Convergence Time to the Ewens Sampling Formula

Joseph C. Watkins

Department of Mathematics, University of Arizona, Tucson, Arizona 85721 USA

Abstract

In this paper, we establish the cutoff phenomena for the discrete time infinite alleles Moran model. If M is the population size and μ is the mutation rate, we find a cutoff time of $\log(M\mu)/\mu$ generations. The stationary distribution for this process in the case of sampling without replacement is the Ewens sampling formula. We show that the bound for the total variation distance from the generation t distribution to the Ewens sampling formula is well approximated by one of the extreme value distributions, namely, a standard Gumbel distribution. Beginning with the card shuffling examples of Aldous and Diaconis and extending the ideas of Donnelly and Rodrigues for the two allele model, this model adds to the list of Markov chains that displays the cutoff phenomenon. Because of the broad use of infinite alleles models, this cutoff sets the time scale of applicability for statistical tests based on the Ewens sampling formula and other tests of neutrality in a number of population genetic studies.

Key Words. Markov chains, infinite alleles Moran model, Ewens sampling formula, cutoff phenomena, Hoppe's urn, lines of descent, extreme value distribution

AMS Classification. 60J10, 60J20, 92D10

1. Introduction. Open a box of playing cards and inspect them after each riffle shuffle. Bayer and Diaconis (1992) prove that features of the original order of the cards can typically be detected after four shuffles, but essentially all evidence of the original order of cards disappears after seven shuffles. Consequently, strategies for a game using this deck should depend on the number of times it is shuffled.

This provides one of several examples of what is known as the cutoff phenomenon for recurrent Markov chains. This phenomenon has been seen in a variety of circumstances, e.g., card shuffling, (Aldous and Diaconis, 1986, 1987, Pemantle, 1989), the Ehrenfest urn model, (Aldous, 1983, and Diaconis and Shahshahani, 1987) random walks on matrix groups, (Diaconis, 1986) the Bernoulli-Laplace diffusion, (Diaconis and Shahshahani, 1987) and the two allele Moran model (Donnelly and Rodrigues, 2000). (See also Rosenthal, 1995, Diaconis, 1996, or Saloff-Coste, 1997.)

The type of analysis applied to genetic data fundamentally depends on understanding if and when a cutoff phenomenon occurs for an appropriate genetic model. If the population has existed for less than the cutoff time, then aspects of the genotypes of the founding population should be visible and any mathematical analysis ought to employ strategies that use this information. On the other hand, if the population has persisted for more generations than the cutoff time, analysis of the genetic data should largely be based on a stationary distribution of a Markov process that models the genetic evolution.

In this manuscript, we establish that the cutoff phenomena holds for one class of the most frequenly used stochastic models in population genetics, neutral infinite alleles models. The main result is a back of the envelop calculation which provides a criterion (26) for the applicability of statistical tests based on a neutral equilibrium. In a separate manuscript, we apply this to investigate the evolution of patrilines in Indonesian communities. (See Lansing et. al., 2007.)

We begin with a time homogeneous Markov chain X on a countable state space S. Probabilities of events are determined from an initial probability measure ν_0 and from the transition matrix

$$T(x,y) = P\{X_{t+1} = y | X_t = x\} = P\{X_{t+1} = y | X_t = x, X_{t-1} = x_{t-1}, \ldots, X_0 = x_0\}.$$

Multistep transition probabilities are determined via matrix multiplication $T^s(x,y) = P\{X_{t+s} = y | X_t = x\}$. Thus, for example, if we set $\nu_t(A) = P\{X_t \in A\}$, then

$$\nu_{t+s}\{y\} = (\nu_t T^s)\{y\} = \sum_{x \in S} \nu_t\{x\} T^s(x,y).$$

In the case in which every state is *accessible* from every other state (i. e., for all x and y, there exists $t \geq 0$ so that $T^t(x,y) > 0$) then the Markov chain is called *irreducible*. The *period* of a state x is defined as $\ell(x) = \gcd\{t > 0; T^t(x,x) > 0\}$ ($\gcd(\emptyset) = 0$). If $\ell(x) = 1$, x is called *aperiodic*. Using the notation P_x to denote probability conditioned on the event $\{X_0 = x\}$, we say that a state x is *recurrent* whenever $P_x\{X_t = x \text{ for some } t > 0\} = 1$ and *transient* otherwise. If the time to return to x has finite mean μ_x, then the state is called *positive recurrent*. Otherwise, a recurrent state is called *null recurrent*. If S is finite, then irreducibility is a sufficient condition for positive recurrence.

For an irreducible Markov chain, every state has the same period and every state in the chain is either positive recurrent, null recurrent, or transient. In the aperiodic case, the asymptotic behavior is described by

$$\lim_{t \to \infty} T^t(x,y). \tag{1}$$

For an irreducible chain in which the states are either transient or null recurrent, this limit is 0. In the positive recurrent case, the limit is $1/\mu_y > 0$. The remarkable connection is that this limit is also $\pi\{y\}$, the unique *stationary probability measure*. In other words, $\pi T = \pi$, π is a left *eigenmeasure* with eigenvalue 1.

Call a Markov chain *ergodic* if it is time homogeneous, irreducible, aperiodic and positive recurrent. From a statistical point of view, if we evaluate an ergodic Markov chain X_t for a large value of t, we must know the rate of convergence in (1) if we want to say that we are sampling from a probability measure near to π. The classical method is to determine a geometric rate of convergence based on an estimation of the second leading eigenvalue.

Seeking an improvement, Aldous and Diaconis (1986) begin by measuring the distance between two probability measures ν and $\tilde{\nu}$ using the *total variation norm*,

$$||\nu - \tilde{\nu}||_{TV} = \sup\{|\nu(B) - \tilde{\nu}(B)|; B \subset S\}. \tag{2}$$

The supremum is achieved on both the sets $\{y; \nu\{y\} > \tilde{\nu}\{y\}\}$ and $\{y; \tilde{\nu}\{y\} > \nu\{y\}\}$. Consequently,

$$||\nu - \tilde{\nu}||_{TV} = \sum_{y \in S} (\nu\{y\} - \tilde{\nu}\{y\}) I_{\{y; \nu\{y\} > \tilde{\nu}\{y\}\}}. \tag{3}$$

Probabilistic techniques are used to obtain an upper bound through the use of *coupling*.

Definition 1. *Let ν and $\tilde{\nu}$ be two probability measures. A* coupling *of ν and $\tilde{\nu}$ is a pair of random variables ζ and $\tilde{\zeta}$ defined on a common probability space such that the marginal distribution of ζ is ν and the marginal distribution of $\tilde{\zeta}$ is $\tilde{\nu}$.*

Theorem 2. *Let ν and $\tilde{\nu}$ be probability measures on a countable state space and let $\mathcal{C}(\nu, \tilde{\nu})$ denote the collection of couplings of ν and $\tilde{\nu}$. Then,*

$$||\nu - \tilde{\nu}||_{TV} = \inf\{P\{\zeta \neq \tilde{\zeta}\}; (P, \zeta, \tilde{\zeta}) \in \mathcal{C}(\nu, \tilde{\nu})\}. \tag{4}$$

Proof. See Lindvall, 1992, page 12, where it is called the *basic coupling inequality* or Levin, Peres, and Wilmer, 2006, Proposition 3.11. □

Aldous and Diaconis introduce a second method to bound this distance using the concept of *strong stationary times*. In this context, we view the Markov chain evolving in association with an increasing family of σ-algebras $\{\mathcal{F}_t; t \geq 0\}$ which is to be regarded as the totality of information available at time step t. In particular, X must be *adapted*, i.e., for each $t \geq 0$ and $A \subset S$, the event $\{X_t \in A\} \in \mathcal{F}_t$. In words, questions concerning the state of the Markov chain at time t can be answered using the information contained in $\mathcal{F}_t$.

In the following development, we shall assume that X is an ergodic $\mathcal{F}_t$-adapted Markov chain on a countable state space S and that X has unique stationary probability measure π.

