



HAL
open science

Fluctuations Analysis of finite discrete Birth and Death Chains with Emphasis on Moran Models with Mutations

Thierry Huillet

► **To cite this version:**

Thierry Huillet. Fluctuations Analysis of finite discrete Birth and Death Chains with Emphasis on Moran Models with Mutations. *Biomathematics*, 2013, 2013, pp.21. 10.1155/2013/939308 . hal-00843285

HAL Id: hal-00843285

<https://hal.science/hal-00843285>

Submitted on 11 Jul 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

FLUCTUATIONS ANALYSIS OF FINITE DISCRETE BIRTH AND DEATH CHAINS WITH EMPHASIS ON MORAN MODELS WITH MUTATIONS

THIERRY E. HUILLET

ABSTRACT. The Moran model is a discrete-time birth and death Markov chain describing the evolution of the number of type 1 alleles in an haploid population with two alleles whose total size N is preserved during the course of evolution. Bias mechanisms such as mutations or selection can affect its neutral dynamics. For the ergodic Moran model with mutations, we get interested into the fixation probabilities of a mutant, the growth rate of fluctuations, the first hitting time of the equilibrium state starting from state $\{0\}$, the first return time to the equilibrium state, the first hitting time of $\{N\}$ starting from $\{0\}$, together with the time needed for the walker to reach its invariant measure, again starting from $\{0\}$. For the last point, an appeal to the notion of Siegmund duality is necessary and a cutoff phenomenon will be made explicit. We are interested into these problems in the large population size limit $N \rightarrow \infty$. The Moran model with mutations includes the heat exchange models of Ehrenfest and Bernoulli-Laplace as particular cases; these were studied from the point of view of the controversy concerning irreversibility (H -theorem) and the recurrence of states.

Running title: Moran Models with Mutations.

Keywords: Moran Models; Mutational and evolutionary processes; Population dynamics; Fixation probabilities; Siegmund dual; hitting and first return times; time to stationarity.

1. INTRODUCTION

In Section 2, we start with generalities on discrete-time birth and death Markov chains with finite state-space, whose transition matrix is of Jacobi type: spectral representation, reversibility versus H -theorem, large deviations of the empirical average, invariant measure, conditions of transience/recurrence, spectral positivity and stochastic monotonicity, distribution of hitting and first return times, Green function. The scale or harmonic function is shown to be of interest to determine the excursions heights would the chain be ergodic. After having fixed the background, we then proceed with the study of specific Moran chains presenting various bias mechanisms.

The Moran chain model is a particular instance of a discrete-time birth and death Markov chain, describing the evolution of the number of type 1 alleles in an haploid population with two alleles whose total size N is preserved during the course of

evolution. Bias mechanisms such as mutations or selection can affect its neutral dynamics¹).

For the ergodic Moran model with mutations [29], we get interested (in Sections 3 and 5) into the fixation probabilities of a mutant, the growth of fluctuations, the first hitting time of the equilibrium state starting from $\{0\}$, the first return time to the equilibrium state, the first hitting time of $\{N\}$ starting from $\{0\}$, together with the time needed for the Moran walker to reach its invariant measure, again starting from $\{0\}$. The detailed growth rates of these quantities are shown to be exactly determined in terms of the mutation probabilities.

- For the first fixation probability point, a sharp threshold phenomenon is exhibited in the following sense: would the mutation probability favoring allele 1 exceed the one favoring allele 2, the fixation probability of a type 1 mutant tends to 1 as N grows large with a $\mathcal{O}(1/N)$ correction. On the contrary, this fixation probability tends to 0 geometrically fast as N grows large. In the critical case when the two mutation probabilities are equal, this fixation probability goes to $1/2$ from below.

Under the additional the presence of selection, the fixation condition can be achieved with probability 1 even if the mutation probability favoring allele 2 exceeds the one favoring allele 1. The condition is that the relative fitness of allele 1 should be large enough with bound given in terms of a ratio of mutation probabilities, translating a mutation-selection balance.

- For the last time to stationarity point, an appeal to the notion of Siegmund duality (requiring stochastic monotonicity) is necessary and a cutoff phenomenon will be made explicit in Section 4, following the general stream of ideas developed in [1], [4], [10] and [11]. The analysis is made simple because the spectral representation of the mutation Moran model is exactly known.

We are interested into these problems in the large population size limit $N \rightarrow \infty$ (i.e. in the thermodynamic limit). From our analysis, it follows that the mutation Moran walker started at $\{0\}$ reaches stationarity in about $\mathcal{O}(N \log N)$ time, far below the time needed to first return to the origin (or excursion's length) which is geometrically large with N . This is because the mutational drift pushes the walker inside the bulk of the state-space, making the reverse move to any of the two boundaries extremely costly.

2. DISCRETE-TIME BIRTH AND DEATH PROCESSES

We start with generalities on discrete birth and death (BD) processes with finite state-space, before particularizing the study to the Moran model under concern which is in this class.

¹There are alternative domains of Science where this model makes sense, namely Economy and Behavioral Sciences. Here, the Markov state X_n takes on the interpretation of the number of agents that belong to the first strategy 1 (say right or left-wing voters) at time n . See [12], [14] and References therein for a recent study with this physical image in mind.

with $1 = t_0 > t_1 > \dots > t_N \geq -1$. The set S constitutes the spectrum of P (see [22], page 78). The quantity $1 - t_1$ is the spectral gap.

Let $\mu(dt) := \sum_{k=0}^N \mu_k \delta_{t_k}$ be the spectral probability measure on $[-1, 1]$ with respect to which $(\mathbf{q}_y(t), y \geq 1)$ are orthogonal, namely:

$$(1) \quad \gamma_y \int_{-1}^1 \mathbf{q}_x(t) \mathbf{q}_y(t) \mu(dt) = \gamma_y \sum_{k=0}^N \mu_k \mathbf{q}_x(t_k) \mathbf{q}_y(t_k) = \delta_{x,y},$$

where $\gamma_y := \frac{\pi_y}{\pi_0} = 1 / \int_{-1}^1 \mathbf{q}_y(t)^2 \mu(dt) = \prod_{z=0}^{y-1} \frac{p_z}{q_{z+1}}$, $y \geq 0$ are the potential coefficients.

These polynomials are important in view of the KMG spectral representation theorem [22]. Indeed, with $\mathbf{q}(t) = (\mathbf{q}_0(t), \mathbf{q}_1(t), \dots, \mathbf{q}_N(t))'$, we have:

$$(2) \quad t\mathbf{q}(t) = P\mathbf{q}(t)$$

and therefore: $t^n \mathbf{q}(t) = P^n \mathbf{q}(t)$. Selecting the row x , multiplying by $\mathbf{q}_y(t)$, integrating with respect to μ and applying the orthogonality relations (1), we get

$$(3) \quad P^n(x, y) := \mathbf{P}_x(X_n = y) = \gamma_y \int_{-1}^1 t^n \mathbf{q}_x(t) \mathbf{q}_y(t) \mu(dt) = \gamma_y \sum_{k=0}^N \mu_k t_k^n \mathbf{q}_x(t_k) \mathbf{q}_y(t_k).$$

Necessarily, $\mu_0 = \pi_0$ because by the ergodic theorem, $\forall x, P^n(x, y) \rightarrow \mu_0 \gamma_y = \frac{\mu_0}{\pi_0} \pi_y = \pi_y$ as $n \uparrow \infty$. From (3), because $\mathbf{q}_0(t) = 1$, we have $\mathbf{P}_0(X_n = 0) = \sum_{k=0}^N \mu_k t_k^n$ and so the generating function of $\mathbf{P}_0(X_n = 0)$ is

$$u_0(z) := \sum_{n \geq 1} z^n \mathbf{P}_0(X_n = 0) = \sum_{k=0}^N \mu_k (1 - z t_k)^{-1} = \int_{-1}^1 \frac{\mu(dt)}{1 - zt},$$

the Stieltjes transform of the spectral measure, satisfying $u_0(1) = \infty$ as a result of $t_0 = 1$. On the other side, let $\phi_0(z) = \sum_{k \geq 1} z^k \mathbf{P}(\tau_{0,0}^* = k)$ be the probability generating function (pgf) of the first return time to zero $\tau_{0,0}^*$ probability. As it can easily be checked by renewal arguments (see [9], page 330), $u_0(z) = 1 + u_0(z) \phi_0(z)$, showing that

$$u_0(z) = \frac{1}{1 - \phi_0(z)} \text{ and } \phi_0(z) = 1 - \frac{1}{u_0(z)} = 1 - \frac{1}{\sum_{k=0}^N \mu_k (1 - z t_k)^{-1}}.$$

We have $\mathbf{P}(\tau_{0,0}^* < \infty) = \phi_0(1) = 1$ and $\phi_0'(1) = \mathbf{E}(\tau_{0,0}^*) = \mu_0^{-1} = \pi_0^{-1}$.

More generally, with $|z| \leq 1$, let

$$\mathbf{g}_z(x, y) = \sum_{n \geq 0} z^n P^n(x, y) = (I - zP)^{-1}(x, y)$$

be the Green potential function of the chain. We let $\mathbf{g}(x, y) := \mathbf{g}_1(x, y)$, the expected sojourn time at y , starting from x . Note $\mathbf{g}_z(0, 0) = u_0(z)$ and $\mathbf{g}_0(x, y) = 1$.

- With $\tau_{x,y} = \inf(n \geq 0 : X_n = y \mid X_0 = x)$ the first **hitting** time of y starting from x , by renewal arguments, the pgf of $\tau_{x,y}$ reads

$$\mathbf{E}(z^{\tau_{x,y}}) = \frac{\mathbf{g}_z(x, y)}{\mathbf{g}_z(x, x)},$$

with $\tau_{x,x} = 0$. We have

$$\mathbf{E}(z^{\tau_{y,x}}) = \frac{\mathbf{g}_z(y,x)}{\mathbf{g}_z(y,y)} = \frac{\gamma_x \mathbf{g}_z(x,x) \mathbf{g}_z(x,y)}{\gamma_y \mathbf{g}_z(y,y) \mathbf{g}_z(x,x)} = \frac{\gamma_x \mathbf{g}_z(x,x)}{\gamma_y \mathbf{g}_z(y,y)} \mathbf{E}(z^{\tau_{x,y}}).$$

With $x < y$, we have:

$$\begin{aligned} \mathbf{P}(\tau_{x,y} > n) &= \mathbf{P}(X_n < y \mid X_0 = x) = \sum_{0 \leq y' < y} \mathbf{P}(X_n = y' \mid X_0 = x) = \sum_{0 \leq y' < y} P^n(x, y') \\ \mathbf{P}(\tau_{y,x} > n) &= \mathbf{P}(X_n > x \mid X_0 = y) = \sum_{N \geq y' > x} \mathbf{P}(X_n = y' \mid X_0 = y) = \sum_{N \geq y' > x} P^n(y, y') \end{aligned}$$

and

$$\begin{aligned} \mathbf{E}(\tau_{x,y}) &= \sum_{n \geq 0} \mathbf{P}(\tau_{x,y} > n) = \sum_{0 \leq y' < y} \mathbf{g}(x, y') \\ \mathbf{E}(\tau_{y,x}) &= \sum_{n \geq 0} \mathbf{P}(\tau_{y,x} > n) = \sum_{N \geq y' > x} \mathbf{g}(y, y'). \end{aligned}$$

- With $\tau_{x,x}^* = \inf(n \geq 1 : X_n = x \mid X_0 = x)$ the first **return** time to x , we have

$$\phi_x(z) := \mathbf{E}(z^{\tau_{x,x}^*}) = \frac{\mathbf{g}_z(x,x) - 1}{\mathbf{g}_z(x,x)} = 1 - \frac{1}{\mathbf{g}_z(x,x)} =: 1 - \frac{1}{u_x(z)}$$

with

$$\mathbf{P}(\tau_{x,x}^* < \infty) = \phi_x(1) = \frac{\mathbf{g}(x,x) - 1}{\mathbf{g}(x,x)}.$$

The probability that the first return time to x is finite is 1 iff $\mathbf{g}(x,x) = \infty$ (a recurrence condition) and this happens iff P (stochastic) is irreducible in which case all states are in the same recurrence class. Else if for some x , $\mathbf{g}(x,x) < \infty$ (a transience condition for x), this probability is smaller than 1 and one possible reason is that absorption occurs in at least one state $\neq x$; all states then but the absorbing ones (typically here, either $\{0\}$ or $\{N\}$, or both endpoints) are in the same transience class and $\mathbf{P}(\tau_{x,x}^* < \infty) \in (0, 1)$ because the walker may or not return to x before hitting one of the absorbing states. So P here is no longer irreducible and this is a particular reducibility case where transience pops in. In the latter transient case however, by removing the rows and columns associated to the absorbing states, we are left with an irreducible substochastic transition matrix of smaller dimension to which the Karlin-McGregor spectral representation applies as well (see [22]). The dominant eigenvalue of the restricted matrix is now strictly smaller than 1.

