presents considerable problems. The solution presented here is not unique and there is no reason for a given simulation to settle on this particular steady state distribution. Within any finite population, it has been demonstrated [23] that if individuals are arbitrarily labeled, then, after sufficient time, the entire population will have descended from individuals with just one of these arbitrary labels. The ansatz used treats the nonterminal mutations as arbitrary labels so the assumption that they are evenly distributed will be upset by finite-sized population dynamics.

F. The solution of a multiple disease Penna model with external death rate and reproductive threshold ages

As in the single mutation case, it is a relatively simple matter to adapt the multiple disease solution to incorporate external death rate and birth thresholds. An upper birth cutoff is not considered as this will only serve to artificially lower $l_{\max}$. We do not spend any time on the derivation here but present the results for those who may consider them of particular interest.

Where $n(l)=[(1-\sigma^l)/\gamma]n(0,l)$ and $\chi_\gamma(l)$ is given by

$$\chi_\gamma(l)=0 \quad \text{for} \quad l<l_B$$

$$\chi_\gamma(l)=\frac{\sigma^{l_B}-\sigma^l}{\gamma} \quad \text{for} \quad l_B\leqslant l, \tag{35}$$

$$\frac{n(l+1)}{n(l)}=\frac{C_{T-1}^{l+1}}{C_{T-1}^{l}}\frac{l-\sigma^{l+1}}{l-\sigma^l} \times\frac{e^{\beta(l-T+1)}-b\chi_\gamma(l)}{e^{\beta(l+1-T+1)}-b\chi_\gamma(l+1)(e^\beta+T-1)}, \tag{36}$$

$$l_{\max}<\frac{\ln\left(\frac{(e^\beta-1)\sigma^{l_B}}{e^\beta\sigma^{-1}-1}\right)}{\ln\sigma}, \tag{37}$$

$$b=\frac{\gamma e^{\beta(l_{\max}-T+1)}}{\sigma^{l_B}-\sigma^{l_{\max}}}. \tag{38}$$

G. A continuum Penna model

The simple Penna model can be reformulated so that rather than considering discrete time steps, time is treated as a continuous variable *t*. As before any subpopulation is characterized by its age *a* and string length *l*, which are no longer constrained to be integers. The birth rate *b* is now the probability of an organism reproducing in unit time. Likewise, the mutation rate *m* is the probability of a mutation occurring in unit string length. The steady state equation for this model is

$$n(0,l)=be^{-\beta l}\int_0^\infty n(x,l)dx+mbe^{-\beta l}\int_0^\infty dx\int_l^\infty dl' n(x,l'). \tag{39}$$

As in the discrete case the sum over ages, now an integral, can be evaluated employing the fact that $n(0,l)$ is constant over time. The integral over ages of $n(x,l)$ can be written as $n(l)$, where $n(l)=l\times n(0,l)$, giving

$$0=be^{-\beta l}n(l)+-\frac{n(l)}{l}+mbe^{-\beta l}\int_l^\infty n(l')dl'. \tag{40}$$

The steady state equation can be rewritten in the form below and the integrals evaluated numerically:

$$\frac{n(l+x)}{n(l)}=\frac{l+x}{l}\frac{e^{\beta l}-bl}{e^{\beta(l+x)}-b(l+x)} \times\exp\left(\int_l^{l+x}\frac{mbl'}{bl'-e^{\beta l'}}dl'\right). \tag{41}$$

As in the discrete Penna model, in the steady state, $n(l_{\max})$ must be self-sustaining and all subpopulations with a lower genetic lifespan partly reliant on mutation. As *l* is no longer constrained to be an integer, the conditions become

$$l_{\max}be^{-\beta l_{\max}}=1, \tag{42}$$

$$(l_{\max}-\delta)be^{-\beta(l_{\max}-\delta)}<1, \tag{43}$$

where δ is arbitrarily small. These conditions give

$$l_{\max}\leqslant\frac{1}{\beta}, \tag{44}$$

$$b=\frac{1}{l_{\max}}e^{\beta l_{\max}}. \tag{45}$$

H. A continuum Penna model with multiple diseases

A multiple disease Penna model can also be reformulated into a continuum case. The mutations are considered to be δ functions [17] and, as in the discrete case, an organism dies once it has accumulated *T* diseases.

