Duality and Intertwining with Applications to discrete Moran Models and related Ones

Thierry Huillet
LPTM, CNRS-UMR 8089 et Université de Cergy-Pontoise
ANR MANEGE, X, March 26.

March 25, 2010

1 Discrete-time BD and Moran model[1]

1.1 Generalities on BD processes with finite state-space

P a $(N+1)^2$ tridiagonal (Jacobi) irreducible stochastic matrix

$$P = \begin{bmatrix} r_0 & p_0 & & & \\ q_1 & r_1 & p_1 & & \\ & \ddots & \ddots & \ddots & \\ & & q_{N-1} & r_{N-1} & p_{N-1} \\ & & & q_N & r_N \end{bmatrix},$$

with $p_0 > 0$, $q_x, p_x > 0$, $x = 1, .., N-1$ and $q_N > 0$, transition matrix of discrete-time Markov chain X_n. Invariant probability measure: $\boldsymbol{\pi}' := (\pi_0, \pi_1, .., \pi_N) : \boldsymbol{\pi}' = \boldsymbol{\pi}' P$. $\pi_y = \pi_0 \prod_{z=0}^{y-1} \frac{p_z}{q_{z+1}} > 0$, $y = 1, .., N$, with π_0 : $\sum_{y=0}^{N} \pi_y = 1$.

[1] based on a joint work with Servet Martinez

P diagonally $\sim$ to the symmetric matrix $[P = D_{\boldsymbol{\pi}}^{-1/2} P_S D_{\boldsymbol{\pi}}^{1/2}]$

$$P_S = \begin{bmatrix} r_0 & \sqrt{p_0 q_1} & & & \\ \sqrt{p_0 q_1} & r_1 & \sqrt{p_1 q_2} & & \\ & \ddots & \ddots & \ddots & \\ & & \sqrt{p_{N-2} q_{N-1}} & r_{N-1} & \sqrt{p_{N-1} q_N} \\ & & & \sqrt{p_{N-1} q_N} & r_N \end{bmatrix},$$

so with real eigenvalues. $D_{\boldsymbol{\pi}} := \text{diag}(\pi_0, .., \pi_N)$. As a symmetric matrix, P_S is diagonalizable by an orthogonal transformation and so P is diagonalizable. Such NN RWs are reversible (detailed balance holds). $P = \overleftarrow{P}$ where $\overleftarrow{P}$ is the transition matrix of the time reversed process, given by $\overleftarrow{P}' = D_{\boldsymbol{\pi}} P D_{\boldsymbol{\pi}}^{-1}$ (P' transpose of P).

KMG spectral theory. RW polynomials $(\mathfrak{q}_y(t); y = 0, .., N)$, $t \in [-1, 1]$, determined by $\mathfrak{q}_0(t) = 1$ and 3-term recurrences:

$$t\mathfrak{q}_0(t) = p_0 \mathfrak{q}_1(t) + r_0 \mathfrak{q}_0(t),$$
$$t\mathfrak{q}_y(t) = p_y \mathfrak{q}_{y+1}(t) + r_y \mathfrak{q}_y(t) + q_y \mathfrak{q}_{y-1}(t), \; y \in \{1, N-1\}.$$

SPECTRUM: $(t_k, k = 0, .., N)$ zeroes of polynomial of degree $N+1 : t \to P_{N+1}(t) = t\mathfrak{q}_N(t) - r_N \mathfrak{q}_N(t) - q_N \mathfrak{q}_{N-1}(t)$, namely:

$$S := \{t_k : \; P_{N+1}(t_k) = 0\},$$

with $1 = t_0 > t_1 > .. > t_N \geq -1$. $1 - t_1$ spectral gap.

$\mu(dt) := \sum_{k=0}^{N} \mu_k \delta_{t_k}$ spectral probability measure on $[-1, 1]$ wr to which $(\mathfrak{q}_y(t), y \geq 1)$ orthogonal:

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

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

KMG spectral representation theorem [*Karlin*]

$$P^n(x, y) := \mathbb{P}_x(X_n = y) = \gamma_y \sum_{k=0}^{N} \mu_k t_k^n \mathfrak{q}_x(t_k) \mathfrak{q}_y(t_k). \tag{2}$$

Ergodic theorem, if $t_N > -1$, $\forall x$, $P^n(x,y) \to \mu_0 \gamma_y = \frac{\mu_0}{\pi_0}\pi_y = \pi_y$ as $n \uparrow \infty$.
Symmetric RW. Suppose $N = 2N_0$. RW X_n on $\{0,..,2N_0\}$ given with

$$P = \begin{bmatrix} 0 & 1 & & & \\ q_1 & 0 & p_1 & & \\ & \ddots & \ddots & \ddots & \\ & & q_{2N_0-1} & 0 & p_{2N_0-1} \\ & & & 1 & 0 \end{bmatrix},$$

transition matrix of some symmetric RW reflected at the boundaries $\{0, 2N_0\}$. KMG theory of P^n: $\mu(dt) := \sum_{k=0}^{2N_0} \mu_k \delta_{t_k}$ is the probability measure on $[-1,1]$ wr to which the polynomial $(\mathfrak{q}_y(t), y \geq 1)$ associated to P are orthogonal. Measure symmetric on $[-1,1]$. In particular, $t_{N_0} = 0$ and $t_{2N_0} = -1$, $t_0 = 1$ are eigenvalues of such Ps.

EXAMPLE*:* Gambler RW X_n : $p_x = p$, $q_x = q$, $x = 1,..,N-1$ $(p+q=1)$ with a pure reflection at the endpoints ($p_0 = q_N = 1$). Invariant measure $\boldsymbol{\pi}$ is a truncated geometric distribution. [*Feller*], p. 438 : $t_k = 2\sqrt{pq}\cos\left(\frac{k\pi}{N}\right)$, $k = 1,.,N-1$, $t_0 = 1$, $t_N = -1$. Spectral gap $\to 1 - 2\sqrt{pq} \neq 0$ as $N \uparrow \infty$ if $p \neq 1/2$. Mass of the spectral measure: $\mu_N = \mu_0 = \pi_0 = (1-(p/q)) / \left[2\left(1-(p/q)^N\right)\right]$, $\mu_k = (1-2\mu_0)/(N-1)$, $k = 1,..,N-1$ (with $\mu_0 = 1/(2N)$ when $p = 1/2$). Boundary effects cause deviation from the uniform measure on $\{0,..N\}$. Orthogonal polynomials involve two sine functions.

Special cases. Let

$$S' = \begin{bmatrix} \sqrt{p_0} & & & & \\ \sqrt{q_1} & \sqrt{p_1} & & & \\ & \ddots & \ddots & & \\ & & \sqrt{q_{N-1}} & \sqrt{p_{N-1}} & \\ & & & \sqrt{q_N} & \sqrt{p_N} \end{bmatrix}$$

sub-diagonal matrix. With $D_{\mathbf{r}} := \text{diag}(r_0, .., r_N)$

$$S'S = \begin{bmatrix} p_0 & \sqrt{p_0 q_1} & & & \\ \sqrt{p_0 q_1} & p_1 + q_1 & \sqrt{p_1 q_2} & & \\ & \ddots & \ddots & \ddots & \\ & & \sqrt{p_{N-2} q_{N-1}} & p_{N-1} + q_{N-1} & \sqrt{p_{N-1} q_N} \\ & & & \sqrt{p_{N-1} q_N} & p_N + q_N \end{bmatrix} =$$

$$I - D_{\mathbf{r}} + \begin{bmatrix} 0 & \sqrt{p_0 q_1} & & & \\ \sqrt{p_0 q_1} & 0 & \sqrt{p_1 q_2} & & \\ & \ddots & \ddots & \ddots & \\ & & \sqrt{p_{N-2} q_{N-1}} & 0 & \sqrt{p_{N-1} q_N} \\ & & & \sqrt{p_{N-1} q_N} & 0 \end{bmatrix}.$$

$$\Rightarrow \ P_S = 2D_{\mathbf{r}} - I + S'S$$

is sum of a diag. matrix and a symmetric positive definite matrix $\Rightarrow$ If holding probabilities $r_x \geq 1/2$, $\forall\ x = 0, .., N$, then $\forall\ \mathbf{z} \in \mathbb{R}^{N+1} \setminus \{\mathbf{0}\}$,

$$\mathbf{z}' P_S \mathbf{z} = \sum_{x=0}^{N} (2r_x - 1) |z_x|^2 + |S\mathbf{z}|^2 > 0 \tag{3}$$

and so P_S and then P is positive definite with real > 0 eigenvalues. Jacobi matrix P has all its principal minors non-negative and is oscillatory (that is totally non-negative and such that P^N is totally positive) with all its minors non-negative, see [*Gantmacher*], p. 99. Oscillatory stochastic matrices have distinct positive eigenvalues, with: $1 = t_0 > t_1 > .. > t_N > 0 \Rightarrow r_x \geq 1/2$ simple sufficient condition for P to be oscillatory but not necessary; full condition is $\mathbf{z}' P_S \mathbf{z} > 0$ for all $\mathbf{z} \in \mathbb{R}^{N+1} \setminus \{\mathbf{0}\}$ or

$$\mathbf{z}' P \mathbf{z} = \sum_{x=0}^{N} r_x z_x^2 + \sum_{x=0}^{N-1} (p_x + q_{x+1})\, z_x z_{x+1} > 0. \tag{4}$$

