, ( ), 1–16 (2003)

Extension of Dale's Moment Conditions with Application to the Wright-Fisher Model*

Kurt Helmes[1] and Richard H. Stockbridge[2]

Abstract

Dale's necessary and sufficient conditions for an array to contain the joint moments for some probability distribution on the unit simplex in $\mathbb{R}^2$ are extended to the unit simplex in $\mathbb{R}^d$. These conditions are then used in a computational method, based on linear programming, to evaluate the stationary distribution for the diffusion approximation of the Wright-Fisher model in population genetics. The computational method uses a characterization of the diffusion through an adjoint relation between the diffusion operator and its stationary distribution. Application of this adjoint relation to a set of functions in the domain of the generator leads to one set of constraints for the linear program involving the moments of the stationary distribution. The extension of Dale's conditions on the moments add another set of lin-

*This research is partially supported by NSF under grant DMS 9803490.
[1]Institute for Operations Research, Humboldt University of Berlin, Berlin, Germany
[2]Department of Mathematical Sciences, University of Wisconsin Milwaukee, Milwaukee, Wisconsin and Department of Statistics, University of Kentucky, Lexington, Kentucky

Copyright © 2000 by Marcel Dekker, Inc. www.dekker.com

ear conditions and the linear program is solved to obtain bounds on numerical quantities of interest. Numerical illustrations are given to illustrate the accuracy of the method.

Key Words: Dale moment conditions, Wright-Fisher model, stationary distribution, linear programming, moments, computational probability.

I. Introduction and model formulation

In this paper, necessary and sufficient conditions on an array to contain the joint moments of a distribution on the unit simplex in $\mathbb{R}^2$ derived by Dale [1] are extended to the unit simplex in $\mathbb{R}^d$. The motivation for this extension is the numerical evaluation, based on linear programming, of the Wright-Fisher model in population genetics.

We take as our starting point the diffusion approximation to the Wright-Fisher genetic model with r alleles and refer the reader to [2], Chapter 10, for an excellent explanation of the original model and the diffusion approximation. We use the notation of [2] so $d = r - 1$. 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 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 [2], letting $X(t) = (X_1(t), \ldots, X_{r-1}(t))$ 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 $f \in \mathcal{D}(A)$.

When $\mu_{ij} > 0$ for $i \neq j$, 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 $f \in \mathcal{D}(A)$,

$$\begin{aligned} E[f(X(0))] &= E\left[f(X(t)) - \int_0^t Af(X(s))\, ds \right] \\ &= E[f(X(t))] - \int_0^t E[Af(X(s))]\, ds \end{aligned}$$

and by the stationarity of the process, it follows that

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

The identity (4), in fact, characterizes the stationary distribution ν (see [2, Theorem 4.9.17]).

Our linear programming approach for the analysis of this diffusion involves utilizing this characterization of ν through (4). The basic idea is to select the collection of test functions

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

and set the conditions (4) for these f as the constraints of a linear program. Using these test functions, the left hand side of (4) becomes a linear combination of the joint moments of ν and the linear program only involves these joint moments.

It is at this point that the extension of Dale's conditions are needed. The array is required to satisfy (4) but need not be an array of joint moments. The linear sufficient conditions are also added as constraints to the 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 [4], [6] 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 [4] computed the moments of the exit time distribution directly, whereas in [6] the Laplace transform of the moments were determined. A similar approach has been applied on stochastic control problems [5] and optimal stopping problems by Röhl [7]. Schwerer [8] uses a linear programming approach involving the moments of a reflected Brownian motion process over an unbounded domain.

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 a special choice of parameters for which the linear program can be immediately solved. A careful examination of the moment conditions for the two-dimensional simplex due to Dale [1] and its extension to higher dimensions is given in Section 3. These conditions are needed for most choices of the parameters. Section 4 then illustrates the accuracy of the LP formulation and includes numerical illustrations for a case in which there are known analytical results and for a choice of parameters for which no analytical results are known.

