Population dynamics in the Penna model

J. B. Coe and Y. Mao

Cavendish Laboratory, Madingley Road, Cambridge CB3 OHE, United Kingdom

We build upon the recent steady-state Penna model solution [J. B. Coe, Y. Mao, and M. E. Cates, Phys. Rev. Lett. **89**, 288103 (2002)] to study the population dynamics within the Penna model. We show that any perturbation to the population can be broken into a collection of modes each of which decay exponentially with its respective time constant. The long time behavior of population is therefore likely to be dominated by the modes with the largest time constants. We confirm our analytical approach with simulation data.

I. INTRODUCTION

Genes, mutation, evolution, and ageing have been topics of intensive research [1–5], particularly after the recent Genome project. In 1995, a bit-string computer simulation model for population evolution was introduced by Penna [6] which successfully encompassed all those elements. The Penna model, essentially a mutation accumulation model, has been so successful that it has rapidly established itself as a major model for population simulations [4]. Recently, analytical solutions have been presented for the steady states of the Penna model [7,8], shedding insights into the interrelationships between various Penna parameters of the simulation model. Here, we build upon this analytical framework to study populations undergoing steady growth or decline and to study the transient behavior of the model when the system is away from its steady-states. Previous attempts [9] to analyze the population dynamics did not have the advantage of the full analytic steady state solution [7,8] and consequently could not fully explore the dynamics of the model. We find that the fluctuations in a population away from steady state can be decomposed into a collection of modes each of which decays exponentially with its respective time constant. The long time behavior of a population will therefore be dominated by the modes with the largest time constants. Our analytical results are confirmed by comparison with simulation results.

II. THE SIMPLE PENNA MODEL

In the Penna model [6], an organism's genome is represented by a bit string. Organisms age in time steps and at each timestep an organism reads the corresponding bit from the bit string (e.g., second site at age 2). If a site contains a 1 the organism develops a disease, and once it has accumulated T diseases it dies. In any time step an organism can reproduce with probability b. The child's bit string is a copy of the parent's with a probability m of each site mutating into a 1. Positive mutations are rare in nature so a 1 mutating into a 0 is forbidden in the model. Variants of the Penna model exist in which an organism can only reproduce between certain ages, and in which there is an external death rate giving each organism a genome-independent chance of death in any time step.

Traditionally [10] the Penna model is implemented com-

putationally and the population controlled by the use of a Verhulst factor [11] which controls either the birth rate or external death rate. The bit string is 32/64 bits long to enable bitwise manipulation on integer-type variables. The finite length bit string is an artifice of simulation and is not an important consideration when approaching the model analytically. A solution to the Penna model in the steady state has been developed and is capable of dealing with a wide range of modifications to the model, namely arbitrary birth and survivability functions [8].

By building upon the steady state solution it is possible to consider dynamic behavior. The simplest form the steady state solution takes is for the simple Penna model in which there is no nongenetic source of death, an organism dies after a single disease ($T = 1$) and can reproduce with equal probability at any point during its life. For simplicity we present our dynamics analysis within this simple Penna model, since it is straightforward to generalize to the case of $T > 1$ [8]. The steady state solution [7] to the simple Penna model is given in brief below.

An organism within the Penna model can be uniquely characterized by its age x and the number of 0s on its bit string before the first 1. The number of 0s determines how long the organism lives and is termed its string-length l. Where $n_j(x,l)$ is the number of organisms with age x and string length l at time step j,

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

(1)

where $e^{-\beta} = 1 - m$ is the probability of avoiding a mutation. In the stationary state $n_{j+1}(x,l) = n_j(x,l)$ and given that an organism with string length l lives for l time steps, the sum over ages of $n(x,l)$ is $ln(0,l)$. The sum over ages is written as $n(l) = \Sigma_x n(x,l)$. Then Eq. (1) simplifies to

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

(2)

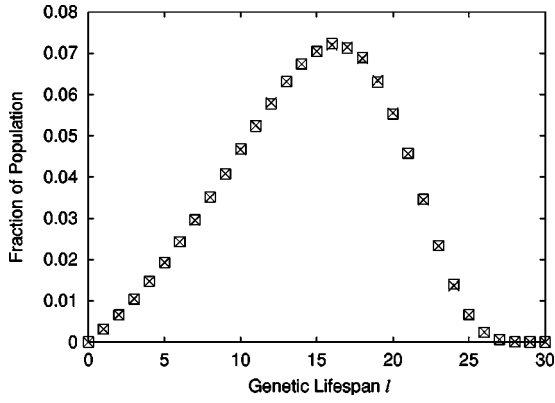This expression can be solved to generate a recursion relation (see Fig. 1).