Definition 3. *A random time τ is called a* strong stationary time *provided that*

1. *For all $t \geq 0$, $\{\tau \leq t\} \in \mathcal{F}_t$, i.e., whether or not the strong stationary time has occurred can be determined from the information available at time t,*
2. *X_τ has distribution π, and*
3. *X_τ and τ are independent.*

Properties 2 and 3 imply that $P_x\{X_s = y, \tau = s\} = P_x\{X_\tau = y, \tau = s\} = \pi\{y\}P_x\{\tau = s\}$. Now, sum on $s \leq t$ to obtain

$$P_x\{X_t = y, \tau \leq t\} = \sum_{s=0}^{t} P_x\{X_t = y, \tau = s\} = \sum_{s=0}^{t} P_x\{X_t = y|\tau = s\}P_x\{\tau = s\} = \pi\{y\}P_x\{\tau \leq t\} \qquad (5)$$

because

$$\begin{aligned} P_x\{X_t = y|\tau = s\} &= \sum_{x\in S} P_x\{X_t = y, X_s = x|\tau = s\} = \sum_{x\in S} P_x\{X_s = x|\tau = s\}P_x\{X_t = y|X_s = x, \tau = s\} \\ &= \sum_{x\in S} P_x\{X_\tau = x|\tau = s\}P_x\{X_t = y|X_s = x\} = \sum_{x\in S} \pi\{x\}T^{t-s}(x, y) = \pi\{y\}. \end{aligned}$$

Their first example of a strong stationary time is described in the "top to random shuffle". Let X be a Markov chain whose state space is the set of permutation on M letters, i.e., the order of the cards in a deck having a total of M cards. We write $X_t(k)$ for the card in the k-th position after t shuffles. The transitions are to take the top card and place it at a random position uniformly distributed in the deck. Note that the uniform probability measure on the $M!$ permutations is a stationary measure. Define

$$\tau = \inf\{t > 0; X_t(1) = X_0(M)\} + 1,$$

the first shuffle after the original bottom card has moved to the top.

Proposition 4. *τ is a strong stationary time.*

Proof. We show that at the time in which we have k cards under the original bottom card ($X_t(M - k) = X_0(M)$), then all $k!$ orderings of these k cards are equally likely.

To establish a proof by induction on k, note that the statement above is obvious in the case $k = 1$. For the case $k = j$, we assume that all $j!$ arrangements of the cards under $X_0(M)$ are equally likely. When an additional card is placed under $X_0(M)$, then each of the $j + 1$ available positions are equally likely to be selected, and thus all $(j + 1)!$ arrangements of the cards under $X_0(M)$ are equally likely. Consequently, the first time that the original bottom card is placed into the deck, then, independent of the value of τ, all $M!$ orderings are equally likely. □

To show the value of strong stationary times, we recall the following.

Definition 5. *The* separation distance

$$s_t = \sup \left\{ 1 - \frac{T^t(x,y)}{\pi\{y\}}; x, y \in S \right\}.$$

Proposition 6. *Let τ be a strong stationary time for X, then*

$$s_t \leq \max_{x \in S} P_x\{\tau > t\}.$$

Proof. For any $x, y \in S$, use equation (5) to see that

$$\begin{aligned} 1 - \frac{T^t(x,y)}{\pi\{y\}} &= 1 - \frac{P_x\{X_t = y\}}{\pi\{y\}} \leq 1 - \frac{P_x\{X_t = y, \tau \leq t\}}{\pi\{y\}} \\ &= 1 - \frac{\pi\{y\} P_x\{\tau \leq t\}}{\pi\{y\}} = 1 - P_x\{\tau \leq t\} = P_x\{\tau > t\}. \end{aligned}$$

□

Theorem 7. *For $t = 0, 1, \ldots$, write $\nu_t(A) = P\{X_t \in A\}$, then $||\nu_t - \pi||_{TV} \leq s_t$.*

Proof. Using identity (3), we have that

$$\begin{aligned} ||\nu_t - \pi||_{TV} &= \sum_{y \in S} (\pi\{y\} - \nu_t\{y\}) I_{\{y; \pi\{y\} > \nu_t\{y\}\}} = \sum_{y \in S} \pi\{y\} \left(1 - \frac{\nu_t\{y\}}{\pi\{y\}}\right) I_{\{y; \pi\{y\} > \nu_t\{y\}\}} \\ &= \sum_{y \in S} \pi\{y\} \sum_{x \in S} \nu_0\{x\} \left(1 - \frac{T^t(x,y)}{\pi\{y\}}\right) I_{\{\pi\{y\} > \nu_t\{y\}\}} \\ &\leq \sum_{y \in S} \pi\{y\} \sum_{x \in S} \nu_0\{x\} s_t = s_t. \end{aligned}$$

□

Now, we combine the proposition and the theorem.

Corollary 8. *Let τ be a strong stationary time for X, then*

$$||\nu_t - \pi||_{TV} \leq \max_{x \in S} P_x\{\tau > t\}.$$

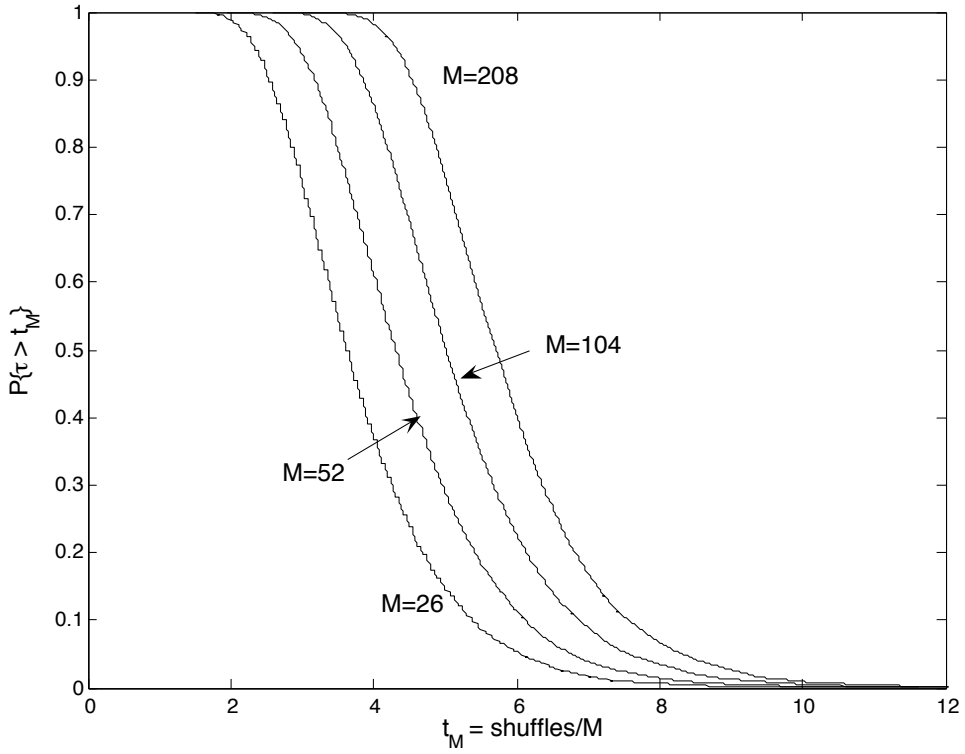

Figure 1: Reading from left to right, the plot of the bound given by the strong stationary time of the total variation distance to the uniform distribution at time $t_M = \text{shuffles}/M$ for a deck of size $M = 26, 52, 104,$ and 208 based on 10,000 simulations of the stopping time τ.

The strong stopping time in Corollary 2 is independent of the intial ordering of cards, thus we have an upper bound on the total variation distance of cards in the deck after t shuffles and the uniform probability measure. $P_x\{\tau > t\}$ is simulated in Figure 1. Note that the horizontal axis is the number of shuffles divided by the number of cards in the deck. The constant linear shift with a doubling of the size of the deck indicates an $M \log M$ dependence on the distribution. Diaconis and Aldous (1986) show that for $c \geq 0$, $P\{\tau > M(\log M + c)\} \leq e^{-c}$.

2. The Ewens Sampling Formula. For more that three decades, the *Ewens sampling formula* has been a major cornerstone in tests of a neutral assumption of evolution. (See Ewens, 1972, Karlin and McGregor, 1972, Ewens, 2000, Durrett, 2002, Tavaré, 2004, Griffiths and Lessard, 2005.) This contribution came soon after the realization through data obtained by electrophoretic methods of the high variability of genetic types and Kimura's (1968) introduction of his neutral theory of molecular evolution. (See also Slatkin, 1994, 1996.) The advancement in the theory of population genetics in 1978 via Kingman's paintbox construction, in 1980 via Griffith's lines of descent process, and in 1982 via Kingman's coalescent led to a variety of new perspectives on and extensions of the sampling formula.

The Ewens sampling formula applies to situations in which

- the sample size m is small compared to the constant haploid population size, M,
- each mutation gives rise to a novel haplotype (*the infinite alleles model*),
- each mutation is selectively neutral, taking place from one generation to the next with probability μ, and
- the population is in equilibrium.

(See Cannings, 1974 and Ethier and Kurtz, 1981.)