For a simple random walk X_n whose $(N+1)^2$ $-$transition matrix P is spectrally non-negative (respectively spectrally positive), $\exists$ a symmetric random walk Y_m on $\{0, .., 2N\}$ (respectively on $\{0, .., 2N+1\}$), reflected at the endpoints, started at an even integer $\{0, 2, 4, .., 2N\}$ (respectively odd integer $\{1, 3, .., 2N+1\}$), such that $\{X_n\} \stackrel{d}{=} \{Y_{2n}/2\}$, (respectively $\{X_n\} \stackrel{d}{=}$

$\{(Y_{2n}-1)/2\}$). [*Whitehurst*, Th. 2.1 finite-dimensional case]. Spectral measure of RW 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 $\uparrow\downarrow$ probabilities that $Y_m\to Y_{m+1}=Y_m\pm 1$ in one step given Y_m is in state y different from the endpoints, $\alpha_y+\beta_y=1$, then:

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

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.

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

Proof: Under our hypothesis indeed,

$$p_x+q_{x+1}=\alpha_{2x}\alpha_{2x+1}+\beta_{2x+2}\beta_{2x+1}<\alpha_{2x}\alpha_{2x+1}+\beta_{2x+1}<1. \tag{5}$$

Condition $\Rightarrow$ X_n stoch. monot.: $\forall\ y\geq 0$ and $n\geq 0$, $\mathbb{P}_x(X_n>y)\uparrow x$. $\triangle$

CONCLUSION: easy to construct spectrally positive RWs by 'squaring' two symmetric random walks, but given a non-symmetric RW, may be difficult to decide whether it is or not spectrally positive because one needs conversely to check whether all the above αs are probabilities.

Scale function and excursion height. Height $H=h\in\{1,..,N-1\}$ for some of excursion (sample paths of X_n between two consecutive visits to state 0). Event realized iff (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. Two events are $\perp$. $\tau_{x,y}$ first hitting time of y starting from x :

$$\mathbb{P}(H=h)=p_0\mathbb{P}(\tau_{1,h}<\tau_{1,0})\,\mathbb{P}(\tau_{h,0}<\tau_{h,h+1}). \tag{6}$$

Assume $X_0=x$. Let $X_{n\wedge\tau_{x,0}}$ denote the random walk stopped when it first hits 0. Scale function φ of this RW makes $M_n:=\varphi\left(X_{n\wedge\tau_{x,0}}\right)$ a martingale.

Function φ important because $\forall x : 0 < x < h \leq N$, with $\tau_x = \tau_{x,0} \wedge \tau_{x,h}$ first hitting time of $\{0, h\}$ starting from x

$$\mathbb{P}\left(X_{\tau_x} = h\right) = \frac{\varphi\left(x\right)}{\varphi\left(h\right)}. \tag{7}$$

$$\mathbb{P}\left(H = h\right) = p_0 \mathbb{P}\left(\tau_{1,h} < \tau_{1,0}\right) \mathbb{P}\left(\tau_{h,0} < \tau_{h,h+1}\right) = \tag{8}$$
$$p_0 \frac{\varphi\left(1\right)}{\varphi\left(h\right)} \left(1 - \frac{\varphi\left(h\right)}{\varphi\left(h+1\right)}\right), \; h \in \{1, .., N-1\}.$$

Note $\mathbb{P}\left(H \geq h\right) = \frac{p_0}{\varphi(h)}$. The event $H = N$ may also occur.

$\mathbb{P}\left(H = N\right) = \frac{p_0}{\varphi(N)}$: probability of a fixation between 2 consecutive extinction events. Compute φ? We wish to have: $\mathbb{E}_x\left(M_{n+1} \mid M_n = y\right) = y$, leading to

$$\varphi\left(x\right) = q_x \varphi\left(x-1\right) + r_x \varphi\left(x\right) + p_x \varphi\left(x+1\right).$$

$\varphi\left(x\right) =: 1 + \sum_{y=1}^{x-1} \psi\left(y\right)$ where $\psi\left(y\right)$ satisfies: $q_y \psi\left(y-1\right) = p_y \psi\left(y\right)$, with $\psi\left(1\right) := 1$. Thus $\psi\left(y\right) = \prod_{z=1}^{y} \frac{q_z}{p_z}$ and:

$$\varphi\left(x\right) = 1 + \sum_{y=1}^{x-1} \prod_{z=1}^{y} \frac{q_z}{p_z}, \; x \geq 1, \; \varphi\left(0\right) := 0. \tag{9}$$

(8) and (9) characterize law of the excursion height of the RW X_n. $\varphi\left(x\right) \uparrow$ with x. (7) and (9): explicit expression of probability of the event $\tau_{x,h} < \tau_{x,0}$:

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

where $\tau_{x,h}$ is the first hitting time of h starting from x, with $0 < x < h \leq N$. In particular, $h = N$, $\mathbb{P}\left(\tau_{x,N} < \tau_{x,0}\right) = \frac{\varphi(x)}{\varphi(N)}$ fixation probability.

1.2 A BD example: the Moran Model

2-allele Moran model with BIAS mechanism p.

$$p\left(u\right) : u \in [0,1] \rightarrow [0,1], \text{ with } 0 \leq p\left(0\right) \text{ and } p\left(1\right) \leq 1 \tag{10}$$

continuous and $q\left(u\right) := 1 - p\left(u\right)$. Moran model X_n:

$$q_x = \frac{x}{N} q\left(\frac{x}{N}\right), \; r_x = \frac{x}{N} p\left(\frac{x}{N}\right) + \left(1 - \frac{x}{N}\right) q\left(\frac{x}{N}\right), \; p_x = \left(1 - \frac{x}{N}\right) p\left(\frac{x}{N}\right). \tag{11}$$

$p_0 = p(0) > 0$ and $q_N = 1 - p(1) > 0$ (mutations) $\Rightarrow$ chain ergodic with invariant distribution $\frac{\pi_y}{\pi_0} = \prod_{x=1}^{y} \frac{p_{x-1}}{q_x}$. Neutral: $p(u) = u$.

Proposition 2 *Let X_n be a Moran process with transition probabilities (11) for some bias p. Then, $\overline{X}_n := N - X_n$ is again a Moran process with transition probabilities of the type (11) but with the new bias $\overline{p}(u) = 1 - p(1-u)$, namely, $\overline{q}_x = p_{N-x} = \frac{x}{N}\overline{q}\left(\frac{x}{N}\right)$, $\overline{p}_x = q_{N-x} = \left(1 - \frac{x}{N}\right)\overline{p}\left(\frac{x}{N}\right)$ where $\overline{q}(u) := 1 - \overline{p}(u)$. 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 $\overline{p}(u) = p(u)$, then the transition matrix of $\overline{X}_n$ coincides with the one of X_n.*

Moran model with mutations.

$$p(u) = (1 - \mu_2)\, u + \mu_1 (1 - u), \tag{12}$$

(μ_1, μ_2) mutation probabilities in $(0, 1]$. $\mu; = \mu_1 + \mu_2 \leq 1 \Rightarrow p$ non-decreasing.

$\mu_1 = \mu_2 = 1$, heat-exchange Bernoulli-Laplace model [*Feller*] as a borderline example but $p(u) = 1 - u$ is strictly decreasing in this case.

$\mu_1 = \mu_2 = 1/2$ then $p(u) = 1/2 \rightarrow$ aperiodic model amenable (through a suitable time substitution) to the Ehrenfest urn model (N even).