FIG. 1. Life-span distribution for a simple Penna model with $l_{max}=30$, $\beta=\frac{1}{30}$. Analytical results ($\times$) are compared with those from simulation ($\square$). Simulation size 10^7.

FIG. 2. Life-span distributions for a Penna model undergoing steady growth with $l_{max}=30$, $r=0.05$ (+) and $l_{max}=30$, $r=-0.05$ ($\triangle$).

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

For any population there is a maximum sustainable string length l_{max}, the stability analysis of which [8,7] leads the stationary-state interdependence of b and β:

$$l_{max} < \frac{1}{1-e^{-\beta}}. \quad (4)$$

$$b = \frac{1}{l_{max}} e^{\beta l_{max}}. \quad (5)$$

For finite-length bit strings, l_{max} of course cannot exceed the bit string itself.

III. DYNAMICS IN THE SIMPLE PENNA MODEL

Penna model dynamics can be divided into three cases.

(a) The birth rate or/and mutation rate are altered from their steady-state values leading to growth or decline of the total population n, and associated changes in the distribution $n(l)$ of the subpopulations with string lengths l.

(b) A change in $n(l)$ from the steady state distribution predicted in Eq. (3) is followed by relaxation back to the steady state.

(c) Within a subpopulation $n(l)$, the distribution of ages $n(x,l)$ can fluctuate with time if the population is not in steady state.

A. Steady growth and decline in the simple penna model

With an unsuitable choice of birth rate, mutation rate, and maximum string length present in a population, stationary-state behavior will not be found. Eventually the behavior of the entire population will be dominated by the growth or decline of the longest l type. The population can exist in a state of steady growth or decline in which the relative sizes of subpopulations remain the same. The governing equation can be written as
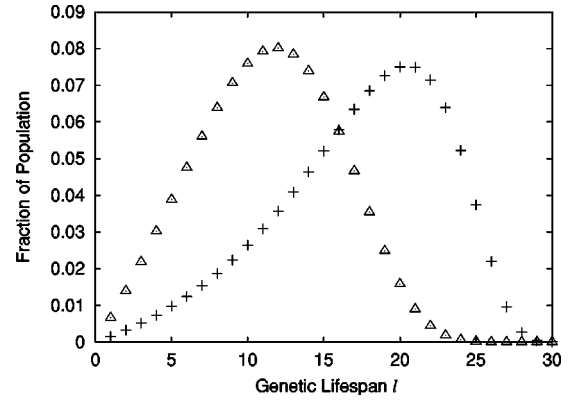
$$n_{j+1}(0,l) = be^{-\beta l} \sum_{x=0}^{\infty} n_j(x,l) + mbe^{-\beta l} \sum_{l'>l}^{\infty} \sum_{x=0}^{\infty} n_j(x,l'). \quad (6)$$

If we label the rate of growth r, then in any time step the number of young produced is $1+r$ times that in the previous time step. Once the population has been growing in this manner for some time, populations at successive iterations are related by $n_{j+1}(x,l) = (1+r)n_j(x,l)$. As in the steady state case, $n(l)$ is defined to be the sum over ages of $n(x,l)$. We define $L_r(l)$ so that $n(l) = L_r(l)n(0,l)$. This leads to a simplified steady-growth equation

$$0 = be^{-\beta l}n_j(l) - \frac{(1+r)}{L_r(l)}n(l) + mbe^{-\beta l}\sum_{l'>l}^{\infty} n_j(l'). \quad (7)$$

where $L_r(l)$ is given by

$$L_r(l) = \frac{1 - \left(\frac{1}{1+r}\right)^l}{1 - \frac{1}{1+r}}. \quad (8)$$

