Published in final edited form as:

Theor Popul Biol. 2014 March ; 92: 30–35. doi:10.1016/j.tpb.2013.11.002.

A PATH INTEGRAL FORMULATION OF THE WRIGHT-FISHER PROCESS WITH GENIC SELECTION

JOSHUA G. SCHRAIBER

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 straight-forward 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, 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].

© 2013 Elsevier Inc. All rights reserved.

Corresponding author: jgschraiber@berkeley.edu, Department of Integrative Biology, UC Berkeley..

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

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 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 a for the case with genic selection. To facilitate 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 $\varphi_a(x, y; t)$ the transition density with genic selection and population-scaled selection coefficient a , standard theory shows that φ satisfies the PDE

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

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 Figure 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 $\alpha = 0$,

$$\begin{aligned} \int_{(0,x)}^{(t,y)} e^{-\frac{\alpha^2}{2} \int_0^t z(1-z) ds} \mathcal{D}z &= \int_{(0,x)}^{(t,y)} \sum_{k=0}^{\infty} (-1)^k \frac{\alpha^{2k}}{2^k k!} \left(\int_0^t z(1-z) ds \right)^k \mathcal{D}z \\ &= \sum_{k=0}^{\infty} (-1)^k \frac{\alpha^{2k}}{2^k k!} \int_{(0,x)}^{(t,y)} \left(\int_0^t z(1-z) ds \right)^k \mathcal{D}z. \end{aligned} \quad (5)$$

In the Appendix, I show that the exchange of the summation and the integral is justified by Fubini's theorem. This is in stark contrast to the case in quantum physics, in which the exchange of the sum and integral is often not justified, leading to zero radius of convergence in the perturbation parameter.

Thus, the task of approximating the transition density with selection is reduced to the task of computing the functional integrals

$$\int_{(0,x)}^{(t,y)} \left(\int_0^t z(1-z) ds \right)^k \mathcal{D}z.$$

Integrals of this form were considered by Nagylaki [1974] and Watterson [1979] although they were focused on the case where the allele is eventually fixed or lost, whereas here we need to consider only those paths that go from x to y in time t . To compute these integrals, it is useful to introduce a diagrammatic method, known as a Feynman diagram [Feynman and Hibbs, 2012, Chorin and Hald, 2006]. Borrowing from the language of physics momentarily, we can regard $V(x) = x(1-x)$ as a *potential energy*, and we can consider the allele frequency being *scattered* by the potential.

The idea can be seen in Figure 2. When the integrand is raised to the k th power, we imagine that the allele frequency changes neutrally until some time s_1 , at which point it interacts with the potential and is scattered. Then, it evolves neutral until time s_2 , at which point it again interacts with the potential and is scattered. This proceeds until the scattering at time s_k , after which the allele frequency evolves neutrally to y at time t . Because the interaction times s_i could have happened at any time between 0 and t and the allele frequency z_i at time s_i is random, we integrate over all times and allele frequencies. For example, we can compute

$$\int_{(0,x)}^{(t,y)} \left(\int_0^t z(1-z) ds \right) \mathcal{D}z = \int_0^t \int_0^1 \phi_0(x, z_1; s_1) z_1 (1-z_1) \phi_0(z_1, y; t-s_1) dz_1 ds_1 \quad (6)$$

and

$$\begin{aligned} & \int_{(0,x)}^{(t,y)} \left(\int_0^t z(1-z) ds \right)^2 \mathcal{D}z \\ &= 2 \int_0^t \int_0^{s_2} \int_0^1 \int_0^1 \phi_0(x, z_1; s_1) z_1 (1-z_1) \phi_0(z_1, z_2; s_2-s_1) z_2 (1-z_2) \phi_0(z_2, y; t-s_2) dz_1 dz_2 ds_1 ds_2, \end{aligned}$$

where the factor of 2 comes from the two orderings in which the scatterings happened. In general, the k th order Feynman diagram will come with a factor of $k!$ to count the number of orderings of the scattering events.

Because we know the neutral transition density, computing the integrals that arise from Feynman diagrams is straightforward. Unfortunately, the neutral transition density is only known as an infinite series and in practice computing the integrals is more difficult. In the Appendix, I show how to achieve efficient computation of these integrals for arbitrary k .