In view of $\gamma_x := \frac{\pi_x}{\pi_0} = 1 / \int_{-1}^1 \mathbf{q}_x(t)^2 \mu(dt)$ and in case of recurrence

$$(4) \quad \mathbf{E}(\tau_{x,x}^*) = \frac{1}{\pi_x}.$$

Equations (2, 3) suggest that, with:

$$(5) \quad r_k(x) = \frac{\mu_k}{\mu_0} \mathbf{q}_x(t_k) \quad \text{and} \quad l_k(y) = \pi_y \mathbf{q}_y(t_k),$$

the vectors $\mathbf{r}_k = (r_k(0), \dots, r_k(x), \dots, r_k(N))'$ and $\mathbf{l}'_k = (l_k(0), \dots, l_k(y), \dots, l_k(N))$ can be chosen to be the right (column) and left (row) eigenvectors of P associated to the eigenvalue t_k for which:

$$(6) \quad P\mathbf{r}_k = t_k\mathbf{r}_k \text{ and } \mathbf{l}'_k P = t_k\mathbf{l}'_k.$$

In particular, $l_0(y) = \pi_y \mathbf{q}_y(1) = \pi_y$ and $r_0(x) = 1$, as required. We have

$$(7) \quad \mathbf{l}'_j \mathbf{r}_k = \delta_{j,k}$$

because $\mathbf{l}'_j P \mathbf{r}_k = t_k \mathbf{l}'_j \mathbf{r}_k = t_j \mathbf{l}'_j \mathbf{r}_k$, with the right normalization. Equation (7) yields another orthogonality relation which is dual to the one in (1), namely:

$$(8) \quad \frac{\mu_k}{\mu_0} \sum_{x=0}^N \pi_x \mathbf{q}_x(t_j) \mathbf{q}_x(t_k) = \delta_{j,k}.$$

With $P_k := \mathbf{r}_k \mathbf{l}'_k$, we now get the decomposition of P into orthogonal projectors which is (3):

$$P = \sum_{k=0}^N t_k P_k \text{ and } P^n = \sum_{k=0}^N t_k^n P_k.$$

Note the biorthogonality relations:

$$(9) \quad \mathbf{l}'_j P \mathbf{r}_k = t_k \delta_{j,k},$$

from which it follows that:

$$\sum_{x,y=0}^N l_j(x) P^n(x,y) r_k(y) = \sum_{x,y=0}^N \pi_x \mathbf{q}_x(t_j) P^n(x,y) \frac{\mu_k}{\mu_0} \mathbf{q}_y(t_k) = t_k^n \delta_{j,k}.$$

There is a matrix version of all this: Calling R (respectively L) the matrix whose k^{th} column (row) is \mathbf{r}_k (respectively \mathbf{l}'_k), (6) reads $PR = RD_{\mathbf{t}}$ and $LP = D_{\mathbf{t}}L$ where $D_{\mathbf{t}} := \text{diag}(\mathbf{t})$, $\mathbf{t} = (t_k, k = 0, \dots, N)$. Since by (7): $LR = I := \text{diag}(\mathbf{1})$, we have $LPR = D_{\mathbf{t}}LR = D_{\mathbf{t}}$ which is (9). Since $RL = I$, $(LPR)^n = D_{\mathbf{t}}^n = LP^nR$. If we let Q be the matrix whose entry (x, k) is $Q(x, k) = \mathbf{q}_x(t_k)$, then $R = \frac{1}{\mu_0} Q D_{\mu}$ and $L = Q' D_{\pi}$ is the matrix formulation of (5). Let us finally check that L and R (that is (5)) are indeed admissible left and right eigenmatrices: The relation $PR = RD_{\mathbf{t}}$ is also $PQD_{\mu} = QD_{\mu}D_{\mathbf{t}} = QD_{\mathbf{t}}D_{\mu}$ which is $PQ = QD_{\mathbf{t}}$ which holds from (2). The relation $LP = D_{\mathbf{t}}L$ is also $Q'D_{\pi}P = D_{\mathbf{t}}Q'D_{\pi}$ or $Q'\overleftarrow{P}' = D_{\mathbf{t}}Q'$ which is $\overleftarrow{P}Q = QD_{\mathbf{t}}$ or again $PQ = QD_{\mathbf{t}}$ because by the reversibility property: $\overleftarrow{P} = P$. If $\mu_0 R^2 = I$ (R is involutive), then $\mu_0 R = L$ which means $Q' = QD_{\mu}D_{\pi}^{-1}$. If in addition $\mu = \pi$, then $Q' = Q$: the duality relations (1) and (8) coincide and this is an exceptional case of self-duality.

The KMG spectral theory therefore gives a representation of the right and left eigenvectors of P associated to its spectrum \mathbf{t} , in terms of the orthogonal polynomials $\mathbf{q}(t)$ evaluated at \mathbf{t} (the Q -matrix).

The interpretation of the right eigenvectors of P is as follows: We have $P^n \mathbf{r}_k = t_k^n \mathbf{r}_k$. Fixing the row x , the left-hand-side of the last equality is $\mathbf{E}_x[r_k(X_n)]$. Therefore:

$$\mathbf{E}_x[r_k(X_n)] = t_k^n r_k(x).$$

Equivalently, if s_k are the positive eigenvalues of $S'S$, we need to check $2r_k - 1 + s_k > 0$ for all k .

For a simple random walk X_n whose $(N+1)^2$ -transition matrix P is spectrally non-negative (respectively spectrally positive), there exists a symmetric random walk Y_m on $\{0, \dots, 2N\}$ (respectively on $\{0, \dots, 2N+1\}$), reflected at the endpoints, started at an even integer $\{0, 2, 4, \dots, 2N\}$ (respectively odd integer $\{1, 3, \dots, 2N+1\}$), such that $\{X_n\} \stackrel{d}{=} \{Y_{2n}/2\}$, (respectively $\{X_n\} \stackrel{d}{=} \{(Y_{2n}-1)/2\}$). This follows simply from adapting [34], Th. 2.1 to our finite-dimensional case. As noted just before, the spectral measure of the random walk Y_m is symmetric on $[-1, 1]$ and by passing to X_n , the spectrum is being folded: If $\sum_{k=0}^{2N} \mu_k \delta_{t_k}$ (respectively $\sum_{k=0}^{2N+1} \mu_k \delta_{t_k}$) is the symmetric spectral measure of Y_m with $t_N = 0$ (respectively $t_N > 0$), then $2 \sum_{k=0}^N \mu_k \delta_{t_k^2}$ is the spectral measure of X_n . Let α_y and β_y be the up and down probabilities that $Y_m \rightarrow Y_{m+1} = Y_m \pm 1$ in one step given Y_m is in state y different from the endpoints, $\alpha_y + \beta_y = 1$, then:

$$q_x = \beta_{2x} \beta_{2x-1}, \quad r_x = \beta_{2x} \alpha_{2x-1} + \alpha_{2x} \beta_{2x+1}, \quad p_x = \alpha_{2x} \alpha_{2x+1}.$$

This, together with $p_0 = \alpha_1$ and $q_N = \beta_{2N-1}$ (respectively $q_N = \beta_{2N}$) allows to determine recursively the transition matrix of Y_m from the one of X_n . From these facts, we get:

Proposition 1. *If a BD chain X_n is spectrally non-negative, then it is stochastically monotone.*

Proof: Under our hypothesis indeed,

$$(13) \quad p_x + q_{x+1} = \alpha_{2x} \alpha_{2x+1} + \beta_{2x+2} \beta_{2x+1} < \alpha_{2x} \alpha_{2x+1} + \beta_{2x+1} < 1,$$

because this is also $\alpha_{2x} < 1$. This condition means that X_n is stochastically monotone (see below the construction of the Siegmund dual of a RW where stochastic monotonicity is needed). See also Lemma 2.4 of [10] for a different proof. Δ

Remark1: Suppose a BD chain X_n with transition matrix P is not spectrally non-negative, so P has strictly negative eigenvalues ($t_N < 0$). With $\rho \in (0, 1)$, consider the lazy modified stochastic matrix

$$P_\rho = (1 - \rho)I + \rho P.$$

The eigenvalues of P_ρ are $t_k^{(\rho)} = 1 - \rho + \rho t_k$, $k = 0, \dots, N$ with the same left and right eigenvectors as the ones of P . One can always choose $\rho \leq (1 - t_N)^{-1}$ small enough so that $t_N^{(\rho)} \geq 0$. For this ρ , P_ρ is now spectrally non-negative and so P_ρ governs a stochastically monotone chain $X_n^{(\rho)}$. The RW $X_n^{(\rho)}$, with shifted spectrum, clearly has the same invariant measure as the one of X_n .

For example, solving (13) with the boundary conditions $p_0 = \alpha_1$ and $q_N = \beta_{2N}$ gives the forward and backward recurrences of the unknown odd and even probability terms in the bulk:

$$\text{For } x = 1, \dots, N-1 : \alpha_{2x+1} = p_x / (1 - q_x / (1 - \alpha_{2x-1})), \quad \alpha_1 = p_0.$$

$$\text{For } x = N-1, \dots, 1 : \alpha_{2x} = p_x / (1 - q_{x+1} / (1 - \alpha_{2(x+1)})), \quad \alpha_{2N} = 1 - q_N.$$

As a conclusion, it is easy to construct spectrally positive RWs by ‘squaring’ two symmetric random walks, but given a non-symmetric RW, it may be difficult to decide whether it is or not spectrally positive because one needs conversely to check whether all the above α s are probabilities. Spectral positivity determines stochastic monotonicity.

2.5. Scale function and excursion height of a RW. Assume the height $H = h \in \{1, \dots, N-1\}$ for some excursion (sample paths of X_n between two consecutive visits to state 0). This event will be realized if and only if (i) downward paths started from h hit state $\{0\}$ before hitting state $h+1$ and (ii) upward paths started at $\{0\}$ first reach 1 (with probability p_0) and then, paths started at 1 hit h without returning to $\{0\}$ again in the intervening time. These two events are independent. Therefore, with $\tau_{x,y}$ the first hitting time of y starting from x :

$$(14) \quad \mathbf{P}(H = h) = p_0 \mathbf{P}(\tau_{1,h} < \tau_{1,0}) \mathbf{P}(\tau_{h,0} < \tau_{h,h+1}), \quad h = 1, \dots, N.$$

Assume $X_0 = x$. Let $X_{n \wedge \tau_{x,0}}$ denote the random walk stopped when it first hits 0. Its transition matrix is P except for its first row which is $(1, 0, 0, \dots, 0)$, translating that state $\{0\}$ is now absorbing. Define the scale (or harmonic) function φ of this random walk as the function which makes $M_n := \varphi(X_{n \wedge \tau_{x,0}})$ a martingale. The function φ is important because, as is well-known, for all $0 < x < h \leq N$, with $\tau_x = \tau_{x,0} \wedge \tau_{x,h}$ the first hitting time of $\{0, h\}$ starting from x

$$(15) \quad \mathbf{P}(X_{\tau_x} = h) = \frac{\varphi(x)}{\varphi(h)}.$$

Using this remark, we get:

$$(16) \quad \mathbf{P}(H = h) = p_0 \mathbf{P}(\tau_{1,h} < \tau_{1,0}) \mathbf{P}(\tau_{h,0} < \tau_{h,h+1}) = p_0 \frac{\varphi(1)}{\varphi(h)} \left(1 - \frac{\varphi(h)}{\varphi(h+1)}\right), \quad h \in \{1, \dots, N-1\}.$$

Note $\mathbf{P}(H \geq h) = \frac{p_0}{\varphi(h)}$ if $\varphi(1) = 1$. The event $H = N$ may also occur, with probability $\mathbf{P}(H = N) = \frac{p_0}{\varphi(N)}$.

The interpretation the event $H = N$ takes on is twofold: either it is the maximal height of an excursion if the state $\{N\}$ remains reflecting, either, if state $\{N\}$ is itself assumed absorbing, $\mathbf{P}(X_{\tau_x} = N) = \frac{\varphi(x)}{\varphi(N)} = \frac{\varphi(x)}{p_0} \mathbf{P}(H = N)$ constitutes the probability of a fixation event at $\{N\}$, starting from state x . We can thus compute the probability of a fixation before extinction at $\{0\}$ occurs. It remains to compute φ (defined up to an arbitrary multiplicative constant). We wish to have: $\mathbf{E}_x(M_{n+1} | M_n = y) = y$, leading to $\varphi(0) = 0$ and

$$\varphi(x) = q_x \varphi(x-1) + r_x \varphi(x) + p_x \varphi(x+1), \quad x \geq 1$$

where $p_x + q_x + r_x = 1$, $x \geq 1$. Imposing $\varphi(1) = 1$ (fixing the arbitrary multiplicative constant), the searched ‘harmonic’ function is $\varphi(x) =: 1 + \sum_{y=1}^{x-1} \psi(y)$ where $\psi(y)$ satisfies: $q_y \psi(y-1) = p_y \psi(y)$. Thus $\psi(y) = \prod_{z=1}^y \frac{q_z}{p_z}$ and:

$$(17) \quad \varphi(x) = 1 + \sum_{y=1}^{x-1} \prod_{z=1}^y \frac{q_z}{p_z}, \quad x \geq 1, \quad \varphi(0) := 0.$$

Equations (16) and (17) characterize the law of the excursion height of the random walk X_n . We note that $\varphi(x)$ increases with x . From (15) and (17), we get an explicit expression of the probability of the event $\tau_{x,h} < \tau_{x,0}$:

$$\mathbf{P}(\tau_{x,h} < \tau_{x,0}) = \mathbf{P}(X_{\tau_x} = h) = \frac{\varphi(x)}{\varphi(h)},$$

where $\tau_{x,h}$ is the first hitting time of h starting from x , with $0 < x < h \leq N$. In particular, if $h = N$ with $\{N\}$ also absorbing, $\mathbf{P}(\tau_{x,N} < \tau_{x,0}) = \frac{\varphi(x)}{\varphi(N)}$ is the fixation probability starting from x , whereas $\mathbf{P}(\tau_{x,0} < \tau_{x,N}) = 1 - \mathbf{P}(\tau_{x,N} < \tau_{x,0})$ is the extinction probability.

3. A BD EXAMPLE: THE MORAN MODEL

The example we have in mind is the 2-allele Moran model with bias mechanism p . Let

$$(18) \quad p(u) : u \in [0, 1] \rightarrow [0, 1], \text{ with } 0 \leq p(0) \text{ and } p(1) \leq 1$$

be continuous and $\bar{p}(u) := 1 - p(u)$.

The Moran model X_n is characterized by the following transition probabilities, given the walker is in state $x \in \{0, \dots, N\}$:

$$(19) \quad q_x = \frac{x}{N} \bar{p}\left(\frac{x}{N}\right), \quad r_x = \frac{x}{N} p\left(\frac{x}{N}\right) + \left(1 - \frac{x}{N}\right) \bar{p}\left(\frac{x}{N}\right), \quad p_x = \left(1 - \frac{x}{N}\right) p\left(\frac{x}{N}\right).$$

For a Moran model, X_n is the amount of, say allele or type 1 individuals, in a population of size N and the reproduction law is obtained while choosing two individuals at random among N , one dying and the other one passing to the next generation while giving birth to an additional individual so as to keep constant the full population size N over the generations. Given then $X_n = x$, the transition $X_n \rightarrow X_n + 1$ ($X_n \rightarrow X_n - 1$) occurs with probability p_x (q_x) translating the fact that the individual bound to die is chosen among type 2 (type 1) individuals with probability $1 - \frac{x}{N}$ ($\frac{x}{N}$) and the one which splits into two new born is chosen among the other type, with probability $p\left(\frac{x}{N}\right)$ (respectively $\bar{p}\left(\frac{x}{N}\right)$), after deforming the neutral frequency $\frac{x}{N}$ by the bias mechanism p . These bias p take into account various evolutionary forces which are the causes of deviation to neutrality (see [28], [15] and [8], for a discussion on various models of utmost interest in population genetics, among which selection and mutations). The classical neutral Moran model is obtained when $p(u) = u$ (no external deformation). Extended Moran models are studied in [21].

