Contents lists available at ScienceDirect







journal homepage: www.elsevier.com/locate/mbs

# Equilibrium distributions and simulation methods for age structured populations



### Fredrik Olsson\*, Ola Hössjer

Department of Mathematics, Division of Mathematical Statistics, Stockholm University, Stockholm, Sweden

#### ARTICLE INFO

Article history: Received 18 December 2014 Revised 3 August 2015 Accepted 5 August 2015 Available online 11 August 2015

Keywords: Age structured population Simulation method Demographic variation Genetic variation Matrix analytic method

#### 1. Introduction

ABSTRACT

A simulation method is presented for the demographic and genetic variation of age structured haploid populations. First, we use matrix analytic methods to derive an equilibrium distribution for the age class sizes conditioned on the total population size. Knowledge of this distribution eliminates the need of a burn-in time in simulations. Next, we derive the distribution of the alleles at a polymorphic locus in various age classes given the allele frequencies in the total population and the age size composition. For the time dynamics, we start by simulating the dynamics for the total population. In order to generate the inheritance of the alleles, we derive their distribution conditionally on the simulated population sizes. This method enables a fast simulation procedure of multiple loci in linkage equilibrium.

© 2015 Elsevier Inc. All rights reserved.

Simulation studies are important as a tool for checking the validity of various assumptions and approximations in population genetic models. Fast and accurate simulation techniques are therefore of interest in order to obtain reliable results. Age structured population models with deterministic growth have been of interest for a long time [4,13,14] and a good overview can be found e.g. in [12] with extended models that account both for demographic and environmental noise. In this paper, we present simulation techniques for an age structured population in a constant environment in which the age class sizes, as well as the total population size, fluctuate stochastically.

We present a discrete time model, where the age composition at each time step is described by means of matrix recursions. Using this model, we derive an approximate distribution for the age composition, given the total population size. By knowing this distribution, the need for a burn-in time is eliminated in simulations since a random draw from this distribution can be used as a starting point.

In population genetics, the genetic information at various loci is important for calculating and estimating e.g. inbreeding and effective population sizes, as reviewed by [2] and [15]. We present a method for simulation of alleles at independent loci given the trajectory of age class sizes. This method can be applied to models in which the

http://dx.doi.org/10.1016/j.mbs.2015.08.003 0025-5564/© 2015 Elsevier Inc. All rights reserved. age class sizes are either constant, or vary stochastically according to some demographic model.

The paper is organized as follows. In Section 2 we define the demographic population model. In the following section we derive an approximate conditional distribution for the age composition, given the total population size. The accuracy of these approximations is tested by means of simulations in Section 4. In Section 5, we present a method for simulation of allele frequencies at loci in linkage equilibrium. A discussion is found in Section 6, derivations in the appendices and a list of the most important notation is given in Table 1.

#### 2. Demographic model

Consider a population divided into J age classes and let

$$\mathbf{N}_t = (N_{t0}, \ldots, N_{tI-1})',$$

with ' referring to vector transposition, be the number of individuals in each age class at time *t*. Let  $Y_{tjh}$  be the number of offspring of an individual *h* in age class *j* at time *t*. For fixed *t* and *j*, all  $Y_{tjh}$ are independent and identically distributed random variables with mean  $b_j$  and variance  $\sigma_j^2$ . The survival  $I_{tjh}$  of an individual *h* from time *t* to t + 1 is Bernoulli distributed with probability  $s_j$  and independent of other individuals' survival. We also allow for a correlation  $\rho_j$  between the number of progeny  $Y_{tjh}$  and survival  $I_{tjh}$ of *h*. Let  $N_{t+1,0j}$  denote the total number of newborns at time *t* of individuals in age class *j*. Then, the dynamics of the population

<sup>\*</sup> Corresponding author. Tel.: +0046 8 16 45 61. *E-mail address:* fredriko@math.su.se (F. Olsson).