3 Results

3.1 Accuracy of the perturbation expansion

A simple error bound can be derived for perturbation expansions [Harlow, 2009]. For the k th-order perturbation expansion, $\phi_\alpha^{(k)}(x, y; t)$, this bound is

$$\left| \phi_\alpha(x, y; t) - \phi_\alpha^{(k)}(x, y; t) \right| \leq \left| \frac{\alpha^{2(k+1)}}{2^{k+1} (k+1)!} \int_{(0,x)}^{(t,y)} \left(\int_0^t z(1-z) ds \right)^{k+1} \mathcal{D}z \right|. \quad (7)$$

As argued in the Appendix, this bound is less than

$$\frac{\alpha^{2(k+1)} t^{k+1}}{8^{k+1} (k+1)!} \phi_0(x, y; t),$$

and when $t < 4$, it approaches 0 as $k \rightarrow \infty$ for any α . Thus, the perturbation expansion converges to the true transition density for any α , provided that t is small enough.

The error bound presented above is rather crude. To get a more informative picture of the accuracy of the perturbation expansion, I compared to simulations. An interesting quantity that sums up the overall accuracy of the perturbation approximation is the time-dependent probability of absorption. This quantity can be calculated analytically as

$$\int_0^1 \phi_\alpha(x, y; t) dy$$

and is easily estimated from simulations. The perturbation method proves to be increasing accurate as more terms are added to the expansion (Figure 3). However, even for moderate values of α , a large number of terms are required for an accurate approximation.

4 Discussion

The Wright-Fisher process with selection is a primary tool for elucidating the impact of natural selection on genetic variation. However, the transient behavior of the process has been difficult to study, with much work focusing on equilibrium aspects, such as the stationary distribution [Wright, 1931] or the site frequency spectrum [Sawyer and Hartl, 1992]. Nonetheless, transient dynamics have an important impact on natural variation and are critical to forming a complete understanding of how natural selection shapes genomes. In this paper, I presented a novel path integral formulation of the Wright-Fisher process with genic selection. This led naturally to a simple perturbation scheme for computing the transition density with weak selection.

The perturbation expansion of the transition density can be understood by using Feynman diagrams (Figure 2). Although the traditional motivation for Feynman diagrams comes from quantum physics [Feynman and Hibbs, 2012], they can be interpreted in a population genetic context. For instance, in the first-order term of the perturbation expansion, an allele begins drifting neutrally. At a time when the allele frequency is z , there is a rate $\frac{\alpha^2}{2}z(1-z)$ of a selective event occurring, which has a natural interpretation as an individual of one allelic type encountering an individual of the other allelic type, weighted by the strength of selection. The strength of selection enters as α^2 because the rate of selective events must be positive. After the selective event occurs, the allele again drifts neutrally. Higher-order terms in the perturbation expansion include more selective events. However, the impact of selective events on an overall neutral trajectory is not sufficient to model the dynamics. The additional factor of $e^{\alpha(y-x)}$ multiplying the entire perturbation expansion arises to account for the deterministic effects of selection [Baibuz et al., 1984]; intuitively, it acts to increase the probability that an allele increases in frequency under positive selection or decreases in frequency under negative selection.

The perturbation scheme described in this paper works best for weak selection. For the weak selection scheme examined here, this method and the method of Song and Steinrücken [2012] perform comparably in terms of time taken to obtain an estimate of the transition density. For stronger selection, other approximation methods, including the Gaussian diffusion approximation [Nagylaki, 1990, Feder et al., 2013] may be useful. Also, the model considered in this paper does not have fully general diploid selection. The path integral approach applies in an equally straightforward fashion to diploid selection, but the details of the mathematics become significantly more complicated. In that case, the orthogonal polynomial method of Song and Steinrücken [2012] may be better suited.

Path integral formulations have been used successfully in population genetics in the past. Rouhani and Barton [1987] made use of a path integral to approximate the probability of shifting between selective optima in the context of quantitative trait evolution. Their approximation scheme, however, was quite different than one that I explored. They assumed relatively strong selection and expanded the path integral around the most likely path between the two selective optima, in contrast to the weak selection perturbative approach taken in this paper. More recently, path integrals have been used to examine fitness flux [Mustonen and Lässig, 2010] and Muller's ratchet [Neher and Shraiman, 2012].

