Analyzing Diffusion Approximations of the Wright-Fisher Model Using Linear Programming[1]

Kurt Helmes[2] and Richard H. Stockbridge[3]

Abstract. A computational method for the analysis of the diffusion approximation to the Wright-Fisher model in population genetics is presented. This paper evaluates the stationary distribution of the diffusion when such a distribution exists and examines the extinction probabilities and mean extinction time when extinction of at least one allele occurs. The computational method uses a characterization of the diffusion through an adjoint relation between the diffusion operator and its stationary distribution or between the diffusion operator and a pair of measures giving the expected occupation of the process and the state when extinction occurs. Application of the adjoint relations to a set of functions in the domain of the generator leads to a set of constraints for linear programs which are solved to obtain bounds on numerical quantities of interest. The accuracy of the method is illustrated on examples for which analytical results are known. The method is also used in cases for which exact solutions are unknown.

Key Words. Wright-Fisher model, stationary distribution, exit time, extinction, linear programming, moments, computational probability.

Subject Classification. *Primary* 60H30. *Secondary* 60G10,90C90,92D10.

1 Introduction and model formulation

The purpose of this paper is to provide a numerical method based on linear programming for computing quantities of interest, such as the mean exit time and the moments of the stationary distribution, for a particular class of diffusion models on the unit simplex. We concentrate on the diffusion approximation to the Wright-Fisher model in population genetics since, for specific choices of the parameters, theoretical results exist for the mean exit time and stationary distribution and thus we can exhibit the accuracy of the linear programming method, but for other parameter selections no analytical results are known. We illustrate the numerical procedure on one-dimensional, two-dimensional and four-dimensional diffusion models.

We take as our starting point the diffusion approximation to the Wright-Fisher genetic model with r alleles and refer the reader to [5, Chapter 10] for an excellent explanation of the original model and the diffusion approximation. Let $K = \{x = (x_1, \ldots, x_{r-1}) \in [0,1]^{r-1} : \sum_{i=1}^{r-1} x_i \leq 1\}$ denote the $(r-1)$-dimensional simplex which identifies the proportions of each allele in the population (the rth proportion being determined by the others). The diffusion

[1] This research is partially supported by NSF under grant DMS 9803490.

[2] Institute for Operations Research, Humboldt University of Berlin, Berlin, Germany, helmes@wiwi.hu-berlin.de.

[3] Department of Mathematical Sciences, University of Wisconsin Milwaukee, Milwaukee, Wisconsin and Department of Statistics, University of Kentucky, Lexington, Kentucky, stockbri@uwm.edu.

operator for the evolution of these proportions is

$$A = \frac{1}{2} \sum_{i,j=1}^{r-1} a_{ij}(x) \frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^{r-1} b_i(x) \frac{\partial}{\partial x_i}, \tag{1}$$

where $a_{ij}(x) = x_i(\delta_{ij} - x_j)$, δ_{ij} being the Kronecker δ, and

$$b_i(x) = -\sum_{j=1}^{r} \mu_{ij} x_i + \sum_{j=1}^{r} \mu_{ji} x_j + x_i \left(\sum_{j=1}^{r} \sigma_{ij} x_j - \sum_{k,l=1}^{r} \sigma_{kl} x_k x_l \right) \tag{2}$$

and we set the domain of the operator to be $\mathcal{D}(A) = C^2(K)$. The coefficients μ_{ij} are related to the mutation rates between alleles ($\mu_{ii} = 0, \ \forall i$) and the coefficients σ_{ij} correspond to selection. As in [5], letting $X(t) = (X_1(t), \dots, X_{r-1}(t))$ denote the proportions of the alleles in the population at time t, the process

$$f(X(t)) - \int_0^t Af(X(s))\, ds \tag{3}$$

is a martingale for each $f \in \mathcal{D}(A)$.

Our linear programming approach for the analysis of this diffusion involves utilizing an adjoint equation involving the generator A and either the stationary distribution of the process, or the occupation measure of the process in the simplex up to the time of extinction and the measure giving the state of the process at extinction. The adjoint equation takes the form (4) when the model has a stationary distribution to be determined and is (22) when the model is such that one or more of the alleles will become extinct.

Consider first the case in which $\mu_{ij} > 0$ for $i \neq j$ so the process X has a stationary distribution on K; denote this distribution by ν. When ν is also the initial distribution of X, $X(t)$ is distributed according to ν for each $t \geq 0$. Since (3) is a martingale for each $f \in \mathcal{D}(A)$,

$$E[f(X(0))] = E\left[f(X(t)) - \int_0^t Af(X(s))\, ds \right] = E[f(X(t))] - \int_0^t E[Af(X(s))]\, ds$$

and by the stationarity of the process, it follows that

$$\int_K Af(x_1, \dots, x_{r-1})\, \nu(dx_1 \times \cdots \times dx_{r-1}) = 0. \tag{4}$$

The identity (4), in fact, characterizes the stationary distribution ν (see [5, Theorem 4.9.17]).

The basic idea of our approach is to judiciously select a finite collection of $f \in \mathcal{D}(A)$ and set the conditions (4) for these f as the constraints of a linear program. A careful selection of objective function then allows for the determination of the quantities of interest. This numerical approach has an advantage over other approaches in that it naturally provides bounds on the quantities of interest and excellent software is readily available.

The authors have applied this numerical method [8, 10] to the analysis of the distribution of the exit time for a variety of processes from a bounded interval in one-dimension and bounded rectangle in two-dimensions. The paper [8] computed the moments of the exit time distribution directly, whereas in [10] the Laplace transform of the moments were determined. A similar approach has been applied on stochastic control problems ([9]) and optimal stopping problems by Röhl [16] and Schwerer [17] uses a linear programming approach involving the moments of a reflected Brownian motion process over an unbounded domain.

Formulating the evolution of the Markov process through occupation measures developed from work by Young [20] for problems in the calculus of variations. The present approach is a special case of the more general situation of formulating stochastic control problems as linear programs over a space of stationary distributions studied in very general settings in [1, 11, 18, 19]. The variable in the infinite-dimensional linear program is a measure on the product of the state and control spaces, and in the case of exit problems, this variable is augmented by a second measure on the exterior of the state space (see [11]).

The linear programming approach to stochastic control at this level of generality had its foundations in the existence of stationary processes corresponding to the measure which satisfies an adjoint equation. This was established in [18] and later extended to the more general adjoint equation for exit problems in [11]. Kurtz also applied these techniques in uncontrolled settings for patchwork and constrained martingale problems ([12, 13]) and, in joint work with Dai, used this in the analysis of Brownian networks [3]. The adjoint equation corresponding to the heavy traffic limit of queueing processes had previously been introduced by Harrison and Reiman [7]. Other papers using linear programming for stochastic control include [14, 15].

The Wright-Fisher process of this paper is *not* controlled. However, the results cited above are applicable to it by taking the control space to consist of a single element. It then follows that the variables are measures solely on the states of the process.

This paper is organized as follows. Section 2 formulates the linear program associated with the stationary distribution of the Wright-Fisher diffusion approximation and computes the moments of the stationary distribution for several choices of parameters for which the process has a stationary distribution. The discussion includes a careful examination of additional linear constraints for the two-dimensional simplex due to Dale [4], and its extension to higher dimensions, which are necessary to ensure that the computed values are consistent with being moments of a distribution. Section 3 then presents the LP formulation used to compute quantities related to extinction for the cases when at least one allele will become extinct in the population. In this paper, we focus on the mean extinction time and the probability of each allele becoming extinct. The section includes numerical illustrations for cases in which there are analytical results and for some choice of parameters for which no analytical results are known. We make brief concluding remarks in section 4 and provide the code for one of the examples in an appendix.