Recalling the first-order steady state equation from the discrete multiple mutation model

$$0=be^{-\beta(l-T+1)}n_T(l)-\frac{n_T(l)}{l} +Tmbe^{-\beta(l-T+1)}\sum_{l'>l}n_T(l'). \tag{46}$$

In the continuum model, where mutations no longer take up a finite length on the string, this will become

$$0=be^{-\beta l}n_T(l)-\frac{n_T(l)}{l}+Tmbe^{-\beta l}\int_{l'>l}n_T(l')dl'. \tag{47}$$

An integral equation can be derived from this using a similar approach as in the single disease continuum case. To evaluate $n(l_T)$, all possible configurations of nondeleterious mutations must be summed over.

In the discrete Penna model this leads to

$$n(l)=C_{T-1}^{l}n_T(l), \tag{48}$$

as each mutation takes up one site on the string. When positioning δ functions, there is no possibility of mutations overlapping as they are of infinitesimal size on the string. In the continuous case, $n(l)$ is given by

$$n(l)=l^{T-1}n_T(l). \tag{49}$$

$n(l_T+x)$ is given by

$$\frac{n(l+x)}{n(l)}=\frac{(l+x)^{T-1}}{l^{T-1}}\frac{l+x}{l}\frac{e^{\beta l}-bl}{e^{\beta(l+x)}-b(l+x)} \times\exp\left(\int_l^{l+x}\frac{mbTl'}{bl'-e^{\beta l'}}dl'\right). \tag{50}$$

The steady state conditions are unchanged from the single disease continuous Penna model as $T>1$ will only affect contribution from mutations that have no effect on the subpopulation with string length $l_{\max}$. The steady state conditions are

$$l_{\max}\leqslant\frac{1}{\beta}, \tag{51}$$

$$b=\frac{1}{l_{\max}}e^{\beta l_{\max}}. \tag{52}$$

III. A PENNA MODEL WITH ARBITRARY BIRTH AND SURVIVABILITY FUNCTIONS

The methods employed to solve the variety of Penna models discussed so far can be generalized to any Penna model where the birth and survival functions are functions of the organisms' age and genetic lifespan. The survival function is defined so that the number of organisms of age x is given by $n(x,l)=f_s(x,l)n(0,l)$. The birth function $b(x,l)$ gives the average number of offsprings produced per time step by an organism of age x and string length l. The steady-state equation can be written as

$$n(0,l)=e^{-\beta l}\sum_{x=0}^{\infty}b(x,l)n(x,l) +me^{-\beta l}\sum_{x=0}^{\infty}\sum_{l'>l}^{\infty}b(x,l')n(x,l'). \tag{53}$$

Defining $n(l)=\Sigma_{x=0}^{\infty}n(x,l)$, where $n(l)$ is the number of organisms with string length l regardless of their age, and employing $\chi(l)$ and $L(l)$, where

$$L(l)=\sum_{x=0}^{\infty}f_s(x,l), \tag{54}$$

$$b\chi(l)=\sum_{x=0}^{\infty}b(x,l)f_s(x,l), \tag{55}$$

the steady state equation can be written as

$$0=be^{-\beta l}\frac{\chi(l)}{L(l)}n(l)-\frac{n(l)}{L(l)}+mbe^{-\beta l}\sum_{l'>l}^{\infty}\frac{\chi(l')}{L(l')}n(l'). \tag{56}$$

$L(l)$ is the expected lifespan of an organism with string length l and $b\chi(l)$ is the expected number of offsprings from an organism of string length l throughout its life. b has been chosen so that throughout this paper the highest birth rate in any time step is b. Using the steady state equation, a general recursion relation can be generated

$$\frac{n(l+1)}{n(l)}=\frac{L(l+1)}{L(l)}\frac{[e^{\beta l}-b\chi(l)]}{[e^{\beta(l+1)}-b\chi(l+1)e^{-\beta}]}. \tag{57}$$

The conditions imposed on $l_{\max}$ to ensure that $n(l_{\max})$ is self-sustaining and that the population remains finite are

$$b\chi(l_{\max})e^{-\beta l_{\max}}=1, \tag{58}$$

$$b\chi(l_{\max}-1)e^{-\beta(l_{\max}-1)}<1. \tag{59}$$

Thus

$$\frac{\chi(l_{\max}-1)}{\chi(l_{\max})}<e^{-\beta}, \tag{60}$$

$$b=\frac{e^{\beta l_{\max}}}{\chi(l_{\max})}. \tag{61}$$

The survival function $f_s(x,l)$ should for physical reasons be a monotonically declining function. There are no constraints on the birth function other than that it must be positive. Through suitable choices of birth and survivability functions, the Penna model can be adapted to model a wider variety of real-life behavior. In previous work [18], we have demonstrated that a subtle modification to the simple Penna model, namely, replacing the step-function survivability with a Fermi function, is capable of producing a mortality plateau. This particular modification does not change the steady-state equation from that of the simple Penna model, though this would not pose any problem to the general solution presented here.