**Table 1**List of notation used in the paper.

Notation	Definition
b <sub>i</sub>	Mean number of offspring for an individual in age class j
li	Probability that an individual survives to age class <i>j</i>
S <sub>i</sub>	Probability that an individual in age class j survives to age class $j + 1$
$\dot{\rho}_i$	Correlation between the number of progeny and survival for an individual in age class j
I <sub>tih</sub>	Survival indicator of individual h in age class j from time t to $t + 1$
Y <sub>tjh</sub>	Number of offspring of individual <i>h</i> in age class <i>j</i> at time <i>t</i>
Nt	Vector containing number of individuals in all age classes at time t
N <sub>ti</sub>	Number of individuals in age class <i>j</i> at time <i>t</i>
N <sub>ti0</sub>	Number of newborns at time t of individuals in age class j
$G_t$	Projection matrix of vital rates for <b>N</b> <sub>t</sub>
g	Expected projection matrix
$\epsilon_t$	Matrix of serially uncorrelated demographic noise for the N-process
λ	Multiplicative growth rate and largest eigenvalue of $m{g}$
и	Vector with components proportional to the center point of the equilibrium age distribution of <b>N</b>
v	Vector with components proportional to the reproductive values
$\tilde{N}_t$	Population size when age classes are weighted by $oldsymbol{v}$
Z <sub>taj</sub>	Number of individuals with allele <i>a</i> in age class <i>j</i> at time <i>t</i>
$Z_{ta}$	Vector containing number of individuals with allele <i>a</i> in all age classes
$\tilde{Z}_{ta}$	Number of individuals with allele $a$ when age classes are weighted by $\mathbf{v}$
$p_{ta}$	Age averaged allele frequency, with respect to $\mathbf{v}$ , of allele $a$ at time $t$

is given by

$$N_{t+1,j+1} = \sum_{h=1}^{N_{tj}} I_{tjh}, \quad j = 0, \dots, J-2,$$
  

$$N_{t+1,0j} = \sum_{h=1}^{N_{tj}} Y_{tjh}, \quad j = 0, \dots, J-1,$$
  

$$N_{t+1,0} = \sum_{j=0}^{J-1} N_{t+1,0j}.$$
(1)

This implies that the length of each age class is the same and it equals one unit of time. Following [17], the time dynamics of the population size can also be described using matrix population models (cf. [1]). Let

$$\boldsymbol{N}_{t+1} = \boldsymbol{G}_t \boldsymbol{N}_t = \boldsymbol{g} \boldsymbol{N}_t + \boldsymbol{\epsilon}_{t+1} \tag{2}$$

where  $G_t$  is a  $J \times J$  projection matrix of vital rates, g is the expected projection or Leslie matrix [13] and  $\epsilon_{t+1}$  is a column vector with  $E(\epsilon_{t+1}|N_t) = 0$  that represents serially uncorrelated demographic noise. Let  $\lambda_0, \ldots, \lambda_{J-1}$  be the complex-valued eigenvalues of g in descending order with respect to their moduli, and let  $g = QAQ^{-1}$  be its Jordan canonical form. The columns (rows) of the matrix  $Q(Q^{-1})$  are the right (left) eigenvectors of g, and A is an upper triangular matrix with  $\lambda_0, \ldots, \lambda_{J-1}$  along the diagonal (see for instance [7]).

The largest eigenvalue  $\lambda = \lambda_0$  of  $\mathbf{g}$ , which is real-valued, positive and unique according to Perron–Frobenius theorem, represents the multiplicative growth rate of the population. The right eigenvector  $\mathbf{u}$ corresponding to  $\lambda$  consists of the stable age distribution and the elements of the left eigenvector  $\mathbf{v}$  are proportional to the age specific reproductive values [6]. It is assumed that the elements of  $\mathbf{u}$  and  $\mathbf{v}$  are normalized so that  $\sum_{j=0}^{J-1} u_j = \sum_{j=0}^{J-1} u_j v_j = 1$ . The age specific reproductive values are of importance for age structured populations. For instance, if they are used as weights when calculating the variance effective population size it is possible to determine the long term genetic drift [5,11,17,20].