Most of our attention will be placed on the last of these four requirements. How long must the population be in place before can we say that the configuration of the population is near equilibrium? The point of view has been taken by a variety of authors, e.g., Griffiths (1979), Watterson (1984), Donnelly (1986), and Donnelly and Tavaré, (1986) in their investigations of ages of alleles and lines of descent. The specific goal here is to use the ideas developed in the introduction to determine a good bound on the total variation distance between the time t distribution of a sample and the Ewens sampling formula. Along the way, we shall encounter a few of the the now classical results of the authors cited above.

The sampling formula gives the distribution of a sample taken from the assumed equilibrium *frequency distribution* or *configuration* of the population. To be precise, let $\mathbf{e}_i$ denote the configuration consisting of a collection of i genetically indistinguishable individuals (called a *haplotype*) and let c_i denote the number of distinct haplotypes represented i times in the population. Then, the permissible configurations are those

$$\mathbf{c} = \sum_i c_i \mathbf{e}_i, \quad \text{for which} \quad \sum_i i c_i = M.$$

Letting b_i denote the number of haplotypes represented i times in a *sample* of m individuals, we have

$$\mathbf{b} = \sum_i b_i \mathbf{e}_i, \quad \text{for which} \quad \sum_i i b_i = m$$

as possible sampling configurations.

The Ewens sampling formula $ESF(m, \theta)$ states that probability that the sample has configuration $\mathbf{b}$ is

$$\frac{m!}{\theta(\theta+1)\cdots(\theta+m-1)} \prod_i \left(\frac{\theta}{i}\right)^{b_i} \frac{1}{b_i!}. \tag{6}$$

The choice of θ is generally approximately equal to $M\mu$.

One simple way to reproduce the Ewens sampling formula is via Hoppe's (1984) urn model. To an urn containing n colored balls, add balls to the urn one by one according to the following rule.

- With probability $\theta/(\theta+n)$, a ball having a new color is added to the urn.
- With probability $n/(\theta+n)$, a ball is chosen at random and returned to the urn. An additional ball of the same color is also placed in the urn.

Then when the urn contains m balls, the configuration has an $ESF(m,\theta)$ distribution. To construct a proof by induction, fix a configuration $\mathbf{b}$ of m balls and consider the configurations of $m-1$ balls that could result in configuration $\mathbf{b}$ upon the addition of a ball. (See Durrett, 2002, Chapter 1, Section 3) Also, note that if the balls have an $ESF(m,\theta)$ distribution and a ball is removed at random, then the remaining $m-1$ balls have an $ESF(m-1,\theta)$ distribution, Consequently, a random sample of balls chosen from configurations distributed according to the Ewens sampling formula itself has the Ewens sampling formula as its distribution.

In a manner similar to Donnelly and Tavaré (1986) the results here are based on a discrete time infinite alleles Moran model. Donnelly and Rodrigues (2000) took the ideas developed in the Introduction and applied them to a continuous time two allele Moran model.

3. The New Alleles Process. The infinite alleles Moran model can be described as follows. Fill an urn with M balls in a variety of colors (called *old alleles*) and consider the following Markov chain U. The graphical representation of a realization of this process is given in Figure 2.

- Choose a ball at random to make the replacement. (the tails of the arrows)
- Choose a second ball *without replacement* to be removed from the population. (the heads of the arrows)
  - With probability $1-\mu$, replace the removed ball with a ball identical to the first ball chosen.
  - With probability μ, replace the removed ball with a ball *striped* with a new color (called *a new allele*). This is displayed by an $\times$ on the arrow in the graphical representation.

Because new colors continue to appear, U is not recurrent. C_t^U, the configuration of U_t, is also a Markov chain. C_t^U has a finite state space, is irreducible, aperiodic, and, thus, positive recurrent. As shown by Kelly (1977 and 1979, Chapter 7), C_t^U has a unique stationary probability measure, $\pi_{M,\mu}^C$. As we shall see, this distribution is $ESF(M,\theta)$, $\theta=(M-1)\mu/(1-\mu)$.

Let N_t denote the number of *striped balls* at time t. Then, $N_0=0$. In the graphical representation, the position at the head of arrow inherits the type at the tail unless a mutation occurs. The changes in the number of striped balls based on the two balls drawn in summarized in Table 1.

Table 1. Transitions in the Markov chain N.

	reproduce, remove			
mutate	S,S	S,U	U,S	U,U
no	0	+1	−1	0
yes	0	+1	0	+1

S - striped ball, U - unstriped ball

Consequently, N is a particular kind of Markov chain, a *birth-death* chain, with transition matrix

$$T_N(m,n)=\begin{cases}\beta_N(m) & \text{if } n=m+1,\\ 1-\beta_N(m)-\delta_N(m) & \text{if } n=m,\\ \delta_N(m) & \text{if } n=m-1.\end{cases}$$

In addition, if the urn has m striped balls, then

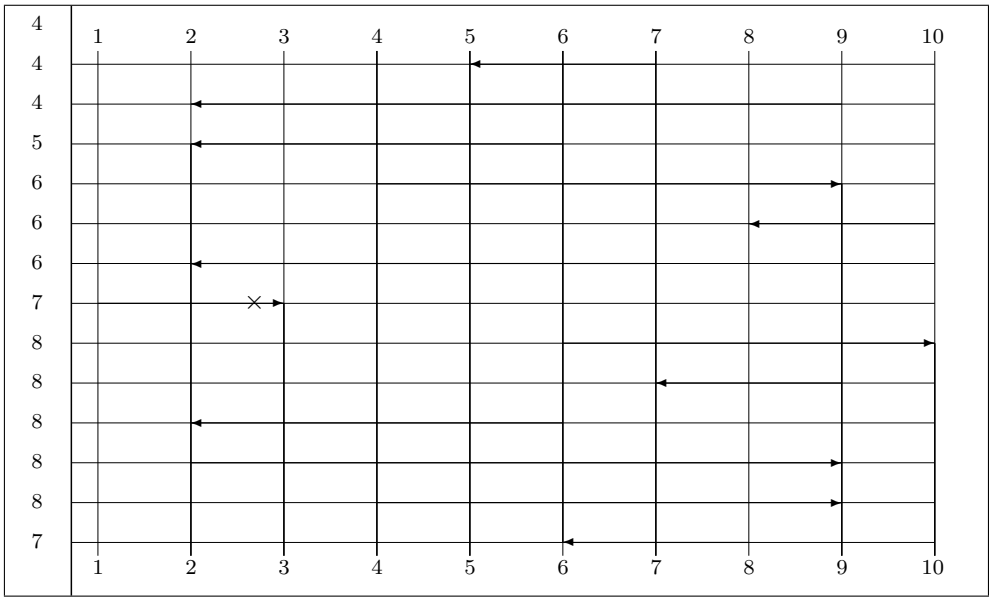

Figure 2: The graphical representation of fourteen time steps of a realization of the urn model process U with $M = 10$ and the evolution moving downward. Balls numbered 4, 5, 6, and 7, are striped at the beginning of the time course displayed. The dark lines indicate balls having new alleles and the numbers on the left give the state of the new allelles process N. The $\times$ shows a mutation that gives rise to a striped ball in column 3.

the choice	has probability
S, S	$\frac{m}{M} \cdot \frac{m-1}{M-1}$
S, U	$\frac{m}{M} \cdot (1 - \frac{m-1}{M-1}) = \frac{m}{M} \cdot \frac{M-m}{M-1}$
U, S	$(1 - \frac{m}{M}) \cdot \frac{m}{M-1} = \frac{M-m}{M} \frac{m}{M-1}$
U, U	$(1 - \frac{m}{M}) \cdot (1 - \frac{m}{M-1}) = \frac{M-m}{M} \cdot \frac{M-m-1}{M-1}$

(7)

Combine Table 1 and (7) to obtain

$$\beta_N(m) = \frac{M-m}{M(M-1)}(m + (M-m-1)\mu), \quad \delta_N(m) = \frac{M-m}{M(M-1)} m(1-\mu).$$

N is a discrete time two allele Moran model with mutations from the unstriped allele type to striped allele type having probability μ and no mutations permitted from striped alleles to unstriped alleles.

Theorem 9. *Let $\mathcal{F}_t$ be the σ-algebra generated by the urn process random variables $U_0, U_1, \ldots, U_t$ and define $\tau_N = \inf\{t \geq 0 : N_t = M\}$, the first time the urn consists entirely of striped balls. Then, τ_N is a strong stationary time.*

Proof. The expression $\{\tau_N \leq t\} \in \mathcal{F}_t$ merely states that by watching the urn process U up to time step t, we can decide whether or not the urn has ever been filled solely with striped balls. Thus, property 1 in the definition of a strong stationary time is satisfied.

The following claim will establish properties 2 and 3.