Assuming $p_0 = p(0) > 0$ and $q_N = 1 - p(1) > 0$, with $y \in \{1, \dots, N\}$, the corresponding Markov chain is ergodic with invariant distribution

$$(20) \quad \frac{\pi_y}{\pi_0} = \prod_{z=0}^{y-1} \frac{p_z}{q_{z+1}} = \binom{N}{y} \prod_{z=0}^{y-1} \frac{p\left(\frac{z}{N}\right)}{\bar{p}\left(\frac{z+1}{N}\right)},$$

where

$$\pi_0 = 1 / \left(1 + \sum_{y=1}^N \binom{N}{y} \prod_{z=0}^{y-1} \frac{p\left(\frac{z}{N}\right)}{\bar{p}\left(\frac{z+1}{N}\right)} \right).$$

We have the obvious facts

Proposition 2. *Let X_n be a Moran process with transition probabilities (19) for some bias p . Then, $\overline{X}_n := N - X_n$ is again a Moran process with transition probabilities of the type (19) but with the new bias $\tilde{p}(u) := 1 - p(1 - u)$, namely, $\overline{q}_x = p_{N-x} = \frac{x}{N} (1 - \tilde{p}(\frac{x}{N}))$, $\overline{p}_x = q_{N-x} = (1 - \frac{x}{N}) \tilde{p}(\frac{x}{N})$. The spectrum of the transition matrix \overline{P} of \overline{X}_n is the same as the one P of X_n . If the bias is such that $\tilde{p}(u) = p(u)$, then the transition matrix of \overline{X}_n coincides with the one of X_n .*

Proof: we have $\overline{P}(x, y) = P(N - x, N - y)$ and \overline{P} is obtained from P applying the permutation $\sigma : (0, 1, \dots, N) \rightarrow (N, \dots, 1, 0)$ reversing the order of the rows and then of the columns. \triangle

3.1. Fixation probability. Forcing the states $\{0, N\}$ to be absorbing, the scale function reads

$$\varphi(0) = 0 \text{ and } \varphi(x) = 1 + \sum_{y=1}^{x-1} \frac{1}{\binom{N-1}{y}} \prod_{z=1}^y \frac{\tilde{p}(\frac{z}{N})}{p(\frac{z}{N})}$$

with $\frac{\varphi(x)}{\varphi(N)}$ being the fixation probability whereas $1 - \frac{\varphi(x)}{\varphi(N)}$ is the extinction probability, starting from x (²). The probability of a fixation before extinction occurs in an excursion reads:

$$\mathbf{P}(H = N) = \frac{p_0}{\varphi(N)} = \frac{p(0)}{1 + \sum_{y=1}^{N-1} \frac{1}{\binom{N-1}{y}} \prod_{z=1}^y \frac{\tilde{p}(\frac{z}{N})}{p(\frac{z}{N})}}.$$

We have

$$\mathbf{P}(H = N) = p_0 \mathbf{P}(\tau_{1,N} < \tau_{1,0})$$

where $\mathbf{P}(\tau_{1,N} < \tau_{1,0}) = 1/\varphi(N)$ is the fixation probability of a mutant-type allele A_1 e.g. the probability that, starting from $X_0 = 1$, the RW hits state $\{N\}$ (fixation of the mutant-type allele 1) before hitting state $\{0\}$ (extinction of allele 1).

3.2. Examples of bias mechanisms. We now give three examples of bias that we shall deal with in the sequel.

Moran model with mutations. A basic bias example is the mutation mechanism

$$(21) \quad p(u) = (1 - \mu_2)u + \mu_1(1 - u),$$

where (μ_1, μ_2) are mutation probabilities in $(0, 1]$. We let $\overline{\mu}_1 := 1 - \mu_1$ and $\overline{\mu}_2 := 1 - \mu_2$, $\mu = \mu_1 + \mu_2$ and $\overline{\mu} = 1 - \mu$.

Schematically, type 1 alleles are being produced whenever type 1 alleles do not mutate to type 2 (with probability $\overline{\mu}_2$) or type 2 alleles mutate to type 1 (with probability μ_1)

$$\overline{\mu}_2 \circlearrowleft A_1 \xrightleftharpoons[\mu_1]{\mu_2} A_2 \circlearrowright \overline{\mu}_1.$$

The transition matrix P of this model has therefore the particular BD structure

$$(22) \quad q_x = \frac{x}{N} \left(\overline{\mu}_1 - \overline{\mu} \frac{x}{N} \right), \quad r_x = 1 - (p_x + q_x), \quad p_x = \left(1 - \frac{x}{N} \right) \left(\mu_1 + \overline{\mu} \frac{x}{N} \right),$$

²For the neutral model with $p(u) = u$, the scale function is $\varphi(x) = x$ and the fixation probability is trivially x/N .

with $p_0 = \mu_1$, $q_N = \mu_2$. When $\mu \neq 1$, we introduce the auxiliary parameters

$$\nu_1 = \mu_1/\bar{\mu}, \nu_2 = \mu_2/\bar{\mu} \text{ and } \nu = \nu_1 + \nu_2 = \mu/\bar{\mu},$$

with $\nu_1 + 1 = \bar{\mu}_2/\bar{\mu}$ and $\nu_2 + 1 = \bar{\mu}_1/\bar{\mu}$.

The drift of this RW is

$$\mathbf{E}(X_{n+1} - X_n | X_n = x) = p_x - q_x = p\left(\frac{x}{N}\right) - \frac{x}{N} = \mu_1 - \mu \frac{x}{N}.$$

The deterministic part of the dynamics of X_n has a stable point at $x = N\mu_1/\mu$, cancelling the drift. When p is non-decreasing, we have $\mu \leq 1$. We note that in the expression of $\tilde{p}(u)$, the roles of μ_1 and μ_2 are interchanged. So $\tilde{p}(u) = p(u)$ if and only if $\mu_1 = \mu_2$ (symmetric mutations).

If $\mu_1 = \mu_2 = 1$, we recover if N is even the heat-exchange Bernoulli-Laplace model [9] as a borderline example but $p(u) = 1 - u$ is strictly decreasing in this case. If $\mu_1 = \mu_2 = 1/2$ then $p(u) = 1/2$ and we obtain an aperiodic model amenable (through a suitable time substitution) to the Ehrenfest urn model provided N is even.

For one-way mutations, $(\mu_1, \mu_2) = (\mu_1, 0)$ or $(0, \mu_2)$ leading to the simpler forms $\mu_1 + u\bar{\mu}_1$ and $\bar{\mu}_2 u$ of $p(u)$ but with $p(1) = 1$ and $p(0) = 0$ respectively, corresponding to $\{N\}$ (respectively $\{0\}$) being absorbing. We avoid this situation because it deserves another treatment due to transience.

Moran model with mutations and selection. Another example could be

$$(23) \quad p(u) = \frac{\mu_1 + u((1+s)\bar{\mu}_2 - \mu_1)}{1 + su},$$

which is the composition of the selection bias mechanism

$$(24) \quad p(u) = \frac{(1+s)u}{1+su} \quad (s > -1)$$

with the mutation mechanism (21). When considering the fixation probability $\mathbf{P}(H = N)$ for the Moran model with this particular mechanism, we obtain a closed-form formula for the fixation probability of a mutant which of course is an interesting issue in genetics.

Cases with positive eigenvalues. We may look for conditions on the mechanism p leading to $r_x \geq \frac{1}{2}$ in which case the RW is spectrally positive. Assume $p(u) : u \in (0, 1) \rightarrow (0, 1)$ is continuous and non-decreasing, with $0 < p(0) \leq p(1) < 1$.

Then, as can easily be checked: $r_x \geq 1/2$ for all x if and only if $p(1/2) = 1/2$ with $p(0) \leq \frac{1}{2} \leq p(1)$. Indeed, imposing $r_x \geq 1/2$ for all x leads to $p(u) \geq 1/2$ if $u \geq 1/2$ and $p(u) \leq 1/2$ if $u \leq 1/2$. So, if p is non-decreasing with $p(1/2) = 1/2$, then $r_x \geq \frac{1}{2}$.

- Take for example the mutation mechanism, now satisfying $0 < \mu_1 \leq \bar{\mu}_2 < 1$. The condition $p(1/2) = 1/2$, leads to $\mu_1 = \mu_2$ (fair mutation). In this case, $p(u) = \mu_1 + (1 - 2\mu_1)u$ with $0 < \mu_1 \leq 1/2$. If $\mu_1 = 1/2$, then $p(u) = \frac{1}{2}$ and $q_x = \frac{x}{2N}$, $r_x = \frac{1}{2}$, $p_x = \frac{1}{2} \left(1 - \frac{x}{N}\right)$.

For example, if $N = 2$:

$$P = \begin{bmatrix} \bar{\mu}_1 & \mu_1 & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & \mu_1 & \bar{\mu}_1 \end{bmatrix}$$

has three distinct real non-negative eigenvalues $\left\{1, \bar{\mu}_1, \frac{1-2\mu_1}{2}\right\}$ ordered in decreasing size. We note however that it is not necessary that $\mu_1 = \mu_2$ for P to be spectrally positive: indeed, for the full Moran model with mutations $\mu_1 \neq \mu_2$, under the sole condition $\mu_1 + \mu_2 \leq 1$ (p is non decreasing), the eigenvalues t_k can be seen to be non-negative. Still, we were not able to find a sharp general condition on p that would lead to a spectrally non-negative Moran chain.

- For the mechanism p of (23) combining mutations with selection, assuming $0 < \mu_1 \leq \bar{\mu}_2 < 1$ to guarantee that p is non-decreasing, the condition $p(1/2) = 1/2$ relates (s, μ_1, μ_2) to give: $s = (\mu_2 - \mu_1) / (1/2 - \mu_2)$ where both $\mu_1, \mu_2 < 1/2$. In this case, the model is spectrally positive.

3.3. Fixation probabilities for the Moran model with mutations and/or selection. We now study the particular Moran model with mutations and/or selection, starting with fixation probabilities.

• **Moran with mutations** (sharp threshold phenomenon): We need to evaluate $\varphi(N) = 1 + \sum_{y=1}^{N-1} \frac{1}{\binom{N-1}{y}} \prod_{z=1}^y \frac{\bar{p}(\frac{z}{N})}{p(\frac{z}{N})}$ for large N and then $\mathbf{P}(\tau_{1,N} < \tau_{1,0}) = \frac{1}{\varphi(N)}$ will follow.

- Suppose first $\mu = 1$ and let $\rho := \frac{\bar{\mu}_1}{\mu_1}$.

Proposition 3. *The fixation probability satisfies the sharp threshold property:*

- (i) If $\mu_1 = 1/2$ ($\rho = 1$), then $\mathbf{P}(\tau_{1,N} < \tau_{1,0}) = \frac{1}{\varphi(N)} \rightarrow 1/2$.
- (ii) If $\mu_1 < 1/2$ ($\rho > 1$), then $\mathbf{P}(\tau_{1,N} < \tau_{1,0}) = \frac{1}{\varphi(N)} \sim \rho^{-N} \rightarrow 0$.
- (iii) If $\mu_1 > 1/2$ ($\rho < 1$), then $\mathbf{P}(\tau_{1,N} < \tau_{1,0}) = \frac{1}{\varphi(N)} \rightarrow 1$.

Proof: With $\rho = \frac{\bar{\mu}_1}{\mu_1}$, we first have the identity for sums of inverse binomial coefficients (see e.g. [27], page 197)

$$\varphi(N) := \sum_{y=0}^{N-1} \frac{\rho^y}{\binom{N-1}{y}} = \frac{N}{(\rho+1) \left(\frac{1}{\rho} + 1\right)^N} \sum_{k=1}^N \frac{(\rho^k + 1) \left(\frac{1}{\rho} + 1\right)^k}{k}.$$

There is therefore a closed-form expression of the fixation probability $\mathbf{P}(\tau_{1,N} < \tau_{1,0}) = \frac{1}{\varphi(N)}$. Observing now

$$\varphi(N) := \sum_{y=0}^{N-1} \frac{\rho^y}{\binom{N-1}{y}} = 1 + \frac{\rho}{N-1} + \frac{\rho^{N-2}}{N-1} + \rho^{N-1} + \sum_{y=2}^{N-3} \frac{\rho^y}{\binom{N-1}{y}},$$

we get:

- (i) $\bar{\mu}_1 = \mu_1$ ($\rho = 1$): $\varphi(N) \rightarrow 2^+$. So: $\mathbf{P}(\tau_{1,N} < \tau_{1,0}) = \frac{1}{\varphi(N)} \rightarrow 1/2$.
- (ii) $\bar{\mu}_1 > \mu_1$ ($\rho > 1$): $\varphi(N) \sim \rho^{N-1}$. So: $\mathbf{P}(\tau_{1,N} < \tau_{1,0}) = \frac{1}{\varphi(N)} \sim \rho^{-N} \rightarrow 0$.
- (iii) $\bar{\mu}_1 < \mu_1$ ($\rho < 1$): $\varphi(N) \rightarrow 1^+$. So: $\mathbf{P}(\tau_{1,N} < \tau_{1,0}) = \frac{1}{\varphi(N)} \rightarrow 1$.

If $\rho = 1$, because $\frac{1}{\binom{N}{y}} \leq \frac{1}{\binom{N}{2}}$ if $y \in \{2, N-2\}$: $2 + 2/N < \varphi(N+1) < 2 + 2/N + 2(N-3)/(N(N-1))$ and $\varphi_1(N+1)$ is very close to $2 + 2/N$. The fixation probability tends to $1/2$ from below.

If $\rho \neq 1$, $1 + \rho^N + \rho/N + \rho^{N-1}/N < \varphi(N+1) < 1 + \rho^N + \rho/N + \rho^{N-1}/N + 2/(N(N-1))\rho^2 \frac{1-\rho^{N-3}}{1-\rho}$ and $\varphi(N+1)$ is very close to ρ^N if $\rho > 1$ and very close to 1 up to a ρ/N factor if $\rho < 1$.

When $\mu_1 < 1/2$, the probability, starting from $\{0\}$ to hit the state $\{N\}$ (fixation) starting from 1 before returning to $\{0\}$ tends to 0 geometrically fast. On the contrary, would $\mu_1 > 1/2$, this probability would converge to 1: mutations producing type 1 individuals are sufficiently large to enhance fixation of a mutant allele with probability 1. If $\mu_1 = 1/2$, the fixation probability approaches $1/2$. Δ

– Suppose now $\mu \neq 1$. With $A_N = N\frac{\bar{\mu}_1}{\mu} = N(\nu_2 + 1)$, $B_N = N\frac{\mu_1}{\mu} = N\nu_1$

$$\begin{aligned} \varphi(N) &= 1 + \sum_{y=1}^{N-1} \frac{1}{\binom{N-1}{y}} \prod_{z=1}^y \frac{\bar{\mu}_1 - \bar{\mu} \frac{z}{N}}{\mu_1 + \bar{\mu} \frac{z}{N}} = 1 + \sum_{y=1}^{N-1} \frac{1}{\binom{N-1}{y}} \prod_{z=1}^y \frac{N\frac{\bar{\mu}_1}{\mu} - z}{N\frac{\mu_1}{\mu} + z} \\ &= 1 + \sum_{y=1}^{N-1} \frac{1}{\binom{N-1}{y}} \frac{\Gamma(A_N)}{\Gamma(A_N - y)} \frac{\Gamma(B_N + 1)}{\Gamma(B_N + 1 + y)} =: 1 + \sum_{y=1}^{N-1} \frac{1}{\binom{N-1}{y}} \psi_N(y). \end{aligned}$$

Proceeding as before, we have

$$1 + \frac{\psi_N(1)}{N-1} + \psi_N(N-1) < \varphi(N) < 1 + \frac{\psi_N(1)}{N-1} + \psi_N(N-1) + O(N^{-2}),$$

showing that $\varphi(N) \underset{N \text{ large}}{\sim} 1 + \frac{\psi_N(1)}{N-1} + \psi_N(N-1)$. The dominating term in $\varphi(N)$ is thus either the one corresponding to $y = 1$ or the one corresponding to $y = N-1$.