#### 3. Distribution of the age composition

Suppose that the reproductively weighted population size at time t is  $\tilde{N}_t = \nu N_t$ . Here, we will derive an approximate age distribution for both the total population as well as for different alleles at a specific chromosomal locus. We show in Appendix A that recursion (2) can be

rewritten as

$$\boldsymbol{N}_{t+1} - \tilde{N}_{t+1}\boldsymbol{u} = \boldsymbol{g}(\boldsymbol{N}_t - \tilde{N}_t\boldsymbol{u}) + \boldsymbol{\Pi}_2\boldsymbol{\epsilon}_{t+1}, \qquad (3)$$

where  $\Pi_2 = \mathbf{Q}\mathbf{I}_2\mathbf{Q}^{-1}$  and  $\mathbf{I}_2 = \text{diag}(0, 1, \dots, 1)$  are  $J \times J$  matrices. Iterating (3) with respect to t we can express the deviation from the stable age distribution  $\tilde{N}_t \mathbf{u}$  at time t as

$$\boldsymbol{N}_t - \tilde{N}_t \boldsymbol{u} = \sum_{\tau=0}^{\infty} \boldsymbol{g}^{\tau} \boldsymbol{\Pi}_2 \boldsymbol{\epsilon}_{t-\tau}.$$
(4)

Following calculations in [3] and [17], the noise covariance matrix is

$$\operatorname{Cov}(\boldsymbol{\epsilon}_t | \boldsymbol{N}_{t-1}) \approx \tilde{N}_{t-1} \boldsymbol{\Sigma},$$
 (5)

where  $\mathbf{\Sigma} = (\Sigma_{ij})$  has non-zero elements given by

$$\Sigma_{00} = \sum_{j=0}^{J-1} u_j \sigma_j^2,$$
  

$$\Sigma_{j+1,j+1} = u_j s_j (1-s_j), \quad j = 0..., J-2,$$
  

$$\Sigma_{0,j+1} = \Sigma_{j+1,0} = u_j \sigma_j \sqrt{s_j (1-s_j)} \rho_j, \quad j = 0..., J-2.$$
(6)

In formula (5), the number of individuals in each class j at time t - 1 is approximated by  $\tilde{N}_{t-1}u_j$ , so that for instance the variance of the total reproductive success of all age j individuals is roughly  $\tilde{N}_{t-1}u_j\sigma_j^2$ . Since  $\{\epsilon_t\}$  are martingale differences, it follows from (4), (5) and the central limit theorem for martingales [8,10] that

$$\begin{aligned} \mathbf{N}_{t} &- \tilde{N}_{t} \boldsymbol{u} | \{ \tilde{N}_{t-\tau-1} \}_{\tau=0}^{\infty} \approx N(\mathbf{0}, \sum_{\tau=0}^{\infty} \tilde{N}_{t-\tau-1} \boldsymbol{g}^{\tau} \, \boldsymbol{\Pi}_{2} \boldsymbol{\Sigma} \, \boldsymbol{\Pi}_{2}^{\prime} (\boldsymbol{g}^{\tau})^{\prime}) \\ &\approx N(\mathbf{0}, \tilde{N}_{t-1} \boldsymbol{V}), \end{aligned}$$
(7)

is a good approximation if the sum does not converge too rapidly, so that many terms contribute. In the last step we assumed that  $\tilde{N}_{t-\tau-1}/\tilde{N}_{t-1} \approx \lambda^{-\tau}$ , so that the covariance matrix is proportional to

$$\boldsymbol{V} = \sum_{\tau=0}^{\infty} \lambda^{-\tau} \boldsymbol{g}^{\tau} \boldsymbol{\Pi}_{2} \boldsymbol{\Sigma} \boldsymbol{\Pi}_{2}^{\prime} (\boldsymbol{g}^{\tau})^{\prime}, \qquad (8)$$

and in order for the sum in (8) to converge, it is necessary that  $|\lambda_1|^2 < \lambda = \lambda_0$ , see [17]. Hence,

$$\boldsymbol{N}_{t}|\tilde{N}_{t},\tilde{N}_{t-1}\approx N(\tilde{N}_{t}\boldsymbol{u},\tilde{N}_{t-1}\boldsymbol{V}), \qquad (9)$$

is the conditional distribution of the age composition given the weighted population size.

Download English Version:

## https://daneshyari.com/en/article/4499956

Download Persian Version:

https://daneshyari.com/article/4499956

Daneshyari.com