The numerical illustrations are given for a Wright-Fisher model with $r = 5$ alleles in which case K is the simplex in $\mathbb{R}^4$. Many numerical methods have difficulty with problems in dimensions strictly greater than 3 due to the "curse of dimensionality". We have chosen a high dimension so as to demonstrate that the linear programming approach gives accurate results for this high dimensional problem.

II. Linear Program for Wright-Fisher Model

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 $\mathcal{D}_\infty$ of (5). In this way, (4) can be expressed in terms of the moments of the distribution. Specifically, fix $k_1, \dots, k_{r-1}$ and consider $f(x_1, \dots, x_{r-1}) = x_1^{k_1} \cdots x_{r-1}^{k_{r-1}}$. Define the joint $(k_1, \dots, k_{r-1})$-moment of ν by

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

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

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

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 $m(k_1, \ldots, k_i + 1, \ldots, k_j + 1, \ldots k_{r-1}) = m(k_1, \ldots, k_i + 2, \ldots, k_{r-1})$.

A. 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 1.

Table 1. Values of $m(k_1, k_2, 0, 0)$ from LP

		k_2					
		0	1	2	3	4	5
	0	1	0.2	0.06666	0.02857	0.01428	0.00793
	1	0.2	0.03333	0.00952	0.00357	0.00158	0.00079
	2	0.06666	0.00952	0.00238	0.00079	0.00031	0.00014
	3	0.02857	0.00357	0.00079	0.00023	8.658e-05	3.607e-05
	4	0.01428	0.00158	0.00031	8.658e-05	2.886e-05	1.11e-05
k_1	5	0.00793	0.00079	0.00014	3.607e-05	1.11e-05	3.964e-06
	6	0.00476	0.00043	7.215e-05	1.665e-05	4.757e-06	1.585e-06
	7	0.00303	0.00025	3.885e-05	8.325e-06	2.22e-06	6.937e-07
	8	0.00202	0.00015	2.22e-05	4.44e-06	1.11e-06	3.264e-07
	9	0.00139	9.99e-05	1.332e-05	2.497e-06	5.876e-07	1.632e-07
	10	0.00099	6.66e-05	8.325e-06	1.469e-06	3.264e-07	8.591e-08

These values agree with the analytic values

$$m(k_1, k_2, 0, 0) = \frac{4!k_1!k_2!}{(k_1 + k_2 + 4)!}.$$

III. 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 [3]) 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-dimensional (and two-dimensional extension to the unit square) Hausdorff moment conditions were used by the authors in [4], [6] 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 [1]. 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

$$\begin{aligned} w(k_1,k_2,k_3) = {} & w(k_1,k_2,k_3-1)-w(k_1+1,k_2,k_3-1) \\ & -w(k_1,k_2+1,k_3-1). \end{aligned}$$

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 $k_3\geq 1$,

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

$$- \int_K x_1^{k_1} x_2^{k_2+1} (1 - x_1 - x_2)^{k_3 - 1} \nu(dx_1 \times dx_2). \tag{12}$$

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*

$$w(0, \ldots, 0) = 1 \qquad \textit{and} \qquad \forall k_1, \ldots, k_{r-1} \in \mathbb{Z}^+$$

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

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 $\mathbb{M} = \{m(\vec{k}) \mid \vec{k} = (k_1, \ldots, k_{r-1}), k_i \in \mathbb{Z}^+\}$ 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

$$p_n \left(\frac{i_1}{n}, \ldots, \frac{i_{r-1}}{n} \right)$$

$$= \binom{n}{i_1,\ldots,i_{r-1}} w\left(i_1,,\ldots,i_{r-1},n-\sum_{l=1}^{r-1} i_l\right) \tag{16}$$

where

$$\binom{n}{i_1,\ldots,i_{r-1}} = \frac{n!}{(\prod_{l=1}^{r-1} i_l!)(n-\sum_{l=1}^{r-1} i_l)!}\,.$$

