Theoretical Population Biology 109 (2016) 63–74

Contents lists available at ScienceDirect

Theoretical Population Biology

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

Genetic drift in populations governed by a Galton–Watson branching process

Conrad J. Burden [a,*], Helmut Simon [b]

[a] Mathematical Sciences Institute, Australian National University, Canberra, Australia
[b] John Curtin School of Medical Research, Australian National University, Canberra, Australia

ARTICLE INFO

Article history:
Received 28 September 2015
Available online 24 March 2016

Keywords:
Genetic drift
Galton–Watson
Branching process
Mitochondrial Eve

ABSTRACT

Most population genetics studies have their origins in a Wright–Fisher or some closely related fixed-population model in which each individual randomly chooses its ancestor. Populations which vary in size with time are typically modelled via a coalescent derived from Wright–Fisher, but use a nonlinear time-scaling driven by a deterministically imposed population growth. An alternate, arguably more realistic approach, and one which we take here, is to allow the population size to vary stochastically via a Galton–Watson branching process.

We study genetic drift in a population consisting of a number of distinct allele types in which each allele type evolves as an independent Galton–Watson branching process. We find the dynamics of the population is determined by a single parameter $\kappa_0 = (2m_0/\sigma^2)\log\lambda$, where m_0 is the initial population size, λ is the mean number of offspring per individual; and σ^2 is the variance of the number of offspring. For $0 \lesssim \kappa_0 \ll 1$, the dynamics are close to those of Wright–Fisher, with the added property that the population is prone to extinction. For $\kappa_0 \gg 1$ allele frequencies and ancestral lineages are stable and individual alleles do not fix throughout the population. The existence of a rapid changeover regime at $\kappa_0 \approx 1$ enables estimates to be made, together with confidence intervals, of the time and population size of the era of mitochondrial Eve.

1. Introduction

The traditional starting point in population genetics for the consideration of genetic drift and related phenomena has been the Wright–Fisher model, often with caveats about its limitations and lack of biological realism (Ewens, 2004). A key constraint on the Wright–Fisher model is that the population is assumed to be of constant size: that is, not only is the mean size of the population constant over time, but the population size does not vary stochastically. The assumption of a constant mean size is often appropriate biologically, on the basis that the population is at a limit constrained by the availability of food, space or other factors, as argued in Kingman (1982a). In other cases, such as the study of human populations or arguably of new species, a model allowing for population growth may be more relevant. The assumption that there is no stochastic influence on overall population size cannot be entirely accurate in relation to any natural population and it is of interest to understand the costs of making such an assumption. For example, it is well accepted that stochastic influences are particularly strong for small population sizes. The differing views between the pioneers of population genetics Ronald Fisher and Sewall Wright on the relative significance of drift and selection are surveyed in Ewens (2004, Section 1.7), with Fisher having emphasised the role of natural selection over stochastic effect in large populations and Wright having given more attention to the effects of drift in relatively small populations. The relative importance of drift (and mutation) and selection on biological populations continues to be debated into the present era as exemplified by Lynch (2007). We will see that in the context of a growing population, the degree of influence of stochastic processes is dependent not only on population size, but rather on a parameter combining population size and growth rate.

In this context, there have been a number of studies addressing by theory or simulation studies the effect of variable and in particular exponential population growth (Slatkin and Hudson, 1991; Keinan and Clark, 2012; Rogers and Harpending, 1992; Rogers, 2014). Exponential population growth may be considered

* Corresponding author.
E-mail addresses: conrad.burden@anu.edu.au (C.J. Burden), helmut.simon@anu.edu.au (H. Simon).