2 Evaluation of the stationary distribution

The goal of this section is the computation of the moments of the stationary distribution of the Wright-Fisher diffusion on the unit simplex so consider the case in which $\mu_{ij} > 0$ for $i \neq j$. Since the adjoint equation (4) characterizes the stationary distribution ν and the distribution ν on the (bounded) simplex is specified by its joint moments, it is sufficient to restrict attention to the countable collection

$$\mathcal{D}_\infty = \{f(x_1, \ldots, x_{r-1}) = x_1^{k_1} \cdots x_{r-1}^{k_{r-1}} : k_1, \ldots, k_{r-1} \in \mathbb{Z}^+\}. \tag{5}$$

In this way, (4) can be expressed in terms of the moments of the distribution. Specifically, fix $k_1, \ldots, k_{r-1}$ and consider $f(x_1, \ldots, x_{r-1}) = x_1^{k_1} \cdots x_{r-1}^{k_{r-1}}$. Define the joint $(k_1, \ldots, k_{r-1})$-moment of ν by

$$m(k_1, \ldots, k_{r-1}) = \int_K x_1^{k_1} \cdots x_{r-1}^{k_{r-1}}\, \nu(dx_1 \times \cdots \times dx_{r-1}). \tag{6}$$

Then applying the generator A of (1) to f, (4) becomes

$$\begin{aligned} 0 \;=\; & \sum_{i=1}^{r-1} \frac{k_i(k_i-1)}{2} \left(m(k_1, \ldots, k_i - 1, \ldots, k_{r-1}) - m(k_1, \ldots, k_{r-1})\right) \\ & - \sum_{i,j=1; i\neq j}^{r-1} (k_i k_j) m(k_1, \ldots, k_{r-1}) - \sum_{i=1}^{r-1} k_i \left(\sum_{i=1}^{r} \mu_{ij}\right) m(k_1, \ldots, k_{r-1}) \\ & + \sum_{i=1}^{r-1} \sum_{j=1}^{r} k_i \mu_{ji} m(k_1, \ldots, k_i - 1, \ldots, k_j + 1, \ldots, k_{r-1}) \\ & + \sum_{i=1}^{r-1} \sum_{j=1}^{r} k_i \sigma_{ij} m(k_1, \ldots, k_j + 1, \ldots, k_{r-1}) \\ & - \sum_{i=1}^{r-1} \sum_{j,l=1}^{r} k_i \sigma_{jl} m(k_1, \ldots, k_j + 1, \ldots, k_l + 1, \ldots k_{r-1}), \end{aligned} \tag{7}$$

where the notation $m(k_1, \ldots, k_i - 1, \ldots, k_j + 1, \ldots, k_{r-1})$ denotes the joint moment having $(k_i - 1)$-moment in the ith variable, $(k_j + 1)$-moment in the jth variable and k_l-moment for the lth variables for all other variables. When $i = j$, $m(k_1, \ldots, k_i - 1, \ldots, k_j + 1, \ldots, k_{r-1}) = m(k_1, \ldots, k_i, \ldots, k_{r-1})$ and $m(k_1, \ldots, k_i + 1, \ldots, k_j + 1, \ldots k_{r-1}) = m(k_1, \ldots, k_i + 2, \ldots, k_{r-1})$.

2.1 Motivating example

To illustrate the approach, consider the simplest example for which analytic results are known. Let $\mu_{ij} = \mu_j$ for $i \neq j, j = 1, \ldots, r$ and set $\sigma_{ij} = 0$ for all i, j. In this case, the stationary distribution ν of the allele proportions $(X_1, \ldots, X_{r-1})$ is a Dirichlet distribution with parameters $(2\mu_1, \ldots, 2\mu_r)$; that is, $(X_1, \ldots, X_{r-1})$ has density on the $(r-1)$-dimensional simplex K given by

$$p(x_1, \ldots, x_{r-1}) = C \left(\prod_{i=1}^{r-1} x_i^{2\mu_i - 1}\right) \left(1 - \sum_{i=1}^{r-1} x_i\right)^{2\mu_r - 1}, \tag{8}$$

where $C = \frac{\Gamma(2\mu_1 + \cdots + 2\mu_r)}{\Gamma(2\mu_1) \cdots \Gamma(2\mu_r)}$ is the normalizing constant. A simple computation shows that the joint $(k_1, \ldots, k_{r-1})$-moment of ν is

$$m(k_1, \ldots, k_{r-1}) = \frac{\Gamma\left(2 \sum_{i=1}^{r} \mu_i\right) \prod_{i=1}^{r-1} \Gamma(2\mu_i + k_i)}{\Gamma\left(2 \sum_{i=1}^{r} \mu_i + \sum_{i=1}^{r-1} k_i\right) \prod_{i=1}^{r-1} \Gamma(2\mu_i)}.$$

Now let $r = 5$. Taking each $\mu_j = 0.5$, the Dirichlet distribution actually gives the uniform distribution on the four-dimensional simplex. In this example, the conditions (7) can be solved recursively for the joint moments.

Considering functions f with $k_1, k_2, k_3, k_4 \leq 20$ in (7) and solving a linear program having these conditions as the constraints, the LP determines the joint moments of ν. Note the fact that the moments can be determined recursively is displayed when the software package CPLEX determines these values exactly using a presolver to reduce the number of undetermined variables in the program. As a result the choice of objective function for the LP has no effect on the optimal solution since there is a unique feasible point.

We illustrate the numerical results with the joint moments $m(k_1, k_2, 0, 0)$ for $k_1 \leq 10$ and $k_2 \leq 5$ in Table 2.1.

Table 2.1: Values of $m(k_1, k_2, 0, 0)$ from LP

		k_2					
		0	1	2	3	4	5
	0	1	0.2	0.0666667	0.0285714	0.0142857	0.00793651
	1	0.2	0.0333333	0.00952381	0.00357143	0.0015873	0.000793651
k_1	2	0.0666667	0.00952381	0.00238095	0.000793651	0.00031746	0.0001443
	3	0.0285714	0.00357143	0.000793651	0.000238095	8.65801e-05	3.6075e-05
	4	0.0142857	0.0015873	0.00031746	8.65801e-05	2.886e-05	1.11e-05
	5	0.00793651	0.000793651	0.0001443	3.6075e-05	1.11e-05	3.96429e-06
	6	0.0047619	0.0004329	7.21501e-05	1.665e-05	4.75715e-06	1.58572e-06
	7	0.0030303	0.000252525	3.885e-05	8.32501e-06	2.22e-06	6.93751e-07
	8	0.0020202	0.0001554	2.22e-05	4.44e-06	1.11e-06	3.26471e-07
	9	0.0013986	9.99001e-05	1.332e-05	2.4975e-06	5.87648e-07	1.63235e-07
	10	0.000999001	6.66001e-05	8.32501e-06	1.46912e-06	3.26471e-07	8.59134e-08

These values agree with the analytic values $m(k_1, k_2, 0, 0) = \frac{4!k_1!k_2!}{(k_1+k_2+4)!}$.

2.2 The Dale conditions

Consider now the general model in which $\mu_{ij} > 0$ for $i \neq j$ (and $\mu_{ii} = 0$) but in which there are no other assumptions on μ_{ij}, and $\sigma_{ij} \neq 0$ for some i, j. It is no longer the case that the adjoint relation (7) can be solved recursively for the joint moments. (Even the case with $\mu_{ij} = \mu_j$ for $i \neq j$, $j = 1, \ldots, r$ cannot be solved recursively when some of the σ_{ij} values are nonzero.)

At this point, we have reformulated the adjoint relation (4) using the joint moments of the distribution ν as the linear conditions in (7) involving the variables $\{m(k_1, \ldots, k_{r-1}) : k_1, \ldots, k_{r-1} \in \mathbb{Z}^+\}$. These conditions on their own, however, do not imply that the collection corresponds to the collection of moments of a measure. It is necessary to impose additional conditions on the collection to ensure that it is the collection of moments of some distribution.

When $r = 2$ so the simplex is actually the unit interval, necessary and sufficient conditions (see [6]) to ensure that the sequence $\{m(0), m(1), m(2), \ldots\}$ are the moments of some distribution ν on $[0, 1]$ are the Hausdorff moment conditions

$$\sum_{j=0}^{n} (-1)^j \binom{n}{j} m(j + k) \geq 0; \tag{9}$$

the necessity following from the observation that $\int x^k (1 - x)^n \, \nu(dx) \geq 0$ for each $k, n \in \mathbb{Z}^+$. The one-dimesional (and two-dimensional extension to the unit square) Hausdorff moment conditions were used by the authors in [8, 10] in the analysis of the exit time distribution for some Markov processes.