Claim. Whenever the urn has n striped balls, then the configuration of these balls has an $ESF(n, \theta)$ distribution, with $\theta = (M-1)\mu/(1-\mu)$.

The claim certainly holds when $n = 1$.

Case 1. The number of striped balls decreases by 1.

In this case, a striped ball is removed at random and an unstriped ball is added. Thus, the distribution of the remaining configuration of striped balls is $ESF(n-1, \theta)$.

Case 2. The number of striped balls increases by 1.

In this case, an unstriped ball is removed and a striped ball is added. The probability that the additional striped ball has a new color resulting from a mutation,

$$\begin{aligned} P\{\text{new striped color}|\text{striped ball added}\} &= \frac{\left(\frac{n}{M}\cdot\frac{M-n}{M-1} + \frac{M-n}{M}\cdot\frac{M-n-1}{M-1}\right)\mu}{\beta_N(n)} = \frac{\frac{M-n}{M(M-1)}(M-1)\mu}{\beta_N(n)} \\ &= \frac{(M-1)\mu}{n+(M-n-1)\mu} = \frac{(M-1)\mu}{(1-\mu)n+(M-1)\mu} = \frac{\theta}{n+\theta}. \end{aligned}$$

Thus the type of ball added follows the rules for Hoppe's urn and the configuration of striped balls has an $ESF(n+1, \theta)$ distribution.

Case 3. The number of striped balls remains constant.

If an unstriped ball is replaced with an unstriped ball, then the distribution of the configuration of striped balls remains $ESF(n, \theta)$.

If a striped ball is replaced by a striped ball, then, as in case 1, after the striped ball is removed, the urn has configurations that follow an $ESF(n-1, \theta)$ distribution.

$$P\{\text{new striped color}|\text{striped ball removed and added}\}$$

$$= \frac{\left(\frac{n}{M}\cdot\frac{n-1}{M-1} + \frac{M-n}{M}\cdot\frac{n}{M-1}\right)\mu}{\frac{n}{M}\cdot\frac{n-1}{M-1} + \frac{M-n}{M}\cdot\frac{n}{M-1}\mu} = \frac{n(n-1+M-n)\mu}{n(n-1+M\mu-n\mu)} = \frac{(M-1)\mu}{(n-1)(1-\mu)+(M-1)\mu} = \frac{\theta}{n-1+\theta}.$$

Again the rules for Hoppe's urn are satisfied and the configuration of striped balls has an $ESF(n, \theta)$ distribution.

Consequently, at time τ_M, independent of the value of τ_M, the configuration of the M striped balls has the $ESF(M, \theta)$. □

In addition, by Corollary 8, if we denote the distribution of configurations of the urn at time step t by ν_t^C, then

$$||\nu_t^C - \pi_{M,\mu}^C||_{TV} \leq \max_n P_n\{\tau_N > t\} = P_0\{\tau_N > t\} \tag{8}$$

Figure 3 displays the same basic behavior on convergence to stationarity for the infinite alleles Moran model as was seen previously in the top to random card shuffle. The horizontal axis measures time in

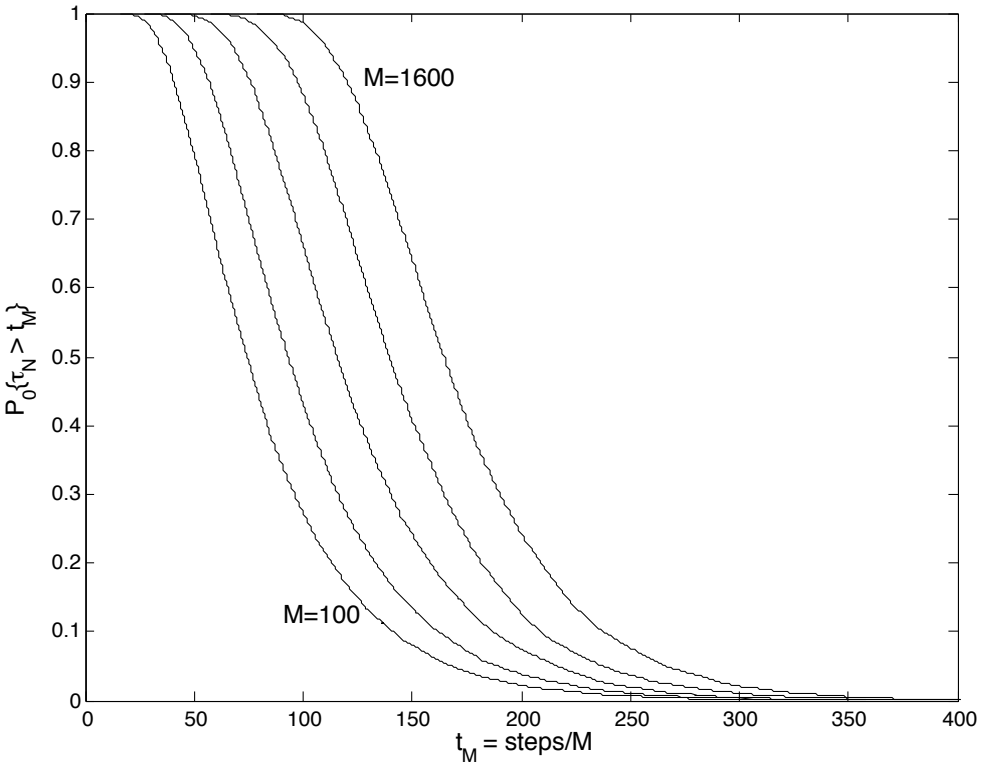

Figure 3: Reading from left to right, the plot $P_0\{\tau_N > t_M\}$ with time $t_M = \text{steps}/M$, measured in generations, for $\mu = 0.025$ and $M = 100, 200, 400, 800$, and 1600 based on 10,000 simulations of the stopping time τ_N.

generations. One generation is the number, M, of time steps in which each individual, on average, reproduces once.

We would like to obtain an analytical expression that will yield the curves in Figure 3. Because the structure of eigenvalues, eigenmeasures, and eigenfunction is difficult to determine for T_N, we will have difficulty in accomplishing this task if we confine ourselves to an analysis of this transition matrix. Following a strategy analagous to Donnelly and Rodrigues (2000), we examine a process dual to the new alleles process.

4. Lines of Descent. Now consider a realization of the entire genetic history of the process introduced above. The dynamics describing the number of striped balls and the configuration of these balls takes a forward in time approach. A backward in time *dual* process introduced by Griffiths (1980) is called the *lines of descent*. To describe this process, begin a sample of size m from the present population and look back at the balls in the urn t time steps in the past. D_t is defined to be the number of balls that are direct ancestors which are identical to those in the present day sample.

The value of introducing the dual process is demonstrated by the following theorem.

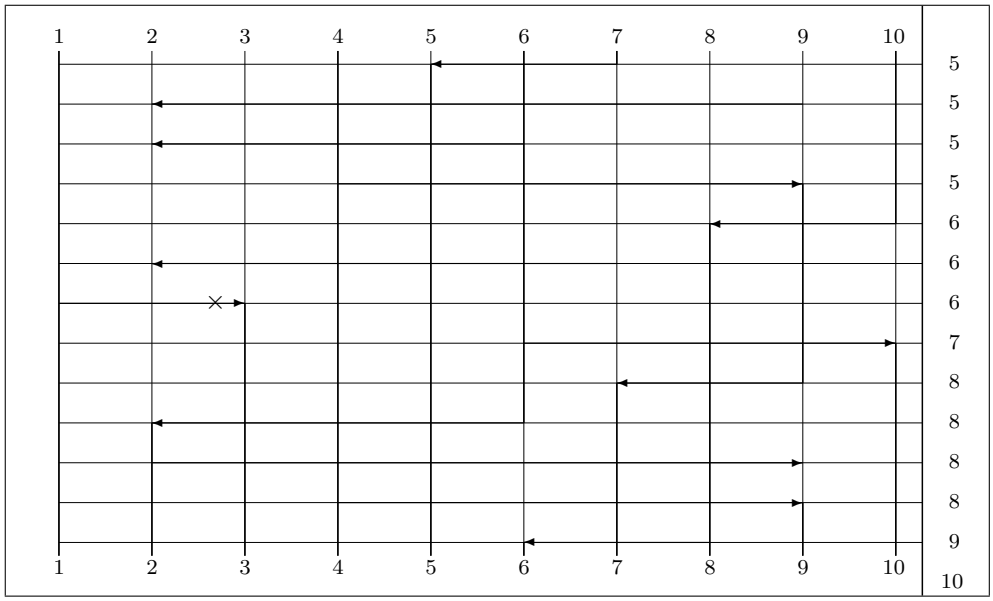

Figure 4: The graphical representation of fourteen time steps of the realization of the urn model process U shown in Figure 2 with $M = 10$. Assuming 10 lines of descent in the present, the numbers on the right give the state of D, the lines of descent process. The $\times$ shows a mutation that gives rise to a striped ball in column 3 which, looking back in time, terminates a line of descent.

