Theoretical Population Biology 92 (2014) 30–35

Contents lists available at ScienceDirect

Theoretical Population Biology

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

A path integral formulation of the Wright–Fisher process with genic selection

Joshua G. Schraiber

Department of Integrative Biology, UC Berkeley, United States

ARTICLE INFO

Article history:
Received 28 October 2013
Available online 21 November 2013

Keywords:
Wright–Fisher
Diffusion theory
Path Integral
Perturbation theory

ABSTRACT

The Wright–Fisher process with selection is an important tool in population genetics theory. Traditional analysis of this process relies on the diffusion approximation. The diffusion approximation is usually studied in a partial differential equations framework. In this paper, I introduce a path integral formalism to study the Wright–Fisher process with selection and use that formalism to obtain a simple perturbation series to approximate the transition density. The perturbation series can be understood in terms of Feynman diagrams, which have a simple probabilistic interpretation in terms of selective events. The perturbation series proves to be an accurate approximation of the transition density for weak selection and is shown to be arbitrarily accurate for any selection coefficient.

1. Introduction

Modern population genetics theory can be broken down into two broad subclasses: forward-in-time, in which the generation-to-generation allele frequency dynamics are tracked, and backward-in-time, in which genealogical relationships are modeled. While forward-in-time models were developed first, the introduction of the coalescent by Kingman (1982) ushered in a revolution in our understanding of neutral genetic variation. The success of the coalescent in providing a simple framework for analyzing neutral loci has inspired a number of attempts to construct a genealogical representation of models with natural selection (Krone and Neuhauser, 1997; Neuhauser and Krone, 1997; Donnelly and Kurtz, 1999). However these models have not been particularly amenable to analysis due to their complicated structure.

The forward-in-time approach remains the most straightforward method for analyzing genetic variation under the combined effects of genetic drift and natural selection. This approach is characterized by the diffusion approximation to the Wright–Fisher model (Ewens, 2004). For many important quantities (such as ultimate fixation probabilities), the diffusion approximation provides a concise, exact analytic expression. These formulas, in terms of common parameters such as the population scaled selection coefficient α, allow for an understanding of how different evolutionary forces impact the dynamics of allele frequency change. Assuming a constant population size, exact analytic results from the diffusion approximation can even be used to estimate the distribution of selection coefficients in the genome (Boyko et al., 2008; Torgerson et al., 2009).

Unfortunately, when both selection and genetic drift affect allele frequency dynamics, there is no simple analytic expression for the transition density of the diffusion (that is, the probability that an allele currently at frequency x is at frequency y after t time units have passed). Recently, interest in the transition density has been fueled by advances in experimental evolution (Kawecki et al., 2012) and ancient DNA (Wall and Slatkin, 2012), leading to the development of numerous methods for estimating the population scaled selection coefficient from allele frequency time series data (Bollback et al., 2008; Malaspinas et al., 2012; Mathieson and McVean, 2013; Feder et al., 2013). Moreover, because the transition density fully characterizes the allele frequency dynamics, many interesting quantities, such as the time-dependent fixation probability, could be calculated once the transition density is known.

While the diffusion approximation allows one to write down a partial differential equation (PDE) that the transition density must satisfy, it has proved challenging to solve in a robust manner either analytically or numerically. Numerical solution of the PDE is, in principle, straightforward by discretization techniques (see Zhao et al. (2013) for a recent approach that accounts for fixations and losses of alleles). However, because the relative importance of drift and selection depend on the allele frequency, the discretization scheme must be chosen wisely. Another drawback of numerical methods is that they can be quite time consuming; in particular, this is what limits the method of Gutenkunst et al. (2009) to 3 populations while using a diffusion approximation to find the site frequency spectrum for demographic inference.

Kimura (1955b) provided an analytical solution to the transitional density with selection, in the form of an eigenfunction decomposition with oblate spheroid wave functions. However, he was unable to compute the eigenvalues exactly, instead resorting to perturbation theory. Motivated by the fact that the

E-mail address: jgschraiber@berkeley.edu.

0040-5809/$ – see front matter
http://dx.doi.org/10.1016/j.tpb.2013.11.002

eigenfunction decomposition of the model with no selection is known, Song and Steinrücken (2012) developed a novel computational method for approximating the transition density analytically. Their method, based on the theory of Hilbert spaces spanned by orthogonal polynomials, is a significant advance and represents the state-of-the-art in terms of finding the transition density with selection. This method still has several limitations, as it needs to be recomputed if a new selection coefficient is chosen; moreover, for certain values of the selection coefficient and dominance parameter, computation times can be long because they were required to use high-precision arithmetic.

In this paper, I present a novel method for approximating the transition density of the Wright–Fisher diffusion with genic selection. This method is based on the theory of path integration, which was introduced by Wiener (1921) for Brownian motion and has found substantial success in applications in quantum mechanics (Feynman, 1948; Feynman and Hibbs, 2012) and quantum field theory (Zee, 2010). The key insight of this approach is to associate every path from x at time 0 to y at time t with a probability, and then integrate over all possible paths to find the transition density. While computing this integral exactly is only possible in the neutral case, I develop a perturbation scheme to approximate it as a power series in α for the case with genic selection. To facilitate the computation of this perturbation expansion, I demonstrate the use of a mnemonic, called Feynman diagrams, to compute the transition density to arbitrary accuracy.