A significant strength of the path integral approach is its adaptability to evolution of a locus with a large number of alleles, each of which corresponds to a phenotypic value. In previous

contexts, such a model has been used to model quantitative trait evolution, called a continuum-of-alleles model [Kimura, 1965]. Earlier approaches to incorporate genetic drift into a continuum-of-alleles model using the theory of measure-valued diffusions [Fleming and Viot, 1979, Ethier and Kurtz, 1987] have made significant advances in understanding the neutral dynamics of such processes [Dawson and Hochberg, 1982, Ethier and Griffiths, 1993, Ethier and Kurtz, 1993, Donnelly and Kurtz, 1996]. However, incorporating selection greatly increases the difficulty of obtaining analytical results (but see Donnelly and Kurtz [1999], Dawson and Feng [2001]). It is possible that a path integral formulation of such a process could lead to a perturbative approach to incorporating selection, in much the same way as the path integral approach has been successful in perturbative quantum field theory [Zee, 2010].

5 Software

A Mathematica program to compute the transition density can be obtained by contacting the author.

Acknowledgments

I thank Anand Bhaskar, Michael Schwemmer, Matthias Steinrucken, Yun Song, Nick Barton and two anonymous reviewers for helpful comments on an earlier version of this manuscript. I am grateful to Kelley Harris and Mason Liang for helpful discussions about the mathematical details of this research. This work was partially supported by NIH grants R01-40282 and T32-00047.

7 Appendix

7.1 Exchanging the order of integration and summation in (5) is justified

To establish this fact, we first need a simple bound on the functional

$$\mathcal{F}_k[X_s] = \frac{\alpha^{2k}}{2^k k!} \left(\int_0^t z(1-z) ds \right)^k$$

Note that, because z represents a frequency, we know that z is bounded between 0 and 1 for all s . Thus,

$$\int_0^t z(1-z) ds \leq \int_0^t \frac{1}{2} \left(1 - \frac{1}{2} \right) ds = \frac{1}{4}t.$$

Therefore,

$$\mathcal{F}_k[z] \leq \frac{\alpha^{2k} t^k}{8^k k!}.$$

Now, to apply the general version of Fubini's theorem, which allows the exchange of the order of integration in general measure spaces [Bogachev, 2007], I must show that

$$\int_{(0,x)}^{(t,y)} \left(\sum_{k=1}^{\infty} \mathcal{F}_k[z] \right) \mathcal{D}z < \infty$$

and

$$\sum_{k=0}^{\infty} \int_{(0,x)}^{(t,y)} \mathcal{F}_k[z] \mathcal{D}z < \infty$$

For the first case, observe that

$$\int_{(0,x)}^{(t,y)} \left(\sum_{k=0}^{\infty} \mathcal{F}_k[z] \right) \mathcal{D}z \leq \int_{(0,x)}^{(t,y)} \left(\sum_{k=0}^{\infty} \frac{\alpha^{2k} t^k}{8^k k!} \right) \mathcal{D}z = \int_{(0,x)}^{(t,y)} e^{\frac{\alpha^2}{8} t} \mathcal{D}z = e^{\frac{\alpha^2}{8} t} \phi_0(x, y; t) < \infty.$$

An extremely similar calculation shows that the second case is true as well.

7.2 Computation of Feynman diagrams

The integrals arising from the Feynman diagrams can only be expressed properly as infinite sums. In practice, it is necessary to truncate these sums after a finite number of terms. In this section, I develop an approach efficiently compute the sums. From Kimura's spectral representation of the transition density, equation (2), and the general form of each term in the perturbation expansion, for example equation (6), it is clear that the k th order Feynman diagram results in a term

$$\begin{aligned} & 4x(1-x) \sum_{i_1, i_2, \dots, i_{k+1}} C_{i_1-1}^{(3/2)}(1-2x) C_{i_{k+1}-1}^{(3/2)}(1-2y) \prod_{j=1}^{k+1} \frac{2i_j+1}{i_j(i_j+1)} \\ & \times \prod_{j=1}^{k-1} \int_0^1 z_j^2 (1-z_j)^2 C_{i_j}^{(3/2)}(1-2z_j) C_{i_{j+1}}^{(3/2)}(1-2z_j) dz_j \\ & \times \int_0^t \int_0^{s_k} \dots \int_0^{s_1} e^{-\left(\lambda_{i_1} s_1 + \sum_{j=2}^k \lambda_{i_j} (s_j - s_{j-1}) + \lambda_{i_{k+1}} (t - s_k)\right)} ds_1 \dots ds_{k-1} ds_k, \end{aligned} \quad (8)$$

with $\lambda_i = i(i+1)/2$, the eigenvalues of Kimura's transition density.