Theorem 10. *Define* $\tau_D = \min\{t \geq 0 : D_t = 0\}$. *If* $N_0 = 0$ *and* $D_0 = M$, *then* τ_D *and* τ_N *have the same distribution.*

Proof. Consider the probability space that consists of the graphical representations described above and illustrated in Figures 2 and 4. That is, independently, for each *integer* t, choose a number from 1 to M at random for the head of the arrow and choose a second number, without replacement, for the tail. On the left side of the grid, place numbers, increasing moving downward, representing the time steps for the new alleles process. On the right side, independent of the placement of the arrows, place 0 across from some positive value, t, for the new alleles process and place numbers for the lines of descent process increasing moving upward.

Consider t_0 time steps of the new alleles process and focus on realizations of the graphical representation that have t_0 across from 0. We will prove the theorem by showing in these realizations $\tau_N > t_0$ if and only if $\tau_D > t_0$.

With $N_0 = 0$ and $D_0 = M$, if $\tau_N > t_0$, then, $N_{t_0} < M$. Consequently, one of the original old alleles survives from time 0 to time t_0. Thus, at least one line of descent survives from t_0 time steps in the past. In other words, $D_{t_0} > 0$ and $\tau_D > t$.

On the other hand, if $\tau_N \leq t_0$, then $N_{t_0} = M$. Thus, all of the old alleles have mutated and looking back in time, none of the lines of descent remain for t_0 time steps. Consequently, $D_{t_0} = 0$ and $\tau_D \leq t_0$. □

The next step is to compute $P\{D_t = 0\} = P\{\tau_D \leq t\}$. Then, by Corollary 8 and Theorem 10, we have an upper bound for the desired total variation distance $||\nu_t - \pi^C_{M,\mu}||_{TV}$.

As we look back one time step, (upwards in Figure 4) the number of lines of descent can decrease by one in one of two ways:

- A ball in the sample resulted from a mutation from the previous generation (indicated by $\times$).
- Two balls in the sample are descended from one in the previous generation.

The changes in the number of lines of descent based on the two balls drawn in summarized in Table 2.

Table 2. Transitions in the Markov chain D.

	parent, offspring			
mutate	S,S	S,U	U,S	U,U
no	-1	0	0	0
yes	-1	0	-1	0

S - sampled, U - unsampled

Thus, D is a *pure-death* chain with transition matrix

$$T_D(m,n) = \begin{cases} 1-\delta_N(m) & \text{if } n=m, \\ \delta_N(m) & \text{if } n=m-1. \end{cases}$$

Returning to the probabilities in (7) and using Table 2, we find that

$$\begin{aligned} \delta_D(m) &= \frac{m}{M}\cdot\frac{m-1}{M-1} + \frac{M-m}{M}\cdot\frac{m}{M-1}\mu = \frac{m}{M(M-1)}(m-1+(M-m)\mu) \\ &= \frac{m}{M(M-1)}((m-1)(1-\mu)+(M-1)\mu) = \frac{m(m-1+\theta)}{M(M-1+\theta)} \end{aligned} \tag{9}$$

because $M-1+\theta = (M-1)/(1-\mu)$.

To determine the (right) eigenfunctions $T_D f = \lambda f$, (See, for example, the appendix in Tavaré, 1984.)

$$T_D(m,m)f(m) + T_D(m,m-1)f(m-1) = \lambda f(m), \quad \text{or} \quad \delta_D(m)f(m-1) = (\lambda - 1 + \delta_D(m))f(m). \tag{10}$$

Fix $\tilde{n}$, then

$$\lambda_{\tilde{n}} = 1 - \delta_D(\tilde{n}) \tag{11}$$

is an eigenvalue. If we set $f^{\tilde{n}}(\tilde{n}) = 1$, then by iterating equation (10), we obtain

$$f^{\tilde{n}}(m) = \begin{cases} 0 & \text{if } m < \tilde{n}, \\ 1 & \text{if } m = \tilde{n}, \\ \prod_{j=\tilde{n}+1}^{m} \frac{\delta_D(j)}{\delta_D(j)-\delta_D(\tilde{n})} & \text{if } m > \tilde{n}. \end{cases} \tag{12}$$

To determine the (left) eigenmeasures $\nu T_D = \lambda\nu$,

$$\nu\{n+1\}T_D(n+1,n) + \nu\{n\}T_D(n,n) = \lambda\nu\{n\}, \quad \text{or} \quad \nu\{n+1\}\delta_D(n+1) = \nu\{n\}(\lambda - 1 + \delta_D(n)). \tag{13}$$

Note that $\lambda_{\tilde{n}} = 1 - \delta_D(\tilde{n})$ is again an eigenvalue. If we set $\nu^{\tilde{n}}\{\tilde{n}\} = 1$, then by iterating equation (13), we obtain

$$\nu^{\tilde{n}}\{n\} = \begin{cases} \prod_{k=n}^{\tilde{n}-1} \frac{\delta_D(k+1)}{\delta_D(k)-\delta_D(\tilde{n})} & \text{if } n < \tilde{n}, \\ 1 & \text{if } n = \tilde{n}, \\ 0 & \text{if } n > \tilde{n}. \end{cases} \tag{14}$$

Let ϵ^n be the probability measure that puts mass 1 at n and zero mass elsewhere. Then, we can write $\nu^{\tilde{n}} = \sum_{n=0}^{\tilde{n}-1} \nu^{\tilde{n}}\{n\}\epsilon^n + \epsilon^{\tilde{n}}$. Consequently, by following the usual row reduction operations for a lower triangular matrix with 1's on the diagonal, we see that

$$\text{span}\{\nu^{\tilde{n}}; 0 \le \tilde{n} \le M\} = \text{span}\{\epsilon^{\tilde{n}}; 0 \le \tilde{n} \le M\}.$$

Note that $\lambda_{\tilde{n}}\nu^{\tilde{n}} \cdot f^{\tilde{m}} = \nu^{\tilde{n}} T_D f^{\tilde{m}} = \lambda_{\tilde{m}}\nu^{\tilde{n}} \cdot f^{\tilde{m}}$. For $\tilde{n} \neq \tilde{m}$, $\lambda_{\tilde{n}} \neq \lambda_{\tilde{m}}$ and consequently, $\nu^{\tilde{n}} \cdot f^{\tilde{m}} = 0$. In addition, from the formulas (12) and (14), $\nu^{\tilde{m}} \cdot f^{\tilde{m}} = 1$.

Proposition 11. *For each $t \ge 0$,*

$$T_D^t(m,n) = \sum_{\tilde{n}} \lambda_{\tilde{n}}^t f^{\tilde{n}}(m)\nu^{\tilde{n}}\{n\}. \tag{15}$$

Proof. For $t = 0$, it suffices to check the action of T_D^0 on elements of the spanning set $\{\nu^{\tilde{m}}; 0 \le \tilde{m} \le M\}$.

$$\sum_m \nu^{\tilde{m}}\{m\} \sum_{\tilde{n}} f^{\tilde{n}}(m)\nu^{\tilde{n}}\{n\} = \sum_{\tilde{n}} \left(\sum_m \nu^{\tilde{m}}\{m\} f^{\tilde{n}}(m) \right) \nu^{\tilde{n}}\{n\} = \sum_{\tilde{n}} \epsilon(\tilde{m},\tilde{n})\nu^{\tilde{n}}\{n\} = \nu^{\tilde{m}}\{n\}$$

yielding the identity matrix. Here $\epsilon(\tilde{m},\tilde{n}) = 1$ if $\tilde{n} = \tilde{m}$ and 0 otherwise.