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 simplify the typing we shall use the following abbreviations throughout the remaining part of the proof:

$$\mathbb{1} := (1,\ldots,1)\in\mathbb{R}^{r-1}$$

$$\vec{\jmath} := (j_1,\ldots,j_{r-1})\in\mathbb{N}_0^{r-1}, \quad \text{etc.}$$

$$n-|\vec{u}| := n-\sum_{l=1}^{r-1} u_l$$

$$\binom{i_r}{\vec{\jmath}} := \binom{i_r}{j_1\ldots j_{r-1}}$$

$$m(\vec{\imath}+\vec{\jmath}) := m(i_1+j_1,\ldots,i_{r-1}+j_{r-1})$$

$$(\vec{\imath},i_r) := (i_1,\ldots,i_{r-1},i_r)\in\mathbb{N}_0^r$$

$$\vec{u}\geq\vec{\imath} \iff u_1\geq i_1,\ldots,u_{r-1}\geq i_{r-1}$$

$$\vec{h}_{\vec{\imath},\vec{v},\vec{\jmath}} := \left(\vec{h}^{(1)},\vec{h}^{(2)},\vec{h}^{(3)}\right) := (\vec{\imath},\vec{v},\vec{\jmath})\in\mathbb{N}_0^{r-1}\times\mathbb{N}_0^{r-1}\times\mathbb{N}_0^{r-1}$$

$$\binom{n}{\vec{h}_{\vec{\imath},\vec{u}-\vec{\imath},\vec{\jmath}}} = \frac{n!}{\left(\prod_{l=1}^{r-1}\left(h_l^{(1)}!\,h_l^{(2)}!\,h_l^{(1)}!\right)\right)\cdot(n-|\vec{u}+\vec{\jmath}|)!}\,.$$

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 $\mathbb{M}$:

$$w(i_1,\ldots,i_{r-1},i_r) = w\Big((\vec{\imath},i_r)\Big)$$

$$= \sum_{\vec{\jmath}\in\mathcal{I}_{i_r}} \binom{i_r}{j} (-1)^{|\vec{\jmath}|}\, m(\vec{\imath}+\vec{\jmath}). \tag{17}$$

Using (17) we have

$$\sum_{\substack{\vec{u}\in\mathcal{I}_n\\ \vec{u}\geq\vec{\imath}}} \left(\prod_{l=1}^{r-1}\binom{u_l}{i_l}\right) p_n\left(\frac{u_1}{n},\ldots,\frac{u_{r-1}}{n}\right)$$

$$= \sum_{\substack{\vec{u}\in\mathcal{I}_n\\ \vec{u}\geq\vec{\imath}}} \sum_{\vec{\jmath}\in\mathcal{I}_{n-|\vec{u}|}} \left(\prod_{l=1}^{r-1}\binom{u_l}{i_l}\right)\binom{n}{\vec{u}}\binom{n-|\vec{u}|}{\vec{\jmath}}(-1)^{|\vec{\jmath}|}\, m(\vec{u}+\vec{\jmath})$$

$$= \sum_{\substack{\vec{u}\in\mathcal{I}_n\\ \vec{u}\geq\vec{\imath}}} \sum_{\vec{\jmath}\in\mathcal{I}_{n-|\vec{u}|}} (-1)^{|\vec{\jmath}|}\binom{n}{\vec{h}_{\vec{\imath},\vec{u},\vec{\jmath}}} m(\vec{u}+\vec{\jmath})$$

$$= \sum_{\substack{\vec{u}\in\mathcal{I}_n\\ \vec{u}\geq\vec{\imath}}} \sum_{\vec{\jmath}\in\mathcal{I}_{n-|\vec{u}|}} \binom{n}{\vec{\imath}}\binom{n-|\vec{\imath}|}{\vec{h}_{\mathbf{1},\vec{u}-\vec{\imath},\vec{\jmath}}}(-1)^{|\vec{\jmath}|}\, m(\vec{u}+\vec{\jmath}) \tag{18}$$