Incorporating the methods used to solve continuous and $T>1$ models in this paper will allow for application of this general method to arbitrary $T>1$ and continuum cases.

IV. EXTRACTING MORTALITY DATA FROM GENETIC LIFESPAN DISTRIBUTIONS

The solutions presented so far have derived relationships for distribution of genetic lifespans of a population. From

genetic lifespan distributions, mortality behavior can be evaluated. Recall that a distribution was determined for $n(l)$, that is, the number of organisms with genetic lifespan l. The required quantity for evaluating mortality rates is $n(0,l)$ and is given by $n(l)/L(l)$.

The number of organisms of genetic lifespan l dying between age x and $x+1$ is given by

$$n(x,l)-n(x+1,l)=n(0,l)[f_s(x,l)-f_s(x+1,l)]. \quad (62)$$

The number of organisms dying at age x is this quantity summed over all genetic lifespans. The fraction of the total population dying at age x is defined as the time-step mortality rate $\mathcal{M}(x)$,

$$\mathcal{M}(x)=\frac{\sum_l n(0,l)[f_s(x,l)-f_s(x+1,l)]}{\sum_l n(0,l)f_s(x,l)}. \quad (63)$$

V. CONCLUSION

We have presented an analytic solution to the steady state Penna model capable of dealing with arbitrary survival and birth functions. We hope this will stimulate further modifications to the Penna model, where suitable choices of birth and survival functions will allow an adapted Penna model to encompass and explain more observed phenomena within age structured populations. In our future work we will study the dynamics of the Penna model and consider more sophisticated complications such as the introduction of a positive mutation rate.

ACKNOWLEDGMENTS

The authors have been assisted and inspired by M. E. Cates, L. Demetrius, R. Farr, M. J. Rutter, and D. Stauffer, who are gratefully acknowledged.

[1] B. Franklin (unpublished).
[2] Opinion, Nature (London) **397**, 89 (1999).
[3] M. Rose, *Evolutionary Biology of Aging* (Oxford University Press, New York, 1991).
[4] B. Charlesworth, *Evolution in Age-Structured Populations*, 2nd ed. (Cambridge University Press, Cambridge, 1994).
[5] C.E. Finch, *Longevity, Senescence and the Human Genome* (University of Chicago Press, Chicago, 1990).
[6] D. Stauffer, *Biological Evolution and Statistical Physics* (Springer, Berlin, 2002).
[7] P.B. Medawar, Mod. Q. **1**, 30 (1946).
[8] P.B. Medawar, *An Unsolved Problem of Biology* (Lewis, London, 1952).
[9] W.D. Hamilton, J. Theor. Biol. **12**, 12 (1966).
[10] S.D. Pletcher and J.W. Curtsinger, Evolution (Lawrence, Kans.) **52**, 2 (1998).
[11] B. Charlesworth, J. Theor. Biol. **210**, 47 (2001).
[12] T.J.P. Penna, J. Stat. Phys. **78**, 1629 (1995).
[13] S. Moss de Oliviera, Physica A **257**, 465 (1998).
[14] T.J.P. Penna and D. Stauffer, Int. J. Mod. Phys. C **6**, 233 (1995).
[15] T.J.P. Penna, S. Moss de Oliviera, and D. Stauffer, Phys. Rev. E **52**, 3309 (1995).
[16] R.M.C. Almeida, S. Moss de Oliviera, and T.J.P. Penna, Physica A **253**, 366 (1998).
[17] R.M.C. Almeida and G.L. Thomas, Int. J. Mod. Phys. C **11**, 1209 (2000).
[18] J.B. Coe, Y. Mao, and M.E. Cates, Phys. Rev. Lett. **89**, 288103 (2002).
[19] P.F. Verhulst, Memoires de l'Academic Royale de Belgique **18**, 1 (1845).
[20] S. Moss de Oliviera, P.M.C. de Oliviera, and D. Stauffer, *Evolution, Money, War and Computers* (Teubner, Stuttgart, 1999).
[21] M. Magdon-Maksymowicz, A.Z. Maksymowicz, and K. Kulakowski, Theory Biosci. **119**, 139 (2000).
[22] P.B. Stacey and W.D. Koenig, *Cooperative Breeding in Birds: Long Term Studies of Ecology and Behaviour* (Cambridge University Press, Cambridge, 1990).
[23] D. Makowiec, J. Dabowski, and M. Groth, Physica A **273**, 169 (1999).