For the induction step,

$$\begin{aligned} T_D^{t+1}(m,n) &= \sum_k T_D^t(m,k)T_D(k,n) = \sum_k \left(\sum_{\tilde{n}} \lambda_{\tilde{n}}^t f^{\tilde{n}}(m)\nu^{\tilde{n}}\{k\} \right) \left(\sum_{\tilde{m}} \lambda_{\tilde{m}} f^{\tilde{m}}(k)\nu^{\tilde{m}}\{n\} \right) \\ &= \sum_{\tilde{n}} \lambda_{\tilde{n}}^t f^{\tilde{n}}(m) \sum_{\tilde{m}} \lambda_{\tilde{m}}\nu^{\tilde{m}}\{n\} \left(\sum_k \nu^{\tilde{n}}\{k\} f^{\tilde{m}}(k) \right) = \sum_{\tilde{n}} \lambda_{\tilde{n}}^t f^{\tilde{n}}(m) \sum_{\tilde{m}} \lambda_{\tilde{m}}\nu^{\tilde{m}}\{n\}\epsilon(\tilde{m},\tilde{n}) \\ &= \sum_{\tilde{n}} \lambda_{\tilde{n}}^{t+1} f^{\tilde{n}}(m)\nu^{\tilde{n}}(n). \end{aligned}$$

□

Corollary 12. *The multistep transition probabilities for the lines of descent process*

$$T_D^t(m,n) = \begin{cases} 0 & \text{if } m < n, \\ \left(1 - \frac{n(n-1+\theta)}{M(M-1+\theta)}\right)^t & \text{if } m = n, \\ \sum_{\tilde{n}=n}^m \left(1 - \frac{\tilde{n}(\tilde{n}-1+\theta)}{M(M-1+\theta)}\right)^t \binom{m}{\tilde{n}} \prod_{j=\tilde{n}+1}^m \frac{j-1+\theta}{j+\tilde{n}-1+\theta} \cdot (-1)^{\tilde{n}-n} \binom{\tilde{n}}{n} \prod_{k=n}^{\tilde{n}-1} \frac{k+\theta}{k+\tilde{n}-1+\theta} & \text{if } m > n. \end{cases}$$

Proof. Note that $f^{\tilde{n}}(m)\nu^{\tilde{n}}\{n\} = 0$ unless $n \le \tilde{n} \le m$. Using the formula for δ_D given in equation (9), we have

$$\delta_D(j) - \delta_D(\tilde{n}) = \frac{j(j-1+\theta) - \tilde{n}(\tilde{n}-1+\theta)}{M(M-1+\theta)} = \frac{(j-\tilde{n})(j+\tilde{n}-1+\theta)}{M(M-1+\theta)}.$$

Thus, substituting into equation (12), we find that

$$f^{\tilde{n}}(m) = \prod_{j=\tilde{n}+1}^{m} \frac{j}{j-\tilde{n}} \cdot \prod_{j=\tilde{n}+1}^{m} \frac{j-1+\theta}{j+\tilde{n}-1+\theta} = \frac{(\tilde{n}+1)\cdots m}{1\cdots(m-\tilde{n})} \cdot \prod_{j=\tilde{n}+1}^{m} \frac{j-1+\theta}{j+\tilde{n}-1+\theta} \tag{16}$$

Similarly, substituting into equation (14), we find that

$$\nu^{\tilde{n}}\{n\} = \prod_{k=n}^{\tilde{n}-1} \frac{k+1}{k-\tilde{n}} \cdot \prod_{k=n}^{\tilde{n}-1} \frac{k+\theta}{k+\tilde{n}-1+\theta} = (-1)^{\tilde{n}-n} \frac{(n+1)\cdots\tilde{n}}{(\tilde{n}-n)\cdots 1} \cdot \prod_{k=n}^{\tilde{n}-1} \frac{k+\theta}{k+\tilde{n}-1+\theta}. \tag{17}$$

Now, substitute these two expressions into equation (15). The eigenvalues are given by (11). □

Now we have a formula for $P_M\{\tau_D \le t\} = P_M\{D_t = 0\} = T_D^t(M,0)$

5. Sampling. Take a sample of m balls from the urn at time step t. Given that all of the balls are striped, we have shown in the proof of Theorem 9 that the distribution, $\pi_{m,\mu}^C$ of the sample will be $ESF(m,\theta)$, with $\theta = (M-1)\mu/(1-\mu)$. Thus, from a statistical point of view, we do not need all of the balls to be striped, only the balls that are to be sampled.

Let $\nu_t^{m,C}$ be the distribution of the configuration of m balls sampled at time t and consider the following two samples. Draw a ball at random. If it is striped, include it in both samples. If it is unstriped, add it only to the first sample. After m draws, the first sample, which we denote ξ, is complete. Its configuration, $C(\xi)$, has distribution $\nu_t^{m,C}$. Now continue drawing until the second sample, $\tilde{\xi}$ has m striped balls. Its configuration $C(\tilde{\xi})$, has distribution $\pi_{m,\mu}^C$. Thus, $(C(\xi), C(\tilde{\xi}))$ is a coupling of $\nu_t^{m,C}$ and $\pi_{m,\mu}^C$

If $N_t = n$, the two samples are the same precisely when all of the first m draws are are from the n striped balls. This happens with probablity

$$\binom{n}{m} \Big/ \binom{M}{m} = \frac{n(n-1)\cdots(m-m+1)}{M(M-1)\cdots(M-m+1)}.$$

Thus, by Theorem 2,

$$||\nu_t^{m,C} - \pi_{m,\mu}^C||_{TV} \le 1 - P_0\{C(\xi) = C(\tilde{\xi})\} \le 1 - P_0\{\xi = \tilde{\xi}\} = 1 - \sum_{n=m}^{M} \frac{n(n-1)\cdots(n-m+1)}{M(M-1)\cdots(M-m+1)} P_0\{N_t = n\}. \tag{18}$$

Note that is the case $m = M$, this sum reduces to $P_0\{N_t = M\} = P_0\{\tau_N \le t\}$ and returns the previous formula (8).

From the point of view of the line of descent process, $\xi \ne \tilde{\xi}$ if and only if at least one among the lines of descent of the m sampled balls extends back more than t time steps. Thus

$$||\nu_t^{m,C} - \pi_{m,\mu}^C|| \le P_m\{\tau_D > t\}.$$

As Figure 5 shows, the impact to the sample size on total variation distance is small.

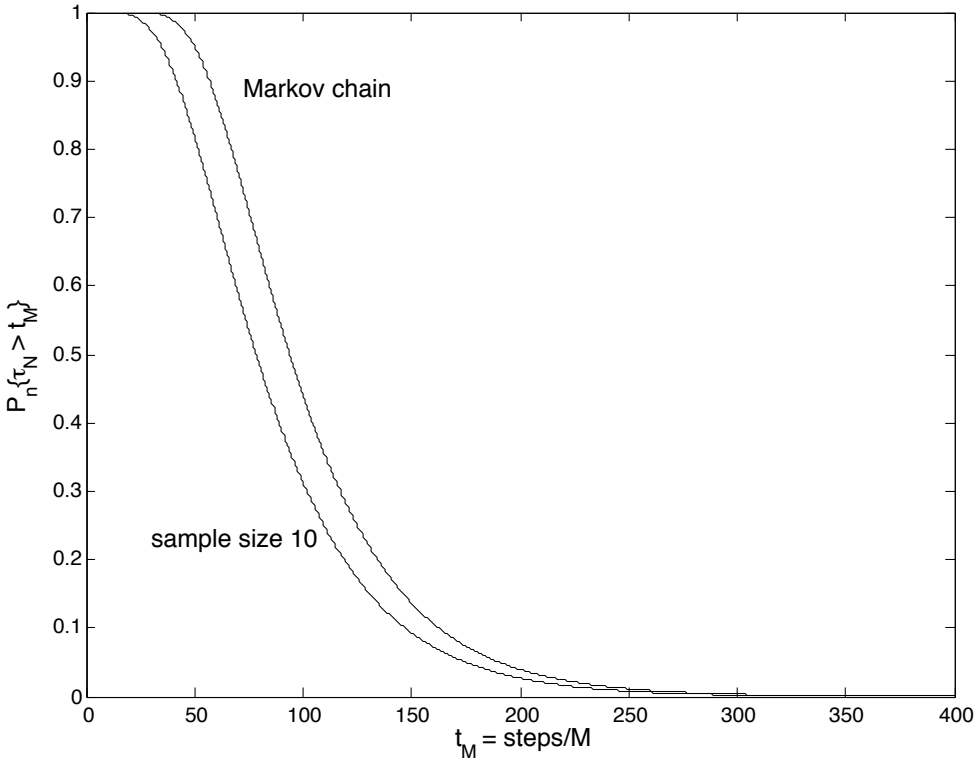

Figure 5: Comparison of $P_{200}\{\tau_D > t_M\}$ and $P_{10}\{\tau_D > t_M\}$ with time measured in generations, for $\mu = 0.025$ and $M = 200$.

6. Cutoff Phenenoma. Diaconis (1996) defines a cutoff phenomena as follows:

Definition 13. *Let $\{T_{(M)}; M \geq 1\}$ be a sequence of Markov chains transition matrices having π_M as their respective stationary distributions. Pick sequences A_M and B_M tending to ∞ as $M \to \infty$ whose ratio B_M/A_M converges to 0. Set $t_M(g)$ to be the integer part of $A_M + gB_M$. Then, the sequence of Markov chains is said to possess a* cutoff phenomena *if*

$$\lim_{M\to\infty} \sup_x ||T_{(M)}^{t_M(g)}(x,\cdot) - \pi_M||_{TV} = c(g)$$

where $\lim_{g\to\infty} c(g) = 0$ *and* $\lim_{g\to-\infty} c(g) = 1$.