The steady-growth equation can be manipulated to give a recursion relation for the relative sizes of successive $n_j(l)$ (see Fig. 2)

$$\frac{n_j(l+1)}{n_j(l)} = \frac{L_r(l+1)}{L_r(l)} \frac{(1+r)e^{\beta l} - bL_r(l)}{(1+r)e^{\beta(l+1)} - bL_r(l+1)e^{-\beta}}. \quad (9)$$

The conditions for steady growth give a limit on the value of the maximum sustainable string length l_{max} and determine the value of the birth rate b:

$$l_{max} < \frac{\ln\left(\frac{1-e^{\beta}}{1-e^{\beta}(1+r)}\right)}{\ln\left(\frac{1}{1+r}\right)}, \quad (10)$$

$$b = \frac{re^{\beta l_{max}}}{1 - \left(\dfrac{1}{1+r}\right)^{l_{max}}}. \tag{11}$$

In the limit of vanishing r, a power series expansion of these expressions will give, to leading order, Eqs. (4) and (5).

B. Subpopulation dynamics in the simple penna model

For subpopulation dynamics we consider the time-step evolution of an arbitrary distribution of $n(l)$ ($l \leqslant l_{max}$) where the birth and mutation rate take their steady-state values. The dynamics of subpopulations within the Penna model can be considered to be that of a series of decay modes. Any subpopulation can be expressed as the sum over contributions from a set of decay modes such that

$$n_j(x,l) = \sum_k A_k n_{k,j}(x,l), \tag{12}$$

where the constants A_k are to be determined. Within each decay mode, labeled by its index k, the population dies away exponentially so that

$$n_{k,j+1}(x,l) = (1-\lambda_k)n_{k,j}(x,l). \tag{13}$$

Within each mode there is a maximum value of l above which $n_{k,j}(l)$ is zero, this is labeled l_k. The mode index k is chosen so that k is the number of nonzero subpopulations within the mode. Once the relationship between $n_j(x,l)$ and $n_{j+1}(x,l)$ has been established Eq. (1) can be considered to be a sum of eigenequations each of which governs the behavior of a given mode. The equation governing the time-step evolution of the longest string length within a mode can be written as

$$n_{k,j+1}(0,l_k) = be^{-\beta l_k} \sum_{x=0}^{\infty} n_{k,j}(x,l_k). \tag{14}$$

Since the time-step evolution of $n_{k,j}(x,l)$ is known from Eq. (13) the sum over ages can be evaluated and λ_k identified as the solution to the equation

$$\lambda_k = 1 - \frac{1 - \left(\dfrac{1}{1-\lambda_k}\right)^{l_k}}{1 - \dfrac{1}{1-\lambda_k}} be^{-\beta l_k}. \tag{15}$$

The characterstic decay time τ_k for any mode is defined as the time taken for the mode to decay to e^{-1} of its initial size (see Fig. 3).

Manipulation of Eq. (1) for an individual mode gives a recursion relation for the relative sizes of subpopulations within a mode (see Fig. 4)

$$\frac{n_{k,j}(l+1)}{n_{k,j}(l)} = \frac{L_k(l+1)}{L_k(l)} \frac{(1-\lambda_k)e^{\beta l} - bL_k(l)}{(1-\lambda_k)e^{\beta(l+1)} - bL_k(l+1)e^{-\beta}}, \tag{16}$$
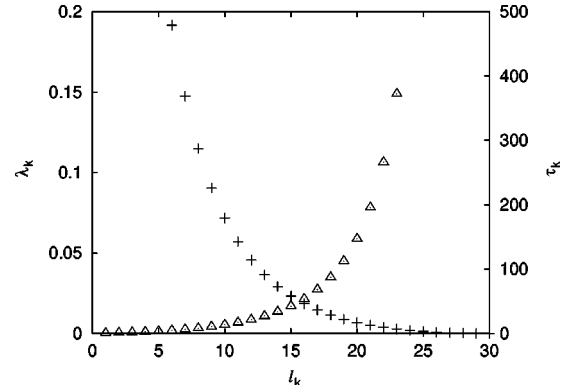

FIG. 3. λ_k (+) and decay time τ_k ($\triangle$) plotted against l_k for a simple Penna model with $l_{max} = 30$, $\beta = \frac{1}{30}$.

where $L_k(l)$ is given by

$$L_k(l) = \frac{1 - \left(\dfrac{1}{1-\lambda_k}\right)^l}{1 - \dfrac{1}{1-\lambda_k}}. \tag{17}$$

Each mode mimics the behavior of a population undergoing steady decline, as given in Eq. (9). Note that the dynamic nature of this model means that the general steady-state solution presented in Ref. [8] does not give a recursion relation of exactly this nature. Naturally in the limit of vanishing λ_k the recursion relation above gives that from the steady-state Penna model.