Moran model with mutations and selection (haploid).

$$p(u) = \frac{\mu_1 + u\left((1+s)(1-\mu_2) - \mu_1\right)}{1 + su} \tag{13}$$

composition (fitness first) of selection bias mechanism $\frac{(1+s)u}{1+su}$ $(s > -1)$ with mutation mechanism (12).

Spectral measure associated to Moran model with general p as in (11) not known in general.

Spectral representation of the Moran model with mutations. Notable exception is the Moran model with positive mutation probabilities when $\mu := \mu_1 + \mu_2 \neq 1$, [*Karlin-McGregor*]. Orthogonal polynomials are the dual Hahn polynomials. Eigenvalues [*Ewens*]

$$t_k = 1 - \frac{k}{N}\left(\mu + \frac{k-1}{N}(1-\mu)\right) \tag{14}$$

depend only on total mutation pressure μ. [NOTE: $\mu > 1$: not spectrally > 0.]

Spectral gap: $1 - t_1 = \frac{\mu}{N}$. Invariant measure is generalized bivariate hypergeometric distribution: $[\alpha_i = N\mu_i / (1-\mu)]$

$$\pi_x = \frac{\binom{-\alpha_1}{x}\binom{-\alpha_2}{N-x}}{\binom{-\alpha_1-\alpha_2}{N}},\ x = 0, .., N$$

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

Spectral measure

$$\mu_k = \frac{2k+\alpha_1+\alpha_2-1}{k+\alpha_1+\alpha_2-1}\binom{N}{k}\frac{(\alpha_2)_N}{(k+\alpha_1+\alpha_2)_N}\frac{(\alpha_1)_k}{(\alpha_2)_k},\ k = 0, .., N \tag{15}$$

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 $t_k = 1 - \frac{k}{N^2}(2N+1-k)$.

Critical line $\mu_1 + \mu_2 = 1$, $p(u) = \mu_1$ is constant $\Rightarrow$ transition probabilities are affine functions of state: $q_x = \overline{\mu}_1 \frac{x}{N}$, $r_x = \overline{\mu}_1 + (2\mu_1 - 1)\frac{x}{N}$, $p_x = \mu_1\left(1-\frac{x}{N}\right)$, $\overline{\mu}_1 := 1 - \mu_1$. Here, $\pi_x = \binom{N}{x}\mu_1^x\overline{\mu}_1^{N-x}$, $\mu_k = \binom{N}{k}\mu_1^k\overline{\mu}_1^{N-k}$ are binomial bin(N, μ_1) −distributed and self-dual and $t_k = 1 - \frac{k}{N}$, independent of μ_1. Dual Hahn polynomials boil down to 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 bin$(N, 1/2)$ distributed.

Cases with positive eigenvalues. Conditions on p leading to $r_x \geq \frac{1}{2}$ in which case the RW is spectrally positive ?. Assume $p(u) : u \in (0,1) \to (0,1)$ 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 iff $p(1/2) = 1/2$ 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 $r_x \geq \frac{1}{2}$.

No NSC on the structure of bias p leading to spectrally-+ Moran.

2 Detailed study of the Siegmund dual of BD chains with application to Moran model

Definition 1 *:[Liggett] Two discrete-time Markov processes* $\left(X_n, \widehat{X}_n; n \geq 0\right)$, *state-spaces* $(\mathcal{X}, \mathcal{Y})$, *possibly with substochastic transition kernels, dual wr to some non-singular duality kernel* $H \geq 0$ *on product space* $\mathcal{X} \times \mathcal{Y}$ *if* $\forall x \in \mathcal{X}$, $\forall y \in \mathcal{Y}$, $\forall n \in \mathbb{N}$:

$$\mathbb{E}_x H\left(X_n, y\right) = \mathbb{E}_y H\left(x, \widehat{X}_n\right). \tag{16}$$

If $(\mathcal{X}, \mathcal{Y}) = \{0, .., N\}^2$ finite and identical, duality kernel is square-matrix and transition matrix of dual process $\widehat{X}$, say $\widehat{P}$ obtained from P by:

$$\widehat{P}' = H^{-1} P H,$$

Note that if $\widehat{P}$ is an $H-$dual to P, then P is an $H'-$dual to $\widehat{P}$. If $H = H'$, $\widehat{P}$ is an $H-$dual to P but also P is an $H-$dual to $\widehat{P}$.

Siegmund example. Siegmund duality kernel SK: $H\left(x, y\right) = \mathbf{1}\left(x \leq y\right)$. If, for a given process X_n a process $\widehat{X}_n$ exists satisfying the above condition, $\widehat{X}_n$ is called the Siegmund dual of X_n. Clearly, in the BD case for X_n the condition is that X_n should be SM in that, for all $y \geq 0$ and $n \geq 0$, $\mathbb{P}_x\left(X_n > y\right)$ $\uparrow$ with x.

For positive recurrent BD processes, and for SK, the transition matrix $\widehat{P}$ of the dual process $\widehat{X}_n$:

$$\widehat{P} = \left[\begin{array}{cccccc} r_0 - q_1 & q_1 & & & & \\ p_1 & \widehat{r}_1 & q_2 & & & \\ & \ddots & \ddots & \ddots & & \\ & & p_{N-1} & \widehat{r}_{N-1} & q_N \\ & & & 0 & 1 \end{array}\right],$$

where $\widehat{r}_y := 1 - (p_y + q_{y+1})$, $y \in \{1, ..., N-1\}$ (and $\widehat{q}_y = p_y$, $y = 1, .., N-1$, $\widehat{p}_y = q_{y+1}$, $y = 0, ..., N-1$). Again the one of a BD process (but not of a

Moran BD process if P is a Moran transition matrix).

$$H=\begin{bmatrix} 1 & 1 & & & 1 \\ 0 & 1 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 0 & 1 & 1 \\ 0 & & & 0 & 1 \end{bmatrix},\ H^{-1}=\begin{bmatrix} 1 & -1 & 0 & & 0 \\ 0 & 1 & -1 & & \\ & \ddots & \ddots & \ddots & 0 \\ & & 0 & 1 & -1 \\ 0 & & & 0 & 1 \end{bmatrix}$$

Dual to exist, ensure $p_y + q_{y+1} \leq 1$ for $y \in \{0, .., N-1\}$ which is a NSC to guarantee the stochastic monotonicity of X_n. We already know that if P is a spectrally non-negative BD matrix, the chain is SM. Here is another sufficient condition relative to the specific ergodic Moran case:

Proposition 3 *Moran model X_n with bias p. If $p(u)$ is non-decreasing, the condition $p_x + q_{x+1} \leq 1$ is fulfilled and so the Siegmund dual exists.*

Structure of $\widehat{P}$: the dual process loses mass at $y = 0$ and is absorbed at $y = N$. Add a coffin state $\partial := \{-1\}$ and let:

$$\widehat{P}_\partial := \begin{bmatrix} 1 & 0 & & & & \\ 1-r_0 & r_0-q_1 & q_1 & & & \\ & p_1 & \widehat{r}_1 & q_2 & & \\ & & \ddots & \ddots & \ddots & \\ & & & p_{N-1} & \widehat{r}_{N-1} & q_N \\ & & & & 0 & 1 \end{bmatrix},$$

Corresponding proper BD chain, ${}_\partial\widehat{X}_n$, now has 2 absorbing states, one at $\{-1\}$, one at $\{N\}$. $\widehat{\varphi}(y)$, $y = -1, 0, 1, ..., N$ scale function of ${}_\partial\widehat{X}_n$, solving $\widehat{P}_\partial\widehat{\varphi} = \widehat{\varphi}$, forcing $\widehat{\varphi}(-1) = 0$. We have:

$$\widehat{\varphi}(-1) = 0,\ \widehat{\varphi}(0) = 1,\ \widehat{\varphi}(y) = \gamma_y^c := \sum_{z=0}^{y} \gamma_z = \frac{1}{\pi_0} \sum_{z=0}^{y} \pi_z. \tag{17}$$

Scale function of ${}_\partial\widehat{X}_n$ is the cum-distribution of the inv- measure of the original process. $\widehat{\tau}_y := \widehat{\tau}_{y,-1} \wedge \widehat{\tau}_{y,N}$ infimum of first hitting time of $\{-1\}$ and $\{N\}$ starting from $y \in \{0, ..., N-1\}$. We have:

$$\mathbb{P}_y\left({}_\partial\widehat{X}_{\widehat{\tau}_y} = N\right) = \frac{\widehat{\varphi}(y)}{\widehat{\varphi}(N)} =: \widehat{\phi}(y) = \frac{\gamma_y^c}{\gamma_N^c} = \pi_y^c \tag{18}$$

where $\pi_y^c = \pi_0 \gamma_y^c$ is the cum- inv- probability distribution of π_x.