With $r := \left(\frac{\bar{\mu}_1^{\nu_1} \mu_1^{\mu_1}}{\bar{\mu}_2^{\nu_2} \mu_2^{\mu_2}} \right)^{1/\bar{\mu}}$, by Stirling formula, we get

$$\begin{aligned} \varphi(N) &\underset{N \text{ large}}{\sim} 1 + \frac{1}{N-1} \frac{A_N - 1}{B_N + 1} + \frac{\Gamma(A_N)}{\Gamma(A_N - N + 1)} \frac{\Gamma(B_N + 1)}{\Gamma(B_N + N)} \\ &\sim 1 + \frac{1}{N-1} \frac{\nu_2 + 1}{\nu_1} + \frac{\Gamma(N(\nu_2 + 1))}{\Gamma(N\nu_2 + 1)} \frac{\Gamma(N\nu_1 + 1)}{\Gamma(N(\nu_1 + 1))} \\ &= 1 + \frac{1}{N-1} \frac{\nu_2 + 1}{\nu_1} + \frac{\nu_1 + 1}{\nu_2 + 1} \frac{\binom{N(\nu_2 + 1)}{N\nu_2}}{\binom{N(\nu_1 + 1)}{N\nu_1}} \\ &\underset{N \text{ large}}{\sim} 1 + \frac{1}{N-1} \frac{\nu_2 + 1}{\nu_1} + \sqrt{\frac{\nu_1 + 1}{\nu_2 + 1}} \left(\frac{(\nu_2 + 1)^{\nu_2 + 1}}{\nu_2^{\nu_2}} \frac{\nu_1^{\nu_1}}{(\nu_1 + 1)^{\nu_1 + 1}} \right)^N \\ &=: 1 + \frac{1}{N-1} \frac{\nu_2 + 1}{\nu_1} + \sqrt{\frac{\nu_1 + 1}{\nu_2 + 1}} r^N. \end{aligned}$$

Now, whenever $\mu_1 < \mu_2$, regardless of whether $\mu > 1$ or $\mu < 1$,

$$(25) \quad r := \frac{(\nu_2 + 1)^{\nu_2 + 1}}{\nu_2^{\nu_2}} \frac{\nu_1^{\nu_1}}{(\nu_1 + 1)^{\nu_1 + 1}} = \left(\frac{\bar{\mu}_1^{\nu_1} \mu_1^{\mu_1}}{\bar{\mu}_2^{\nu_2} \mu_2^{\mu_2}} \right)^{1/\bar{\mu}} > 1.$$

Indeed, let $H(\mu_1) = -(\mu_1 \log \mu_1 + \bar{\mu}_1 \log \bar{\mu}_1)$. The condition $r > 1$ is also:

$$\frac{1}{\bar{\mu}} (H(\mu_2) - H(\mu_1)) > 0.$$

If $\mu_1 < \mu_2$, then $H(\mu_2) - H(\mu_1) < 0$ iff $\mu_2 > \bar{\mu}_1$ (or $\bar{\mu} < 0$) and then $\frac{1}{\bar{\mu}} (H(\mu_2) - H(\mu_1)) > 0$.

If $\mu_1 < \mu_2$, then $H(\mu_2) - H(\mu_1) > 0$ iff $\mu_2 < \bar{\mu}_1$ (or $\bar{\mu} > 0$) and then $\frac{1}{\bar{\mu}} (H(\mu_2) - H(\mu_1)) > 0$.

We obtained:

Proposition 4. *When $\mu \neq 1$, the fixation probability also satisfies a sharp threshold property:*

(i) *If $\mu_1 < \mu_2$, the fixation probability tends to 0 geometrically fast because $\varphi(N) \underset{N \text{ large}}{\sim} \sqrt{\frac{\nu_1+1}{\nu_2+1}} r^N = \sqrt{\frac{\bar{\mu}_2}{\bar{\mu}_1}} r^N$, leading to $\mathbf{P}(\tau_{1,N} < \tau_{1,0}) = \frac{1}{\varphi(N)} \underset{N \text{ large}}{\sim} \sqrt{\frac{\bar{\mu}_1}{\bar{\mu}_2}} r^{-N}$. The fixation probability of a mutant tends to 0, algebraically fast at rate r given by (25).*

(ii) *If $\mu_1 > \mu_2$, regardless of whether $\mu > 1$ or $\mu < 1$, $r < 1$. Thus $\varphi(N) \underset{N \text{ large}}{\sim} 1 + \frac{1}{N} \frac{\nu_2+1}{\nu_1} = 1 + \frac{1}{N} \frac{\bar{\mu}_1}{\mu_1}$ leading to $\mathbf{P}(\tau_{1,N} < \tau_{1,0}) = \frac{1}{\varphi(N)} \underset{N \text{ large}}{\sim} 1 / \left(1 + \frac{1}{N} \frac{\bar{\mu}_1}{\mu_1}\right) \rightarrow 1$ from below.*

(iii) *If $\mu_1 = \mu_2$ (balanced mutations), $\nu_1 = \nu_2$, $r = 1$ and therefore: $\varphi(N) \underset{N \text{ large}}{\sim} 2 + \frac{1}{N} \frac{\bar{\mu}_1}{\mu_1} \rightarrow 2_+$: the fixation probability tends to 1/2 from below.*

• **For the Moran model with selection** mechanism $p(u) = \frac{(1+s)u}{1+su}$ as in (24), gene A_1 (respectively A_2) has fitness $w_1 = 1 + s$ (respectively $w_2 = 1$) so that, with $w := w_1/w_2 = 1 + s$ the mutant relative fitness

$$\varphi(x) = 1 + \sum_{y=1}^{x-1} \frac{1}{\binom{N-1}{y}} \prod_{z=1}^y \frac{\bar{p}\left(\frac{z}{N}\right)}{p\left(\frac{z}{N}\right)} = 1 + \sum_{y=1}^{x-1} w^{-y} = \frac{1 - w^{-x}}{1 - w^{-1}},$$

showing that: $\mathbf{P}(\tau_{1,N} < \tau_{1,0}) = \frac{1}{\varphi(N)} = \frac{1-w^{-1}}{1-w^{-N}}$. Whenever $s > 0$ ($s < 0$), the mutant gene A_1 is selectively advantageous (deleterious) compared to A_2 . When $w^{-1} < 1$ ($s > 0$), $\mathbf{P}(\tau_{1,N} < \tau_{1,0})$ tends to $1 - w^{-1}$ whereas when $w^{-1} > 1$ ($s < 0$), this probability tends to 0 exponentially fast like $(w^{-1} - 1) w^N$. When $s = 0$ (neutral case), $\mathbf{P}(\tau_{1,N} < \tau_{1,0}) = 1/N$ tends to 0 much slower.

In this pure selection case, these results are well-known (see e.g. (3.66) page 109 of [8] and also [1] in [30]).

• **For the Moran model with bias** mechanism (23) involving both **mutation and selection**

$$\varphi(x) = 1 + \sum_{y=1}^{x-1} \frac{1}{\binom{N-1}{y}} \prod_{z=1}^y \frac{\bar{p}\left(\frac{z}{N}\right)}{p\left(\frac{z}{N}\right)} = 1 + \sum_{y=1}^{x-1} \frac{1}{\binom{N-1}{y}} \psi_N(y)$$

where, with $\tilde{\mu} = (1+s)\bar{\mu}_2 - \mu_1 = \bar{\mu} + s\bar{\mu}_2$, $A_N = N \frac{\bar{\mu}_1}{\tilde{\mu}}$, $B_N = N \frac{\mu_1}{\tilde{\mu}}$, $\psi_N(y) = \frac{\Gamma(A_N)}{\Gamma(A_N-y)} \frac{\Gamma(B_N+1)}{\Gamma(B_N+1+y)}$.

From the preceding analysis, we conclude that with

$$(26) \quad r := \left(\frac{\bar{\mu}_1^{\bar{\mu}_1}}{(\bar{\mu}_2(1+s))^{\bar{\mu}_2(1+s)}} \frac{\mu_1^{\mu_1}}{(1-\bar{\mu}_2(1+s))^{1-\bar{\mu}_2(1+s)}} \right)^{1/[\bar{\mu}+s\bar{\mu}_2]},$$

$$\varphi(N) \underset{N \text{ large}}{\sim} 1 + \frac{1}{N-1} \frac{\bar{\mu}_1}{\mu_1} + \sqrt{(1+s) \frac{\bar{\mu}_2}{\bar{\mu}_1}} r^N.$$

We thus obtained:

Proposition 5. *For the Moran model with bias mechanism (23):*

(i) *If $w = 1 + s < \frac{\bar{\mu}_1}{\bar{\mu}_2}$, the fixation probability tends to 0 geometrically fast because $\varphi(N) \underset{N \text{ large}}{\sim} \sqrt{w \frac{\bar{\mu}_2}{\bar{\mu}_1}} r^N$, leading to $\mathbf{P}(\tau_{1,N} < \tau_{1,0}) = \frac{1}{\varphi(N)} \underset{N \text{ large}}{\sim} \frac{1}{\sqrt{w \frac{\bar{\mu}_2}{\bar{\mu}_1}}} r^{-N}$. The fixation probability of a mutant tends to 0, algebraically fast at rate r now given by (26).*

(ii) *If $w > \frac{\bar{\mu}_1}{\bar{\mu}_2}$, $\varphi(N) \underset{N \text{ large}}{\sim} 1 + \frac{1}{N} \frac{\bar{\mu}_1}{\mu_1}$ leading to $\mathbf{P}(\tau_{1,N} < \tau_{1,0}) = \frac{1}{\varphi(N)} \underset{N \text{ large}}{\sim} 1 / \left(1 + \frac{1}{N} \frac{\bar{\mu}_1}{\mu_1}\right) \rightarrow 1$ from below.*

(iii) *In the critical regime when $w = \frac{\bar{\mu}_1}{\bar{\mu}_2}$ and then $\varphi(N) \underset{N \text{ large}}{\sim} 2 + \frac{1}{N} \frac{\bar{\mu}_1}{\mu_1}$: the fixation probability tends again to 1/2.*

Proof: Regardless of whether $\mu < 1$ or $\mu > 1$, the extinction condition $r > 1$ is achieved whenever

$$\frac{1}{\bar{\mu} + s\bar{\mu}_2} (H(1 - \bar{\mu}_2(1+s)) - H(\mu_1)) > 0.$$

By the above arguments, this leads to the following mutation-selection balance condition: $\mu_1 < 1 - \bar{\mu}_2(1+s)$ which is also

$$w := 1 + s < \frac{\bar{\mu}_1}{\bar{\mu}_2}.$$

The selective advantage of allele A_1 (its relative fitness) should not be too large and the upper bound is given by the ratio of the mutation probabilities $\bar{\mu}_1/\bar{\mu}_2$. Note that $\mu_1 < \mu_2$ (respectively $\mu_1 > \mu_2$) $\Rightarrow \bar{\mu}_1/\bar{\mu}_2 > 1$ (respectively $\bar{\mu}_1/\bar{\mu}_2 < 1$). In the presence of selection, the extinction condition can be achieved even if $\mu_1 > \mu_2$ provided $s < 0$ is small enough. When $\mu_1 = \mu_2$, the extinction of the mutant allele occurs if it is selectively deleterious: $s < 0$. \triangle

To the best of the author's knowledge, the results displayed in Propositions 3 – 5 seem to be new.

3.4. Spectral representation and invariant measure of the Moran model with mutations. Except for some exceptional special cases, the spectral measure associated to the Moran model with general bias mechanism p as in (19) is not known. One of the notable exceptions is the Moran model with positive mutation probabilities μ_1, μ_2 ; see [23], [24] and [17]. The transition matrix P of this model has therefore the BD structure (22).

For the Moran model with mutations, the orthogonal polynomials coincide with the dual Hahn polynomials $R_x(\lambda_k) := H_k(x)$ where, with $\lambda_k = k(k + N\nu - 1)$

$$(27) \quad H_k(x) = {}_3F_2 \left(\left(\begin{matrix} -k, k + N\nu - 1, -x \\ N\nu_1, -N \end{matrix} \right); 1 \right), \quad k = 0, \dots, N$$

are the classical Hahn or Hahn-Eberlein polynomials. We note that (Hahn case) $\mu < 1 \Rightarrow \nu_1, \nu_2 > 0$ with $\nu_1, \nu_2 \rightarrow +\infty$ ($\mu \rightarrow 1^-$) and $\nu_1, \nu_2 \rightarrow 0$ ($\mu \rightarrow 0^+$) whereas (Hahn-Eberlein case) $\mu > 1 \Rightarrow \nu_1, \nu_2 < -1$ with $\nu_1, \nu_2 \rightarrow -\infty$ ($\mu \rightarrow 1^+$) and $\nu_1, \nu_2 \rightarrow -1$ ($\mu \rightarrow 2^-$).

The case $\mu < 1$ ($\mu > 1$) corresponds to a weak (strong) mutation regime.

As a truncated hypergeometric series, the polynomials $R_x(\lambda_k)$ are of degree k in x . In this case, thanks to the 3-term recurrence satisfied by the $H_k(x)$, the eigenvalues of the transition matrix (22) are

$$(28) \quad t_k = 1 - \frac{\lambda_k}{N(N + N\nu)} = 1 - \frac{k}{N} \left(\mu + \frac{k-1}{N} \bar{\mu} \right),$$

depending only on the total mutation pressure μ . In particular, $1 - t_1 = \frac{\mu}{N}$ is the spectral gap and $t_N = \frac{\bar{\mu}}{N}$ is the smallest eigenvalue (with $t_N < 0$ if $\mu > 1$).

- In the weak mutation regime $\mu < 1$, the RW is spectrally positive with

$$0 < t_N = \frac{\bar{\mu}}{N} < \dots < t_1 = 1 - \frac{\mu}{N} < t_0 = 1.$$

- In the strong mutation regime $1 < \mu < 2$, the RW is not spectrally positive with

$$t_N = \frac{\bar{\mu}}{N} < \dots < 0 < \dots < t_1 = 1 - \frac{\mu}{N} < t_0 = 1.$$

When N gets large, there is a fixed (independent of N) number of negative eigenvalues. Only when $\mu = 2$ do we have a number of negative eigenvalues of order $\sqrt{N/2}$.