When $r = 3$ so the simplex is the triangle in two dimensions having vertices $(0, 0)$, $(1, 0)$ and $(0, 1)$, necessary and sufficient conditions for the collection $\{m(k_1, k_2) : k_1, k_2 \in \mathbb{Z}^+\}$ to be the moments of a distribution ν on K were derived by Dale [4]. These conditions are most easily expressed in terms of iterated differences of the moments, which we now describe. For $k_1, k_2 \in \mathbb{Z}^+$, define

$$w(k_1, k_2, 0) = m(k_1, k_2)$$

and for $k_1, k_2, k_3 \in \mathbb{Z}^+$, with $k_3 \geq 1$, define

$$w(k_1, k_2, k_3) = w(k_1, k_2, k_3 - 1) - w(k_1 + 1, k_2, k_3 - 1) - w(k_1, k_2 + 1, k_3 - 1).$$

The Dale conditions require, for $k_1, k_2, k_3 \in \mathbb{Z}^+$

$$w(0, 0, 0) = 1 \qquad \text{and} \qquad w(k_1, k_2, k_3) \geq 0; \tag{10}$$

the necessity follows from the observations that

$$\int_K x_1^{k_1} x_2^{k_2} (1 - x_1 - x_2)^{k_3} \nu(dx_1 \times dx_2) \geq 0 \tag{11}$$

for each $k_1, k_2, k_3 \geq 0$, and additionally when $k_3 \geq 1$,

$$\begin{aligned} \int_K x_1^{k_1} x_2^{k_2} (1 - x_1 - x_2)^{k_3} \nu(dx_1 \times dx_2) \;&=\; \int_K x_1^{k_1} x_2^{k_2} (1 - x_1 - x_2)^{k_3 - 1} \nu(dx_1 \times dx_2) \\ &\quad - \int_K x_1^{k_1+1} x_2^{k_2} (1 - x_1 - x_2)^{k_3 - 1} \nu(dx_1 \times dx_2) \\ &\quad - \int_K x_1^{k_1} x_2^{k_2+1} (1 - x_1 - x_2)^{k_3 - 1} \nu(dx_1 \times dx_2). \end{aligned} \tag{12}$$

Dale's conditions can be extended to the $(r - 1)$-dimensional simplex.

Theorem 2.1. A collection $\{m(k_1, \ldots, k_{r-1}) : k_1, \ldots, k_{r-1} \in \mathbb{Z}^+\}$ are the joint moments of some distribution ν on the $(r - 1)$-dimensional simplex K (see (6)) if and only if the conditions

$$w(0, \ldots, 0) = 1 \qquad \textit{and} \qquad w(k_1, \ldots, k_{r-1}, k_r) \geq 0 \tag{13}$$

are satisfied, where

$$w(k_1, \ldots, k_{r-1}, 0) = m(k_1, \ldots, k_{r-1}) \tag{14}$$

and for $k_r \in \mathbb{Z}^+$ with $k_r \geq 1$,

$$\begin{aligned} w(k_1, \ldots, k_{r-1}, k_r) \;&=\; w(k_1, \ldots, k_{r-1}, k_r - 1) \\ &\quad - \sum_{i=1}^{r-1} w(k_1, \ldots, k_i + 1, \ldots, k_{r-1}, k_r - 1). \end{aligned} \tag{15}$$

Proof. The necesssity for the higher dimension follows analagously from observations as in (11) and (12).

Now assume that the collection $\mathbb{M} = \{m(k_1, \ldots, k_{r-1}) : k_1, \ldots, k_{r-1} \in \mathbb{Z}^+\}$ satisfies (13), (14) and (15). For each $n \geq 0$, define a discretization K_n of the simplex K by

$$K_n = K \bigcap \left\{ \left(\frac{i_1}{n}, \ldots, \frac{i_{r-1}}{n} \right) : i_1, \ldots, i_{r-1} \in \mathbb{Z}^+ \right\}.$$

Note, in particular, that $0 \leq i_1, \ldots, i_{r-1} \leq n$ and $i_1 + \cdots + i_{r-1} \leq n$; denote this collection of $(r-1)$-tuples by $\mathcal{I}_n$. Again for each n, for $(\frac{i_1}{n}, \ldots, \frac{i_{r-1}}{n}) \in K_n$ define

$$p_n \left(\frac{i_1}{n}, \ldots, \frac{i_{r-1}}{n} \right) = \left(\begin{array}{c} n \\ i_1, \ldots, i_{r-1} \end{array} \right) w \left(i_1, \ldots, i_{r-1}, n - \sum_{l=1}^{r-1} i_l \right) \tag{16}$$

where

$$\left(\begin{array}{c} n \\ i_1, \ldots, i_{r-1} \end{array} \right) = \frac{n!}{(\prod_{l=1}^{r-1} i_l!)(n - \sum_{l=1}^{r-1} i_l)!}.$$

Observe that by their definition in conjunction with (13), $p_0(0, \ldots, 0) = 1$ and $p_n(i_1, \ldots, i_{r-1}) \geq 0$ for every $0 \leq i_1, \ldots, i_{r-1} \leq n$ with $i_1 + \cdots + i_{r-1} \leq n$. To establish an important identity ((18) below) it is helpful to express $w(i_1, \ldots, i_r)$ in terms of the elements of the set $\mathbb{M}$:

$$\begin{aligned} & w(i_1, \ldots, i_{r-1}, i_r) \\ = & \sum_{\substack{(j_1, \ldots, j_{r-1}) : \\ j_1 + \cdots + j_{r-1} \leq i_r}} (-1)^{\sum_{l=1}^{r-1} j_l} \left(\begin{array}{c} i_r \\ j_1 \; \cdots \; j_{r-1} \end{array} \right) m(i_1 + j_1, \ldots, i_{r-1} + j_{r-1}) \end{aligned} \tag{17}$$

A tedious but straightforward calculation shows that

$$\begin{aligned} & \sum_{(u_1, \ldots, u_{r-1}) \in \mathcal{I}_n} \left(\prod_{l=1}^{r-1} \binom{u_l}{i_l} \right) p_n \left(\frac{u_1}{n}, \ldots, \frac{u_{r-1}}{n} \right) \\ = & \sum_{(u_1, \ldots, u_{r-1}) \in \mathcal{I}_n} \sum_{\substack{(j_1, \ldots, j_{r-1}) : \\ j_1 + \cdots + j_{r-1} \leq n - \sum_{l=1}^{r-1} u_l}} (-1)^{\sum_{l=1}^{r-1} j_l} \left(\prod_{l=1}^{r-1} \binom{u_l}{i_l} \right) \\ & \cdot \left(\begin{array}{c} n \\ i_1, \ldots, i_{r-1} \end{array} \right) \left(\begin{array}{c} n - \sum_{l=1}^{r-1} u_l \\ j_1, \; \cdots, \; j_{r-1} \end{array} \right) m(u_1 + j_1, \ldots, u_{r-1} + j_{r-1}) \\ = & \sum_{(u_1, \ldots, u_{r-1}) \in \mathcal{I}_n} \sum_{\substack{(j_1, \ldots, j_{r-1}) : \\ j_1 + \cdots + j_{r-1} \leq n - \sum_{l=1}^{r-1} u_l}} (-1)^{\sum_{l=1}^{r-1} j_l} \\ & \cdot \left(\begin{array}{c} n \\ i_1, \cdots, i_{r-1}, (u_1 - i_1), \cdots, (u_{r-1} - i_{r-1}), j_1, \cdots, j_{r-1} \end{array} \right) m(u_1 + j_1, \ldots, u_{r-1} + j_{r-1}) \\ = & \left(\begin{array}{c} n \\ i_1 \; \cdots \; i_{r-1} \end{array} \right) m(i_1, \ldots, i_{r-1}), \end{aligned} \tag{18}$$

where the summations over $(u_1, \ldots, u_{r-1}) \in \mathcal{I}_n$ are also subject to the restriction that each $u_l \geq i_l$ and

$$\begin{pmatrix} n \\ i_1, \cdots, i_{r-1}, (u_1 - i_1), \cdots, (u_{r-1} - i_{r-1}), j_1, \cdots, j_{r-1} \end{pmatrix}$$
$$= \frac{n!}{\prod_{l=1}^{r-1} (i_l!(u_l - i_l)! j_l!) \cdot (n - \sum_{l=1}^{r-1} (u_l + j_l))!} \, .$$

Of particular interest is the case in which $i_1 = \cdots = i_{r-1} = 0$, in which case (17) implies

$$\sum_{(u_1/n, \ldots, u_{r-1}/n) \in K_n} p_n \left(\frac{u_1}{n}, \ldots, \frac{u_{r-1}}{n} \right) = 1$$

and thus

$$\mathbb{P}_n = \left\{ p_n \left(\frac{u_1}{n}, \ldots, \frac{u_{r-1}}{n} \right) : \left(\frac{u_1}{n}, \ldots, \frac{u_{r-1}}{n} \right) \in K_n \right\}$$

is a probability measure on $K_n \subset K$. $\mathbb{P}_n$ is also a probability measure on the simplex K.