2. Methods

2.1. Partial differential equation formulation

Here I review some preliminaries about the Wright–Fisher diffusion that will prove useful in the following. Denoting by $\phi_\alpha(x, y; t)$ the transition density with genic selection and population-scaled selection coefficient α, the standard theory shows that ϕ satisfies the PDE

$$\frac{\partial}{\partial t}\phi_\alpha(x,y;t) = \frac{1}{2}\frac{\partial^2}{\partial y^2}\left\{y(1-y)\phi_\alpha(x,y;t)\right\} - \alpha\frac{\partial}{\partial y}\left\{y(1-y)\phi_\alpha(x,y;t)\right\}, \quad (1)$$

with the initial condition $\phi_\alpha(x, y, 0) = \delta(x - y)$ where $\delta(\cdot)$ is the usual Dirac delta function (Ewens, 2004).

Kimura (1955a) found that for the case $\alpha = 0$, the transition density admits an eigenfunction decomposition,

$$4x(1-x)\sum_{i=1}^{\infty}\frac{2i+1}{i(i+1)}C_{i-1}^{(3/2)}(1-2x)C_{i-1}^{(3/2)}(1-2y)e^{-\frac{1}{2}i(i+1)t}, \quad (2)$$

where the $C_i^\lambda(z)$ are the Gegenbauer polynomials.

2.2. Path integral formulation

The path integral formulation begins by defining a probability density functional, which assigns a probability density to any path from x to y. Then, the total transition probability from x to y is computed by integrating this density over all paths from x to y.

This probability density functional can be developed intuitively by considering the "short-time transition densities". Standard theory for diffusion processes shows that when $\delta t \ll 1$, we can approximate

$$\phi_\alpha(x, y; \delta t) \approx \frac{1}{\sqrt{2\pi x(1-x)\delta t}}\exp\left\{\frac{(y-(x+\alpha x(1-x)))^2}{2x(1-x)\delta t}\right\}.$$

A naive approach might be to attempt to approximate the probability density of a path by dividing the interval $[0, t]$ into n intervals of length δt. Then we approximate with the probability of the so-called "zig–zag path",

$$\mathcal{P}[z] \approx \prod_{i=1}^{n}\phi_\alpha(z_{i-1}, z_i; \delta t)dz_i,$$

with $z_i = z(i\delta t)$. However, this fails for a variety of reasons, in particular the dependence of the diffusion coefficient on the current allele frequency (Graham, 1977; Dürr and Bach, 1978). Instead, I compute the relative probability density function for a path with selection compared to a neutral path. This functional, which can be rigorously derived using Girsanov's theorem (Rogers and Williams, 2000) can be intuitively developed as

$$\begin{aligned}\mathcal{G}[z] &\approx \frac{\prod_{i=1}^{n}\phi_\alpha(z_{i-1}, z_i; \delta t)dz_i}{\prod_{i=1}^{n}\phi_0(z_{i-1}, z_i; \delta t)dz_i}\\ &\approx \frac{\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi z_{i-1}(1-z_{i-1})\delta t}}\exp\left\{\frac{(z_i-(z_{i-1}+\alpha z_{i-1}(1-z_{i-1})))^2}{2z_{i-1}(1-z_{i-1})\delta t}\right\}dz_i}{\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi z_{i-1}(1-z_{i-1})\delta t}}\exp\left\{\frac{(z_i-z_{i-1})^2}{2z_{i-1}(1-z_{i-1})\delta t}\right\}dz_i}\\ &= \exp\left\{\alpha\sum_{i=1}^{n}(z_i - z_{i-1}) - \frac{\alpha^2}{2}\sum_{i=1}^{n}z_i(1-z_i)\delta t\right\}.\end{aligned}$$

Thus, as $\delta t \downarrow 0$ and $n \uparrow \infty$ such that $n\delta t = t$, we have

$$\mathcal{G}[z] = \exp\left\{\alpha(y-x) - \frac{\alpha^2}{2}\int_0^t z(1-z)ds\right\}, \quad (3)$$

in which the time dependence of z is suppressed for notational convenience. Now, we can write the transition density as the integral over all *neutral* Wright–Fisher paths of the relative probability of that path with selection,

$$\phi_\alpha(x, y; t) = \int_{(0,x)}^{(t,y)} e^{\alpha(y-x)-\frac{\alpha^2}{2}\int_0^t z(1-z)ds}\mathcal{D}z \quad (4)$$

where $\mathcal{D}z$ is the measure on path space induced by the neutral Wright–Fisher process.

The path integral (4) can be understood as depicted in Fig. 1. Paths from x to y can be drawn from the neutral Wright–Fisher path measure, $\mathcal{D}z$. For each path, the functional $\mathcal{G}[\cdot]$ is evaluated (panel a). This results in a one-dimensional probability distribution for values of $\mathcal{G}$ (panel b). Then, the mean value of the distribution of $\mathcal{G}$ can be computed, which is equal to the transition density.

2.3. Perturbation approximation

I now show how to approximate the transition density using a perturbation expansion. Note that the first term in the exponential of (4) is independent of the path, and hence we focus on the path integral

$$\int_{(0,x)}^{(t,y)} e^{-\frac{\alpha^2}{2}\int_0^t z(1-z)ds}\mathcal{D}z.$$

We begin by expanding the exponential in a Taylor series about $\alpha = 0$,