Invariant measure. Furthermore, with $(a)_x := \Gamma(a+x)/\Gamma(a)$ the rising factorial of a , the invariant probability measure is the generalized bivariate hypergeometric distribution:

$$(29) \quad \pi_x = \binom{N}{x} \frac{(N\nu_1)_x (N\nu_2)_{N-x}}{(N\nu)_N}, \quad x = 0, \dots, N$$

which may also be recast as

$$(30) \quad \pi_x = \frac{\binom{-N\nu_1}{x} \binom{-N\nu_2}{N-x}}{\binom{-N\nu}{N}}, \quad x = 0, \dots, N$$

with the convention $\binom{-a}{x} = \{-a\}_x / x!$, $\{-a\}_x = -a(-a-1)\dots(-a-x+1)$, the falling factorials of $-a$ with $\{-a\}_x = (-1)^x (a)_x$.

The mean and variance of π_x are

$$m_\infty := N \frac{\nu_1}{\nu} = N \frac{\mu_1}{\mu} \quad \text{and} \quad \sigma_\infty^2 := N^2 \frac{\nu_1 \nu_2}{\nu^2} \frac{1 + \nu}{N\nu + 1} = \frac{N^2}{N\mu + \bar{\mu}} \frac{\mu_1 \mu_2}{\mu^2}.$$

Upon scaling, the law of $U_\infty := X_\infty/N$ is very much peaked around its mean $\frac{\mu_1}{\mu}$, with density

$$f_N(u) = N\pi_{Nu} = \frac{e^{NV(u)}}{\int_0^1 du e^{NV(u)}}, \quad u \in (0, 1)$$

where $V(u) = \log W(u)$ and, by Stirling formula

$$W(u) \stackrel{(\mu \leq 1)}{=} \frac{(\nu_1 + u)^{\nu_1 + u}}{\nu_1^{\nu_1} u^u} \frac{(\nu_2 + 1 - u)^{\nu_2 + 1 - u}}{\nu_2^{\nu_2} (1 - u)^{1 - u}} \frac{\nu^\nu}{(\nu + 1)^{\nu + 1}} \quad \text{if } \nu_1, \nu_2 > 0$$

$$W(u) \stackrel{(\mu \geq 1)}{=} \frac{(-\nu_1)^{-\nu_1}}{u^u (-\nu_1 - u)^{-\nu_1 - u}} \frac{(-\nu_2)^{-\nu_2}}{(1 - u)^{1 - u} (-\nu_2 - 1 + u)^{-\nu_2 - 1 + u}} \frac{(-\nu - 1)^{-\nu - 1}}{(-\nu)^{-\nu}} \quad \text{if } \nu_1, \nu_2 < -1$$

$$W(u) \stackrel{(\mu = 1)}{=} \frac{\mu_1^u \bar{\mu}_1^{1 - u}}{u^u (1 - u)^{1 - u}}.$$

We have $V'(u) = 0$ at $u_* = \frac{\nu_1}{\nu} = \frac{\mu_1}{\mu}$ and $V''(u_*) = -\frac{\nu^2}{1 + \nu} \left(\frac{1}{\nu_1} + \frac{1}{\nu_2} \right) = -\frac{\mu^3}{\mu_1 \mu_2} < 0$.

A saddle point estimate gives

$$\int_0^1 du e^{NV(u)} = e^{NV(u_*)} \sqrt{\frac{2\pi}{N|V''(u_*)|}} \quad \text{and so } f_N(u) \sim \sqrt{\frac{N|V''(u_*)|}{2\pi}} e^{-N(V(u_*) - V(u))}.$$

• When both $\nu_1, \nu_2 > 0$ ($\mu < 1$) the distribution (29) is the negative hypergeometric distribution which can be obtained as a beta($N\nu_1, N\nu_2$) mixture of the binomial distribution $\binom{N}{x} p^x (1 - p)^{N - x}$ with parameter $p \in [0, 1]$, namely:

$$\pi_x = \binom{N}{x} \int_0^1 p^x (1 - p)^{N - x} \frac{\Gamma(N\nu)}{\Gamma(N\nu_1)\Gamma(N\nu_2)} p^{N\nu_1 - 1} (1 - p)^{N\nu_2 - 1} dp,$$

with (by Stirling formula)

$$\begin{aligned} \pi_0 &= \frac{\Gamma(N\nu)}{\Gamma(N\nu_1)\Gamma(N\nu_2)} \int_0^1 p^{N\nu_1 - 1} (1 - p)^{N(\nu_2 + 1) - 1} dp = \frac{\Gamma(N\nu)}{\Gamma(N\nu_2)} \frac{\Gamma(N(\nu_2 + 1))}{\Gamma(N(\nu + 1))} \\ &\sim \sqrt{\frac{\nu_2}{\nu}} \frac{\nu + 1}{\nu_2 + 1} \left(\frac{(\nu_2 + 1)^{\nu_2 + 1}}{\nu_2^{\nu_2}} \frac{\nu^\nu}{(\nu + 1)^{\nu + 1}} \right)^N =: \sqrt{\frac{\mu_2}{\mu \bar{\mu}_1}} \rho^N, \end{aligned}$$

displaying exponential decay ($\rho = \left(\frac{\bar{\mu}_1 \mu^\mu}{\mu_2^{\mu_2}} \right)^{1/\bar{\mu}} < 1$). Note that, with

$$\bar{\rho} := \frac{(\nu_1 + 1)^{\nu_1 + 1}}{\nu_1^{\nu_1}} \frac{\nu^\nu}{(\nu + 1)^{\nu + 1}} = \left(\frac{\bar{\mu}_2 \mu^\mu}{\mu_1^{\mu_1}} \right)^{1/\bar{\mu}} < 1$$

it holds that

$$\begin{aligned} \pi_N &= \frac{\Gamma(N\nu)}{\Gamma(N\nu_1)\Gamma(N\nu_2)} \int_0^1 p^{N(\nu_1 + 1) - 1} (1 - p)^{N\nu_2 - 1} dp \\ &= \frac{\Gamma(N\nu)}{\Gamma(N\nu_1)} \frac{\Gamma(N(\nu_1 + 1))}{\Gamma(N(\nu + 1))} \sim \sqrt{\frac{\mu_1}{\mu \bar{\mu}_2}} \bar{\rho}^N. \end{aligned}$$

Note that both ρ and $\bar{\rho} < 1$ because, with $h(x) := -x \log x$, $\mu < 1 \Rightarrow h(\mu) > h(\mu_2) - h(\bar{\mu}_1)$ and $h(\mu) > h(\mu_1) - h(\bar{\mu}_2)$.

Furthermore, the rate r (25) appearing in the fixation probability is $r = \rho/\bar{\rho}$.

The case of symmetric mutations $\mu_1 = \mu_2$ (with $\mu < 1$) is obtained while setting $\nu_1 = \nu_2 = \nu/2$.

- When both $\nu_1, \nu_2 < -1$ ($2 > \mu > 1$) the distribution π_x in (30) constitutes the classical (generalized) hypergeometric distribution arising in sampling without replacement.

By Stirling formula, for all $b > a > 0$ (exponential growth):

$$(31) \quad \binom{Nb}{Na} \sim \frac{1}{\sqrt{2\pi(b-a)N/b}} \left[\frac{b^b}{a^a (b-a)^{b-a}} \right]^N.$$

This shows that, with $\rho < 1$ and $\bar{\rho} < 1$

$$\begin{aligned} \pi_0 &= \frac{\binom{-N\nu_2}{-N\nu}}{\binom{-N\nu}{-N\nu}} \sim \sqrt{\frac{\nu_2(\nu+1)}{\nu(\nu_2+1)}} \left[\frac{(-\nu_2)^{-\nu_2}}{(-\nu_2-1)^{-\nu_2-1}} \frac{(-\nu-1)^{-\nu-1}}{(-\nu)^{-\nu}} \right]^N =: \sqrt{\frac{\mu_2}{\mu\mu_1}} \rho^N \\ \pi_N &= \frac{\binom{-N\nu_1}{-N\nu}}{\binom{-N\nu}{-N\nu}} \sim \sqrt{\frac{\nu_1(\nu+1)}{\nu(\nu_1+1)}} \left[\frac{(-\nu_1)^{-\nu_1}}{(-\nu_1-1)^{-\nu_1-1}} \frac{(-\nu-1)^{-\nu-1}}{(-\nu)^{-\nu}} \right]^N =: \sqrt{\frac{\mu_1}{\mu\mu_2}} \bar{\rho}^N \end{aligned}$$

Note that again

$$\rho = \left(\frac{\bar{\mu}_1^{\mu_1} \mu^{\mu}}{\mu_2^{\mu_2}} \right)^{1/\bar{\mu}} \quad \text{and} \quad \bar{\rho} = \left(\frac{\bar{\mu}_2^{\mu_2} \mu^{\mu}}{\mu_1^{\mu_1}} \right)^{1/\bar{\mu}},$$

with ρ and $\bar{\rho} < 1$ because, with $h(x) := -x \log x$, $\mu > 1 \Rightarrow h(\mu) < h(\mu_2) - h(\bar{\mu}_1)$ and $h(\mu) < h(\mu_1) - h(\bar{\mu}_2)$.

- When both $\nu_1, \nu_2 = -1$ ($\mu_1 = \mu_2 = 1$), we get the standard hypergeometric distribution $\pi_x = \binom{N}{x} \binom{N}{N-x} / \binom{2N}{N}$, with $\pi_0 = \pi_N = \frac{1}{\binom{2N}{N}} \sim \sqrt{\pi N} 4^{-N}$.

The case of symmetric mutations $\mu_1 = \mu_2$ (with $\mu > 1$) is obtained while setting $\nu_1 = \nu_2 = \nu/2$.

- When $\mu_1 = \mu_2 = 1$ (Bernoulli-Laplace), the transition probabilities read: $q_x = \left(\frac{x}{N}\right)^2$, $r_x = 2\frac{x}{N}\left(1 - \frac{x}{N}\right)$, $p_x = \left(1 - \frac{x}{N}\right)^2$. Here, $\pi_x = \binom{N}{x} \binom{N}{N-x} / \binom{2N}{N}$ (the standard hypergeometric distribution), $\mu_k = \frac{2N+1-2k}{2N+1-k} \binom{N}{k} / \binom{2N-k}{N}$ and $t_k = 1 - \frac{k}{N^2} (2N+1-k)$.

The expected return time to $\{0\}$ is $2^{2N}/\sqrt{\pi N}$ which is huge when N is large, whereas the expected return time to $\{N/2\}$ is of order $\sqrt{\pi N}/2$, much smaller.

- On the critical line $\mu_1 + \mu_2 = 1$, $p(u) = \mu_1$ is constant and so the transition probabilities become affine functions of the state: $q_x = \bar{\mu}_1 \frac{x}{N}$, $r_x = \bar{\mu}_1 + (2\mu_1 - 1) \frac{x}{N}$, $p_x = \mu_1 \left(1 - \frac{x}{N}\right)$. Here, $\pi_x = \binom{N}{x} \mu_1^x \bar{\mu}_1^{N-x}$, $\mu_k = \binom{N}{k} \mu_1^k \bar{\mu}_1^{N-k}$ are binomial $\text{bin}(N, \mu_1)$ -distributed and self-dual and $t_k = 1 - \frac{k}{N}$, independent of μ_1 (this RW is spectrally non-negative). The Hahn polynomials boil down to $H_k(x) = {}_2F_1\left(\begin{matrix} -k, -x \\ -N \end{matrix}; \mu_1^{-1}\right)$, $k = 0, \dots, N$, a class of Krawtchouk polynomials. When $\mu_1 = 1/2$ (the lazy Ehrenfest urn), the holding probabilities are $r_x = 1/2$ and both π_x and μ_k are symmetric $\text{bin}(N, 1/2)$ distributed.

Spectral measure when $\mu \neq 1$.

The spectral probability measure is

$$(32) \quad \mu_k = \frac{2k + N\nu - 1}{k + N\nu - 1} \binom{N}{k} \frac{(N\nu_2)_N}{(k + N\nu)_N} \frac{(N\nu_1)_k}{(N\nu_2)_k}, \quad k = 0, \dots, N$$

and the orthogonality relation reads:

$$(33) \quad \sum_{k=0}^N \mu_k R_x(\lambda_k) R_y(\lambda_k) = \frac{(N\nu_2)_N}{N!} \frac{x!(N-x)!}{(N\nu_1)_x (N\nu_2)_{N-x}} \delta_{x,y} = \frac{\pi_0}{\pi_x} \delta_{x,y}.$$

Notice, in our notations: $\mathbf{q}_x(t_k) = R_x(N(N+N\nu)(1-t_k)) = R_x(N^2(1-t_k)/(\bar{\mu}))$.

In particular

$$\begin{aligned} \mathbf{q}_x(t_1) &= : H_1(x) = 1 - \frac{\nu}{N\nu_1}x \\ \mathbf{q}_x(t_2) &= : H_2(x) = 1 - Ax + Bx^2 \end{aligned}$$

where

$$A = \frac{N\nu + 1}{N\nu_1} \frac{2N\nu_1 + 2 + \nu_2 - \nu_1}{(N-1)(N\nu_1+1)}, \quad B = \frac{(N\nu+1)(N\nu+2)}{N^2(N\nu_1+1)(N-1)\nu_1}.$$

Evolution of the conditional mean and variance of X_n given $X_0 = x$.

$\mathbf{q}_x(t_1)$ and $\mathbf{q}_x(t_2)$ being right eigenvectors of P ,

$$\begin{aligned} \mathbf{E}_x H_1(X_n) &= t_1^n H_1(x) \\ \mathbf{E}_x H_2(X_n) &= t_2^n H_2(x) \end{aligned}$$

leading, by virtue of (10), to

Proposition 6. *For the Moran model with mutations*

$$\begin{aligned} \mathbf{E}_x(X_n) &= m_\infty (1 - t_1^n H_1(x)) \xrightarrow{n \rightarrow \infty} m_\infty = N \frac{\nu_1}{\nu} = N \frac{\mu_1}{\mu} \\ \sigma_x^2(X_n) &= \frac{Am_\infty}{B} (1 - t_1^n H_1(x)) - m_\infty^2 (1 - t_1^n H_1(x))^2 - \frac{1}{B} + \frac{1}{B} t_2^n H_2(x) \xrightarrow{n \rightarrow \infty} \sigma_\infty^2. \end{aligned}$$

This shows how the mean and variance of X_n varies, during the course of evolution, as a function of n and the initial condition $X_0 = x$ before reaching the mean and variance of the limit law. Concerning the mean value

$$\mathbf{E}_x(X_n) - m_\infty = t_1^n (x - m_\infty)$$

and depending on $x > m_\infty$ or $x < m_\infty$, $\mathbf{E}_x(X_n)$ approaches monotonically the limiting value m_∞ from above or from below. Note however that for all initial conditions x ,

$$|\mathbf{E}_x(X_n - m_\infty)| \text{ is a decreasing (negentropic) function of } n.$$

Let us now consider the variance. Noting that $\frac{A}{B} - 2m_\infty = (\mu_1 - \mu_2)(2 - \mu) / (\mu - 1)$, rearranging the terms,

$$\sigma_x^2(X_n) = \sigma_\infty^2 - m_\infty H_1(x) \frac{(\mu_2 - \mu_1)(2 - \mu)}{\bar{\mu}} t_1^n - m_\infty^2 H_1(x)^2 t_1^{2n} + \frac{1}{B} H_2(x) t_2^n$$

with $\frac{(\mu_2 - \mu_1)(2 - \mu)}{\bar{\mu}} > 0$ if $(\mu_2 > \mu_1 \text{ and } \mu < 1)$ or $(\mu_2 < \mu_1 \text{ and } \mu > 1)$ and $t_2 < t_1^2 < t_1 < 1$. The corrective term to the limiting variance is a linear combination

of three terms with geometric decay at different rates. For large n , the discrepancy between $\sigma_x^2(X_n)$ and σ_∞^2 is indeed controlled by the slower t_1^n term. In general, the variance does not increase monotonically towards its limiting value, starting from $\sigma_x^2(X_0 = x) = 0$.