Doob $h-$transform. New transition matrix $\widetilde{P}_\partial$ by:

$$\widetilde{P}_\partial(x,y) = \frac{\pi_y^c}{\pi_x^c} \widehat{P}_\partial(x,y), \; x,y \in \{-1,0,...,N\}^2. \tag{19}$$

$$\widetilde{P}_\partial = \begin{bmatrix} 1 & 0 & & & & & \\ 0 & r_0 - q_1 & p_0 + q_1 & & & & \\ & \frac{\pi_0^c}{\pi_1^c} p_1 & \widehat{r}_1 & \frac{\pi_2^c}{\pi_1^c} q_2 & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & \frac{\pi_{N-2}^c}{\pi_{N-1}^c} p_{N-1} & \widehat{r}_{N-1} & \frac{\pi_N^c}{\pi_{N-1}^c} q_N \\ & & & & 0 & 1 \end{bmatrix},$$

state $\{-1\}$ becomes isolated and disconnected. Deleting the line and row $\{-1\}$ of $\widetilde{P}_\partial$, we get a stochastic matrix, call it $\widetilde{P}$, of a process $\widetilde{X}_n$ on $\{0,...,N\}$ which is ${}_\partial\widehat{X}_n$ conditioned to first hit state $\{N\}$ before state $\{-1\}$. State 0 of this conditioned BD process now is partially reflecting whereas the remaining absorbing state, say a, is $a = \{N\}$.

$\widetilde{P}$ and P are intertwined through a stochastic link.

Proposition 4 (i) *Matrices $\widetilde{P}$ and P are similar (with the same eigenvalues)*

$$\widetilde{P} = \Lambda P \Lambda^{-1}. \tag{20}$$

Link Λ is: $\Lambda(\widetilde{x},x) = \frac{\pi_x}{\pi_{\widetilde{x}}^c} \mathbf{1}(x \leq \widetilde{x})$, [$\Lambda = D_{\widehat{\boldsymbol{\phi}}}^{-1} H' D_{\boldsymbol{\pi}}$], lower-triangular stochastic *matrix. For all $n \geq 0$*

$$\Lambda(\widetilde{x},x) = \mathbb{P}\left(X_n = x \mid \widetilde{X}_n = \widetilde{x}\right) \tag{21}$$

and $\boldsymbol{\pi}_n' = \widetilde{\boldsymbol{\pi}}_n' \Lambda$ where $\boldsymbol{\pi}_n = \mathbb{P}_{\boldsymbol{\pi}_0}(X_n = \cdot)$ and $\widetilde{\boldsymbol{\pi}}_n = \mathbb{P}_{\widetilde{\boldsymbol{\pi}}_0}\left(\widetilde{X}_n = \cdot\right)$.

(ii) *The link Λ satisfies*

$$\Lambda(N,x) = \pi_x, \; x = 0,..,N. \tag{22}$$

(iii) $\widetilde{\boldsymbol{\pi}}_0' = \boldsymbol{\pi}_0' = \mathbf{e}_0' := (1, 0, ..., 0)$ *are admissible initial distributions of the chains* $\widetilde{X}_n$ *and* X_n*, satisfying*

$$\boldsymbol{\pi}_0' = \widetilde{\boldsymbol{\pi}}_0'\Lambda. \tag{23}$$

(iv) $\widetilde{P}$ *is* K*-dual to* P *:*

$$\widetilde{P}' = K^{-1}PK \tag{24}$$

where $K(x,y) = \frac{1}{\pi_y^c}\mathbf{1}(x \leq y)$.

Strong stationary time [*Diaconis-Fill*]. The intertwining construction shows that original +-recurrent BD chain X_n with transition matrix P may also be viewed as the output (through the link Λ) of a dual hidden Markov chain $\widetilde{X}_n$ with transition matrix $\widetilde{P}$. Once $\widetilde{X}_n$ hits its absorbing state $\{N\}$, the RW X_n is distributed like $\boldsymbol{\pi}$, provided both X_n and $\widetilde{X}_n$ were started at 0. There exists a bivariate Markov chain $\left(\widetilde{X}, X\right)$ with transition kernel:

$$P\left(\left(\widetilde{x}, x\right), \left(\widetilde{y}, y\right)\right) = \frac{P(x,y) \cdot \widetilde{P}(\widetilde{x}, \widetilde{y}) \cdot \Lambda(\widetilde{y}, y)}{(\Lambda P)(\widetilde{x}, y)}\mathbf{1}_{(\Lambda P)(\widetilde{x},y)>0} \tag{25}$$

where $\widetilde{x} \in \{x, x \pm 1\}$, $\widetilde{y} \in \{y, y \pm 1\}$. We have: $(\Lambda P)(\widetilde{x}, y) > 0$ iff $y \leq \widetilde{x} + 1$. With $x \in \{0, ..., N-1\}$

$$\widetilde{\tau}_{\widetilde{x}_0,N} = \inf\left(n : \widetilde{X}_n = N \mid \widetilde{X}_0 = \widetilde{x}_0\right) \tag{26}$$

first hitting time of N of $\widetilde{X}_n$, starting from state $\widetilde{x}_0 \in \{0, ..., N-1\}$. $\widetilde{\tau}_{0,N}$ $\rightarrow$ information on the speed of convergence of the law of original process X_n to its inv- measure (is a SST in the sense of Diaconis and Fill). (20, 21, 22, 23) $\Rightarrow$ $\widetilde{\tau}_{0,N}$ is a SST of X_n in that $X_{\widetilde{\tau}_{0,N}} \overset{d}{\sim} \boldsymbol{\pi}$ and is $\perp$ of $\widetilde{\tau}_{0,N}$ (see [*Diaconis-Fill*] Theorems 2.4 and 2.17 or [*Fill*] Theorem 2.1). Equivalently (see [*Aldous-Diaconis*], Prop. 3.2), it holds that:

$$\text{sep}(\boldsymbol{\pi}_n, \boldsymbol{\pi}) \leq \mathbb{P}(\widetilde{\tau}_{0,N} > n) \leq \mathbb{E}(\widetilde{\tau}_{0,N})/n \tag{27}$$

where $\pi_n(\cdot) = P^n(0, \cdot)$ is the law of X_n started at 0, $\boldsymbol{\pi}$ inv- measure. In (27), the separation discrepancy: $\text{sep}(\boldsymbol{\pi}_n, \boldsymbol{\pi}) := \sup_y [1 - \pi_n(y)/\pi_y]$. Satisfies $\text{sep}(\boldsymbol{\pi}_n, \boldsymbol{\pi}) \geq \|\boldsymbol{\pi}_n - \boldsymbol{\pi}\|_{TV}$ where $\|\boldsymbol{\pi}_n - \boldsymbol{\pi}\|_{TV} = \frac{1}{2}\sum_y |\pi_n(y) - \pi_y|$ is the total variation distance between $\boldsymbol{\pi}_n$ and $\boldsymbol{\pi}$.

From (20, 23), there is a unique 'witness' state say $d=N$ such that either $\widetilde{\pi}_n(N)=0$ or $\widetilde{\pi}_n(N)>0 \Rightarrow \pi_n(d)=\widetilde{\pi}_n(N)\pi_d>0$ showing that this random time is stochastically the smallest since the first inequality in (27) turns out to be an equality (see Remark 2.39 of [*Diaconis-Fill*] and Proposition 13 below).