The integrals over allele frequencies can be done exactly by using the properties of the Gegenbauer polynomials. First,

$$\int_0^1 x(1-x) C_i^{(3/2)}(1-2x) C_j^{(3/2)}(1-2x) dx = -\frac{1}{32} \int_{-1}^1 (1-z^2)^2 C_i^{(3/2)}(z) C_j^{(3/2)}(z) dz \quad (9)$$

after making the substitution $z = 1 - 2x$. This puts the Gegenbauer polynomials on their natural domain, $[-1, 1]$. Now, multiplying through by one of the factors of $(1 - z^2)$,

$$\int_{-1}^1 (1-z^2)^2 C_i^{(3/2)}(z) C_j^{(3/2)}(z) dz = \int_{-1}^1 (1-z^2) C_i^{(3/2)}(z) C_j^{(3/2)}(z) dz - \int_{-1}^1 (1-z^2) z C_i^{(3/2)}(z) z C_j^{(3/2)}(z) dz.$$

The first term can be recognized as the orthogonality condition for the Gegenbauer polynomials and hence,

$$\int_{-1}^1 (1-z^2) C_i^{(3/2)}(z) C_j^{(3/2)}(z) dz = \delta_{i,j} \frac{2(i+1)(i+2)}{3+2i}.$$

To simplify the second term, use the recurrence relation for the Gegenbauer polynomials to find that,

$$z C_i^{(3/2)}(z) = \frac{1}{3+2i} \left((i+1) C_{i+1}^{(3/2)}(z) + (i+2) C_{i-1}^{(3/2)}(z) \right).$$

Substituting and multiplying through yields

$$\begin{aligned} & \int_{-1}^1 (1-z^2) z C_i^{(3/2)}(z) z C_j^{(3/2)}(z) dz \\ &= \frac{1}{(2i+3)(2j+3)} \\ & \times \left((i+1)(j+1) \int_{-1}^1 (1-z^2) C_{i+1}^{(3/2)}(z) C_{j+1}^{(3/2)}(z) dz \times (i+1)(j+2) \int_{-1}^1 (1-z^2) C_{i+1}^{(3/2)}(z) C_{j-1}^{(3/2)}(z) dz \times (i+2)(j+1) \right). \end{aligned}$$

Again, these integrals can be simplified using the orthogonality of the Gegenbauer polynomials to finally see that the integral in (9) equals

$$-\frac{1}{32} \left(\delta_{i,j} \frac{2(i+1)(i+2)}{3+2i} + \frac{1}{(2i+3)(2j+3)} \left(\delta_{i,j} (i+1)^2 \frac{2(i+2)(i+3)}{5+2i} + \delta_{i,j-2} (i+1)(i+4) \frac{2(i+2)(i+3)}{5+2i} + \delta_{i,j+2} (i+2)(i-1) \frac{i(i+1)}{1+2i} \right) \right).$$

An important consequence of this fact is that the many of the terms in the sum (8) are equal to zero.

The integrals over the intermediate times can also be evaluated exactly, although I have not been able to find a general formula. In this case, it is straight-forward to precompute the integral for all possible sets of equal indices and then substitute into the sum.

References

- Baibuz VF, Yu Zitserman V, Drozdov AN. Diffusion in a potential field: Path-integral approach. *Physica A: Statistical Mechanics and its Applications*. 1984; 127(1):173–193.
- Bogachev, Vladimir I. Measure theory, volume 1. Springer; 2007.
- Bollback, Jonathan P.; York, Thomas L.; Nielsen, Rasmus. Estimation of 2nes from temporal allele frequency data. *Genetics*. 2008; 179(1):497–502. [PubMed: 18493066]
- Boyko, Adam R.; Williamson, Scott H.; Indap, Amit R.; Degenhardt, Jeremiah D.; Hernandez, Ryan D.; Lohmueller, Kirk E.; Adams, Mark D.; Schmidt, Steffen; Sninsky, John J.; Sunyaev, Shamil R., et al. Assessing the evolutionary impact of amino acid mutations in the human genome. *PLoS genetics*. 2008; 4(5):e1000083. [PubMed: 18516229]
- Chorin, Alexandre Joel; Hald, Ole H. Stochastic tools in mathematics and science, volume 1. Springer; 2006.
- Dawson, Donald A.; Feng, Shui. Large deviations for the fleming-viot process with neutral mutation and selection, ii. *Stochastic processes and their applications*. 2001; 92(1):131–162.
- Dawson, Donald A.; Hochberg, Kenneth J. Wandering random measures in the flemingviot model. *The Annals of Probability*. 1982:554–580.
- Donnelly, Peter; Kurtz, Thomas G. A countable representation of the fleming-viot measure-valued diffusion. *The Annals of Probability*. 1996; 24(2):698–742.
- Donnelly, Peter; Kurtz, Thomas G. Genealogical processes for fleming-viot models with selection and recombination. *Annals of Applied Probability*. 1999:1091–1148.
- Dürr, Detlef; Bach, Alexander. The onsager-machlup function as lagrangian for the most probable path of a diffusion process. *Communications in Mathematical Physics*. 1978; 60(2):153–170.