Proposition 7. *If for instance starting from $x < N\mu_1/\mu = m_\infty$ (which is $H_1(x) > 0$), depending on $\frac{(\mu_2 - \mu_1)(2 - \mu)}{\mu} > 0$ or not, $\sigma_x^2(X_n)$ approaches its limiting value σ_∞^2 from below or from above, respectively. When $\frac{(\mu_2 - \mu_1)(2 - \mu)}{\mu} < 0$, there is a time $n = n_*(x)$ at which the fluctuation $\sigma_x^2(X_n)$ is maximum and this maximum exceeds σ_∞^2 . The conclusions should be reversed would $x > N\mu_1/\mu = m_\infty$ (or $H_1(x) < 0$).*

It can also be shown ([31], page 47) that, for $m \geq 0$

$$\text{cov}_x(X_n, X_{n+m}) = t_1^m \sigma_x^2(X_n),$$

meaning a geometric asymptotic decay of the auto-covariance also controlled by the second largest eigenvalue t_1 .

Evolution of heterozygosity. Let

$$\mathfrak{h}(X_n) = X_n(N - X_n) =: X_n \bar{X}_n$$

denote the heterozygosity of the chain. This function is the largest when $X_n = N/2$ i.e. when both type population sizes X_n and \bar{X}_n are the closest. The exact evolution of $\mathbf{E}_x \mathfrak{h}(X_n)$ and $\sigma_x^2(\mathfrak{h}(X_n))$ can be derived in a similar way, using (10). For the variance $\sigma_x^2(\mathfrak{h}(X_n))$, $\mathfrak{q}_x(t_4) =: H_4(x)$, as a degree-4 polynomial in x , is needed. It can be shown that $\mathbf{E}_x \mathfrak{h}(X_n)$ tends monotonically to $\sigma_\infty^2 \mu(N-1) \sim N^2 \frac{\mu_1 \mu_2}{\mu^2}$, geometrically fast at rate t_1 , as $n \rightarrow \infty$. The variance tends to 0 geometrically fast as $n \rightarrow \infty$ and it is maximal at some intermediate time $n = n_*(x)$.

Note that, with $\mathfrak{g}(X_n) = X_n - \bar{X}_n = 2X_n - N$ denoting the population gap,

$$\mathbf{E}_x \mathfrak{g}(X_n) = 2m_\infty(1 - t_1^n H_1(x)) - N$$

and this function, just like $\mathbf{E}_x(X_n) - m_\infty$, is not always (i.e. independently of x) decreasing with n .

Increase of the conditional entropy given $X_0 = x$ (H -theorem):

Let $\mathbf{P}(X_n | X_0 = x) = P^n(x, X_n)$, $n \geq 1$. For any initial distribution $\mathbf{P}(X_0 = x)$, with

$$H_n := -\mathbf{E} \log \mathbf{P}(X_n | X_0) = -\sum_{x=0}^N \mathbf{P}(X_0 = x) \mathbf{E} \log \mathbf{P}(X_n | X_0 = x), \quad n \geq 1 \text{ and } H_0 = 0$$

the Shannon conditional entropy of the chain, we have [18]

$$0 \leq H_{n+1} - H_n \leq H_n - H_{n-1}.$$

The first inequality shows that H_n is non-decreasing. It converges to its limiting maximal value, which is the Shannon entropy of the invariant measure. The second inequality shows that H_n is n -concave: the speed of increase of entropy decreases as time passes by. In particular, letting $H_n(x) := -\mathbf{E} \log \mathbf{P}(X_n | X_0 = x)$, for all fixed starting point x , $H_0(x) = 0$ and

$$0 \leq H_{n+1}(x) - H_n(x) \leq H_n(x) - H_{n-1}(x).$$

The increase of entropy is a manifestation of irreversibility apparently in conflict with the reversibility property of all ergodic BD chains and their recurrence of states.

Empirical average and large deviation ([33]).

Given $X_0 = x_0$, let

$$M_n^U := \frac{1}{n} S_n^U := \frac{1}{n} \sum_{m=1}^n U(X_m)$$

denote the empirical average of some bounded measurable output function U of the (ergodic) chain. Examples of interesting U s are $U(x) = x$ in which case M_n^U is directly the empirical average of the X s, but also $U(x) = \mathbf{1}(x = y)$ in which case S_n^U is the fraction of time spent by the walker at state y (its local time). With \mathbf{e}'_{x_0} the row vector which is 0 everywhere but at position x_0 where it is 1 and $\mathbf{1}$ a column vector of 1s, we have

$$\begin{aligned} \mathbf{E} \left(e^{-\beta S_n^U} \mid X_0 = x_0 \right) &= \sum_{x_1, \dots, x_n} e^{-\beta \sum_{m=1}^n U(x_m)} \prod_{m=1}^n P(x_{m-1}, x_m) \\ &= \sum_{x_1, \dots, x_n} \prod_{m=1}^n P(x_{m-1}, x_m) e^{-\beta U(x_m)} = \mathbf{e}'_{x_0} P_\beta^n \mathbf{1}, \end{aligned}$$

where, with \mathbf{w}_β a column weight-vector with x entry $\mathbf{w}_\beta(x) = e^{-\beta U(x)}$, $P_\beta := P D_{\mathbf{w}_\beta}$. By Perron-Frobenius theorem, we have

$$-\frac{1}{n} \log \mathbf{E} \left(e^{-\beta S_n^U} \mid X_0 = x_0 \right) \xrightarrow{n \rightarrow \infty} -\log \lambda(\beta) =: p(\beta),$$

where $\lambda(\beta)$ is the dominant principal eigenvalue of P_β and $p(\beta)$ a concave function of $\beta \in \mathbb{R}$, with $p(0) = 0$. Thus

$$\frac{1}{n} \log \mathbf{P} \left(\frac{1}{n} \sum_{m=1}^n U(X_m) \in d\alpha \right) \xrightarrow{n \rightarrow \infty} f(\alpha) d\alpha$$

where $f(\alpha) = \inf_\beta (\alpha\beta - p(\beta)) \leq 0$ is the concave Legendre transform of the pressure function $p(\beta)$, giving the large deviation rate function of M_n^U . Clearly,

$$\frac{1}{n} \sum_{m=1}^n U(X_m) \xrightarrow[n \rightarrow \infty]{a.s.} \alpha_* := \mathbf{E}(U(X_\infty)),$$

which is the value where $f(\alpha)$ is maximum ($\beta_* = f'(\alpha_*) = 0$), equal to 0.

4. A DETAILED STUDY OF THE SIEGMUND DUAL OF BD CHAINS WITH AN APPLICATION TO THE MUTATION MORAN MODEL

In this Section, we shall illustrate the power of the duality/intertwining relationship by considering the simplest Siegmund dual of a birth and death chain, with the mutation Moran birth and death example in mind. We also address here the question of computing the strong stationary time distribution that helps quantifying the ‘distance’ to equilibrium of the original positive recurrent BD process.

We already know that if P is a spectrally non-negative BD matrix, the chain is stochastically monotone. Here is another sufficient condition relative to the specific ergodic Moran case:

Proposition 8. *Consider the Moran model X_n with bias p satisfying $0 < p(0)$, $p(1) < 1$. If $p(u)$ is non-decreasing, the condition $p_x + q_{x+1} \leq 1$ is fulfilled (the Moran chain is stochastically monotone) and so the Siegmund dual exists.*

Proof: We first need to guarantee $1 - p_0 - q_1 \geq 0$ corresponding to the row $x = 0$ of \widehat{P} . But this is true because $p(1/N) \geq p(0) \geq Np(0) - (N-1)$. Next, when $x = N$, the inequality clearly holds because $p_N = q_{N+1} = 0$. Finally, for $x \in \{1, \dots, N-1\}$

$$\begin{aligned} p_x + q_{x+1} &= p(x/N) - xp(x/N)/N + (x+1)/N - (x+1)p((x+1)/N)/N \\ &\leq p(x/N)(1 - x/N - (x+1)/N) + (x+1)/N \\ &= [p(x/N)(N - 2x - 1) + x + 1]/N \leq 1 \end{aligned}$$

because

$$\begin{aligned} p(x/N) &\leq (1 - x/(N-1))/(1 - 2x/(N-1)) \text{ if } x \in \left\{0, \dots, \left\lfloor \frac{N-1}{2} \right\rfloor\right\} \\ p(x/N) &\geq (1 - x/(N-1))/(1 - 2x/(N-1)) \text{ if } x \in \left\{\left\lceil \frac{N-1}{2} \right\rceil, \dots, N-1\right\}. \end{aligned}$$

\triangle

For the mutation Moran model in the weak mutation regime $0 < \mu < 1$, $p(u)$ is non-decreasing and therefore this chain is stochastically monotone. Although the condition that $p(u)$ is non-decreasing is a sufficient condition to guarantee that the chain is stochastically monotone, it is not necessary.

Proposition 9. *Consider the Moran model X_n with mutation bias p in the strong mutation regime $2 > \mu \geq 1$. Although $p(u)$ is not non-decreasing, the condition $p_x + q_{x+1} \leq 1$ is still fulfilled would N be large enough (in which case the Moran chain with strong mutations is still stochastically monotone). And then the Siegmund dual exists.*

Proof: When $2 > \mu > 1$, $p(u) = \mu_1 + \bar{\mu}u$ is monotone decreasing. We need to check that in that case also $p_x + q_{x+1} \leq 1$ when x varies from 0 to $N-1$, as soon as N is large enough. After some elementary algebra,

$$p_x + q_{x+1} - 1 = -2\bar{\mu} \left(\frac{x}{N}\right)^2 + \left(2\bar{\mu}_1 - \mu - 2\frac{\bar{\mu}}{N}\right) \left(\frac{x}{N}\right) - \left(\bar{\mu}_1 - \frac{\bar{\mu}_1}{N} + \frac{\bar{\mu}}{N^2}\right)$$

The minimum if this convex degree-2 polynomial is attained at the point $x = N \left(2\bar{\mu}_1 - \mu - 2\frac{\bar{\mu}}{N}\right) / (4\bar{\mu})$ so, when N is large, to the left (right) of $x = N/2$ if $\mu_2 > \mu_1$ (if $\mu_1 > \mu_2$) with a negative value of this minimum. The maximum of $p_x + q_{x+1} - 1$ is thus attained either at $x = N-1$ (at $x = 0$), with $p_{N-1} + q_N - 1 = -\bar{\mu}_2 + \frac{1}{N} \left(\bar{\mu}_2 - \frac{\bar{\mu}}{N}\right)$ (respectively $p_0 + q_1 - 1 = -\bar{\mu}_1 + \frac{1}{N} \left(\bar{\mu}_1 - \frac{\bar{\mu}}{N}\right)$). These maxima are non-positive for all N sufficiently large so that $N^2 - N > \max\left(-\frac{\bar{\mu}}{\mu_2}, -\frac{\bar{\mu}}{\mu_1}\right)$. Then, given a Moran model with $2 > \mu > 1$, for all N large enough both $p_{N-1} + q_N - 1 < 0$ and $p_0 + q_1 - 1 < 0$.

Proof: (i) From (37)

$$\tilde{P} = D_{\pi^c}^{-1} \hat{P} D_{\pi^c} \text{ where } D_{\pi^c} = \text{diag}(\pi_0^c, \dots, \pi_N^c).$$

The random walk X_n being reversible, $P' = D_{\pi} P D_{\pi}^{-1}$. Thus:

$$\tilde{P} = D_{\pi^c}^{-1} H' P' H'^{-1} D_{\pi^c} = D_{\pi^c}^{-1} H' D_{\pi} P D_{\pi}^{-1} H'^{-1} D_{\pi^c}.$$

As a result, $\Lambda = D_{\pi^c}^{-1} H' D_{\pi}$ with the right entries $\Lambda(\tilde{x}, x) = \frac{\pi_x}{\pi_{\tilde{x}}} \mathbf{1}(x \leq \tilde{x})$, satisfying $\sum_{x \leq \tilde{x}} \frac{\pi_x}{\pi_{\tilde{x}}} = 1$.

(ii) The last row of Λ is given by $\Lambda(N, x) = \pi_x$ so that once \tilde{X}_n hits state N , the law of X_n is π .

(iii) The first row of Λ is $(1, 0, \dots, 0)$ so that $\mathbf{e}'_0 = \mathbf{e}'_0 \Lambda$.

(iv) Indeed, $\tilde{P}' = D_{\pi^c} \hat{P}' D_{\pi^c}^{-1} = D_{\pi^c} H^{-1} P H D_{\pi^c}^{-1} := K^{-1} P K$ with $K = H D_{\pi^c}^{-1}$ with the right entries. Δ

So \tilde{P} (as the composition of the Siegmund dual of P with a Doob-transform) can be obtained from P either from (i) through a stochastic link Λ or from (iv) through the duality kernel K . This is because BD chains such as the Moran model are reversible.

Strong stationary time. The intertwining construction shows that the original positive recurrent BD chain X_n with transition matrix P may also be viewed as the output (through the link Λ) of an intertwined hidden Markov chain \tilde{X}_n with transition matrix \tilde{P} . Once \tilde{X}_n hits its absorbing state $\{N\}$, the RW X_n is distributed like π , provided both X_n and \tilde{X}_n were started at 0. Furthermore, there exists a bivariate Markov chain (\tilde{X}, X) with transition kernel:

$$(44) \quad P((\tilde{x}, x), (\tilde{y}, y)) = \frac{P(x, y) \cdot \tilde{P}(\tilde{x}, \tilde{y}) \cdot \Lambda(\tilde{y}, y)}{(\Lambda P)(\tilde{x}, y)} \mathbf{1}_{(\Lambda P)(\tilde{x}, y) > 0}$$

where $\tilde{x} \in \{x, x \pm 1\}$, $\tilde{y} \in \{y, y \pm 1\}$. We have: $(\Lambda P)(\tilde{x}, y) > 0$ iff $y \leq \tilde{x} + 1$. We can check that $\sum_y \sum_{\tilde{y}} P((\tilde{x}, x), (\tilde{y}, y)) = 1$ as required for a stochastic matrix.