$$= \binom{n}{\vec{\imath}} \sum_{\substack{\vec{k}\in\mathcal{I}_n\\ \vec{k}\geq\vec{\imath}}} \binom{n-|\vec{\imath}|}{\vec{k}-\vec{\imath}} \sum_{\vec{\jmath}\in\mathcal{I}_{n-|\vec{u}|}} \binom{|\vec{k}-\vec{u}|}{\vec{h}_{\mathbf{1},\vec{k}-\vec{\imath}-\vec{\jmath},\vec{\jmath}}}(-1)^{|\vec{\jmath}|}\, m(\vec{k}) \tag{19}$$

$$= \binom{n}{\vec{\imath}} \sum_{\substack{\vec{k}\in\mathcal{I}_n\\ \vec{k}\geq\vec{\imath}}} \binom{n-|\vec{\imath}|}{\vec{k}-\vec{\imath}}(1+\cdots+1-1-\cdots-1)^{|\vec{k}-\vec{\imath}|}\, m(\vec{k}) \tag{20}$$

$$= \binom{n}{\vec{\imath}} m(\vec{\imath}); \tag{21}$$

the term $(1+\cdots+1-1-\cdots-1)$ in the second to last expression consists of $r-1$ terms of $+1$ and $r-1$ terms of -1 and a careful examination of the summation in the third to last expression shows that, in the second to last expression, $(1+\cdots+1-1-\cdots-1)^0 = 1$. Of particular interest is the case in which $i_1 = \cdots = i_{r-1} = 0$, in which case (18) 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

$$\begin{aligned}
m(\vec{\imath}) &= \binom{n}{\vec{\imath}}^{-1} E\left[\prod_{l=1}^{r-1}\binom{nX_l^{(n)}}{i_l}\right] \\
&= \frac{n^{|\vec{\imath}|}(n-|\vec{\imath}|)!}{n!}\, E\left[\prod_{l=1}^{r-1}\prod_{k_l=0}^{i_l}\left(X_l^{(n)} - \frac{k_l}{n}\right)\right] \\
&\to E\left[\prod_{l=1}^{r-1} X_l^{i_l}\right]
\end{aligned}$$

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 $\mathbb{P}$.

IV. Numerical examples

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.

Example 4.1. Consider a modification of the model in Section (A) 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

$$\begin{aligned} &p(x_1, x_2, x_3, x_4) \\ &\quad = C x_1^{2\mu_1 - 1} x_2^{2\mu_2 - 1} x_3^{2\mu_3 - 1} x_4^{2\mu_4 - 1} (1 - x_1 - x_2 - x_3 - x_4)^{2\mu_5 - 1} \\ &\qquad \cdot e^{s(x_1, x_2, x_3, x_4)} \end{aligned} \tag{22}$$

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

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} & \sigma_{15} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} & \sigma_{24} & \sigma_{25} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} & \sigma_{34} & \sigma_{35} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_{44} & \sigma_{45} \\ \sigma_{51} & \sigma_{52} & \sigma_{53} & \sigma_{54} & \sigma_{55} \end{pmatrix}, \qquad \text{and}$$

$$\tilde{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ 1 - x_1 - x_2 - x_3 - x_4 \end{pmatrix},$$

and C is a normalizing constant.

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

$$\Sigma = \begin{pmatrix} 1 & 0.5 & 0.5 & 0.5 & 0.5 \\ 0.5 & 0 & 0 & 0 & 0 \\ 0.5 & 0 & 0 & 0 & 0 \\ 0.5 & 0 & 0 & 0 & 0 \\ 0.5 & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{23}$$