Since K is compact, the collection $\{\mathbb{P}_n : n \in \mathbb{Z}^+\}$ is tight and there exists at least one probability measure $\mathbb{P}$ that is a weak limit of a subsequence of $\{\mathbb{P}_n\}$. Without loss of generality, we assume that the entire sequence converges: $\mathbb{P}_n \Rightarrow \mathbb{P}$.

Now let $X^{(n)} = (X_1^{(n)}, \ldots, X_{r-1}^{(n)})$ be a random vector having distribution $\mathbb{P}_n$ and let X be a random vector having distribution $\mathbb{P}$. Then (18) and the weak convergence of $\mathbb{P}_n$ to $\mathbb{P}$ implies

$$\begin{aligned} m(i_1, \ldots, i_{r-1}) &= \begin{pmatrix} n \\ i_1 \ \cdots \ i_{r-1} \end{pmatrix}^{-1} E\left[\prod_{l=1}^{r-1} \begin{pmatrix} nX_l^{(n)} \\ i_l \end{pmatrix} \right] \\ &= \frac{n^{\sum_{l=1}^{r-1} i_l} (n - \sum_{l=1}^{r-1} i_l)!}{n!} \, E\left[\prod_{l=1}^{r-1} \prod_{k_l=0}^{i_l} \left(X_l^{(n)} - \frac{k_l}{n} \right) \right] \\ &\to E\left[\prod_{l=1}^{r-1} X_l^{i_l} \right] \end{aligned}$$

and hence $\mathbb{M}$ is the set of joint moments of the distribution $\mathbb{P}$ on K. Note that since the collection $\mathbb{M}$ characterizes the distribution, $\mathbb{P}$ is, in fact, unique and moreover, the entire sequence $\mathbb{P}_n$ does converge weakly to $\mathbb{P}$.

2.3 Bounds on the joint moments

The inclusion of the Dale conditions implies that the variables $\{m(k_1, \ldots, k_{r-1})\}$ are the joint moments of some distribution. In order to numerically solve a linear program, however, it is necessary to limit the analysis to a finite subset and require the adjoint relation (7) to be satisfied only for this finite collection of variables. The result of doing this is that the feasible points $\{m(k_1, \ldots, k_{r-1})\}$ no longer need to be the moments of a distribution.

The *key observation,* however, is that the set of feasible points contains the (finite subset of) the moments of the stationary distribution. This containment enables both upper and lower bounds to be determined on the values of the moments. By selecting a particular moment as the objective function of a linear program with the adjoint conditions (7) and the Dale conditions (13) as constraints, running a minimization procedure will provide a lower bound and a maximization procedure will give an upper bound.

2.4 Numerical examples

Example 2.4.1. Consider a modification of the model in Section (2.1) in which $r = 5$ and $\mu_{ij} = \mu_j$ for $i \neq j$, $j = 1, \ldots, 5$, only this time we require $\sigma_{ij} \neq 0$ for some i, j (along with the symmetry conditions $\sigma_{ij} = \sigma_{ji}$ for all i, j). Again under these conditions the stationary distribution can be analytically determined. The distribution is absolutely continuous with respect to Lebesgue measure on the simplex and has density

$$p(x_1, x_2, x_3, x_4) = C x_1^{2\mu_1 - 1} x_2^{2\mu_2 - 1} x_3^{2\mu_3 - 1} x_4^{2\mu_4 - 1} (1 - x_1 - x_2 - x_3 - x_4)^{2\mu_5 - 1} e^{s(x_1, x_2, x_3, x_4)} \quad (19)$$

where $s(x_1, x_2, x_3, x_4) = \tilde{x}^T \Sigma \tilde{x}$ in which

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} & \sigma_{15} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} & \sigma_{24} & \sigma_{25} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} & \sigma_{34} & \sigma_{35} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_{44} & \sigma_{45} \\ \sigma_{51} & \sigma_{52} & \sigma_{53} & \sigma_{54} & \sigma_{55} \end{pmatrix}, \quad \text{and} \quad \tilde{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ 1 - x_1 - x_2 - x_3 - x_4 \end{pmatrix},$$

and C is a normalizing constant.

Again, let $\mu_j = 0.5$ for $j = 1, \ldots, 5$ and let

$$\Sigma = \begin{pmatrix} 1 & 0.5 & 0.5 & 0.5 & 0.5 \\ 0.5 & 0 & 0 & 0 & 0 \\ 0.5 & 0 & 0 & 0 & 0 \\ 0.5 & 0 & 0 & 0 & 0 \\ 0.5 & 0 & 0 & 0 & 0 \end{pmatrix}. \quad (20)$$

It then follows that each $2\mu_i - 1 = 0$ and $s(x_1, x_2, x_3, x_4) = x_1$ in (19) so the density is $p(x_1, x_2, x_3, x_4) = e^{x_1}$ and the moments can be exactly determined.

Table 2.2 displays the exact values of the joint (x_1, x_2)-moments up to order 5 as well as the values obtained by maximizing and minimizing linear programs having constraints given by (7) and (13) with objective functions consisting of each joint moment. The linear programs were run using up to the sixth moment in each variable.

Example 2.4.2. We now consider a modification of the model in Example 2.4.2 by removing the condition that $\mu_{ij} = \mu_j$ for $i \neq j$ and all j. Specifically select

$$\begin{pmatrix} \mu_{11} & \mu_{12} & \mu_{13} & \mu_{14} & \mu_{15} \\ \mu_{21} & \mu_{22} & \mu_{23} & \mu_{24} & \mu_{25} \\ \mu_{31} & \mu_{32} & \mu_{33} & \mu_{34} & \mu_{35} \\ \mu_{41} & \mu_{42} & \mu_{43} & \mu_{44} & \mu_{45} \\ \mu_{51} & \mu_{52} & \mu_{53} & \mu_{54} & \mu_{55} \end{pmatrix} = \begin{pmatrix} 0 & 2 & 0.5 & 0.5 & 0.5 \\ 3 & 0 & 0.5 & 0.5 & 0.5 \\ 0.5 & 0.5 & 0 & 1 & 0.5 \\ 0.5 & 0.5 & 1 & 0 & 0.5 \\ 0.5 & 0.5 & 0.5 & 0.5 & 0 \end{pmatrix}$$

and let Σ be as in (20). *The stationary distribution for this choice of parameters is not known.* Nevertheless, the linear programming approach gives bounds on the moments of the stationary distribution. Table 2.3 displays the bounds for the joint (x_1, x_2)-moments up to order 5 obtained by solving the LPs having constraints (7) and (13) with objective functions consisting of each joint moment individually. The LPs were run using moments up to order six in each variable.

Table 2.2. Values of $m(k_1, k_2, 0, 0)$ from LP

k_1		k_2 = 0	1	2	3	4	5
0	max	1	0.1927437	0.0625848	0.0263031	0.0129544	0.0076051
	exact	1	0.1927437	0.0625848	0.0263031	0.0129544	0.0071103
	min	1	0.1927436	0.0625848	0.0263031	0.0129544	0.0065777
1	max	0.2290259	0.0362818	0.0099787	0.0036328	0.0015769	0.0008325
	exact	0.2290254	0.0362818	0.0099787	0.0036328	0.0015769	0.0007734
	min	0.2290249	0.0362816	0.0099786	0.0036328	0.0015769	0.0006959
2	max	0.0839003	0.0113354	0.0027131	0.0008733	0.0003395	0.0001667
	exact	0.0838984	0.0113350	0.0027130	0.0008733	0.0003395	0.0001507
	min	0.0838964	0.0113348	0.0027129	0.0008733	0.0003395	0.0001342
3	max	0.0385699	0.0045542	0.0009667	0.0002794	9.856e-05	4.939e-05
	exact	0.0385584	0.0045529	0.0009664	0.0002792	9.839e-05	3.992e-05
	min	0.0385481	0.0045512	0.0009660	0.0002790	9.822e-05	2.750e-05
4	max	0.0204125	0.0021457	0.0004101	0.0001083	3.592e-05	3.055e-05
	exact	0.0203429	0.0021312	0.0004080	0.0001070	3.452e-05	1.291e-05
	min	0.0202700	0.0021280	0.0004058	0.0001059	3.311e-05	0
5	max	0.0123896	0.0011815	0.0002101	5.602e-05	2.838e-05	1.274e-05
	exact	0.0118376	0.0011624	0.0001939	4.662e-05	1.387e-05	4.810e-06
	min	0.0113047	0.0010463	0.0001774	3.488e-05	0	0

