On Markov intertwining relations and primal conditioning

Marc Arnaudon[(1)], Koléhè Coulibaly-Pasquier[(2)] and Laurent Miclo[(3)*]

[(1)] Institut de Mathématiques de Bordeaux
Université de Bordeaux, CNRS and INP Bordeaux
marc.arnaudon@math.u-bordeaux.fr

[(2)] Institut Élie Cartan de Lorraine
Université de Lorraine and CNRS
kolehe.coulibaly@univ-lorraine.fr

[(3)*] Toulouse School of Economics
Institut de Mathématiques de Toulouse
miclo.math.cnrs.fr@gmail.com

Abstract

Given an intertwining relation between two finite Markov chains, we investigate how it can be transformed by conditioning the primal Markov chain to stay in a proper subset. A natural assumption on the underlying link kernel is put forward. The three classical examples of discrete Pitman, top-to-random shuffle and absorbed birth-and-death chain intertwinings serve as illustrations.

Keywords: Algebraic and probabilistic Markov intertwining relations, subset conditioning, Doob transform, discrete Pitman intertwining, top-to-random shuffle, birth-and-death Markov chains, strong stationary times.

MSC2020: primary: 60J10, secondary: 15B51.

Competing Interests Statement: The work of L.M. was supported by the grants ANR-17-EURE-0010 and AFOSR-22IOE016.

Data Availability Statement: The used data only consist in the mathematical concepts fully available in the text.

*Corresponding author

1 Introduction: Markov intertwining relations

Markov intertwinings are a kind of commutation or weak similarity relation from a **dual** Markov process toward a **primal** Markov process, enabling us to transfer information between them. They were introduced by Rogers and Pitman [13] to give an alternative proof of Pitman's relation [12] between the Brownian motion and the Bessel-3 process. Subsequently, they were used by Aldous and Diaconis [1] in their analysis of the convergence to equilibrium of the top-to-random shuffle. Nowadays they appeared in various domains in Probability Theory, see for instance the review by Pal and Shkolnikov [11]. Our goal here is to investigate when it is possible, and how, to transform a given Markov intertwining relation when the primal Markov chain is conditioned to stay in a proper subset of its state space.

Our framework is that of a finite state V with discrete time. Our primary objects are thus Markov kernels $P \coloneqq (P(x,x'))_{x,x'\in V}$ and corresponding Markov chains $X \coloneqq (X(n))_{n\in\mathbb{Z}_+}$ taking values in V, standing for the primal Markov chains. The dual Markov chains $Y \coloneqq (Y(n))_{n\in\mathbb{Z}_+}$ take values in another finite state space W and we denote $Q \coloneqq (Q(y,y'))_{y,y'\in W}$ their Markov kernels.

Let Λ be a Markov kernel from W to V, which is called the **link**, we say that an **algebraic Λ-intertwining relation** holds from Q to P, when

$$Q\Lambda \quad = \quad \Lambda P \tag{1}$$

Such a relation from Q to P is always satisfied for some Λ, since it is sufficient to take

$$\forall\ y \in W, \qquad \Lambda(y,\cdot) \quad \coloneqq \quad \pi$$

where π is an invariant probability measure for P (which exists in the finite setting).

So to be able to transfer informations between X and Y, some further assumptions are needed, we are to give a general example below. Nevertheless for the general purpose of this paper, we will just assume that an intertwining relation is given.

We will also be interested in a more precise kind of intertwining relations, enabling to couple X and Y. As above, consider Markov kernels P, Q and Λ, respectively from V to V, from W to W and from W to V. We furthermore assume that for any $x,x' \in V$ such that $P(x,x') > 0$, we are given a Markov kernel on W, say $K_{x,x'} \coloneqq (K_{x,x'}(y,y'))_{y,y'\in W}$. We deduce a Markov kernel $K \coloneqq (K((x,y),(x',y')))_{(x,y),(x',y')\in V\times W}$ on $V \times W$ given by

$$\forall\ (x,y),(x',y') \in V \times W, \qquad K((x,y),(x',y')) \quad \coloneqq \quad P(x,x')K_{x,x'}(y,y') \tag{2}$$

Note that whatever $(x,y) \in V \times W$, the V-marginal of $K((x,y),\cdot)$ is $P(x,\cdot)$. Indeed we compute

$$\begin{aligned}\forall\ x' \in V, \qquad \sum_{y'\in W} K((x,y),(x',y')) \quad &= \quad P(x,x') \sum_{y'\in W} K_{x,x'}(y,y') \\ &= \quad P(x,x')\end{aligned}$$