BD chains absorbed at N: pgf of $\widetilde{\tau}_{0,N}\geq N$ is [*Keilson and Fill*]:

$$\mathbb{E}\left(u^{\widetilde{\tau}_{0,N}}\right)=\prod_{k=1}^{N}\frac{(1-t_k)\,u}{1-t_k u},\ u\in[0,1] \tag{28}$$

where $-1<t_k<+1$, $k=1,...,N$ are the $N\neq$ eigenvalues of both $\widetilde{P}$ and P, avoiding $t_0=1$.

$$\mathbb{P}\left(\widetilde{\tau}_{0,N}>n\right)=\sum_{l=1}^{N}\prod_{k\neq l}\frac{1-t_k}{t_l-t_k}t_l^n,\ n\geq N-1. \tag{29}$$

Thus, $t_1^{-n}\mathbb{P}\left(\widetilde{\tau}_{0,N}>n\right)\to_{n\uparrow\infty}\prod_{k=2}^{N}\frac{1-t_k}{t_1-t_k}$; $\widetilde{\tau}_{0,N}$ has geometric tails with exponent t_1.

$$\mathbb{E}\left(\widetilde{\tau}_{0,N}\right)=\sum_{k=1}^{N}\left(1-t_k\right)^{-1}\ \text{ and} \tag{30}$$

$$\sigma^2\left(\widetilde{\tau}_{0,N}\right)=\sum_{k=1}^{N}\left(1-t_k\right)^{-2}-\sum_{k=1}^{N}\left(1-t_k\right)^{-1}. \tag{31}$$

Note since t_1 is the dominant eigenvalue

$$\sigma^2\left(\widetilde{\tau}_{0,N}\right)\leq\frac{\mathbb{E}\left(\widetilde{\tau}_{0,N}\right)}{1-t_1}. \tag{32}$$

If eigenvalues t_k are ≥ 0, then $\widetilde{\tau}_{0,N}\overset{d}{=}\sum_{k=1}^{N}\tau_k$ where the τ_ks are independent with $\tau_k\overset{d}{\sim}$ geom$(1-t_k)$ on $\{1,2,...\}$. When the eigenvalues t_k are not all positive, not obvious that the above expression (28) of $\mathbb{E}\left(u^{\widetilde{\tau}_{0,N}}\right)$ is indeed a pgf but it is. Assuming $t_N<...<t_{l+1}<0\leq t_l<...<t_1<t_0=1$, (28) interprets as:

$$\widetilde{\tau}_{0,N}-\sum_{k=l+1}^{N}b_k\overset{d}{=}\sum_{k=1}^{l}\tau_k,$$

where $b_k\overset{d}{\sim}$ bernoulli$(1/\left(1-t_k\right))$, $\tau_k\overset{d}{\sim}$ geom$(1-t_k)$ and $\widetilde{\tau}_{0,N}$ are all mutually $\perp$.

Proposition 5 (*A-F*) *Suppose a Siegmund dual exists for a finite state-space ergodic BD chain X_n. Then there exists a Markov chain $\widetilde{X}_n$, intertwined with X_n, with $\{N\}$ as an absorbing state and fully described in Proposition 4. The random time $\widetilde{\tau}_{0,N}$ is a fastest strong stationary time for X_n whose law is characterized either by (28) or (29) involving the spectrum of either P or $\widetilde{P}$, the transition matrices governing the 2 processes.*

Computing the mean and variance of $\widetilde{\tau}_{0,N}$. If t_k are known explicitly. In this case, compute $\mathbb{E}(\widetilde{\tau}_{0,N})$ and $\sigma^2(\widetilde{\tau}_{0,N})$ and find conditions under which

$$\mathbb{E}(\widetilde{\tau}_{0,N}) \to \infty \text{ and } \sigma^2\left(\frac{\widetilde{\tau}_{0,N}}{\mathbb{E}(\widetilde{\tau}_{0,N})}\right) \to 0 \text{ as } N \uparrow \infty. \tag{33}$$

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

Example: Moran model with mutations, with $\mu := \mu_1 + \mu_2$, $\overline{\mu} := 1 - \mu$, because the eigenvalues t_k are known leading to: $1 - t_k = \frac{k}{N}\left(\mu + \overline{\mu}\frac{k-1}{N}\right)$.

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

we easily get

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

showing that $\sigma^2(\widetilde{\tau}_{0,N}/\mathbb{E}(\widetilde{\tau}_{0,N})) \sim (\log N)^{-2} \to 0$. Gumbel weak limit law:

$$\frac{\widetilde{\tau}_{0,N} - \frac{N}{\mu}\log N}{\frac{N}{\mu}} \overset{d}{\to} X \overset{d}{\sim} e^{-\left(x + e^{-x}\right)},\ x \in \mathbb{R}.$$

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

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

where $c(\theta) \to_{\theta\uparrow\infty} 0$ and $c(\theta) \to_{\theta\uparrow-\infty} 1$.

Expected mixing time is $\mu_N \sim \frac{N}{\mu} \log N$ whereas spectral gap is $1 - t_1 = \frac{\mu}{N}$, the product of the 2 of which tends to ∞. Recalling $\sigma^2 \left(\widetilde{\tau}_{0,N} \right) \leq \frac{\mu_N}{1-t_1}$, $\sigma^2 \left(\widetilde{\tau}_{0,N} / \mu_N \right) = \mu_N^{-2} \sigma^2 \left(\widetilde{\tau}_{0,N} \right) \leq 1 / \left(\left(1 - t_1 \right) \mu_N \right)$, the condition $\left(1 - t_1 \right) \mu_N \to \infty$ is a sufficient condition for $\sigma^2 \left(\widetilde{\tau}_{0,N} / \mu_N \right) \to 0$. If this holds, the contribution of $\sum_{k=2}^{N} \left(1 - t_k \right)^{-1}$ to μ_N dominates the lead term $\left(1 - t_1 \right)^{-1}$. (see [*Diaconis-Saloff-Coste*]).

However, in general, the t_k are not known. How to compute differently $\mathbb{E} \left(\widetilde{\tau}_{0,N} \right)$ and $\sigma^2 \left(\widetilde{\tau}_{0,N} \right)$?. Use representation of the Green function in terms of the scale function of the RW.

3 Some extensions

Chain X_n ergodic. Previous construction extends to a wider class of problems than the birth and death (Moran or not) model associated with the Siegmund kernel. In the context of population genetics, the first model one may think of is the Wright-Fisher model with bias $p(u)$ for which

$$P(x,y) = \binom{N}{y} p \left(\frac{x}{N} \right)^y \left(1 - p \left(\frac{x}{N} \right) \right)^{N-y}. \tag{34}$$

Model with binomial transition probabilities not reversible, nor is it in the BD class.

However: The matrix P is TP [*Karlin*] in the sense that for all $\mathbf{x}_q \equiv (x_1, .., x_q)$ with $1 \leq x_1 < .. < x_q \leq N - 1$ and $\mathbf{y}_q \equiv (y_1, .., y_q)$ with $1 \leq y_1 < .. < y_q \leq N - 1$, it has all its minors > 0:

$$\det \left[P \left(\mathbf{x}_q, \mathbf{y}_q \right) \right] > 0.$$

$P(x,y)$ may be written as: $P(x,y) = \phi(x) W(x,y) \psi(y)$ with $\phi(x), \psi(y) > 0$ and $W(x,y)$ a TP kernel:

$$P(x,y) = \left(1 - p \left(\frac{x}{n} \right) \right)^N \left[\frac{p \left(\frac{x}{N} \right)}{1 - p \left(\frac{x}{N} \right)} \right]^y \binom{N}{y}$$

where kernel $W(x,y) = \left[\frac{p\left(\frac{x}{n} \right)}{1 - p\left(\frac{x}{n} \right)} \right]^y \equiv e^{y \tau(x)}$ is TP because $x \to \tau(x)$ is $\uparrow$ since $u \to p(u)$ is $\uparrow$. Therefore, under this assumption, P is spectrally > 0.

$\boldsymbol{\pi}$ invariant measure associated to P:

$$\pi_x = \frac{(I-P)_{x,x}}{\sum_{x=0}^{N}(I-P)_{x,x}},\ x=0,..,N$$

where $(I-P)_{x,x}$ is cofactor of the $(x,x)-$entry of the matrix $I-P$

Fix up how intertwining operates in this more general setting, extending the main steps of the Siegmund dual construction for birth and death chains.

State-spaces: $(\mathcal{X},\mathcal{Y})=\{0,...,N\}^2$. Consider $G\geq 0$ on $(\mathcal{X},\mathcal{Y})$ non-singular. Assume a single state $a\in\mathcal{Y}$ such that $G(x,a)=$ Constant, for all $x\in\mathcal{X}$. Define H by $H(x,y)=G(x,y)/\max_x G(x,y)$. Then, $H\geq 0$ and $H(x,a)=1$, for all $x\in\mathcal{X}$ and $H(x,y)\in[0,1]$, for all (x,y). $\mathbf{e}_a=\left(0,..,0,\overset{a}{1},0,..0\right)'$: $H\mathbf{e}_a=\mathbf{1}$.

P stochastic transition matrix of some ergodic Markov chain X_n on $\mathcal{X}$ with invariant probability measure $\boldsymbol{\pi}>0$.