- Ethier, SN.; Kurtz, TG. In Stochastic methods in biology. Springer; 1987. The infinitely-many-alleles model with selection as a measure-valued diffusion.; p. 72-86.
- Ethier, Stewart N.; Griffiths, RC. The transition function of a Fleming-viot process. *The Annals of Probability*. 1993;1571–1590.
- Ethier, Stewart N.; Kurtz, Thomas G. Fleming-viot processes in population genetics. *SIAM Journal on Control and Optimization*. 1993; 31(2):345–386.
- Ewens, Warren J. Theoretical introduction. Vol. 27. Springer; 2004. *Mathematical population genetics: I.*
- Feder, Alison; Kryazhimskiy, Sergey; Plotkin, Joshua B. Identifying signatures of selection in genetic time series. *arXiv preprint arXiv:1302.0452*. 2013
- Feynman, Richard P.; Hibbs, Albert R. Quantum mechanics and path integrals. Emended edition. DoverPublications. com; 2012.
- Feynman, Richard Phillips. Space-time approach to non-relativistic quantum mechanics. *Reviews of Modern Physics*. 1948; 20(2):367.
- Fleming, Wendell H.; Viot, Michel. Some measure-valued markov processes in population genetics theory. *Indiana Univ. Math. J.* 1979; 28(5):817–843.
- Graham, Robert. Path integral formulation of general diffusion processes. *Zeitschrift für Physik B Condensed Matter*. 1977; 26(3):281–290.
- Gutenkunst, Ryan N.; Hernandez, Ryan D.; Williamson, Scott H.; Bustamante, Carlos D. Inferring the joint demographic history of multiple populations from multidimensional snp frequency data. *PLoS genetics*. 2009; 5(10):e1000695. [PubMed: 19851460]
- Harlow, Daniel. A simple bound on the error of perturbation theory in quantum mechanics. *arXiv preprint arXiv:0905.2466*. 2009
- Kawecki, Tadeusz J.; Lenski, Richard E.; Ebert, Dieter; Hollis, Brian; Olivieri, Isabelle; Whitlock, Michael C. Experimental evolution. *Trends in ecology & evolution*. 2012
- Kimura, Motoo. Solution of a process of random genetic drift with a continuous model. *Proceedings of the National Academy of Sciences of the United States of America*. 1955a; 41(3):144. [PubMed: 16589632]
- Kimura, Motoo. In *Cold Spring Harbor Symposia on Quantitative Biology*. Vol. 20. Cold Spring Harbor Laboratory Press; 1955b. Stochastic processes and distribution of gene frequencies under natural selection.; p. 33-53.
- Kimura, Motoo. A stochastic model concerning the maintenance of genetic variability in quantitative characters. *Proceedings of the National Academy of Sciences of the United States of America*. 1965; 54(3):731. [PubMed: 5217452]
- Kingman, John FC. The coalescent. *Stochastic processes and their applications*. 1982; 13(3):235–248.
- Krone, Stephen M.; Neuhauser, Claudia. Ancestral processes with selection. *Theoretical population biology*. 1997; 51(3):210–237. [PubMed: 9245777]
- Malaspina, Anna-Sapfo; Malaspina, Orestis; Evans, Steven N.; Slatkin, Montgomery. Estimating allele age and selection coefficient from time-serial data. *Genetics*. 2012; 192(2):599–607. [PubMed: 22851647]
- Mathieson, Iain; McVean, Gil. Estimating selection coefficients in spatially structured populations from time series data of allele frequencies. *Genetics*. 2013; 193(3):973–984. [PubMed: 23307902]
- Mustonen, Ville; Lässig, Michael. Fitness flux and ubiquity of adaptive evolution. *Proceedings of the National Academy of Sciences*. 2010; 107(9):4248–4253.
- Nagylaki, Thomas. The moments of stochastic integrals and the distribution of sojourn times. *Proceedings of the National Academy of Sciences*. 1974; 71(3):746–749.
- Nagylaki, Thomas. Models and approximations for random genetic drift. *Theoretical Population Biology*. 1990; 37(1):192–212.
- Neher, Richard A.; Shraiman, Boris I. Fluctuations of fitness distributions and the rate of mullers ratchet. *Genetics*. 2012; 191(4):1283–1293. [PubMed: 22649084]
- Neuhauser, Claudia; Krone, Stephen M. The genealogy of samples in models with selection. *Genetics*. 1997; 145(2):519–534. [PubMed: 9071604]