Table 2.3. Bounds on the Values of $m(k_1, k_2, 0, 0)$ from LP

k_1		k_2 = 0	1	2	3	4	5
0	max	1	0.1712548	0.0425506	0.0131836	0.0047600	0.0021360
	min	1	0.1712546	0.0425504	0.0131834	0.0047568	0.0016055
1	max	0.2527244	0.0456259	0.0113314	0.0034436	0.0012238	0.0005410
	min	0.2527241	0.0456255	0.0113309	0.0034426	0.0011850	0.0003830
2	max	0.0872044	0.0157279	0.0038019	0.0011166	0.0003851	0.0001700
	min	0.0872027	0.0157259	0.0038005	0.0011107	0.0003636	0.0001100
3	max	0.0359544	0.0063569	0.0014827	0.0004192	0.0001400	6.310e-05
	min	0.0359430	0.0063478	0.0014787	0.0004133	0.0001287	2.690e-05
4	max	0.0167553	0.0028873	0.0006506	0.0001770	5.883e-05	5.029e-05
	min	0.0166578	0.0028493	0.0006338	0.0001682	4.875e-05	0
5	max	0.0089794	0.0015198	0.0003363	9.641e-05	4.818e-05	2.160e-05
	min	0.0080040	0.0012834	0.0002666	5.468e-05	0	0

3 Mean extinction time and extinction probabilities

We now consider the model in which $\mu_{ij} = 0$ for all $i \neq j$ for at least one j and thus one or more of the alleles will become extinct in the population. Let τ denote the random time at which extinction occurs for the first such allele. Note that this corresponds to the first time that the diffusion process hits the boundary of the simplex. In this section, we evaluate the mean of τ and the probabilities of the alleles becoming extinct. We assume $E[\tau] < \infty$.

Since (3) is a martingale for each $f \in \mathcal{D}(A)$, an application of the optional sampling theorem implies

$$E\left[f(X(\tau)) - \int_0^\tau Af(X(s))\,ds\right] = E[f(X(0))]. \tag{21}$$

Let ν_τ denote the distribution of $X(\tau)$ and ν_{occ} be the expected occupation measure of X up to the time τ defined by

$$\nu_{occ}(G) = E\left[\int_0^\tau I_G(X(s))\,ds\right], \qquad \forall \text{ Borel sets } G \subset K$$

and notice that by taking $G = K$, $E[\tau] = \nu_{occ}(K)$; that is the mean exit time is given by the total mass of the occupation measure. The identity (21) can be written in terms of ν_τ and ν_{occ} as

$$\int_{\partial K} f\,d\nu_\tau - \int_K Af\,d\nu_{occ} = \int f d\nu_0, \qquad \forall f \in \mathcal{D}(A), \tag{22}$$

where ν_0 denotes the initial distribution of $X(0)$ on K. (For simplicity of notation in the sequel, assume $X(0) = (\tilde{x}_1, \ldots, \tilde{x}_{r-1})$.)

In [11], it is shown in a controlled setting that for each ν_{occ} and ν_τ satisfying (22) there is a process X and a stopping time τ for which (3) is satisfied for $t \le \tau$ and τ is essentially the first exit time of X. Thus (22) characterizes the occupation measure ν_{occ} and exit distribution ν_τ of the Wright-Fisher diffusion process having generator A. The relation (22) is the adjoint relation between the generator A of the Wright-Fisher diffusion and the measures ν_τ and ν_{occ}.

As in the analysis of the stationary distribution, ν_τ and ν_{occ} are measures on bounded sets so their respective joint moments uniquely characterize them. Restrict the set of test functions to $\mathcal{D}_\infty$ defined in (5) and define the two moment sets, respectively, by

$$m_\tau(k_1, \ldots, k_{r-1}) = \int_{\partial K} x_1^{k_1} \cdots x_{r-1}^{k_{r-1}}\, \nu_\tau(dx_1 \times \cdots \times dx_{r-1}) \tag{23}$$

and

$$m_{occ}(k_1, \ldots, k_{r-1}) = \int_K x_1^{k_1} \cdots x_{r-1}^{k_{r-1}}\, \nu_{occ}(dx_1 \times \cdots \times dx_{r-1}). \tag{24}$$

For $f \in \mathcal{D}_\infty$, the adjoint relation (22) is

$$\begin{aligned}
\tilde{x}_1^{k_1} \cdots \tilde{x}_{r-1}^{k_{r-1}} &= m_\tau(k_1, \ldots, k_{r-1}) \\
&\quad - \sum_{i=1}^{r-1} \frac{k_i(k_i-1)}{2} \left(m_{occ}(k_1, \ldots, k_i - 1, \ldots, k_{r-1}) - m_{occ}(k_1, \ldots, k_{r-1})\right) \\
&\quad + \sum_{i,j=1; i\neq j}^{r-1} (k_i k_j) m_{occ}(k_1, \ldots, k_{r-1}) + \sum_{i=1}^{r-1} k_i \left(\sum_{j=1}^{r} \mu_{ij}\right) m_{occ}(k_1, \ldots, k_{r-1}) \\
&\quad - \sum_{i=1}^{r-1} \sum_{j=1}^{r} k_i \mu_{ji} m_{occ}(k_1, \ldots, k_i - 1, \ldots, k_j + 1, \ldots, k_{r-1})
\end{aligned}$$

$$-\sum_{i=1}^{r-1}\sum_{j=1}^{r} k_i\sigma_{ij}m_{occ}(k_1,\ldots,k_j+1,\ldots,k_{r-1}) \tag{25}$$
$$+\sum_{i=1}^{r-1}\sum_{j,l=1}^{r} k_i\sigma_{jl}m_{occ}(k_1,\ldots,k_j+1,\ldots,k_l+1,\ldots k_{r-1}).$$

Again, the adjoint relation (25) does not imply that the collections $\{m_\tau(k_1,\ldots,k_{r-1}) : k_1,\ldots,k_{r-1} \in \mathbb{Z}^+\}$ and $\{m_{occ}(k_1,\ldots,k_{r-1}) : k_1,\ldots,k_{r-1} \in \mathbb{Z}^+\}$ are the moments of some distributions ν_τ and ν_{occ},respectively. It is necessary to include the Dale conditions (13) on each collection. Notice, in particular, that the boundary of the $(r-1)$-dimensional simplex consists of $(r-2)$-dimensional simplices so the measure ν_τ can be split into measures on each $(r-2)$-dimensional face, and thus the variables $\{m_\tau(k_1,\ldots,k_{r-1}) : k_1,\ldots,k_{r-1} \in \mathbb{Z}^+\}$ can also be split into variables on each $(r-2)$-dimensional face. These variables then need to satisfy the $(r-2)$-dimensional Dale conditions.

Additional information is contained in the distribution ν_τ and these collections of moments on the faces. Since each face corresponds to the extinction of an allele, the mass of ν_τ on each face gives the extinction probability for the corresponding allele. These probabilities are therefore provided by the zeroth order moments of ν_τ on each face.

3.1 Numerical examples

Example 3.1.1. Consider the model of $r = 2$ alleles with $\mu_1 = \mu_2 = 0$, so the Wright-Fisher diffusion process evolves in the unit interval in one-dimension and may exit at either endpoint (corresponding to the extinction of one of the alleles from the population). The extinction probability for the second allele is given analytically (see equation (2.32) of [5, Chapter 10]) as a function of the initial proportion x_0 of the first allele by

$$x_0 \overset{\varphi_0}{\longmapsto} \begin{cases} x_0, & \sigma = 0 \\ \dfrac{1-e^{-2\sigma x_0}}{1-e^{-2\sigma}}, & \sigma \neq 0 \end{cases}$$

where $\sigma_1 = -\sigma_2 = \sigma$. Table 3.1 compares the numerical results with the analytical values for the exit probability of the process at the point 1 (the probability of extinction of the second allele) in the case that $\sigma_1 = -\sigma_2 = \sigma = 10$. The values are computed using 20 and 35 moments.