It then follows that each $2\mu_i - 1 = 0$ and $s(x_1, x_2, x_3, x_4) = x_1$ in (22) 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 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 4.2. We now consider a modification of the model in Example 4.1 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 (23). *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 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. Values of $m(k_1, k_2, 0, 0)$ from LP

			k_2 0	1	2	3	4	5
k_1	0	max	1	0.19274	0.06258	0.02630	0.01295	0.00760
		exact	1	0.19274	0.06258	0.02630	0.01295	0.00711
		min	1	0.19274	0.06258	0.02630	0.01295	0.00657
	1	max	0.22902	0.03628	0.00997	0.00363	0.00157	0.00083
		exact	0.22902	0.03628	0.00997	0.00363	0.00157	0.00077
		min	0.22902	0.03628	0.00997	0.00363	0.00157	0.00069
	2	max	0.08390	0.01133	0.00271	0.00087	0.00033	0.00016
		exact	0.08389	0.01133	0.00271	0.00087	0.00033	0.00015
		min	0.08389	0.01133	0.00271	0.00087	0.00033	0.00013
	3	max	0.03856	0.00455	0.00096	0.00027	9.85e-05	4.93e-05
		exact	0.03855	0.00455	0.00096	0.00027	9.83e-05	3.99e-05
		min	0.03854	0.00455	0.00096	0.00027	9.82e-05	2.75e-05
	4	max	0.02041	0.00214	0.00041	0.00010	3.59e-05	3.05e-05
		exact	0.02034	0.00213	0.00040	0.00010	3.45e-05	1.29e-05
		min	0.02027	0.00212	0.00040	0.00010	3.31e-05	0
	5	max	0.01238	0.00118	0.00021	5.60e-05	2.83e-05	1.27e-05
		exact	0.01183	0.00116	0.00019	4.66e-05	1.38e-05	4.81e-06
		min	0.01130	0.00104	0.00017	3.48e-05	0	0

Table 3. Bounds on the Values of $m(k_1, k_2, 0, 0)$ from LP

			k_2 0	1	2	3	4	5
k_1	0	max	1	0.17125	0.04255	0.01318	0.00476	0.00213
		min	1	0.17125	0.04255	0.01318	0.00475	0.00160
	1	max	0.25272	0.04562	0.01133	0.00344	0.00122	0.00054
		min	0.25272	0.04562	0.01133	0.00344	0.00118	0.00038
	2	max	0.08720	0.01572	0.00380	0.00111	0.00038	0.00017
		min	0.08720	0.01572	0.00380	0.00111	0.00036	0.00011
	3	max	0.03595	0.00635	0.00148	0.00041	0.00014	6.31e-05
		min	0.03594	0.00634	0.00147	0.00041	0.00012	2.69e-05
	4	max	0.01675	0.00288	0.00065	0.00017	5.88e-05	5.02e-05
		min	0.01665	0.00284	0.00063	0.00016	4.87e-05	0
	5	max	0.00897	0.00151	0.00033	9.64e-05	4.81e-05	2.16e-05
		min	0.00800	0.00128	0.00026	5.46e-05	0	0

References

1. Dale, A.I. Two-dimensional Moment Problems. Math. Scientist. **1987**, *12*, 21-29.
2. Ethier, S.N.; Kurtz, T.G. *Markov Processes: Characterization and Convergence*; John Wiley & Sons, Inc.: New York, 1986.
3. Feller, W. *An Introduction to Probability Theory and its Applications*; John Wiley & Sons, Inc.: New York, 1965; Vol. 2.
4. Helmes, K.; Röhl, S.; Stockbridge, R.H. Computing Moments of the Exit Time Distribution for Markov Processes by Linear Programming. Operations Research **2001**, *49*, 516-530.
5. Helmes, K.; Stockbridge, R.H. Numerical Comparison of Controls and Verification of Optimality for Stochastic Control Problems. J. Optim. Th. Appl. **2000**, *106*, 107-127.
6. Helmes, K.; Stockbridge, R.H. Numerical Evaluation of Resolvents and Laplace Transforms of Markov Processes. Mathematical Methods of Operations Research **2001**, *53*, 309-331.
7. Röhl, S. *Ein linearer Programmierungsansatz zur Lösung von Stopp- und Steuerungsproblemen*, Dissertation; Humboldt-Universität zu Berlin: Berlin, Germany, 2001; 1-114.
8. Schwerer, E. A linear programming approach to the steady-state analysis of reflected Brownian motion. Stoch. Models **2001**, *17*, 341-368.