http://dx.doi.org/10.1016/j.tpb.2016.03.002
0040-5809/

as the simplest model extending the assumption of constant population size. Two approaches are evident. The first is by an extension of the Wright–Fisher model to the case of variable population size based on transformation of the time variable to reflect the local rate of change of population size (Griffiths and Tavaré, 1994). This approach retains the assumption that overall population size is deterministic, and that the Wright–Fisher model describes the relationship of one generation to its successor. This approach is conducive to efficient simulation and to a considerable degree of analytic study. However, it potentially leads to inaccurate estimation of population parameters due to underestimation of the variability in population size, particularly in the early stages of population growth (Stadler et al., 2015). An alternative approach (O'Connell, 1995; Cyran and Kimmel, 2010) is to use the theory of branching processes, and in particular Galton–Watson processes, to model populations evolving according to a common probability distribution for the number of offspring on the basis that this distribution is the same for the entire population and in each generation.

It is of interest to compare the effects of these population models (constant-size Wright–Fisher; deterministic growth; and fully stochastic growth) across various population types. As we shall see the key population parameter to be considered depends on both population size and growth rate.

The structure of this paper is as follows. The stochastic model used throughout, which is equivalent to that of O'Connell (1995), is set out in Section 2. Section 3 contains a comparison of the model with Wright–Fisher for the case of zero mean population growth, with emphasis on fixation probabilities and times. The case of supercritical growth is analysed in Section 4. Implications for the loss or otherwise of heterozygosity under different parameter regimes are considered in Section 5. An application to the estimation of the time elapsed since the life of mitochondrial Eve and human population size during her lifetime is given in Section 6. Conclusions are drawn in Section 7, and an Appendix is devoted to critiquing O'Connell's analysis of the mitochondrial Eve problem.

2. The model

We consider a population of $M(t)$ haploid individuals which are assumed to reproduce in discrete, non-overlapping generations $t = 0, 1, 2, \ldots$. We further assume that the population contains n allele types and that the number of copies of type i within the population is $Y_i(t)$, thus $M(t) = \sum_{i=1}^{n} Y_i(t)$. The alleles are assumed to be neutral with respect to selection and no mutation is included in the model in the current paper. The central tenet of the model is an assumption that the number of offspring per individual in any generation is given by a set of identically and individually distributed (i.i.d.) random variables $S_\alpha^{(i)}$, $\alpha = 1, \ldots, Y_i(t)$, whose common distribution is denoted by a generic non-negative integer valued random variable S with mean and variance

$$E(S) = \lambda, \qquad \mathrm{Var}\,(S) = \sigma^2, \tag{1}$$

and finite moments to all higher orders. Thus

$$Y_i(t+1) = \sum_{\alpha=1}^{Y_i(t)} S_\alpha^{(i)}, \tag{2}$$

and if the $Y_i(0)$ are mutually independent, then $Y_i(t)$ are mutually independent at all subsequent times t. The standard formula for the mean of the sum of a random number of random variables gives $E(Y_i(t+1)) = \lambda E(Y_i(t))$. Given initial conditions $Y_i(0) = y_{i0}$ and $M(0) = m_0 = \sum_{i=1}^{n} y_{i0}$, it follows that

$$E(Y_i(t)) = y_{i0}\lambda^t, \qquad E(M(t)) = m_0\lambda^t. \tag{3}$$

The $Y_i(t)$ represent n independent Galton–Watson branching processes evolving in parallel. This model differs from the canonical Wright–Fisher and Cannings exchangeable (Cannings, 1974) models of population genetics primarily in allowing the mean growth rate λ to take values other than unity, and in not constraining the sum of the random variables $S_\alpha^{(i)}$ to a constant value. In the population genetics context, the most widely used adaptation of these models assumes a deterministic rather than stochastic growth rate (Slatkin and Hudson, 1991; Griffiths and Tavaré, 1994) in the context of the Kingman coalescent (Kingman, 1982b). As we shall see, this assumption leads to an underestimate of the variability in a population, particularly in the early stages of population growth, when population size or, more precisely the parameter $(2m_0/\sigma^2)\log\lambda$, is small (see Stadler et al., 2015 for a related discussion).