We would like that K is such that if $(X(0),Y(0))$ is a $V \times W$-random variable satisfying that the conditional law of $X(0)$ knowing $Y(0)$ is given by $\Lambda(Y(0),\cdot)$ and if $(X(1),Y(1))$ is obtained from $(X(0),Y(0))$ through the Markov kernel K, then:

- the conditional law of $Y(1)$ knowing $Y(0)$ is given by $Q(Y(0),\cdot)$,
- the conditional law of $X(1)$ knowing $Y(0)$ and $Y(1)$ is given by $\Lambda(Y(1),\cdot)$.

This amounts to asking that

$$\forall\ y,y' \in W, \forall\ x' \in V, \qquad \sum_{x\in V} \Lambda(y,x)K((x,y),(x',y')) \quad = \quad Q(y,y')\Lambda(y',x') \tag{3}$$

It follows from these considerations that we are only interested in couples $(x, y) \in V \times W$ which are such that $\Lambda(y, x) > 0$. Call $V \ltimes W$ the set of such couples (x, y). Denote

$$\forall\, x \in V, \qquad W_x \;\coloneqq\; \{y \in W \,:\, (x, y) \in V \ltimes W\} \tag{4}$$

For any $x, x' \in V$ with $P(x, x') > 0$, the kernel $K_{x,x'}$ has just to be defined from W_x to $W_{x'}$.

Relation (3) is called a **probabilistic Λ-intertwining relation** from Q to P. By summing it with respect to $y' \in W$, we get the algebraic Λ-intertwining relation from Q to P: $\Lambda P = Q\Lambda$.

Conversely Diaconis and Fill [4] show that it always possible to associate a probabilistic Λ-intertwining relation to an algebraic Λ-intertwining relation.

The above coupling of $(X(0), X(1))$ and $(Y(0), Y(1))$ can be extended by iteration over time into a whole coupling of X and Y, still under the assumption that the conditional law of $X(0)$ knowing $Y(0)$ is given by $\Lambda(Y(0), \cdot)$. This coupling has the following properties for the conditional laws

$$\forall\, n \in \mathbb{Z}_+, \qquad \mathcal{L}(Y(\llbracket 0, n\rrbracket)|X) \;=\; \mathcal{L}(Y(\llbracket 0, n\rrbracket)|X(\llbracket 0, n\rrbracket)) \tag{5}$$

$$\forall\, n \in \mathbb{Z}_+, \qquad \mathcal{L}(X(n)|Y(\llbracket 0, n\rrbracket)) \;=\; \Lambda(Y(n), \cdot) \tag{6}$$

where $X(\llbracket 0, n\rrbracket)$ stands for the truncated process $(X(m))_{m\in\llbracket 0,n\rrbracket}$ and similarly for Y, with $\llbracket m, n\rrbracket \coloneqq \{m, m+1, ..., n\}$, for any $m \leqslant n \in \mathbb{Z}$ (later, we will also use the convention $\llbracket n\rrbracket \coloneqq \llbracket 1, n\rrbracket$ for $n \in \mathbb{N}$).

Relation (6) is a probabilistic counter-part of (1). Relation (5) admits the important consequence that a stopping time for Y is also a stopping time for X, up to enriching its filtration with independent randomness. It has also the practical advantage that Y can be constructed from X in an adapted way. This is important since in the traditional uses of Markov intertwining relations, X is given and we want to construct Y satisfying (5) and (6).

As a motivation, let us give such a classical application of intertwining relations coming from Diaconis and Fill [4]. In addition to a probabilistic Λ-intertwining relation from Q to P, assume that P admits a unique invariant measure π and that Q admits an absorbing point, say $\infty \in W$, such that $\{\infty\}$ is the unique minimal recurrence class of Q (said otherwise, the Dirac mass δ_∞ is the unique invariant measure of Q). From the algebraic intertwining relation, we deduce that $\Lambda(\infty, \cdot) = \pi$. Finally assume furthermore that the respective initial laws of X and Y satisfy $\mathcal{L}(X(0)) = \mathcal{L}(Y(0))\Lambda$, then it is possible to couple X and Y as above.

Denote τ the absorption time of Y:

$$\tau \;\coloneqq\; \min\{n \in \mathbb{Z}_+ \,:\, Y(n) = \infty\}$$

which is a.s. finite due to the above assumptions on ∞. Then τ is a **strong stationary time** for X, namely a finite stopping time (relatively to a filtration generated by X and some independent randomness), such that τ and X_τ are independent and X_τ is distributed according to π.