$n_{k,j}(l)$ retains its meaning from the steady-state analysis as a sum of $n_{k,j}(x,l)$ over ages x. As in the steady-state analysis, $L_k(l)$ gives the sum over ages: $n_{k,j}(l) = L_k(l)n_{k,j}(0,l)$. The decay mode for which $l_k = l_{max}$ has a decay rate of 0 ($\lambda_k = 0$) and is the steady-state solution.

An arbitrary distribution of $n_j(l)$ can be broken down into a sum over decay modes. Decomposition of an $n_j(l)$ distribution into decay modes can be done using a "top down" approach: The largest value of l for which $n_j(l) \neq 0$ gives the l_k value for the largest mode, A_k is then chosen so that
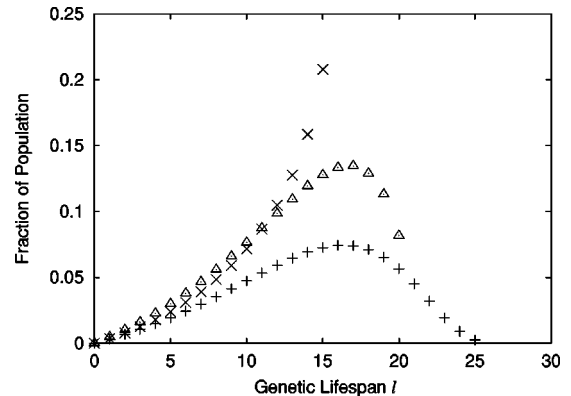

FIG. 4. Sizes of subpopulations within decay modes for $\beta = \frac{1}{30}$, $l_{max} = 30$ with $l_k = 25$ (+), $l_k = 20$ ($\triangle$) and $l_k = 15$ ($\times$). These plots have been rescaled (they are not all normalized to 1) to plot them all on the same axes.

$n_j(l_k) = A_k n_{k,j}(l_k)$. Having determined A_k, A_{k-1} follows as: $n_j(l_k-1) = A_k n_{k,j}(l_k-1) + A_{k-1} n_{k-1,j}(l_k-1)$. This process is repeated, accounting for all contributions from higher modes at each l, until all A_k are determined. Any distribution of $n_j(l)$ [$n(l)=0$ for $l > l_{max}$] can be uniquely broken down into decay modes.

C. Age distribution dynamics in the simple Penna model

The dynamic and steady state behavior of the simple Penna model is dependent on the distribution of ages taking a particular form. An arbitrary distribution of $n(x,l)$ will have its own dynamic behavior and must be considered as a separate case. In the simple Penna model the evolution of $n(x,l)$ over time without contributions from mutation can be dealt with by a Leslie matrix [12] of rank l acting on a vector $n(x,l)$ where the vector components correspond to ages. For $n(x,l)$ the allowed ages are $0,1,2,3,\ldots,l-1$ and the time step evolution can be described by the following matrix equation:

$$\begin{pmatrix} n(0,l) \\ n(1,l) \\ n(2,l) \\ n(3,l) \\ \vdots \end{pmatrix}_{j+1} = A \begin{pmatrix} n(0,l) \\ n(1,l) \\ n(2,l) \\ n(3,l) \\ \vdots \end{pmatrix}_j , \qquad (18)$$

where the matrix A is given by

$$\begin{pmatrix} be^{-\beta l} & be^{-\beta l} & be^{-\beta l} & be^{-\beta l} & \cdots \\ 1 & 0 & 0 & 0 & \cdots \\ 0 & 1 & 0 & 0 & \cdots \\ 0 & 0 & 1 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} . \qquad (19)$$