With $x \in \{0, \dots, N-1\}$, we shall let

$$(45) \quad \tilde{\tau}_{\tilde{x}_0, N} = \inf \left(n : \tilde{X}_n = N \mid \tilde{X}_0 = \tilde{x}_0 \right)$$

be the first hitting time of $\{N\}$ of \tilde{X}_n , starting from state $\tilde{x}_0 \in \{0, \dots, N-1\}$. The random time $\tilde{\tau}_{0, N}$ gives some information on the speed of convergence of the law of the original process X_n to its invariant measure (is a strong stationary time in the sense of Diaconis and Fill, [4]). The facts (39, 40, 41, 42) indeed guarantee that $\tilde{\tau}_{0, N}$ is a strong stationary time of X_n in the sense that $X_{\tilde{\tau}_{0, N}} \stackrel{d}{\sim} \pi$ and is independent of $\tilde{\tau}_{0, N}$ (see [4] Theorems 2.4 and 2.17 or [10] Theorem 2.1). Equivalently (see [1], Prop. 3.2), it holds that:

$$(46) \quad \text{sep}(\pi_n, \pi) \leq \mathbf{P}(\tilde{\tau}_{0, N} > n) \leq \mathbf{E}(\tilde{\tau}_{0, N}) / n$$

where $\pi_n(\cdot) := P^n(0, \cdot)$ is the law of X_n started at 0, π its invariant measure. In (46), the separation discrepancy is defined by: $\text{sep}(\pi_n, \pi) := \sup_y [1 - \pi_n(y) / \pi_y]$. It satisfies $\text{sep}(\pi_n, \pi) \geq \|\pi_n - \pi\|_{TV}$ where $\|\pi_n - \pi\|_{TV} = \frac{1}{2} \sum_y |\pi_n(y) - \pi_y|$ is the total variation distance between π_n and π .

Furthermore, from (39, 42), there is a unique ‘witness’ state say $d = N$ such that either $\tilde{\pi}_n(N) = 0$ or $\tilde{\pi}_n(N) > 0 \Rightarrow \pi_n(d) = \tilde{\pi}_n(N) \pi_d > 0$ showing that this random time is stochastically the smallest since the first inequality in (46) turns out to be an equality (see Remark 2.39 of [4] and Proposition 11 below).

In the context of BD chains absorbed at N , the probability generating function (pgf) of $\tilde{\tau}_{0,N} \geq N$ is [25], [10]:

$$(47) \quad \mathbf{E} \left(z^{\tilde{\tau}_{0,N}} \right) = \prod_{k=1}^N \frac{(1-t_k)z}{1-t_k z}, \quad z \in [0, 1]$$

where $-1 < t_k < +1$, $k = 1, \dots, N$ are the N distinct eigenvalues of both \tilde{P} and P , avoiding $t_0 = 1$. The formula (47) also reads

$$\mathbf{E} \left(z^{\tilde{\tau}_{0,N}} \right) = \sum_{l=1}^N A_l \frac{z(1-t_l)}{1-t_l z}$$

where

$$A_l = \prod_{k \neq l} \frac{1-t_k}{t_l - t_k}$$

and

$$(48) \quad \mathbf{P}(\tilde{\tau}_{0,N} > n) = \sum_{l=1}^N \prod_{k \neq l} \frac{1-t_k}{t_l - t_k} t_l^n, \quad n \geq N-1.$$

Thus, $t_1^{-n} \mathbf{P}(\tilde{\tau}_{0,N} > n) \rightarrow_{n \uparrow \infty} \prod_{k=2}^N \frac{1-t_k}{t_1 - t_k}$ and $\tilde{\tau}_{0,N}$ has geometric tails with exponent t_1 . We have:

$$(49) \quad \mathbf{E}(\tilde{\tau}_{0,N}) = \text{trace} \left(I - \tilde{P} \right)^{-1} = \sum_{k=1}^N (1-t_k)^{-1} \text{ and}$$

$$(50) \quad \sigma^2(\tilde{\tau}_{0,N}) = \sum_{k=1}^N t_k (1-t_k)^{-2} = \sum_{k=1}^N (1-t_k)^{-2} - \sum_{k=1}^N (1-t_k)^{-1}.$$

Note since t_1 is the dominant eigenvalue

$$(51) \quad \sigma^2(\tilde{\tau}_{0,N}) \leq \frac{\mathbf{E}(\tilde{\tau}_{0,N})}{1-t_1}.$$

When the eigenvalues t_k are non-negative, then $\tilde{\tau}_{0,N} \stackrel{d}{=} \sum_{k=1}^N \tau_k$ where the τ_k s are independent with $\tau_k \stackrel{d}{\sim} \text{geom}(1-t_k)$ the geometric distribution with success parameter $1-t_k$ on $\{1, 2, \dots\}$. When the eigenvalues t_k are not all positive, it is not obvious that the above expression (47) of $\mathbf{E}(z^{\tilde{\tau}_{0,N}})$ is indeed a pgf but it turns out that this is the case. Assuming $t_N < \dots < t_{l+1} < 0 \leq t_l < \dots < t_1 < t_0 = 1$, (47) takes on the interpretation of:

$$\tilde{\tau}_{0,N} = \sum_{k=l+1}^N b_k \stackrel{d}{=} \sum_{k=1}^l \tau_k,$$

where $b_k \stackrel{d}{\sim} \text{bernoulli}(1/(1-t_k))$, $\tau_k \stackrel{d}{\sim} \text{geom}(1-t_k)$ and $\tilde{\tau}_{0,N}$ are all mutually independent. We can summarize these results that should mainly be attributed to [1], [4], [10] and [11] as follows:

Proposition 11. *Suppose a Siegmund dual exists for a finite state-space ergodic BD chain X_n . Then there exists a Markov chain \tilde{X}_n , intertwined with X_n , with $\{N\}$ as an absorbing state and fully described in Proposition 10. The random time $\tilde{\tau}_{0,N}$ is a fastest strong stationary time for X_n whose law is characterized either by (47) or (48) involving the spectrum of either P or \tilde{P} , the transition matrices governing the two processes.*

Remark 3: When the RW is symmetric, except for $t_N = -1$ and maybe $\{0\}$, the eigenvalues can be paired because if $t_k \in (0, 1)$ is an eigenvalue, then $t_{N-k} = -t_k$ is one also. The pgf reads

$$\begin{aligned} \mathbf{E}\left(z^{\tilde{\tau}_{0,N}}\right) &= \frac{2z^2}{1+z} \prod_{k=1}^{N_0-1} \frac{(1-t_k^2)z^2}{1-t_k^2 z^2}, \text{ if } N = 2N_0 \\ \mathbf{E}\left(z^{\tilde{\tau}_{0,N}}\right) &= \frac{2z}{1+z} \prod_{k=1}^{N_0} \frac{(1-t_k^2)z^2}{1-t_k^2 z^2}, \text{ if } N = 2N_0 + 1 \end{aligned}$$

depending on whether $\{0\}$ is or is not an eigenvalue.

Computing the mean and variance of $\tilde{\tau}_{0,N}$. Suppose that the t_k are known explicitly. In this case, it is possible to compute $\mathbf{E}(\tilde{\tau}_{0,N})$ and $\sigma^2(\tilde{\tau}_{0,N})$ and find conditions under which

$$(52) \quad \mathbf{E}(\tilde{\tau}_{0,N}) \rightarrow \infty \text{ and } \sigma^2\left(\frac{\tilde{\tau}_{0,N}}{\mathbf{E}(\tilde{\tau}_{0,N})}\right) \rightarrow 0 \text{ as } N \uparrow \infty.$$

If this is the case, then $\frac{\tilde{\tau}_{0,N}}{\mathbf{E}(\tilde{\tau}_{0,N})} \rightarrow 1$ in probability and $\lfloor \mathbf{E}(\tilde{\tau}_{0,N})/2 \rfloor$ is expected to be a cutoff time for X_n started at $\{0\}$.

We now apply this construction to the particular Moran model under study.

The case of the Moran model with mutations. The eigenvalues t_k are known, leading to: $1 - t_k = \frac{k}{N}(\mu + \bar{\mu} \frac{k-1}{N})$. If we scale the characteristic time $\tilde{\tau}_{0,N}$ to N , the summations for large N can be replaced by integrals, leading to the estimate

$$\mu_N \sim N \int_0^1 \frac{dx}{(x + 1/N)(\mu + \bar{\mu}x)} = \frac{N^2}{N\mu - \bar{\mu}} \left(\int_0^1 \frac{dx}{x + 1/N} - \bar{\mu} \int_0^1 \frac{dx}{\mu + \bar{\mu}x} \right),$$

and similarly for $\sigma^2(\tilde{\tau}_{0,N})$. From this, we easily get

Proposition 12. *For the Moran model with mutations with $0 < \mu < 2$, whenever the Siegmund dual exists*

$$\mu_N \sim \frac{N}{\mu} (\log N + \log \mu) \text{ and } \sigma^2(\tilde{\tau}_{0,N}) \sim \left(\frac{N}{\mu}\right)^2,$$

showing that $\sigma^2(\tilde{\tau}_{0,N}/\mathbf{E}(\tilde{\tau}_{0,N})) \sim (\log N)^{-2} \rightarrow 0$.

We have

$$A_l = \prod_{k \neq l} \frac{1-t_k}{t_l-t_k} = \left(\prod_{k \neq l} \left(1 - \frac{l(N\mu + (l-1)\bar{\mu})}{k(N\mu + (k-1)\bar{\mu})} \right) \right)^{-1}$$

and the tail distribution of $\tilde{\tau}_{0,N}$ is

$$(53) \quad \mathbf{P}(\tilde{\tau}_{0,N} > n) = \sum_{l=1}^N \prod_{k \neq l} \frac{1-t_k}{t_l-t_k} t_l^n, \quad n \geq N-1.$$

Remark 4: When $\mu = 2$, the Bernoulli-Laplace RW fails to be stochastically monotone and so the above construction also fails as the chain has no Siegmund dual. However, as was shown in [2], similar conclusions on the fastest strong stationary time holds for the embedded continuous-time version of the Bernoulli-Laplace chain.

Proposition 13. *We have the Gumbel limit law*

$$\tilde{\tau}_{0,N}^{Scaled} := \frac{\tilde{\tau}_{0,N} - \frac{N}{\mu} \log N}{\frac{N}{\mu}} \xrightarrow{d} X \stackrel{d}{\sim} e^{-(x+e^{-x})}, \quad x \in \mathbb{R}.$$

Proof: Indeed, using $t_k/(1-t_k) \sim N/(k\mu)$ to the dominant order in N

$$\begin{aligned} \mathbf{E}\left(e^{it\tilde{\tau}_{0,N}^{Scaled}}\right) &\sim N^{-it} \prod_{k=1}^N \frac{1-t_k}{1-t_k(1+it\mu/N)} \sim N^{-it} \prod_{k=1}^N \frac{k}{k-it} \\ &\rightarrow \Gamma(1-it) \quad \text{as } N \uparrow \infty \end{aligned}$$

which is the Fourier transform of $e^{-(x+e^{-x})}$, $x \in \mathbb{R}$. Δ

With $n_N(\theta) = \left\lfloor \frac{N}{2\mu} (\log N + \theta) \right\rfloor$, then

$$\left\| P^{n_N(\theta)}(0, \cdot) - \pi \right\|_{TV} \xrightarrow{N \uparrow \infty} c(\theta)$$

where $c(\theta) \rightarrow_{\theta \uparrow \infty} 0$ and $c(\theta) \rightarrow_{\theta \downarrow -\infty} 1$. The expected mixing time is $\mu_N \sim \frac{N}{\mu} \log N$ whereas the spectral gap is $1-t_1 = \frac{\mu}{N}$, the product of the 2 of which tends to ∞ . Recalling $\sigma^2(\tilde{\tau}_{0,N}) \leq \frac{\mu_N}{1-t_1}$, $\sigma^2(\tilde{\tau}_{0,N}/\mu_N) = \mu_N^{-2} \sigma^2(\tilde{\tau}_{0,N}) \leq 1/((1-t_1)\mu_N)$, the condition $(1-t_1)\mu_N \rightarrow \infty$ is a sufficient condition for $\sigma^2(\tilde{\tau}_{0,N}/\mu_N) \rightarrow 0$. If this holds, the contribution of $\sum_{k=2}^N (1-t_k)^{-1}$ to μ_N dominates the lead term $(1-t_1)^{-1}$ (see [3] and [5] for recent developments and precisions).

5. RELATED TRANSITION TIMES FOR THE MORAN MODEL WITH MUTATIONS

Consider the ergodic Moran birth and death Markov chain X_n with mutations on the state-space $\{0, \dots, N\}$. By (4), the mean return time to state $\{0\}$ is $\mathbf{E}(\tau_{0,0}^*) = \frac{1}{\pi_0}$.

Thus, with $\rho = \left(\frac{\bar{\mu}_1 \mu^\mu}{\mu_2^{\mu_2}} \right)^{1/\mu} < 1$, regardless of $\mu < 1$ or $2 > \mu > 1$

$$\mathbf{E}(\tau_{0,0}^*) \sim \sqrt{\frac{\mu \bar{\mu}_1}{\mu_2}} (1/\rho)^N \quad \text{as } N \rightarrow \infty.$$

When $\mu = 1$, $\mathbf{E}(\tau_{0,0}^*) = \bar{\mu}_1^{-N} = \mu_2^{-N}$ ($= 2^N$ if $\mu_1 = \mu_2 = 1/2$).

If $\mu = 2$ ($\mu_1 = \mu_2 = 1$), $\mathbf{E}(\tau_{0,0}^*) = 1/\pi_0 = \binom{2N}{N} \sim 4^N / \sqrt{\pi N}$.

We wish now to estimate the expected time it takes for X_n to move from one end of the state-space to the other, that is from state $\{0\}$ to $\{N\}$. Let $\tau_{0,N}$ be the random time of such a sweep⁽⁴⁾. We shall prove the following estimation of its mean value growing geometrically fast:

Proposition 14. *With $\bar{\rho} = \left(\frac{\mu_2 \mu^\mu}{\mu_1}\right)^{1/\mu} < 1$, regardless of $\mu < 1$ or $2 > \mu > 1$, it holds that*

$$\mathbf{E}(\tau_{0,N}) \sim \frac{1}{\mu_2} \sqrt{\frac{\mu \mu_2}{\mu_1}} (1/\bar{\rho})^N \text{ as } N \rightarrow \infty.$$

When $\mu = 1$, $\mathbf{E}(\tau_{0,N}) \sim \frac{\mu_1^{-N}}{\mu_2}$ (= 2^{N+1} in the lazy Ehrenfest case: $\mu_1 = \mu_2 = 1/2$).
When $\mu = 2$, $\mathbf{E}(\tau_{0,N}) \sim 4^N / \sqrt{\pi N}$.