In particular, we get from Proposition 1.10 of Diaconis and Fill [4] the following bound on the convergence to equilibrium for X in separation:

$$\forall\, n \in \mathbb{Z}_+, \qquad \mathfrak{s}(\mathcal{L}(X(n)), \pi) \;\leqslant\; \mathbb{P}[\tau > n] \tag{7}$$

where the separation discrepancy $\mathfrak{s}(\mu, \nu)$ between two probability measures μ and ν defined on the same measurable space is defined by

$$\mathfrak{s}(\mu, \nu) \;\coloneqq\; \operatorname*{ess\,sup}_{\nu} \left(1 - \frac{d\mu}{d\nu}\right)$$

(where $\frac{d\mu}{d\nu}$ stands for the Radon-Nikodym density of μ with respect to ν).

Thus given a primal ergodic Markov chain X, one is looking for dual absorbed Markov chains Y as above and leading to relevant bounds (7).

As already mentioned, the purpose of this paper is to investigate the behavior of a Markov intertwining under the conditioning of the primal Markov X to stay in a proper subset A of V. When X is ergodic (i.e. irreducible and aperiodic) and A is connected with respect to the underlying graph (whose edge set corresponds to positive probability transitions for P), this question is related to the comparison of the convergence to equilibrium of X with that of the conditioned Markov chain, but we will not address this aspect here. The case of algebraic intertwining relations is treated in the next section, while Section 3 deals with probabilistic intertwining relations.

Sections 4, 5 and 6 respectively consider three examples: the discrete version of the Pitman intertwining between the Brownian motion and the Bessel-3 process (the usual random walk on $\mathbb{Z}$ is conditioned to stay in a finite segment), the top-to-random shuffle (the card initially at the bottom of the deck is conditioned to stay in the first half of the deck after reaching it) and absorbed birth-and-death Markov chains conditioned not to be absorbed. Finally an appendix recalls the relations between conditioning to stay in a proper subset and Doob transforms in the framework of finite state space and discrete time.

2 Algebraic intertwinings

We start by dealing with the simpler case of algebraic intertwinings. The following arguments will serve as a guide for the probabilistic intertwinings, even if a posteriori the algebraic case can be formally deduced from the probabilistic one.

Recall that we are given a transition matrix P on a finite set V. Let A be a proper subset of V, not reduced to a singleton, such that the restriction P_A of P to $A \times A$ is irreducible (i.e. A is P-connected). By Perron-Frobenius theorem, we can find a function $h : V \to \mathbb{R}_+$, positive on A and vanishing on $V \backslash A$, and a number $\theta \in (0, 1]$, such that

$$\forall\ x \in A, \qquad P[h](x) \;=\; \theta h(x) \tag{8}$$

The number θ and the function h are called the **largest Dirichlet eigenvalue** and the **Doob function**.

On A, consider the transition matrix $\widetilde{P}$ defined by

$$\forall\ x, y \in A, \qquad \widetilde{P}(x, y) \;\coloneqq\; \frac{P(x, y)h(y)}{\theta h(x)} \tag{9}$$

It corresponds to the Markov chain conditioned to stay in A, see Appendix A.

Let be given another transition matrix Q on a finite set W, as well as a transition matrix Λ from W to V such that the algebraic Λ-intertwining (1) from Q to P holds.

Our goal here is to present some conditions enabling to deduce an intertwining for $\widetilde{P}$.

Define

$$B \;\coloneqq\; \{y \in W \,:\, \Lambda[A](y) = 1\}$$

where we assume that $B \neq \varnothing$, and consider the transition kernel $\widetilde{\Lambda}$ from B to A given by

$$\forall\ y \in B, \forall\ x \in A, \qquad \widetilde{\Lambda}(y, x) \;\coloneqq\; \frac{\Lambda(y, x)h(x)}{\Lambda[h](y)}$$

Denote

$$\bar{B} \;\coloneqq\; \{y \in W \,:\, \exists\ z \in B \text{ with } Q(z, y) > 0\}$$

and furthermore assume that there exists a non-negative kernel G from $\bar{B}$ to B such that

$$\forall\ y \in \bar{B}, \forall\ x \in A, \qquad \Lambda(y, x) \;=\; \sum_{y' \in B} G(y, y')\Lambda(y', x) \tag{10}$$