We require eigenvalue solutions to this so that

$$n(x,l)_{j+1} = \xi n(x,l)_j , \qquad (20)$$

where ξ are the eigenvalues of the matrix A. These eigenvalues can be identified as nonunity solutions to the polynomial

$$\xi^{l+1} - (be^{-\beta l}+1)\xi^l + be^{-\beta l} = 0. \qquad (21)$$

The large eigenvalues give decay at roughly the rate predicted from the $n(l)$ analysis and oscillation around this decay. The small eigenvalues play no significant part over large time scales. ξ_{max} the largest eigenvalue, is related to the decay rate of the decay mode $n_{kj}(l)$ by $\lambda_k = 1 - \xi_{max}$. If $l \ll 1/\lambda_k$ then the long term behavior of any perturbation in $n(l)$ will be dictated by a series of decay modes.

IV. ANALYSIS OF COMPUTATIONAL DYNAMICS

A Simple Penna model simulation was run with b and β chosen to give $l_{max}=30$. The population was initialized with $n(x,l)=0$ with the exception of a spike at $n(0,25)$ where 10^5 organisms were created. The simulation was then al-
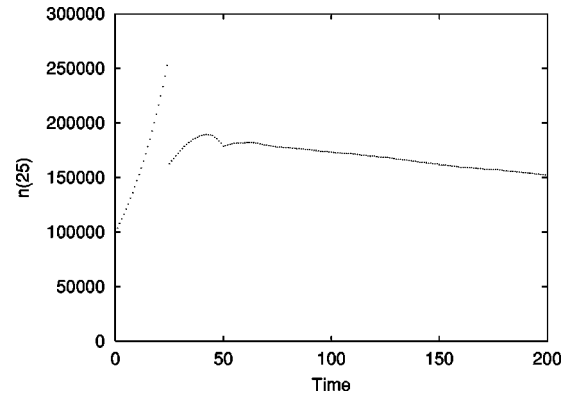

FIG. 5. A plot of $n(25)$ against time from simulation results. The population is initialized with $n(0,25)=10^5$ and all other $n(x,l)$ set to zero. The sudden drop in population at 25 time steps arises because all of the initial population die once they reach age 25. After 100 iterations the noise induced by age distribution dynamics has essentially disappeared and the population dies away exponentially with time constant τ_k of 832.4.

lowed to run generating data for $n(l)$ at each iteration. As the maximum life-span in the population is less than l_{max} the population will eventually decay away to nothing. The spike was chosen as the initial distribution as any initial distribution can be considered to be a sum of such spikes.

Initially the dynamics of $n(25)$ will be dominated by age distribution dynamics and cannot be explained in terms of modes. After 100 iterations this noise has all but disappeared and the subpopulation can be seen to decay exponentially as predicted by Eq. (13) (see Fig. 5). All other subpopulations initially grow to later enter a period of exponential decay (see Fig. 6).

Considering the population after 100 iterations, modes with $\tau_k < 100$ will play an insignificant part in the long term dynamics of any subpopulation. After large number of time steps t, each subpopulation, to a good approximation, can be represented by the highest few modes (with largest decay times):
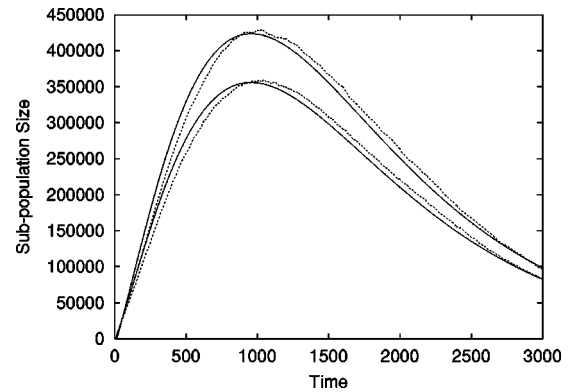

FIG. 6. Analytical results for the decay of $n(20)$ and $n(10)$ are compared with those from simulation. The upper graphs correspond to $n(20)$, the lower to $n(10)$. Analytical results are given by the solid lines, simulation results by the dotted lines. The analytical results are obtained using only the highest four modes.

$$n_t(l) \approx A_{25} n_{25,0}(l) e^{-t/\tau_{25}} + A_{24} n_{24,0}(l) e^{-t/\tau_{24}}$$
$$+ A_{23} n_{23,0}(l) e^{-t/\tau_{23}} + A_{22} n_{22,0}(l) e^{-t/\tau_{22}} \ldots \tag{22}$$

The notation used is $n_{k,t}(l)$, where k is the mode index, t is the number of time steps after the 100th iteration. A_k are mode coefficients [determined from the distribution of $n(l)$ at the 100*th* iteration].