Time reversal. $\overleftarrow{P}$ stochastic transition matrix of backward (reversed in time) Markov chain

$$\overleftarrow{P}' = D_{\boldsymbol{\pi}} P D_{\boldsymbol{\pi}}^{-1}. \tag{35}$$

If chain reversible (as in the Moran nearest-neighbor RW model), this step is not necessary because $\overleftarrow{P}=P$.

Lemma 6 $\overleftarrow{P}$ *is dual to* P *with respect to the diagonal duality kernel* $D_{\boldsymbol{\pi}}^{-1}$.

When $\overleftarrow{P}=P$, we have self-duality (reversibility).

$H-$**dual.** With H defined as above, suppose the duality relation

$$\widehat{P}' = H^{-1}\overleftarrow{P}H \tag{36}$$

defines a substochastic matrix $\widehat{P}\geq 0$, which is then $H-$dual to $\overleftarrow{P}$.

Remark: This is a key-point: for given P decide for which H, $\widehat{P}$ behaves well. Also, for each specific case study, identify the states which are mass-defective for $\widehat{P}$ in terms of the structure of H.

a absorb: $\left(\widehat{P}\mathbf{1}\right)_a=\mathbf{e}_a'\widehat{P}\mathbf{1}=\mathbf{1}'\widehat{P}'\mathbf{e}_a=\mathbf{1}'H^{-1}\overleftarrow{P}H\mathbf{e}_a=\mathbf{1}'\mathbf{e}_a=1$ and $\left(\widehat{P}\mathbf{1}\right)_x<1$ for at least one $x\neq a$. Duality

$$\mathbb{E}_x H\left(\overleftarrow{X}_n,y\right)=\mathbb{E}_y H\left(x,\widehat{X}_n\right),$$

so that H is within the dual space of $\overleftarrow{P}$. If such $\widehat{P}$ exists, then :

Proposition 7 *It holds that*

$$\widehat{P} = SPS^{-1}, \tag{37}$$

where $S = H'D_{\boldsymbol{\pi}}$. P is $S-$similar to $\widehat{P}$ (with the same eigenvalues). $S\mathbf{1}$ is a right eigenvector to $\widehat{P}$ associated to the unit eigenvalue. We have $S\mathbf{1} = H'\boldsymbol{\pi}$ so that $(S\mathbf{1})_a = 1$ and $(S\mathbf{1})_x < 1$ when $x \neq a$. In particular S is substochastic.

Thus, if the $H-$dual $\widehat{P}$ of $\overleftarrow{P}$ is substochastic, P is similar to $\widehat{P}$ the similarity transform being itself substochastic. We have

$$\mathbf{e}_a'\widehat{P} = \mathbf{e}_a'SPS^{-1} = (H\mathbf{e}_a)' D_{\boldsymbol{\pi}}PS^{-1} = \boldsymbol{\pi}'PD_{\boldsymbol{\pi}}^{-1}H^{-1\prime} = \left(H^{-1}1\right)' = \mathbf{e}_a'$$

and so $\{a\}$ is absorbing for $\widehat{P}$.

Coffin state and scale function. Enlarged stochastic matrix

$$\widehat{P}_\partial = \begin{bmatrix} 1 & \mathbf{0}' \\ \mathbf{1}-\widehat{P}\mathbf{1} & \widehat{P} \end{bmatrix},$$

by adding an extra coffin state, say $\partial := \{-1\}$. $\widehat{P}_\partial$ transition matrix of a proper Markov chain ${}_\partial X_n$ on $\{\partial = -1, 0, 1, .., N\}$ now with the 2 absorbing states $\{\partial, a\}$.

Let $\widehat{\varphi}(y)$, $y = -1, 0, 1, ..., N$ be a scale function of ${}_\partial\widehat{X}_n$, solving $\widehat{P}_\partial\widehat{\boldsymbol{\varphi}} = \widehat{\boldsymbol{\varphi}}$, and imposing $\widehat{\varphi}(\partial) = 0$. Note $\widehat{P}\widehat{\boldsymbol{\phi}} = \widehat{\boldsymbol{\phi}}$ where $\widehat{\boldsymbol{\phi}}$ is the restriction of $\widehat{\boldsymbol{\varphi}}$ to $\{0, 1, .., N\}$ and a solution is, up to a constant, $\widehat{\boldsymbol{\phi}} = S\mathbf{1} = H'\boldsymbol{\pi} > \mathbf{0}$ (because $\widehat{P}S = SP$, P is stochastic and $\boldsymbol{\pi} > \mathbf{0}$). Then, $\widehat{\boldsymbol{\phi}}$ is maximal at $y = a$ and $\widehat{\phi}(a) = 1$.

Let $\widehat{\tau}_x := \widehat{\tau}_{x,\partial} \wedge \widehat{\tau}_{x,a}$ infimum of the first hitting times of $\partial = \{-1\}$ and $\{a\}$ starting from $x \in \{0, ..., N\} \setminus \{a\}$. $\widehat{\varphi}$ scale function for ${}_\partial X_n$ [*Dynkin*]:

$$\mathbb{P}\left({}_\partial\widehat{X}_{\widehat{\tau}_x} = a\right) = \frac{\widehat{\varphi}(x)}{\widehat{\varphi}(a)} =: \widehat{\phi}(x). \tag{38}$$

$\widehat{\phi}(x)$ interprets as the prob. that ${}_\partial X_n$ gets absorbed at $\{a\}$ before $\{\partial\}$ when started at x.

Doob transform and conditioning. Define new transition matrix $\widetilde{P}_\partial$ by:

$$\widetilde{P}_\partial = D_{\widehat{\boldsymbol{\varphi}}}^{-1} \widehat{P}_\partial D_{\widehat{\boldsymbol{\varphi}}}. \tag{39}$$

$\widetilde{P}_\partial \mathbf{1} = \mathbf{1}$ and $\widetilde{P}_\partial$ stochastic. State $\{-1\}$ becomes isolated and disconnected. Deleting the line and row $\{-1\}$ of $\widetilde{P}_\partial$, we get a stochastic matrix, call it $\widetilde{P}$, of a process $\widetilde{X}_n$ on $\{0, ..., N\}$ which corresponds to ${}_\partial\widehat{X}_n$ conditioned to first hit the state $\{a\}$ before the state $\{\partial\}$. The state $\{a\}$ remains the single absorbing state of the reduced RW $\widetilde{X}_n$.

Proposition 8 *We have*

$$\widetilde{P} = D_{\widehat{\boldsymbol{\phi}}}^{-1} \widehat{P} D_{\widehat{\boldsymbol{\phi}}} = \Lambda P \Lambda^{-1} \tag{40}$$

where Λ *is the stochastic link:*

$$\Lambda = D_{\widehat{\boldsymbol{\phi}}}^{-1} S = D_{\widehat{\boldsymbol{\phi}}}^{-1} H' D_{\boldsymbol{\pi}}, \tag{41}$$

satisfying $\Lambda \mathbf{1} = D_{\widehat{\boldsymbol{\phi}}}^{-1} H' \boldsymbol{\pi} = D_{H'\boldsymbol{\pi}}^{-1} H' \boldsymbol{\pi} = \mathbf{1}$. *So* $\widetilde{X}_n$ *and* X_n *are* $\Lambda-$*intertwined if in addition* $\pi_0' = \widetilde{\pi}_0' \Lambda$. *Note that*

$$\Lambda (a, x) = \mathbf{e}_a' \Lambda \mathbf{e}_x = (H \mathbf{e}_a)' D_{\boldsymbol{\pi}} \mathbf{e}_x = \mathbf{1}' D_{\boldsymbol{\pi}} \mathbf{e}_x = \pi_x, \tag{42}$$

so the row $\Lambda (a, \cdot)$ *coincides with* $\boldsymbol{\pi}'$.

P and $\widetilde{P}$ share the same eigenvalues.

Proposition 9 $\widetilde{P}$ *and* $\overleftarrow{P}$ *are* $K-$*duals with*

$$\widetilde{P}' = K^{-1} \overleftarrow{P} K, \text{ and } K = H D_{\widehat{\boldsymbol{\phi}}}^{-1}, \tag{43}$$

but $\widetilde{P}$ *and* P *are not duals in general (they are* $\Lambda-$*intertwined), unless* $P = \overleftarrow{P}$, *i.e. when detailed balance holds for the reversible chain* X_n.