Remark 1 For $y \in B$, we can take $G(y, \cdot) = \delta_y$, the Dirac mass at y, so the above assumption can be restricted to the elements $y \in \bar{B} \backslash B$, namely the outside boundary of B (relatively to Q). $\square$

Summing (10) over $x \in A$, and taking into account that $\Lambda(y', A) = 1$ for $y' \in B$, we get

$$\forall\, y \in \bar{B}, \qquad \sum_{y' \in B} G(y, y') \;=\; \Lambda(y, A) \tag{11}$$

Example 1 A particular case of the above situation is when there exist two mappings $\varphi : \bar{B} \to B$ and $g : \bar{B} \to \mathbb{R}_+$ such that

$$\forall\, y \in \bar{B}, \forall\, y' \in B, \qquad G(y, y') \;=\; g(y)\delta_{\varphi(y)}(y')$$

The mapping g is necessarily given by

$$\forall\, y \in \bar{B}, \qquad g(y) \;=\; \Lambda(y, A)$$

due to (11).

The subset-valued situation provides such an example. Assume P irreducible and let π be its invariant probability measure, it gives a positive weight to all points of V. Then we take W to be the set of non-empty subsets of V and Λ is given by

$$\forall\, y \in W, \forall\, x \in V, \qquad \Lambda(y, x) \;=\; \begin{cases} \frac{\pi(x)}{\pi(y)} & , \text{if } x \in y \\ 0 & , \text{otherwise} \end{cases} \tag{12}$$

In this situation we can even define φ and g on the whole set W:

$$\begin{aligned} B &= \{y \in W \,:\, \Lambda[A](y) = 1\} \\ &= \{y \in W \,:\, \pi[A \cap y]/\pi[y] = 1\} \\ &= \{y \in W \,:\, y \subset A\} \end{aligned}$$

$$\forall\, y \in W, \qquad \varphi(y) \;=\; \begin{cases} y \cap A & , \text{if } y \cap A \neq \varnothing \\ A & , \text{otherwise} \end{cases}$$

$$\forall\, y \in W, \qquad g(y) \;=\; \begin{cases} \frac{\pi(\varphi(y))}{\pi(y)} & , \text{if } y \cap A \neq \varnothing \\ 0 & , \text{otherwise} \end{cases}$$

(the definition of $\varphi(y)$ when $y \cap A = \varnothing$ is arbitrary, since then $g(y) = 0$).

In Section 6, we will encounter a measure-valued instead of a subset-valued case where the kernel Λ is not of the form (12).

$\square$

We construct the Markov kernel $\widetilde{Q}$ on B given by

$$\forall\, y, y' \in B, \qquad \widetilde{Q}(y, y') \;\coloneqq\; \frac{1}{Z(y)} \sum_{z \in \bar{B}} Q(y, z) G(z, y') \Lambda[h](y') \tag{13}$$

with the normalizing constant

$$\forall\, y \in B, \qquad Z(y) \;\coloneqq\; \sum_{z \in \bar{B}, y' \in B} Q(y, z) G(z, y') \Lambda[h](y')$$

Note that $\widetilde{Q}$ is well-defined, namely that $Z(y)>0$. Indeed, we have for $y\in B$,

$$\begin{aligned}
Z(y) &= \sum_{z\in\bar{B},\, y'\in B,\, x\in A} Q(y,z)G(z,y')\Lambda(y',x)h(x)\\
&= \sum_{z\in\bar{B},\, x\in A} Q(y,z)\Lambda(z,x)h(x)\\
&= Q[\Lambda[h]](y)\\
&= \Lambda[P[h]](y)\\
&= \Lambda[\mathbb{1}_A P[h]](y)\\
&= \Lambda[\mathbb{1}_A \theta h](y)\\
&= \theta\Lambda[h](y) \qquad (14)\\
&> 0
\end{aligned}$$

Example 1 continued In the set-valued situation of Example 1 we get

$$\forall\ y,y'\in B,\qquad \widetilde{Q}(y,y') = \frac{\Lambda[h](y')\pi(y')}{\theta\Lambda[h](y)}\sum_{z'\subset V\setminus A}\frac{Q(y,y'\cup z')}{\pi(y')+\pi(z')} \tag{15}$$

and

$$\forall\ y',y''\in B,\ z'\subset V\setminus A,\qquad G(y'\cup z',y'') = \frac{\pi(y')}{\pi(y')+\pi(z')}\delta_{y'}(y''). \tag{16}$$