Taking more modes into account will give a more accurate picture so that the behavior of any subpopulation can be described exactly (in the absence of age dynamic induced noise) by

$$n_t(l) = \sum_{k=1}^{25} A_k n_{k,0}(l) e^{-t/\tau_k}. \tag{23}$$

V. CONCLUSION

As a natural extension to our work on the steady state Penna model we have considered and obtained analytical expressions for the various forms of dynamics which can be displayed by a simple Penna model. Our approach can be applied to the governing equations derived in Ref. [8] to analyze the dynamics of a variety of Penna models, such as those with external death rates and birth cutoffs. For multiple disease Penna models ($T>1$) the breakdown of the population into decay modes remains valid so long as there is no bias in the distribution of nonterminal mutations within bit strings.

The decay coefficients λ_k and relationship between $n_{k,j}(l)$ within modes for models with $T>1$ are given as they may be of particular interest. Each λ_k satisfies

$$\lambda_k = 1 - \frac{1 - \left(\dfrac{1}{1-\lambda_k} \right)^{l_k}}{1 - \dfrac{1}{1-\lambda_k}} b e^{-\beta(l_k - T + 1)}. \tag{24}$$

The recursion relation between successive subpopulations within a mode is

$$\frac{n_{k,j}(l+1)}{n_{k,j}(l)} = \frac{C_{T-1}^{l+1}}{C_{T-1}^l} \frac{L_k(l+1)}{L_k(l)}$$
$$\times \frac{(1-\lambda_k) e^{\beta(l-T+1)} - b L_k(l)}{(1-\lambda_k) e^{\beta(l+1-T+1)} - b L_k(l+1) e^{-\beta}}, \tag{25}$$

where $L_k(l)$ is unchanged from the single-mutation case and given by Eq. (17).

The dynamic behavior we have considered is for unregulated populations where there is no Verhulst factor and steady state is obtained by suitable choice of birth rate and mutation rate. Dynamics in the presence of a Verhulst factor is considerably more complicated as any change in a subpopulation will act to change the birth (or death) rate, affecting all other subpopulations. The modes, which in the absence of a Verhulst factor are independent, become coupled through a shared birth (or death) rate. The resulting series of coupled, nonlinear difference equations cannot be generally studied using the decay modes analysis we have presented. However, depending how the Verhulst factor is implemented there may be regimes in which the coupling between decay modes is sufficiently weak that they may be treated as independent.

[1] M. R. Rose, *The Evolutionary Biology of Ageing* (Oxford University Press, Oxford, 1991).

[2] S. M. de Oliveira, P. M. C. de Oliveira, and D. Stauffer, *Evolution, Money, War and Computers* (Teubner, Stuttgart-Leipzig, 1999).

[3] B. Charlesworth, *Evolution in Age-Structured Populations* (Cambridge University Press, Cambridge, 1994).

[4] D. Stauffer, *Biological Evolution and Statistical Physics* (Springer, Berlin, 2002).

[5] S.M. Jazwinski, Science **273**, 54 (1996).

[6] T.J.P. Penna, J. Stat. Phys. **78**, 1629 (1995).

[7] J.B. Coe, Y. Mao, and M.E. Cates, Phys. Rev. Lett. **89**, 288103 (2002).

[8] J.B. Coe and Y. Mao, Phys. Rev. E **67**, 061909 (2003).

[9] A.F.R. de Toledo Piza, Physica A **242**, 195 (1997).

[10] T.J.P. Penna and D. Stauffer, Int. J. Mod. Phys. **6**, 233 (1995).

[11] P. F. Verhulst, Nouv. mm. de l'Academie Royale des Sci. et Belles-Lettres de Bruxelles **18**, 1 (2845).

[12] P.H. Leslie, Biometrika **33**, 183 (1945).