To make a start in determining the sequences A_M and B_M given in the definition, note that if $M \geq 2\tilde{n}$, then from (16), we have that

$$f^{\tilde{n}}(M) = \binom{M}{\tilde{n}} \frac{(\tilde{n}+\theta)\cdots(M-1+\theta)}{(2\tilde{n}+\theta)\cdots(M+\tilde{n}-1+\theta)} = \binom{M}{\tilde{n}} \frac{(\tilde{n}+\theta)\cdots(2\tilde{n}-1+\theta)}{(M+\theta)\cdots(M+\tilde{n}-1+\theta)} = \binom{M}{\tilde{n}} \prod_{j=0}^{\tilde{n}-1} \frac{\tilde{n}+j+\theta}{M+j+\theta}. \tag{19}$$

In addition, for the product term in (19)

$$\lim_{M\to\infty}\prod_{j=0}^{\tilde{n}-1}\frac{\tilde{n}+j+\theta}{M+j+\theta}=\lim_{M\to\infty}\prod_{j=0}^{\tilde{n}-1}\frac{(\tilde{n}+j)(1-\mu)+(M-1)\mu}{(M+j)(1-\mu)+(M-1)\mu}=\lim_{M\to\infty}\prod_{j=0}^{\tilde{n}-1}\frac{(\tilde{n}+j)(1-\mu)+(M-1)\mu}{(j+1)(1-\mu)+(M-1)}=\mu^{\tilde{n}}. \tag{20}$$

Also, using (17),

$$\nu^{\tilde{n}}\{0\}=(-1)^{\tilde{n}}\prod_{k=0}^{\tilde{n}-1}\frac{k/\theta+1}{(k+\tilde{n}-1)/\theta+1}.$$

Consequently, for each $\tilde{n}$,

$$\lim_{M\to\infty}\nu^{\tilde{n}}\{0\}=(-1)^{\tilde{n}}. \tag{21}$$

The eigenvalue term

$$\lambda_{\tilde{n}}^{t}=\left(1-\frac{\tilde{n}(\tilde{n}-1+\theta)}{M(M-1+\theta)}\right)^{t}\approx\exp\left(-\frac{\tilde{n}(\tilde{n}-1+\theta)t}{M(M-1+\theta)}\right)=\exp\left(-\frac{\tilde{n}((\tilde{n}-1)(1-\mu)+(M-1)\mu)t}{M(M-1)}\right). \tag{22}$$

Looking at the approximation of the term $\tilde{n}=1$ involving the second leading eigenvalue we want to set $t_M(g)$ so that

$$|\lambda_1^{t_M(g)}f^1(M)\nu^1\{0\}|\approx\exp\left(-\frac{\mu t_M(g)}{M}\right)M\mu\approx\exp(-g).$$

Solving we have the following:

Theorem 14. *Set $t_M(g)$ to be the integer part of $M(\log(M\mu)+g)/\mu$. Then*

$$\lim_{M\to\infty}P_M\{\tau_M>t_M(g)\}=1-\exp(-e^{-g}). \tag{23}$$

Proof. We shall work with $T^{t_M(g)}(M,0)=P_M\{\tau_M\le t_M(g)\}$. Using equations (19), (20), and (22), we find that

$$f^{\tilde{n}}(M)\approx\frac{(M\mu)^{\tilde{n}}}{\tilde{n}!},\quad \lambda_{\tilde{n}}^{t_M(g)}\approx\exp\left(-\frac{\tilde{n}(\tilde{n}-1)(1-\mu)(\log(M\mu)+g)/\mu}{M-1}-\tilde{n}g\right)/(M\mu)^{\tilde{n}}.$$

Now, combine with (21) to see that for $\tilde{n}\le M/2$,

$$\lim_{M\to\infty}\lambda_{\tilde{n}}^{t_M(g)}f^{\tilde{n}}(M)\nu^{\tilde{n}}\{0\}=\frac{(-1)^{\tilde{n}}e^{-\tilde{n}g}}{\tilde{n}!}. \tag{24}$$

For all $\tilde{n}\ge 0$, $|\nu^{\tilde{n}}\{0\}|\le 1$ and $f^{\tilde{n}}(M)\le M^{\tilde{n}}/\tilde{n}!$. Because $1-x\le\exp(-x)$,

$$\lambda_{\tilde{n}}^{t_M(g)}\le\exp\left(-\frac{\tilde{n}(\tilde{n}-1)(1-\mu)(\log(M\mu)+g)/\mu}{M-1}-\tilde{n}g\right)/(M\mu)^{\tilde{n}}\le\exp\left(-\tilde{n}g\right)/(M\mu)^{\tilde{n}}.$$

provided that $\log(M\mu)+g\ge 0$ or $M\ge\exp(-g)/\mu$.

Thus, for all $\tilde{n}\ge 0$

$$|\lambda_{\tilde{n}}^{t_M(g)}\nu^{\tilde{n}}\{0\}f^{\tilde{n}}(M)|\le\frac{(e^{-g}/\mu)^{\tilde{n}}}{\tilde{n}!}.$$

To establish (23), let $\epsilon > 0$ and choose $\tilde{m}$ so that

$$\left|\sum_{\tilde{n}=0}^{\tilde{m}} \frac{(-1)^{\tilde{n}} e^{-\tilde{n}g}}{\tilde{n}!} - \exp(-e^{-g})\right| < \frac{\epsilon}{3} \quad \text{and} \quad \sum_{\tilde{n}=\tilde{m}+1}^{M} |\lambda_{\tilde{n}}^{t_M(g)} \nu^{\tilde{n}}\{0\} f^{\tilde{n}}(M)| \leq \sum_{\tilde{n}=\tilde{m}+1}^{\infty} \frac{(e^{-g}/\mu)^{\tilde{n}}}{\tilde{n}!} < \frac{\epsilon}{3}.$$

Now, using the limit (24), choose $\tilde{M} > 2\tilde{m}$ so that $M > \tilde{M}$ implies that

$$\left|\sum_{\tilde{n}=0}^{\tilde{m}} \left(\lambda_{\tilde{n}}^{t_M(g)} f^{\tilde{n}}(M) \nu^{\tilde{n}}\{0\} - \frac{(-1)^{\tilde{n}} e^{-\tilde{n}g}}{\tilde{n}!}\right)\right| < \frac{\epsilon}{3}.$$

Consequently,

$$\begin{aligned} |T^{t_g}(M,0) - \exp(-e^{-g})| &\leq \left|\sum_{\tilde{n}=\tilde{m}+1}^{M} \lambda_{\tilde{n}}^{t_M(g)} \nu^{\tilde{n}}\{0\} f^{\tilde{n}}(M)\right| + \left|\sum_{\tilde{n}=0}^{\tilde{m}} \left(\lambda_{\tilde{n}}^{t_M(g)} f^{\tilde{n}}(M) \nu^{\tilde{n}}\{0\} - \frac{(-1)^{\tilde{n}} e^{-\tilde{n}g}}{\tilde{n}!}\right)\right| \\ &+ \left|\sum_{\tilde{n}=0}^{\tilde{m}} \frac{(-1)^{\tilde{n}} e^{-\tilde{n}g}}{\tilde{n}!} - \exp(-e^{-g})\right| < \epsilon \end{aligned}$$

and (23) follows. □

Remark 15. *For the infinite alleles Moran model and the Ewens sampling formula, we have a cutoff with* $A_M = M \log(M\mu)/\mu$, $B_M = M/\mu$ *and* $c(g) = 1 - \exp(-e^{-g})$. *The distribution function* $1 - c(g)$ *is called the* type 1 extreme value distribution for maxima *or* the standard Gumbel distribution. *To see one way that this distribution approximates the maximum of random variables, let* $\{\eta_n; n \geq 1\}$ *be independent exponential random variables with mean one. Then,* $P\{\eta_n \leq g\} = 1 - \exp(-g)$. *Thus,*

$$P\{\max_{1 \leq n \leq M} \eta_i - \log M \leq g\} = (1 - \exp(-g - \log M))^M = \left(1 - \frac{1}{M} e^{-g}\right)^M \to \exp(-e^{-g})$$

as $M \to \infty$

Transforming the limit in (23) into time t_M *in generations, we find that*

$$P_M\{\tau_D > t_M\} \approx 1 - \exp\left(-M\mu e^{-\mu t_M}\right). \tag{25}$$

If we use the Gumbel distribution estimate to choose a time $\tilde{t}_M(p)$ *to ensure that* $||\nu_{t_M} - \pi_{M,\mu}^C|| \leq p$, *then*

$$\tilde{t}_M(p) \geq -\frac{1}{\mu} \log\left(-\frac{1}{M\mu} \log(1-p)\right). \tag{26}$$

Acknowledgements. The author would like to acknowledge Steve Lansing whose questions about the time scale necessary for the use of the Ewens sampling formula led to this investigation and Brian Hallmark whose simulation results privided insights into the nature of the cutoff phenomenon. Finally, the author acknowledges the National Science Foundation support through its Human and Social Dynamics Program grant SES-0725470.