Let (U_n) be a iid uniform sequence generating $\widetilde{X}_n$. As a Markov chain with transition matrix $\widetilde{P}$, the dynamics of $\widetilde{X}_n$ is given by $\widetilde{X}_{n+1} = f\left(\widetilde{X}_n, U_{n+1}\right)$ with:

$$\widetilde{X}_{n+1} = \sum_{y=0}^{N} y \mathbf{1}\left(U_{n+1} \in \left[\widetilde{P}_c\left(\widetilde{X}_n, y-1\right), \widetilde{P}_c\left(\widetilde{X}_n, y\right)\right]\right),$$

where $\widetilde{P}_c(x,y) = \sum_{z=0}^{y} \widetilde{P}(x,z)$. Given $\widetilde{X}_n = x$, $\widetilde{X}_{n+1} = y$ with probability $\widetilde{P}_c(x,y) - \widetilde{P}_c(x,y-1) = \widetilde{P}(x,y)$.

If intertwining holds, then $\exists$ sequence (V_n) of iid uniform random variables, $\perp$ of (U_n) generating $\widetilde{X}_n$, and a measurable function h such that, for each n, $X_n = h\left(\widetilde{X}_n, V_n\right)$:

$$X_n = \sum_{x=0}^{N} x \mathbf{1}\left(V_n \in \left[\Lambda_c\left(\widetilde{X}_n, x-1\right), \Lambda_c\left(\widetilde{X}_n, x\right)\right]\right)$$

where $\Lambda_c(\widetilde{x}, x) = \sum_{y=0}^{x} \Lambda(\widetilde{x}, y)$ is the cum- $\Lambda-$kernel. Given $\widetilde{X}_n = \widetilde{x}$, $X_n = x$ with probability $\Lambda_c(\widetilde{x}, x) - \Lambda_c(\widetilde{x}, x-1) = \Lambda(\widetilde{x}, x)$. With $\boldsymbol{\pi}_n'$ and $\widetilde{\boldsymbol{\pi}}_n'$ the row probabilities of X_n and $\widetilde{X}_n$, for each $n \geq 0$, we thus have $\boldsymbol{\pi}_n' = \widetilde{\boldsymbol{\pi}}_n'\Lambda$. In particular, if $\widetilde{X}_0 \overset{d}{\sim} \widetilde{\pi}_0$, then $X_0 \overset{d}{\sim} \pi_0$ where $\pi_0' = \widetilde{\pi}_0'\Lambda$.

REMARK: Not necessary for intertwining construction to hold that $\widehat{P}$ is a substochastic matrix. $\widehat{P} \geq 0$ is enough. [*H-Martinez*].

COUPLING: For each n, joint stochastic transition matrix:

$$P\left(\left(\widetilde{X}_{n+1} = \widetilde{y}, X_{n+1} = y\right) \mid \left(\widetilde{X}_n = \widetilde{x}, X_n = x\right)\right) =$$

$$\frac{P(x,y) \cdot \widetilde{P}(\widetilde{x}, \widetilde{y}) \cdot \Lambda(\widetilde{y}, y)}{(\Lambda P)(\widetilde{x}, y)} \mathbf{1}\left((\Lambda P)(\widetilde{x}, y) > 0\right).$$

If intertwining: original ergodic Markov chain X_n, governed by P, may be viewed as a random output of the Markov process $\widetilde{X}_n$ governed by $\widetilde{P} = \Lambda P \Lambda^{-1}$ and absorbed at a single state $\{a\}$. Setup reminiscent of filtering theory with $\widetilde{X}_n$ the hidden process and X_n the observable. Peculiarity of intertwining construction: X_n is a Markov output which is itself Markov. Interpret $\widetilde{X}_n$?

Sharpness. Consider two processes $\widetilde{X}_n$ and X_n intertwined through a stochastic link Λ. Interpretation of the link $\Lambda(\widetilde{x}, x) = \mathbb{P}\left(X_n = x \mid \widetilde{X}_n = \widetilde{x}\right)$ for all $n \geq 0$ and so: $\boldsymbol{\pi}_n' = \widetilde{\boldsymbol{\pi}}_n'\Lambda$, $n \geq 0$. Sharpness result alluded to in Remark 2.39 of [*D-F*] p. 1495.

Proposition 10 *Suppose there is a state d of X_n such that $\Lambda \mathbf{e}_d = \pi_d \mathbf{e}_a$. Then $\widetilde{X}_n$ is a sharp dual to X_n in that, given $\widetilde{X}_0 \overset{d}{\sim} \widetilde{\pi}_0$ and $X_0 \overset{d}{\sim} \pi_0$ where*

$\pi_0' = \widetilde{\pi}_0'\Lambda$, *with* $\boldsymbol{\pi}_n = P^n(X_0,\cdot)$, *then:* *sep*$(\boldsymbol{\pi}_n,\boldsymbol{\pi}) = \mathbb{P}\left(\widetilde{\tau}_{\widetilde{X}_0,a} > n\right) < 1$, $n > n_+$ *for some entrance time* $n_+ \geq 0$ *in the absorbing state* a.

Proof: Minimum of $\frac{\pi_n(x)}{\pi_x}$ attained at $x = d$ with $\min_x \frac{\pi_n(x)}{\pi_x} = \widetilde{\pi}_n(a) \Rightarrow$ sep$(\boldsymbol{\pi}_n,\boldsymbol{\pi}) = 1 - \widetilde{\pi}_n(a) = \mathbb{P}\left(\widetilde{\tau}_{\widetilde{X}_0,a} > n\right)$ holds for $n > n_+$ with $n_+ = \inf(n : \widetilde{\pi}_n(a) > 0)$, the first entrance time of $\widetilde{X}_n$ within state a. Before time n_+, $\widetilde{\pi}_n(a) = 0$ and so sep$(\boldsymbol{\pi}_n,\boldsymbol{\pi}) = \mathbb{P}\left(\widetilde{\tau}_{\widetilde{X}_0,a} > n\right) = 1$. $\triangle$

Non-zero entries of Λ are the ones of H' : [recall $\Lambda = D_{\widehat{\boldsymbol{\phi}}}^{-1}S = D_{\widehat{\boldsymbol{\phi}}}^{-1}H'D_{\boldsymbol{\pi}}$]

$$\Lambda = \begin{bmatrix} & & & 0 & \\ & & & 0 & \\ \pi_0 & \pi_1 & \cdots & \pi_d & \pi_N \\ & & & 0 & \\ & & & \vdots & \\ & & & 0 & \end{bmatrix}$$

Initial conditions. If $\mathbf{e}_b'\Lambda = \mathbf{e}_c'$ for some singleton states $(b,c) \Rightarrow \widetilde{X}_0 \overset{d}{\sim} \delta_b$ and $X_0 \overset{d}{\sim} \delta_c$ is an admissible atomic distribution for the initial conditions.

Further examples of kernels of potential interest. Duality kernels for which $H(x,a) = 1$, $\forall\ x \in \mathcal{X} = \{0,...,N\}$ and for some a. Inverses H^{-1} known explicitly : useful to decide whether for given P, the $H-$dual $\widehat{P}$ of $\overleftarrow{P}$ defines a substochastic matrix. If true, problem of *interpreting* the chain with transition matrix $\widetilde{P}$ remains challenging and open problem.

- Siegmund kernels. $G(x,y) = H(x,y) = \mathbf{1}(x \leq y)$. $a = \{N\}$ and $d = \{N\}$. $G(x,y) = H(x,y) = \mathbf{1}(x \geq y)$. $a = \{0\}$ and $d = \{0\}$.

- Pascal kernel. $G(x,y) = \binom{x+y}{y}$, $H(x,y) = \binom{x+y}{y}/\binom{N+y}{y}$. We have $G = LL'$ where L is defined by $L(x,y) = \binom{x}{y}$, $x \geq y$. The Pascal matrix has no zero entries: no expected sharpness.

- Hypergeometric kernel. $G(x,y) = \binom{N-x}{y}$, $H(x,y) = \binom{N-x}{y}/\binom{N}{y}$ satisfying $H = H'$. $a = \{0\}$ and $d = \{N\}$. [Also $H(x,y) = \binom{x}{y}/\binom{N}{y}$].