Table 3.1. Extinction Probability as a function of x_0

$\sigma = 10$; $M = 20$ and $M = 35$

	$M = 20$ moments			$M = 35$ moments		
initial value x_0	lower bound	exact value	upper bound	lower bound	exact value	upper bound
0.10	0.855396	0.864665	0.874617	0.864665	0.864665	0.864665
0.20	0.974084	0.981684	0.989759	0.981684	0.981684	0.981684
0.30	0.996259	0.997521	0.999449	0.997521	0.997521	0.997521
0.40	0.999481	0.999665	0.999986	0.999785	0.999665	0.999665
0.50	0.999929	0.999955	1.000000	0.999955	0.999955	0.999955
0.60	0.999990	0.999994	1.000000	0.999994	0.999994	0.999994
0.70	0.999999	0.999999	1.000000	0.999999	0.999999	0.999999
0.80	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
0.90	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000

The mean extinction time of the second allele for the (general) case $\mu_1 = 0$, $\mu_2 > 0$, $\sigma_1, \sigma_2 \in \mathbb{R}$ as a function of the initial proportion x_0 of the first allele is given by (see [5]),

$$x_0 \mapsto E_{x_0}[\tau] = 2 \int_{x_0}^{1} y^{-2\mu_2} e^{-\lambda(y)} \int_{0}^{y} \frac{z^{2\mu_2 - 1} e^{\lambda(z)}}{1-z} \, dz \, dy,$$

where $\lambda(y) = \sigma_1 y^2 + \sigma_2 (1-y)^2$. For the special cases $\mu_2 = 0$, $\mu_2 = 0$ and $\sigma_1 = -\sigma_2 = \sigma$, respectively, this formula changes to

$$E_{x_0}[\tau] = 2 \int_{0}^{x_0} e^{-\lambda(y)} \int_{y}^{1/2} \frac{e^{\lambda(z)}}{z(1-z)} \, dz \, dy - 2\varphi_0(x_0) \int_{0}^{1} e^{-\lambda(y)} \int_{y}^{1/2} \frac{e^{\lambda(z)}}{z(1-z)} \, dz \, dy \tag{26}$$

and

$$E_{x_0}[\tau] = -2 \left[x_0 \log(x_0) + (1 - x_0) \log(1 - x_0) \right],$$

respectively.

Table 3.2 reports the results for the mean extinction time as a function of the number of moments used for the case $\mu_2 = 0$, $\sigma_1 = -\sigma_2 = 1$. If, as for the computations for Table 3.1 we only use monomials as test functions, the gap between the maximum value and the minimum value of the associated LPs might be large. To reduce the gap we have included additional test functions like $x \log(x)$, $(1-x)\log(1-x)$ and $\log(x)$, and used their Taylor series approximations up to order M as well as substitutions like $\log(10^{-7})$ for $\log(0)$ in our programs.

Table 3.2. Bounds on the mean extinction time as a function of the moments

$\mu_1 = \mu_2 = 0$, $\sigma_1 = -\sigma_2 = 1$, $x_0 = 0.5$

number of moments	lower bound	exact value	upper bound
10	1.258551	1.293240	1.369422
20	1.285232	1.293240	1.307876
30	1.289563	1.293240	1.299367
40	1.290855	1.293240	1.296671
50	1.291369	1.293240	1.295419
60	1.291853	1.293240	1.294976
70	1.292097	1.293240	1.294466
80	1.292412	1.293240	1.294412
90	1.292592	1.293240	1.294182
100	1.292527	1.293240	1.294090

Tables 3.3 and 3.4 report the results for $E_{x_0}[\tau]$ as a function of x_0 for the following two cases: $\mu_1 = 0$, $\mu_2 = 0$ and $\sigma_1 = -\sigma_2 = 1$ (case I), and $\mu_1 = 0$, $\mu_2 = 1$ and $\sigma_1 = 1$ and $\sigma_2 = 2$ (case II). Whereas for case I, Maple and Mathematica have no difficulties evaluating formula (26), neither program can directly deal with case II. The numbers in the column labeled NEF are the solutions of an approximation of the boundary value problem for which $x_0 \mapsto E_{x_0}[\tau]$ is the solution. The numbers in the column "estimated values" are the average values of the upper and lower bounds. The accuracy of the estimates decreases for larger sigma parameters and increases should these parameter values become smaller.

Table 3.3. Mean extinction time as a function of x_0

$\mu_1 = \mu_2 = 0,\ \sigma_1 = -\sigma_2 = 1,\ M = 100$

initial value x_0	lower bound	exact value	estimated value	upper bound
0.1	0.750470	0.752053	0.751667	0.752865
0.2	1.101586	1.103172	1.102855	1.104124
0.3	1.272638	1.273869	1.273683	1.274728
0.4	1.325367	1.326332	1.326250	1.327134
0.5	1.292527	1.293240	1.293308	1.294090
0.6	1.192493	1.193147	1.193221	1.193948
0.7	1.034307	1.034934	1.035021	1.035736
0.8	0.817125	0.817530	0.817739	0.818352
0.9	0.520805	0.521171	0.521351	0.521897

Table 3.4. Mean extinction time as a function of x_0

$\mu_1 = 0,\ \mu_2 = 1,\ \sigma_1 = 1,\ \sigma_2 = 2,\ M = 100$

initial value x_0	lower bound	NEF	estimated value	upper bound
0.1	2.0507	2.1185	2.0887	2.1266
0.2	1.9413	1.9867	1.9663	1.9913
0.3	1.8039	1.8328	1.8200	1.8361
0.4	1.6359	1.6574	1.6479	1.6599
0.5	1.4477	1.4618	1.4557	1.4637
0.6	1.2402	1.2477	1.2445	1.2488
0.7	1.0091	1.0154	1.0127	1.0163
0.8	0.7566	0.7609	0.7590	0.7615
0.9	0.4649	0.4667	0.4660	0.4670

Example 3.1.2. To illustrate the power but also the limitations of the LP-method we close with the numerical analysis of a case with three alleles ($r = 3$) so the diffusion operates in the two-dimensional simplex. The tables below are representative of the results obtained for a variety of parameters.

Table 3.5 reports the moments up to $M = 12$ of the boundary distributions for the case of three alleles and $\mu_{ij} = 0$, $\sigma_{ij} = 0$ for all i, j. The numbers nicely reveal the symmetry of the distributions on the boundary if, at the beginning of the evolution, the proportions of the alleles are the same, i.e. $x_0 = y_0 = 1/3$. We have included one additional test function, $f(x, y) = x\ln(x) + y\ln(y) + (1-x-y)\ln(1-x-y)$, along with the monomials up to order 12. The use of this function has been suggested by the analysis of the one-dimensional model.

Table 3.5. Moments of extinction probabilities

$\mu_{12} = \mu_{21} = \sigma_{ij} = 0,\ M = 100$

M	moments for $x = 0$	moments for $x + y = 1$	moments for $y = 0$
0	0.333333	0.333335	0.333331
1	0.166666	0.166668	0.166666
2	0.102935	0.120856	0.10717
3	0.0687223	0.0930412	0.0774225
4	0.0469631	0.0715403	0.0571901
5	0.0325845	0.0549768	0.0417154
6	0.0229829	0.0421709	0.029809
7	0.0163253	0.0321397	0.0208761
8	0.0114141	0.0240968	0.014322
9	0.00755308	0.0174522	0.00955211
10	0.00441248	0.0118128	0.0059716
11	0.00189479	0.006982	0.0029858
12	0	0.00295978	0

Table 3.6 reports bounds for the mean extinction time of at least one allele for the same set of parameters. To narrow the gap between the maximum and minimum values of the corresponding LP problems we have used three additional test functions, $f(x,y) = \ln(x+y)$, $f(x,y) = \ln(1-x)$ and $f(x,y) = \ln(1-y)$. To implement the associated constraints we have used the truncated power series of these functions.

In the two-dimensional case the choice of largest moment M becomes critical and a balance needs to be struck between accuracy and computing time. For instance, if $M = 20$ each LP run requires approximately 25,000 Simplex iterations. We have used CPLEX versions 6 and 7 on SUN worstations and modern PCs.