- Rogers, L Chris G.; Williams, David. Diffusions, Markov processes and martingales: Volume 2, Itô calculus, volume 2. Cambridge university press; 2000.
- Rouhani S, Barton N. Speciation and the shifting balance in a continuous population. *Theoretical Population Biology*. 1987; 31(3):465–492.
- Sawyer, Stanley A.; Hartl, Daniel L. Population genetics of polymorphism and divergence. *Genetics*. 1992; 132(4):1161–1176. [PubMed: 1459433]
- Schraiber, Joshua G.; Griffiths, Robert C.; Evans, Steven N. Analysis and rejection sampling of wright-fisher diffusion bridges. *Theoretical Population Biology*. 2013; 89:64–74. [PubMed: 24001410]
- Song, Yun S.; Steinrücken, Matthias. A simple method for finding explicit analytic transition densities of diffusion processes with general diploid selection. *Genetics*. 2012; 190(3):1117–1129. [PubMed: 22209899]
- Torgerson, Dara G.; Boyko, Adam R.; Hernandez, Ryan D.; Indap, Amit; Hu, Xiaolan; White, Thomas J.; Sninsky, John J.; Cargill, Michele; Adams, Mark D.; Bustamante, Carlos D., et al. Evolutionary processes acting on candidate cis-regulatory regions in humans inferred from patterns of polymorphism and divergence. *PLoS genetics*. 2009; 5(8):e1000592. [PubMed: 19662163]
- Wall, Jeffrey D.; Slatkin, Montgomery. Paleopopulation genetics. *Annual Review of Genetics*. 2012; 46:635–649.
- Watterson GA. Estimating and testing selection: the two-alleles, genic selection diffusion model. *Advances in Applied Probability*. 1979:14–30.
- Wiener, Norbert. The average of an analytic functional and the brownian movement. *Proceedings of the National Academy of Sciences of the United States of America*. 1921; 7(10):294. [PubMed: 16576610]
- Wright, Sewall. Evolution in mendelian populations. *Genetics*. 1931; 16(2):97. [PubMed: 17246615]
- Zee, Anthony. Quantum field theory in a nutshell. Princeton University Press. 2010
- Zhao, Lei; Yue, Xingye; Waxman, David. Complete numerical solution of the diffusion equation of random genetic drift. *Genetics*. 2013

Figure 1.

Path integrals. In panel a, neutral Wright-Fisher paths starting at $x = 0.2$ and ending at $y = 0.7$ after 0.1 time units have passed are sampled using the rejection-sampling method of Schraiber et al. [2013]. Paths are colored according to $\mathcal{G}[z]$ with $\alpha = 5$, with darker paths corresponding to larger $\mathcal{G}[z]$ values. In panel b, the density of $\mathcal{G}[z]$ is plotted. The path integral estimate of the transition density for a Wright-Fisher process with $\alpha = 5$ to go from 0.2 to 0.7 in 0.1 time units is the mean value of the density in panel b, indicated by the vertical line.

Figure 2.

Feynman diagrams. Feynman diagrams are used to evaluate the integrals that show up in the perturbation expansion. The allele starts at time 0 and frequency x , evolving neutrally until time s_1 , when it has frequency z_1 and is perturbed by natural selection. It then evolves to time s_2 and allele frequency z_2 , at which point it is again perturbed by natural selection. This continues until the final perturbation at time s_k and frequency z_k , after which it evolves neutrally to time t and frequency y .

Figure 3. Accuracy of the perturbation expansion. The perturbation expansion is compared to simulations of the Wright-Fisher diffusion. An allele with starting frequency $x = .2$ evolves under genetic drift and natural selection for $t = .1$ with a variety of selection coefficients. The probability that the allele was not absorbed is then computed. Dots show the values from simulations while lines indicate successively higher orders of perturbation expansion.