EXAMPLE (WF): $(P, \boldsymbol{\pi})$ reversible ergodic Moran model with mechanism p, $H(x,y) = \binom{N-x}{y}/\binom{N}{y}$, we can interpret the $H-$dual in terms of a multi-sex backward process [*H-Moehle*], provided $p : [0,1] \to [0,1]$ is such that $q = 1 - p$ is CM $[(-1)^k q^{(k)}(u) \geq 0]$. In [*H-Moehle*] for the Moran model with CM mechanism: $\left(\widehat{P}\mathbf{1}\right)_{a=0} = 1$ and $0 < \left(\widehat{P}\mathbf{1}\right)_x = 1 - \frac{x}{N}p(0) < 1$ for all $x \neq 0$ if $p(0) \in (0,1)$. When $p(0) \neq 0$, all states but $a = 0$ of $\widehat{P}$ are mass-defective.

$\widetilde{P}$, as a normalized version of $\widehat{P}$, transition matrix of skip-free to the left RW that can be described from [*H-M*]. Note : $H(x,N) = 0$ for all $x \neq 0$ so that $\mathbf{e}_N'\Lambda = \mathbf{e}_0'$. $\Lambda\mathbf{e}_N = \pi_N\mathbf{e}_0$ ($a = 0$ and $d = N$) $\Rightarrow \widetilde{\tau}_{N,0} > 0$ stochast. smallest time at which $X_{\widetilde{\tau}_{N,0}} \overset{d}{\sim} \boldsymbol{\pi}$ given $X_0 = 0$ and $\widetilde{X}_0 = N$.

Time $\widetilde{\tau}_{N,0}$ to reach 0 starting from N of the skip-free to the left RW $\widetilde{P}$ (with same spectrum as P) $\overset{d}{=} \widetilde{\tau}_{0,N}$ to reach N starting from 0 of the Siegmund dual RW of same Moran model, namely like (28). In accordance with Theorem 1.2 of [*Fill*]: for a skip-free to the right Markov chain absorbed at N, the law of the time it takes to hit N starting from 0 is given by *K-F* (28).

- Vandermonde kernel. $G(x,y) = x^y$, $H(x,y) = (x/N)^y$. We have $G = LU$ where $L(x,y) = \binom{x}{y}$, $x \geq y$ and $U(x,y) = x!S_{y,x}$, $S_{y,x}$, 2nd kind Stirling numbers. No sharpness.

RK: Ergodic chain governed by P not reversible. Start defining the $H-$dual of P, without appealing first to $\overleftarrow{P}$, miss the idea of a link between $\widetilde{P}$ and P. However, get a similar link between $\widetilde{P}$ and $\overleftarrow{P}$: with H defined as above, suppose

$$\widehat{P}' = H^{-1}PH \tag{44}$$

defines directly a substochastic matrix $\widehat{P} \geq 0$, which is $H-$dual to P, so with $\left(\widehat{P}\mathbf{1}\right)_a = 1$ and $\left(\widehat{P}\mathbf{1}\right)_x < 1$ for at least one $x \neq a$. Using $\overleftarrow{P}' = D_{\boldsymbol{\pi}}PD_{\boldsymbol{\pi}}^{-1}$, we get

$$\widehat{P} = S\overleftarrow{P}S^{-1}, \tag{45}$$

where $S = H'D_{\boldsymbol{\pi}}$. Define the new $\widehat{\boldsymbol{\phi}}$ by: $\widehat{P}\widehat{\boldsymbol{\phi}} = \widehat{\boldsymbol{\phi}}$ for this new $\widehat{P}$. Applying the same Doob transform, this $\widehat{P}$ leads to

$$\widetilde{P} = D_{\widehat{\boldsymbol{\phi}}}^{-1} \widehat{P} D_{\widehat{\boldsymbol{\phi}}} = \Lambda \overleftarrow{P} \Lambda^{-1}, \tag{46}$$

expressing a stochastic link $\Lambda = D_{\widehat{\boldsymbol{\phi}}}^{-1} H' D_{\boldsymbol{\pi}}$ now between $\widetilde{P}$ and $\overleftarrow{P}$ or between the hidden process $\widetilde{X}_n$ and the observable $\overleftarrow{X}_n$ which now is the time-reversed of X_n.

EXAMPLES: (i) assume the bias $p(u)$ appearing in the Wright-Fisher matrix P (34) is such that q is CM, with $p(0) > 0$. Then, using again the hypergeometric duality kernel $H(x, y) = \binom{N-x}{y} / \binom{N}{y}$, it was shown in $[H]$ that the $H-$dual $\widehat{P}$ to P in (44) defines a substochastic matrix $\Rightarrow$ Corresponding $\widetilde{P}$ is $\Lambda-$linked to $\overleftarrow{P}$.

(ii) If $p(u)$ in (34) does not contain mutation effects $\Rightarrow$ chain governed by P transient with 2 absorbing states $\{0, N\} \rightarrow$ substochastic matrix Q obtained from P by deleting its first/last lines and columns. New WF chain has state-space $\{1, .., N-1\}$. Consider ergodic $Q-$process conditioned to never hit boundaries in remote future, governed by

$$\mathcal{P} = \rho^{-1} D_{\boldsymbol{\psi}}^{-1} Q D_{\boldsymbol{\psi}}$$

where $Q\boldsymbol{\psi} = \rho\boldsymbol{\psi}$, $\boldsymbol{\psi} > \mathbf{0}$, $\rho =$Spectral radius of Q. Apply duality-intertwining theory to this new matrix $\mathcal{P}$, using (?) hypergeometric DK $H(x, y) = \binom{x-1}{y-1} / \binom{N-2}{y-1}$, $1 \leq y \leq x \leq N-1$. $\triangle$

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.; Fill, J. A. Strong stationary times via a new form of duality. Ann. Probab. 18, no. 4, 1483–1522, (1990).

[3] Diaconis, P.; Saloff-Coste, L. Separation cut-offs for birth and death chains. Ann. Appl. Probab. 16 (2006), no. 4, 2098–2122.

[4] Diaconis, P.; Shahshahani, M. Time to reach stationarity in the Bernoulli-Laplace diffusion model. SIAM J. Math. Anal. 18 (1987), no. 1, 208–218.

[5] 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.

[6] Ewens, W. J. *Mathematical population genetics. I. Theoretical introduction.* Second edition. Interdisciplinary Applied Mathematics, 27. Springer-Verlag, New York, 2004.

[7] Feller, W. *An introduction to probability theory and its applications,* Vol. **1**. John Wiley and Sons, Third Edition, New York, 1968.

[8] Fill, J. A. The passage time distribution for a birth-and-death chain: Strong stationary duality gives a first stochastic proof. arxiv.org/abs/0707.4042. Journal of Theoretical Probability, Volume 22, Number 3, 543-557, 2009.

[9] Fill, J. A. On hitting times and fastest strong stationary times for skip-free chains. Journal of Theoretical Probability, Volume 22, Number 3, 587-600, 2009.

[10] 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.

[11] Gillespie, J. H. *The Causes of Molecular Evolution.* New York and Oxford: Oxford University Press, 1991.

[12] Gladstien, K. The characteristic values and vectors for a class of stochastic matrices arising in genetics. SIAM J. Appl. Math. 34, no. 4, 630–642, (1978).

[13] 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).

[14] Huillet, T., Moehle, M. Duality and asymptotics for a class of non-neutral discrete Moran models. Journal of Applied Probability 46(3), 866–893, hal-00356083, (2010).

[15] Huillet, T., Martinez, S. Duality and Intertwining for discrete Markov kernels: a relation and examples. preprint: hal-00401732, (2009).

[16] Karlin, S.; McGregor, J. Random walks. Illinois J. Math. 3, 66–81, (1959).

[17] Karlin, S.; McGregor, J. On a genetics model of Moran. Proc. Cambridge Philos. Soc. 58, 299–311, (1962).

[18] Keilson, J. *Markov chain models - rarity and exponentiality.* Applied Mathematical Sciences, 28. Springer-Verlag, New York-Berlin, 1979.

[19] Liggett, T. M. *Interacting particle systems*. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], 276. Springer-Verlag, New York, 1985.

[20] Maruyama, T. *Stochastic problems in population genetics.* Lecture Notes in Biomathematics, 17. Springer-Verlag, Berlin-New York, 1977.

[21] Nagylaki, T. Anecdotal, historical and critical commentaries on Genetics. Gustave Malécot and the transition from classical to modern population genetics. Edited by James F. Crow and William F. Dove, Genetics, 122, 253-268, (1989).

[22] Whitehurst, T. An application of orthogonal polynomials to random walks. Pacific J. Math. 99, no. 1, 205–213, (1982).