Table 3.6. Bounds on the mean exit time

$\mu_{12} = \mu_{21} = \sigma_{ij} = 0$, $M = 100$

M	lower bound	upper bound	estimated value
13	0.530103	0.575364	0.552734
14	0.533645	0.575364	0.554505
15	0.536699	0.575364	0.556031
16	0.539007	0.575364	0.557186
17	0.541503	0.575364	0.558433
18	0.542556	0.575365	0.558961
19	0.544815	0.575365	0.560090
20	0.546042	0.575365	0.560703

4 Concluding remarks

In this paper, we have illustrated the accuracy of linear programming in determining the mean extinction time and extinction probabilities for the diffusion approximation of the Wright-Fisher model in population genetics, and in computing the moments of the stationary distribution. The examples have included one-dimensional, two-dimensional and four-dimensional models. The method relies on adjoint equations (appropriate to the quantities of interest) which are determined from the martingale formulation of the dynamics of the diffusion. This approach naturally provides both lower and upper bounds on the moments of the distributions. An important issue which remains open is an inverse problem in which, for example, the density of the stationary distribution is approximated based on the computed moments. We expect that having both upper and lower bounds on the moments will be quite valuable since clear error estimates are provided.

5 Appendix

AMPL code for the stationary distribution of the Wright-Fisher model having 5 alleles

```
#
option cplex_options 'feasibility=1.0e-9 advance=0';
#
# Here are the Parameters
#
param M default 6 ;            # M denotes the highest moment in each variable
#
param mu11:=0  ;  param mu12:=2  ;  param mu13:=0.5;  param mu14:=0.5;  param mu15:=0.5;
param mu21:=3  ;  param mu22:=0  ;  param mu23:=0.5;  param mu24:=0.5;  param mu25:=0.5;
param mu31:=0.5;  param mu32:=0.5;  param mu33:=0  ;  param mu34:=1  ;  param mu35:=0.5;
```

```
param mu41:=0.5;  param mu42:=0.5;  param mu43:=1  ;  param mu44:=0  ;  param mu45:=0.5;
param mu51:=0.5;  param mu52:=0.5;  param mu53:=0.5;  param mu54:=0.5;  param mu55:=0  ;

param sg11:=1  ;  param sg12:=0.5;  param sg13:=0.5;  param sg14:=0.5;  param sg15:=0.5;
param sg21:=0.5;  param sg22:=0  ;  param sg23:=0  ;  param sg24:=0  ;  param sg25:=0  ;
param sg31:=0.5;  param sg32:=0  ;  param sg33:=0  ;  param sg34:=0  ;  param sg35:=0  ;
param sg41:=0.5;  param sg42:=0  ;  param sg43:=0  ;  param sg44:=0  ;  param sg45:=0  ;
param sg51:=0.5;  param sg52:=0  ;  param sg53:=0  ;  param sg54:=0  ;  param sg55:=0  ;

#
#
set IM := -2..M ;  %# Indizes i bezeichnen die Differenzenordnung
set I := 0..M ;  %# Indizes i bezeichnen die Differenzenordnung
set I1M := 1..M ; %# hat programmtechnischen Grund
set I2M := 2..M ; %# hat programmtechnischen Grund


set QUAD := { IM, IM ,IM, IM  } ;       %# Definitionsbereich fuer Variable y
set QUINT:= { IM, IM ,IM, IM, IM  } ;   %# Definitionsbereich fuer Variable y
set PAIRS := { I ,I } ;         %# Definitionsbereich fuer Variable y

#
# Definition of the variables
#
var y { QUAD }  ;
var v {QUINT}  ;
var w {QUINT}  ;

#
# The objective function

  maximize perform_max:
     y[3,5,0,0];

  minimize perform_min:
     y[3,5,0,0];

#

subject to prob :

y[0,0,0,0]  = 1;

subject to equation1 {r in 0..M-2,s in 0..M-2,t in 0..M-2,u in 0..M-2}:
                # the diffusion term for the Wright-Fisher model
  (1/2)* r*(r-1)*(y[r-1,s,t,u] - y[r,s,t,u]) + 1/2*s*(s-1)*(y[r,s-1,t,u] - y[r,s,t,u] )
 + (1/2)* t*(t-1)*(y[r,s,t-1,u] - y[r,s,t,u]) + (1/2)* u*(u-1)*(y[r,s,t,u-1] - y[r,s,t,u])
-r*s*y[r,s,t,u] - r*t*y[r,s,t,u] - r*u*y[r,s,t,u] - s*t*y[r,s,t,u] - s*u*y[r,s,t,u] - t*u*y[r,s,t,u]

                        # zusaetzlicher Drifterm
- r*(mu12+mu13+mu14+mu15+mu51)*y[r,s,t,u]
+ r*(mu21-mu51)*y[r-1,s+1,t,u] + r*(mu31-mu51)*y[r-1,s,t+1,u] + r*(mu41-mu51)*y[r-1,s,t,u+1]
+ r*mu51*y[r-1,s,t,u]

- s*(mu21+mu23+mu24+mu25+mu52)*y[r,s,t,u]
+ s*(mu12-mu52)*y[r+1,s-1,t,u] + s*(mu32-mu52)*y[r,s-1,t+1,u] + s*(mu42-mu52)*y[r,s-1,t,u+1]
+ s*mu52*y[r,s-1,t,u]

- t*(mu31+mu32+mu34+mu35+mu53)*y[r,s,t,u]
+ t*(mu13-mu53)*y[r+1,s,t-1,u] + t*(mu23-mu53)*y[r,s+1,t-1,u] + t*(mu43-mu53)*y[r,s,t-1,u+1]
+ t*mu53*y[r,s,t-1,u]

- u*(mu41+mu42+mu43+mu45+mu54)*y[r,s,t,u]
+ u*(mu14-mu54)*y[r+1,s,t,u-1] + u*(mu24-mu54)*y[r,s+1,t,u-1] + u*(mu34-mu54)*y[r,s,t+1,u-1]
+ u*mu54*y[r,s,t,u-1]

                # the terms involving the sigmas in the drift terms
                      #b_1
```

```
+    r*(sg15-sg55)*y[r,s,t,u]
+    r*(sg15+sg25+sg51+sg52-sg12-sg21-2*sg55)*y[r+1,s+1,t,u]
+    r*(sg15+sg35+sg51+sg53-sg13-sg31-2*sg55)*y[r+1,s,t+1,u]
+    r*(sg15+sg45+sg51+sg54-sg14-sg41-2*sg55)*y[r+1,s,t,u+1]
+    r*(sg25+sg35+sg52+sg53-sg23-sg32-2*sg55)*y[r,s+1,t+1,u]
+    r*(sg25+sg45+sg52+sg54-sg24-sg42-2*sg55)*y[r,s+1,t,u+1]
+    r*(sg35+sg45+sg53+sg54-sg34-sg43-2*sg55)*y[r,s,t+1,u+1]
+    r*(sg15+sg51-sg11-sg55)*y[r+2,s,t,u]
+    r*(sg25+sg52-sg22-sg55)*y[r,s+2,t,u]
+    r*(sg35+sg53-sg33-sg55)*y[r,s,t+2,u]
+    r*(sg45+sg54-sg44-sg55)*y[r,s,t,u+2]
+    r*(sg11+2*sg55-2*sg15-sg51)*y[r+1,s,t,u]
+    r*(sg12+2*sg55-sg15-sg25-sg52)*y[r,s+1,t,u]
+    r*(sg13+2*sg55-sg15-sg35-sg53)*y[r,s,t+1,u]
+    r*(sg14+2*sg55-sg15-sg45-sg54)*y[r,s,t,u+1]

                          #b_2
+    s*(sg25-sg55)*y[r,s,t,u]
+    s*(sg15+sg25+sg52+sg51-sg12-sg21-2*sg55)*y[r+1,s+1,t,u]
+    s*(sg15+sg35+sg51+sg53-sg13-sg31-2*sg55)*y[r+1,s,t+1,u]
+    s*(sg15+sg45+sg51+sg54-sg14-sg41-2*sg55)*y[r+1,s,t,u+1]
+    s*(sg25+sg35+sg52+sg53-sg23-sg32-2*sg55)*y[r,s+1,t+1,u]
+    s*(sg25+sg45+sg52+sg54-sg24-sg42-2*sg55)*y[r,s+1,t,u+1]
+    s*(sg35+sg45+sg53+sg54-sg34-sg43-2*sg55)*y[r,s,t+1,u+1]
+    s*(sg15+sg51-sg11-sg55)*y[r+2,s,t,u]
+    s*(sg25+sg52-sg22-sg55)*y[r,s+2,t,u]
+    s*(sg35+sg53-sg33-sg55)*y[r,s,t+2,u]
+    s*(sg45+sg54-sg44-sg55)*y[r,s,t,u+2]
+    s*(sg21+2*sg55-sg15-sg25-sg51)*y[r+1,s,t,u]
+    s*(sg22+2*sg55-2*sg25-sg52)*y[r,s+1,t,u]
+    s*(sg23+2*sg55-sg25-sg35-sg53)*y[r,s,t+1,u]
+    s*(sg24+2*sg55-sg25-sg45-sg54)*y[r,s,t,u+1]

                          #b_3
+    t*(sg35-sg55)*y[r,s,t,u]
+    t*(sg15+sg25+sg52+sg51-sg12-sg21-2*sg55)*y[r+1,s+1,t,u]
+    t*(sg15+sg35+sg51+sg53-sg13-sg31-2*sg55)*y[r+1,s,t+1,u]
+    t*(sg15+sg45+sg51+sg54-sg14-sg41-2*sg55)*y[r+1,s,t,u+1]
+    t*(sg25+sg35+sg52+sg53-sg23-sg32-2*sg55)*y[r,s+1,t+1,u]
+    t*(sg25+sg45+sg52+sg54-sg24-sg42-2*sg55)*y[r,s+1,t,u+1]
+    t*(sg35+sg45+sg53+sg54-sg34-sg43-2*sg55)*y[r,s,t+1,u+1]
+    t*(sg15+sg51-sg11-sg55)*y[r+2,s,t,u]
+    t*(sg25+sg52-sg22-sg55)*y[r,s+2,t,u]
+    t*(sg35+sg53-sg33-sg55)*y[r,s,t+2,u]
+    t*(sg45+sg54-sg44-sg55)*y[r,s,t,u+2]
+    t*(sg31+2*sg55-sg15-sg35-sg51)*y[r+1,s,t,u]
+    t*(sg32+2*sg55-sg25-sg35-sg52)*y[r,s+1,t,u]
+    t*(sg33+2*sg55-2*sg35-sg53)*y[r,s,t+1,u]
+    t*(sg34+2*sg55-sg35-sg45-sg54)*y[r,s,t,u+1]

                          #b_4
+    u*(sg45-sg55)*y[r,s,t,u]
+    u*(sg15+sg25+sg52+sg51-sg12-sg21-2*sg55)*y[r+1,s+1,t,u]
+    u*(sg15+sg35+sg51+sg53-sg13-sg31-2*sg55)*y[r+1,s,t+1,u]
+    u*(sg15+sg45+sg51+sg54-sg14-sg41-2*sg55)*y[r+1,s,t,u+1]
+    u*(sg25+sg35+sg52+sg53-sg23-sg32-2*sg55)*y[r,s+1,t+1,u]
+    u*(sg25+sg45+sg52+sg54-sg24-sg42-2*sg55)*y[r,s+1,t,u+1]
+    u*(sg35+sg45+sg53+sg54-sg34-sg43-2*sg55)*y[r,s,t+1,u+1]
+    u*(sg15+sg51-sg11-sg55)*y[r+2,s,t,u]
+    u*(sg25+sg52-sg22-sg55)*y[r,s+2,t,u]
+    u*(sg35+sg53-sg33-sg55)*y[r,s,t+2,u]
+    u*(sg45+sg54-sg44-sg55)*y[r,s,t,u+2]
+    u*(sg41+2*sg55-sg15-sg45-sg51)*y[r+1,s,t,u]
+    u*(sg42+2*sg55-sg25-sg45-sg52)*y[r,s+1,t,u]
+    u*(sg43+2*sg55-sg35-sg45-sg53)*y[r,s,t+1,u]
+    u*(sg44+2*sg55-2*sg45-sg54)*y[r,s,t,u+1]

= 0;
```

```
##############    The Dale Conditions   #######################
#
#
#      Defining the initial values of v
#
subject to diff_dale0 {r in 0..M, s in 0..M, t in 0..M, u in 0..M}:

        v[0,r,s,t,u] = y[r,s,t,u];

#
#      Defining the iterative differences of v
#
subject to diff_dale1 {t1 in I1M, r in 0..(M-t1),s in 0..(M-t1),t in 0..(M-t1),u in 0..(M-t1)}:

        v[t1,r,s,t,u] = v[t1-1,r,s,t,u] - v[t1-1,r+1,s,t,u] - v[t1-1,r,s+1,t,u]
                                        - v[t1-1,r,s,t+1,u] - v[t1-1,r,s,t,u+1] ;

#
#      The Dale conditions
#
subject to haussdorff_dale1 {t1 in 0..M, r in 0..(M-t1), s in 0..(M-t1), t in 0..(M-t1), u in 0..(M-t1)}:

        v[t1,r,s,t,u] >= 0;
```