Proof: Let indeed $\tau_{x,x+1}$ be the random time to first hit the state $\{x+1\}$ starting from the state $\{x\}$. Depending on whether the move starting in x is up, down or no move, $\tau_{x,x+1}$ is either 1 or $1 + \tau'_{x,x+1}$ (with $\tau'_{x,x+1}$ a statistical copy of $\tau_{x,x+1}$) or $1 + \tau_{x-1,x} + \tau'_{x,x+1}$. If we let μ_x be the mean value of $\tau_{x,x+1}$, we thus get $\mu_x = p_x + r_x(1 + \mu_x) + q_x(1 + \mu_{x-1} + \mu_x)$, $x \geq 2$. This leads to the recurrence ($\mu_0 = 1/p_0$):

$$\mu_x = \frac{q_x}{p_x} \mu_{x-1} + \frac{1}{p_x}, \quad x \geq 1.$$

This recurrence can be solved explicitly to give

$$\mu_x = \frac{1}{p_x \pi_x} \sum_{y=0}^x \pi_y =: \frac{\pi_x^c}{p_x \pi_x}.$$

Thus

$$\mathbf{E}(\tau_{0,N}) = \sum_{x=0}^{N-1} \mu_x = \sum_{x=0}^{N-1} \frac{1}{p_x \pi_x} \sum_{y=0}^x \pi_y = \sum_{x=0}^{N-1} \frac{1}{p_x} \sum_{y=0}^x \prod_{z=y}^{x-1} \frac{q_{z+1}}{p_z}.$$

Looking at this sum formula, one expects that its leading term is $\frac{1}{p_{N-1} \pi_{N-1}} \sum_{y=0}^{N-1} \pi_y$ because this is where $\sum_{y=0}^x \pi_y$ is the largest and π_x the smallest. Observing $\pi_{N-1} p_{N-1} = \pi_N q_N$ and $q_N = \mu_2$, $\sum_{y=0}^{N-1} \pi_y = 1 - \pi_N \rightarrow 1$, this estimation leads, recalling $\pi_N \sim \sqrt{\frac{\mu_1}{\mu \mu_2}} \bar{\rho}^N$, to

$$\mathbf{E}(\tau_{0,N}) \sim \frac{1}{p_{N-1} \pi_{N-1}} \sim \frac{1}{\mu_2 \pi_N} \sim \frac{1}{\mu_2} \sqrt{\frac{\mu \mu_2}{\mu_1}} (1/\bar{\rho})^N,$$

and this is indeed true because the remaining terms do not contribute significantly, see [6]. When $\mu = 1$,

$$\mathbf{E}(\tau_{0,N}) \sim \frac{1}{\mu_2 \pi_N} = \frac{\mu_1^{-N}}{\mu_2}.$$

When $\mu = 2$ (Laplace-Bernoulli), $\mathbf{E}(\tau_{0,N}) \sim \frac{1}{\mu_2 \pi_N} = \binom{2N}{N}$.

⁴Note that the expected fixation time of a wild-type allele is given by $\mathbf{E}(\tau_{1,N}) = \mathbf{E}(\tau_{0,N}) - 1/p_0$ as a result of $\tau_{0,N} \stackrel{d}{=} \tau_{1,N} + G$, the sum of $\tau_{1,N}$ and an independent geometrically distributed rv $G \geq 1$ with success probability p_0 .

Remark: Regardless of whether $\mu \leq 1$ or $2 > \mu > 1$, $\rho > \bar{\rho}$ iff $\mu_2 > \mu_1$. In this case, the limiting growth rate of the expected value of $\tau_{0,N}$ dominates the one of $\tau_{0,0}^*$. On the contrary, $\mu_1 > \mu_2 \Rightarrow \rho < \bar{\rho}$.

Note that if $\mu_1 = \mu_2 = \mu/2$ (symmetric Moran model), $\rho = \bar{\rho}$ and $\left(\frac{\mathbf{E}(\tau_{0,N})}{\mathbf{E}(\tau_{0,0}^*)}\right)^{1/N} \rightarrow 1$ as $N \rightarrow \infty$, whereas $\frac{\mathbf{E}(\tau_{0,N})}{\mathbf{E}(\tau_{0,0}^*)} \rightarrow 2/\mu$ (> 1 if $\mu < 2$, $= 1$ if $\mu = 2$).

The physical image of the situation is the following: It takes a small $\mathcal{O}(N)$ amount of time to move from $\{0\}$ to the stable equilibrium state $\{Nu_*\}$ because the drift is favorable to this displacement. Once at state $\{Nu_*\}$, X_n returns rapidly and very often to $\{Nu_*\}$ (it takes a $\mathcal{O}(\sqrt{N})$ time to do so) and coming back either to $\{0\}$ or to $\{N\}$ is hard because the drift now pushes X_n inside the domain $\{0, \dots, N\}$...As a result, both $\mathbf{E}(\tau_{0,N})$ and $\mathbf{E}(\tau_{0,0}^*)$ are, comparatively, geometrically very large, with rates computed just before. Recalling it takes about $(N \log N)/\mu$ steps for the chain to reach maximum entropy under the ‘sep-distance’, starting from $\{0\}$, although the return to $\{0\}$ will eventually occur, the time it takes is so huge that such returns will not be observed: In other words, the chain will reach rapidly the invariant measure inside a single excursion. Let us finally check these last statements.

First return time to equilibrium state. Indeed, from (30), using (31)

$$\pi_{Nu_*} = \frac{\binom{-N\nu_1}{Nu_*} \binom{-N\nu_2}{N(1-u_*)}}{\binom{-N\nu}{N}} \sim \frac{1}{\sqrt{2\pi\mu N}}.$$

Thus, by (4), $\mathbf{E}(\tau_{Nu_*,Nu_*}^*) \sim \sqrt{2\pi\mu N} = \mathcal{O}(\sqrt{N})$ and the return time to the equilibrium state is relatively small.

First hitting time of the equilibrium state, starting from $\{0\}$. In this case, we have

$$\mathbf{E}(\tau_{0,Nu_*}) = \sum_{x=0}^{Nu_*-1} \frac{1}{p_x} \sum_{y=0}^x \prod_{z=y}^{x-1} \frac{q_{z+1}}{p_z}.$$

If we scale this characteristic time to N , the summations for large N can be replaced by integrals, leading to the estimate

$$\mathbf{E}(\tau_{0,Nu_*}) \sim N^2 \int_0^{u_*} \frac{dv}{(1-v)(\mu_1 + \bar{\mu}v)} \int_0^v du e^{N(V(u)-V(v))}.$$

Recalling

$$\int_0^v du e^{NV(u)} \sim e^{NV(u_*)} \sqrt{\frac{\pi}{2N|V''(u_*)|}} \text{ if } v \leq u_*,$$

two successive appeals to the saddle point method shows that $\mathbf{E}(\tau_{0,Nu_*}) \sim \frac{\pi N^2}{2N|V''(u_*)|} \frac{\mu^2}{\mu_1 \mu_2} = \frac{\pi}{2\mu} N$, which is indeed $\mathcal{O}(N)$.

Concerning the related first hitting time of state Nu_* , starting from $Nu_* - 1$ (one step below), we have

$$\mathbf{E}(\tau_{Nu_*-1, Nu_*}) \sim \frac{N\mu^2}{\mu_1\mu_2} \int_0^{u_*} du e^{N(V(u)-V(u_*))} \sim \frac{\mu^2}{\mu_1\mu_2} \sqrt{\frac{\pi N}{2|V''(u_*)|}} = \sqrt{\frac{\pi\mu N}{2\mu_1\mu_2}},$$

showing that $\mathbf{E}(\tau_{Nu_*-1, Nu_*})$ is also $\mathcal{O}(\sqrt{N})$, in accordance with $\mathbf{E}(\tau_{Nu_*, Nu_*}^*) = \mathcal{O}(\sqrt{N})$.

Acknowledgments: T.H. acknowledges partial support from the ANR Modélisation Aléatoire en Écologie, Génétique et Évolution (ANR-Manège- 09-BLAN-0215 project) and from the MME-DII Center of Excellence (Modèles Mathématiques et Économiques de la Dynamique, de l' Incertitude et des Interactions, ANR-11-LABX-0023-01 project).

REFERENCES

- [1] Aldous, D.; Diaconis, P. Strong uniform times and finite random walks. *Adv. in Appl. Math.* 8, no. 1, 69–97, (1987).
- [2] Diaconis, P.; Shahshahani, M. Time to reach stationarity in the Bernoulli-Laplace diffusion model. *SIAM J. Math. Anal.* 18 (1987), no. 1, 208-218.
- [3] Diaconis, P. The cutoff phenomenon in finite Markov chains. *Proc. Nat. Acad. Sci. U.S.A.* 93 (1996), no. 4, 1659-1664.
- [4] Diaconis, P.; Fill, J. A. Strong stationary times via a new form of duality. *Ann. Probab.* 18, no. 4 (1990), 1483–1522.
- [5] Diaconis, P.; Saloff-Coste, L. Separation cut-offs for birth and death chains. *Ann. Appl. Probab.* 16 (2006), no. 4, 2098–2122.
- [6] Dunham, B. Fluctuation theory for Moran's genetics model. *J. Math. Anal. Appl.* 210 (1997), no. 2, 777-789.
- [7] Dynkin, E. B. *Markov processes*. Vols. I, II. Translated with the authorization and assistance of the author by J. Fabius, V. Greenberg, A. Maitra, G. Majone. Die Grundlehren der Mathematischen Wissenschaften, Bände 121, 122 Academic Press Inc., Publishers, New York; Springer-Verlag, Berlin-Göttingen-Heidelberg, 1965.
- [8] Ewens, W. J. *Mathematical population genetics. I. Theoretical introduction*. Second edition. *Interdisciplinary Applied Mathematics*, 27. Springer-Verlag, New York, 2004.
- [9] Feller, W. *An introduction to probability theory and its applications*, Vol. 1. John Wiley and Sons, Third Edition, New York, 1968.
- [10] Fill, J. A. The passage time distribution for a birth-and-death chain: Strong stationary duality gives a first stochastic proof. *J. Theoret. Probab.* 22 (2009), no. 3, 543-557.
- [11] Fill, J. A. On hitting times and fastest strong stationary times for skip-free chains. *J. Theoret. Probab.* 22 (2009), no. 3, 587-600.
- [12] Flajolet, P.; Huillet, T. Analytic combinatorics of the Mabinogion urn. Fifth Colloquium on Mathematics and Computer Science, 549-571, *Discrete Math. Theor. Comput. Sci. Proc.*, AI, Assoc. Discrete Math. Theor. Comput. Sci., Nancy, (2008).
- [13] Gantmacher, F. R. *Théorie des matrices*. Tome 2: Questions spéciales et applications. (French) Traduit du Russe par Ch. Sarthou. Collection Universitaire de Mathématiques, No. 19 Dunod, Paris, 1966.
- [14] Garibaldi, U.; Viarengo, P. Exchangeability and non-self-averaging. *Journal of Economic Interaction and Coordination*. Vol. 7. (2012), 2, 181-195.
- [15] Gillespie, J. H. *The Causes of Molecular Evolution*. New York and Oxford: Oxford University Press, 1991.
- [16] Gladstien, K. The characteristic values and vectors for a class of stochastic matrices arising in genetics. *SIAM J. Appl. Math.* 34, no. 4 (1978), 630–642.

- [17] Grünbaum, F. A. Random walks and orthogonal polynomials: some challenges. Probability, geometry and integrable systems, 241-260, Math. Sci. Res. Inst. Publ., 55, Cambridge Univ. Press, Cambridge, (2008).
- [18] Horibe, Y. On the increase of conditional entropy in Markov chains. Transactions of the Tenth Prague Conference on Information Theory, Statistical Decision Functions, Random Processes, Vol. A (Prague, 1986), 391-396, Reidel, Dordrecht, (1988).
- [19] Huillet, T. A Duality Approach to the Genealogies of Discrete Non-Neutral Wright-Fisher Models. Journal of Probability and Statistics, vol. 2009, Article ID 714701, 22 pages, (2009).
- [20] Huillet, T.; Möhle, M. Duality and asymptotics for a class of non-neutral discrete Moran models. J. of Appl. Probab. 46(3) (2009), 866-893.
- [21] Huillet, T.; Möhle, M., On the extended Moran model and its relation to coalescents with multiple collisions. Theor. Pop. Biol., (2012). Online first. DOI: 10.1016/j.tpb.2011.09.004.
- [22] Karlin, S.; McGregor, J. Random walks. Illinois J. Math. 3 (1959), 66-81.
- [23] Karlin, S.; McGregor, J. The Hahn polynomials, formulas and an application. Scripta Math. 26 (1961), 33-46.
- [24] Karlin, S.; McGregor, J. On a genetics model of Moran. Proc. Cambridge Philos. Soc. 58 (1962), 299-311.
- [25] Keilson, J. *Markov chain models - rarity and exponentiality*. Applied Mathematical Sciences, 28. Springer-Verlag, New York-Berlin, 1979.
- [26] Liggett, T. M. *Interacting particle systems*. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], 276. Springer-Verlag, New York, 1985.
- [27] Mansour, T. Combinatorial identities and inverse binomial coefficients. Adv. in Appl. Math. 28 (2002), no. 2, 196-202.
- [28] Maruyama, T. *Stochastic problems in population genetics*. Lecture Notes in Biomathematics, 17. Springer-Verlag, Berlin-New York, 1977.
- [29] Moran, P. A. P. Random processes in genetics. Proc. Cambridge Philos. Soc. 54 (1958), 60-71.
- [30] Sella, G.; Hirsh, A. E. The application of statistical physics to evolutionary biology. PNAS (2005), 102, no 27, 9541-9546.
- [31] Seneta, E. Characterization by orthogonal polynomial systems of finite Markov chains. Probability, statistics and seismology. J. Appl. Probab. 38A (2001), 42-52.
- [32] Siegmund, D. The equivalence of absorbing and reflecting barrier problems for stochastically monotone Markov processes. Ann. Probab. 4, (1976), no. 6, 914-924.
- [33] Varadhan, S. R. S. Large deviations. Ann. Probab. 36 (2008), no. 2, 397-419.
- [34] Whitehurst, T. An application of orthogonal polynomials to random walks. Pacific J. Math. 99, no. 1 (1982), 205-213.

LABORATOIRE DE PHYSIQUE THÉORIQUE ET MODÉLISATION, CNRS-UMR 8089 ET UNIVERSITÉ DE CERGY-PONTOISE, 2 AVENUE ADOLPHE CHAUVIN, 95302, CERGY-PONTOISE, FRANCE, E-MAIL: THIERRY.HUILLET@U-CERGY.FR