The continuum approximation of a Galton–Watson branching process via a forward Kolmogorov diffusion equation has been analysed in detail by Feller et al. (1951) and is summarised by Bailey (1964) and Cox and Miller (1978). Here we summarise the derivation using the formal method given in Chapter 4 of Ewens (2004). For large initial population sizes m_0 and y_{i0}, consider the substitutions

$$\tau = t/m_0, \qquad X_i(\tau) = Y_i(t)/m_0, \qquad x_{i0} = y_{i0}/m_0. \tag{4}$$

Set $\delta X_i(\tau) = (Y_i(t+1) - Y_i(t))/m_0$ and $\delta\tau = 1/m_0$. In general, if

$$\begin{aligned} &E(\delta X_i(\tau)|X_i(\tau) = x) = a(x)\delta\tau + o(\delta\tau),\\ &E(\delta X_i(\tau)^2|X_i(\tau) = x) = b(x)\delta\tau + o(\delta\tau),\\ &E(\delta X_i(\tau)^k|X_i(\tau) = x) = o(\delta\tau), \quad k \geq 3, \end{aligned} \tag{5}$$

for finite functions $a(x)$ and $b(x)$ as $\delta t \to 0$, then the forward Kolmogorov equation for the density of the distribution of $X_i(\tau)$ takes the form

$$\frac{\partial f_{X_i}}{\partial\tau} = -\frac{\partial}{\partial x}\left(a(x)f_{X_i}(x,\tau)\right) + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left(b(x)f_{X_i}(x,\tau)\right). \tag{6}$$

In the current case, the limit is obtained by simultaneously taking $m_0 \to \infty, \lambda \to 1$ in such a way that

$$\alpha = m_0\log\lambda, \tag{7}$$

and σ^2 remain fixed. One obtains

$$\begin{aligned} E(\delta X_i(\tau)|X_i(\tau) = x) &= \frac{1}{m_0}(\lambda - 1)xm_0\\ &= \alpha x\,\delta\tau + o(\delta\tau), \end{aligned} \tag{8}$$

and

$$\begin{aligned} E(\delta X_i(\tau)^2|X_i(\tau) = x) &= \mathrm{Var}\,(\delta X_i(\tau)|X_i(\tau) = x)\\ &\quad + E(\delta X_i(\tau)|X_i(\tau) = x)^2\\ &= \frac{1}{m_0^2}\sigma^2 xm_0 + O\left(\frac{1}{m_0^2}\right)\\ &= \sigma^2 x\,\delta\tau + o(\delta\tau). \end{aligned} \tag{9}$$

Comparing with the general form Eq. (5) yields the forward Kolmogorov equation for the density function $f_{X_i}(x,\tau)$,

$$\frac{\partial f_{X_i}}{\partial\tau} = -\alpha\frac{\partial}{\partial x}(xf_{X_i}) + \frac{1}{2}\sigma^2\frac{\partial^2}{\partial x^2}(xf_{X_i}). \tag{10}$$

The solution for $\alpha \neq 0$ and initial condition $f_{X_i}(x,0) = \delta(x - x_{i0})$ is (see Bailey, 1964, Eqs. (14.49) and (14.53))

$$\begin{aligned} f_{X_i}(x,\tau) &= \frac{2\alpha}{\sigma^2(e^{\alpha\tau} - 1)}\left(\frac{x_{i0}e^{\alpha\tau}}{x}\right)^{\frac{1}{2}}\exp\left\{\frac{-2\alpha(x_{i0}e^{\alpha\tau} + x)}{\sigma^2(e^{\alpha\tau} - 1)}\right\}\\ &\quad \times I_1\left(\frac{4\alpha(x_{i0}xe^{\alpha\tau})^{\frac{1}{2}}}{\sigma^2(e^{\alpha\tau} - 1)}\right) + \delta(x)p_0(\tau), \end{aligned} \tag{11}$$