References

[1] A. G. Bhatt and V. S. Borkar, Occupation measures for controlled Markov processes: Characterization and optimality, *Ann. Probab.*, **24** (1996), 1531–1562.

[2] J.G. Dai and J.M. Harrison, Steady-state analysis of RBM in a rectangle: Numerical methods and a queueing application, *Annals of Appl. Probab.,* **1** (1991), 16–35.

[3] J.G. Dai and T.G. Kurtz, The sufficiency of the basic adjoint relationship, unpublished manuscript.

[4] A.I. Dale, Two-dimensional Moment Problems, *Math. Scientist.* **12** (1987), 21-29.

[5] S.N. Ethier and T.G. Kurtz, *Markov Processes: Characterization and Convergence*, Wiley, New York (1986).

[6] W. Feller, *An Introduction to Probability Theory and its Applications*, Vol. 2, Wiley, New York (1965).

[7] J.M. Harrison and M.I. Reiman, On the distribution of multidimensional reflected Brownian motion, *SIAM J. Appl. Math,* **41** (1981), 345–361.

[8] K. Helmes, S. Röhl, and R.H. Stockbridge, Computing Moments of the Exit Time Distribution for Markov Processes by Linear Programming, *Operations Research,* **49** (2001), 516–530.

[9] K. Helmes, R.H. Stockbridge, Numerical Comparison of Controls and Verification of Optimality for Stochastic Control Problems", *J. Optim. Th. Appl.,* **106** (2000), 107-127.

[10] K. Helmes and R.H. Stockbridge, Numerical Evaluation of Resolvents and Laplace Transforms of Markov Processes, *Mathematical Methods of Operations Research,* **53** (2001), 309–331.

[11] T.G. Kurtz and R.H. Stockbridge, Existence of Markov Controls and Characterization of Optimal Markov Controls, *SIAM J. Control. Optim.,* **36** (1998), 609–653.

[12] T.G. Kurtz, Martingale problems for constrained Markov processes, in *Recent Advances in Stochastic Calculus*, J.S. Baras, V. Mirelli, eds., Springer–Verlag, New York, 1990.

[13] T.G. Kurtz, A control formulation for constrained Markov processes, in *Mathematics of Random Media*, Lectures in Applied Mathematics,**27**, 1991.

[14] T.G. Kurtz and R.H. Stockbridge, Martingale problems and linear programs for singular control, *Thirty-seventh annual Allerton conference on communication, control, and computing (Monticello, Ill., 1999)*, 11-20, Univ. Illinois, Urbana-Champaign, Ill.

[15] M.S. Mendiondo and R.H. Stockbridge, Long-term average control of a local time process, to appear in *Markov processes and controlled Markov chains,* Kluwer.

[16] S. Röhl, Ein linearer Programmierungsansatz zur Lösung von Stopp- und Steuerungsproblemen, Ph.D. Dissertation, Humboldt-Universität zu Berlin, Berlin, Germany, 2001.

[17] E. Schwerer, A linear programming approach to the steady-state analysis of reflected Brownian motion, *Stoch. Models,* **17** (2001), 341–368.

[18] R.H. Stockbridge, Time-average control of martingale problems: Existence of a stationary solution, *Ann. Probab.*, **18** (1990), 190–205.

[19] R.H. Stockbridge, Time-average control of martingale problems: A linear programming formulation, *Ann. Probab.*, **18** (1990), 206–217.

[20] L.C. Young, On approximation by polygons in the calculus of variations, *Proc. Royal Society, Series A*, **141** (1933), 325–341.