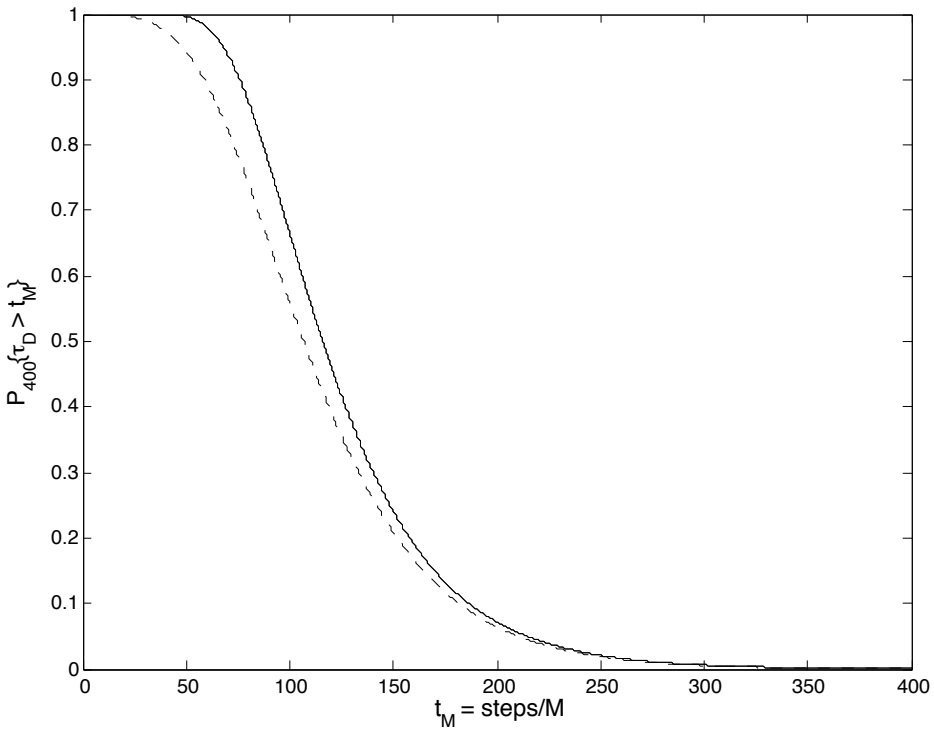

Figure 6: Comparison of $P_{400}\{\tau_D > t_M\}$ (solid line) and its approximating Gumbel distribution (dashed line) as given in (25). Time t_M is measured in generations, $\mu = 0.025$ and $M = 400$.

References

Aldous, D. 1983. Random walks on finite groups and rapidly mixing Markov chains. *Séminaire de Probabilités XVII*, *Lecture Notes in Mathematics*, **986**, 243-297.

Aldous, D. and Diaconis, P., 1986. Shuffling cards and stopping times. *Amer. Math. Monthly* **93**, 333-348.

Aldous, D. and Diaconis, P., 1987. Strong uniform times and finite random walks. *Adv. Appl. Math.* **8**, 69-97.

Bayer, B., and Diaconis, P., 1992. Trailing the dovetail shuffle to its lair. *Ann. Appl. Probab.* **2**, 294-313.

Cannings, C., 1974. The latent roots of certain Markov chains arising in genetics: A new approach. 1. Haploid models. *Adv. Appl. Probab.* **6**, 260-290.

Diaconis, P., 1988. *Group Representations in Probability and Statistics*. Institute of Mathematical Statistics, Lecture Series **11** Hayward, California.

Diaconis, P., 1996. The cutoff phenomena in finite Markov chains. *Proceedings of the National Academy of Sciences* **931**, 1659-1664.

Diaconis, P., and Shahshahani, M., 1987. Time to stationarity in the Bernouli-Laplace diffusion model. *SIAM J. Math. Anal.* **18**, 208-218.

Donnelly, P., 1986. Partition structures, Polya urns, the Ewens sampling formula, and the ages of alleles. *Theoretical Population Biology* **30**, 271-288.

Donnelly, P., and E. Rodrigues, 2000. Convergence to stationarity in the Moran model. *Adv. Appl. Probab.* **37**, 705-717.

Donnelly, P., and Tavaré, S., 1986. The ages of alleles and a coalescent. *Adv. Appl. Probab.* **18**, 1-19.

Durrett, R., 2002. *Probability Models for DNA Sequence Evolution*. Springer Verlag, New York.

Ethier, S. N., and Kurtz, T. G., 1981. The infinitely-many-neutral-alleles diffusion model. *Adv. Appl. Probab.* **13**, 429-452.

Ewens, W. J., 1972. The sampling theory of selectively neutral alleles. *Theoretical Population Biology* **3**, 87-112.

Ewens, W. J., 2000. *Mathematical Population Genetics. 1. Theoretical Introduction*. Springer Verlag, New York.

Griffiths, R. C., 1979. Exact sampling distributions form the infinite neutral alleles model. *Adv. Appl. Probab.* **11**, 326-354.

Griffiths, R. C., 1980. Lines of descent in the diffusion approximation of neutral Wright-Fisher Models. *Theoretical Population Biology* **17**, 37-50.

Griffiths, R. C. and Lessard S., 2005. Ewens' sampling formula and related formulae: combinatorial proofs, extensions to variable population size and applications to the age of alleles. *Theoretical Population Biology* **68**, 167-177.

Hoppe, F. M., 1984. Polya-like urns and the Ewens' sampling formula. *J. Math. Biol.* **20**, 91-94.

Karlin, S., and McGregor, J., 1972. Addendum to a paper of Ewens. *Theoretical Population Biology* **3**, 113-116.

Kelly, F., 1977. Exact results for the Moran neutral allele model. *Adv. Appl. Probab.* **9**, 197-201.

Kelly, F., 1979. *Reversibility and Stochastic Networks*. Wiley, Chichester.

Kimura, M., 1968. Evolutionary rate at the molecular level. *Nature* **217**, 624-626.

Kingman, J. F. C., 1978. Random partitions in population genetics. *Proc. R. Soc. London Ser. A* **361**, 1-20.

Kingman, J. F. C., 1982. The coalesent. *Stochastic Process. Appl.* **13**, 235-248.

Lansing, J. S., Watkins, J. C., Hallmark, B., Cox, M. P., Karafet, T. M., Sudoyo, H., and Hammer, M. F.,2007 Male dominance rarely yields evolutionary benefits. (preprint)

Levin, D. A., Peres, Y., and Wilmer, E. L., 2006. *Markov Chains and Mixing Times* (on-line manuscript http://www.oberlin.edu/markov/).

Lindvall, T., 1992. *Lectures on the Coupling Method*. Wiley, New York.

Pemantle, R. (1989) Randomization time for the overhand shuffle. *J. Theor. Prob.* **2**, 37-49.

Rosenthal, J. S. (1995) Convergence rates for Markov Chains. *SIAM Review* **37**, 387-405.

Saloff-Coste, L., 1997. Lectures on finite Markov chains. Lectures on Probability Theory and Statistics. École d'Été de Probabilité de Saint-Flour. Vol. XXVI. In: Bernard, P. (Ed.) *Lecture Notes in Mathematics* **1665**, 301-413.

Slatkin, M., 1994. An exact test for neutrality based on the Ewens sampling formula. *Genetical Research* **64** 71-74. (correction *Genetical Research* **68** 259-260.)

Tavaré, S., 1984. Lines of descent and genealogical processes and their applications in population genetics. *Theoretical Population Biology* **26**, 119-164.

Tavaré, S., 2004. Ancestral inference in population genetics. Lectures on Probability Theory and Statistics. École d'Été de Probabilité de Saint-Flour. Vol. XXXI. In: Picard, J. (Ed.) *Lecture Notes in Mathematics* **1837**, 1-188.

Watterson, G. A., 1984 Lines of descent and the coalescent. *Theoretical Population Biology* **26**, 77-92.