□

Theorem 1 *The algebraic* $\widetilde{\Lambda}$*-intertwining from* $\widetilde{Q}$ *to* $\widetilde{P}$ *holds.*

Proof

Consider a test function f on A. We have to check that

$$\forall\ y\in B,\qquad \widetilde{Q}\widetilde{\Lambda}[f](y) = \widetilde{\Lambda}\widetilde{P}[f](y)$$

Let us start with the r.h.s., we have

$$\begin{aligned}
\widetilde{\Lambda}\widetilde{P}[f](y) &= \frac{1}{\Lambda[h](y)}\Lambda[\mathbb{1}_A h\widetilde{P}[f]](y)\\
&= \frac{1}{\theta\Lambda[h](y)}\Lambda[\mathbb{1}_A P[hf]](y)\\
&= \frac{1}{\theta\Lambda[h](y)}\Lambda[P[hf]](y)\\
&= \frac{1}{\theta\Lambda[h](y)}Q[\Lambda[hf]](y)
\end{aligned}$$

Let us compute the last factor. For any $y\in B$, we have

$$\begin{aligned}
Q[\Lambda[hf]](y) &= \sum_{z\in\bar{B},\, x\in A} Q(y,z)\Lambda(z,x)h(x)f(x)\\
&= \sum_{z\in\bar{B},\, y'\in B,\, x\in A} Q(y,z)G(z,y')\Lambda(y',x)h(x)f(x)\\
&= \sum_{y'\in B,\, z\in\bar{B}} Q(y,z)G(z,y')\Lambda[hf](y')\\
&= Z(y)\sum_{y'\in B}\widetilde{Q}(y,y')\frac{\Lambda[hf](y')}{\Lambda[h](y')}\\
&= Z(y)\widetilde{Q}[\widetilde{\Lambda}[f]](y)
\end{aligned}$$

We have thus proven that

$$\widetilde{\Lambda}\widetilde{P}[f](y) = \frac{Z(y)}{\theta\Lambda[h](y)}\widetilde{Q}[\widetilde{\Lambda}[f]](y)$$

and the desired result follows from (14). ■

Denote $\widehat{Q}$ the kernel on B given by

$$\forall\ y, y' \in B, \qquad \widehat{Q}(y, y') := \sum_{z\in\bar{B}} Q(y, z)G(z, y')$$

This kernel is sub-Markovian, since we compute that for any $y \in B$, taking (11) into account,

$$\begin{aligned}
\sum_{y'\in B} \widehat{Q}(y, y') &= \sum_{y'\in B,\, z\in\bar{B}} Q(y, z)G(z, y') \\
&= \sum_{z\in\bar{B}} Q(y, z)\Lambda(z, A) \\
&\leqslant \sum_{z\in\bar{B}} Q(y, z) \\
&= 1
\end{aligned}$$

To transform $\widehat{Q}$ into a Markov kernel, consider a point $\widehat{\infty}$ not belonging to B and denote $\widehat{B} := B \sqcup \{\widehat{\infty}\}$. Extend $\widehat{Q}$ into a Markov kernel on $\widehat{B}$ by taking

$$\begin{aligned}
\forall\ y \in B, \qquad \widehat{Q}(y, \widehat{\infty}) &= 1 - \widehat{Q}(y, B) \\
\widehat{Q}(\widehat{\infty}, \widehat{\infty}) &= 1
\end{aligned}$$

Consider the mapping $H : \widehat{B} \to \mathbb{R}_+$ given by

$$\forall\ y \in \widehat{B}, \qquad H(y) := \begin{cases} \Lambda[h](y) & \text{, if } y \in B \\ 0 & \text{, if } y = \widehat{\infty} \end{cases}$$

From (13) and (14), we have

$$\begin{aligned}
\forall\ y, y' \in B, \qquad \widetilde{Q}(y, y') &= \frac{1}{Z(y)}\widehat{Q}(y, y')H(y') \\
&= \frac{\widehat{Q}(y, y')H(y')}{\theta H(y)}
\end{aligned}$$

which is similar to (9). Thus if $\widehat{Q}_B$, the restriction of $\widehat{Q}$ to $B \times B$, is irreducible, according to Appendix A, $\widetilde{Q}$ is the conditioning of $\widehat{Q}$ to stay in B. But in situations such as the motivation presented in the introduction, $\widehat{Q}_B$ will not be irreducible, because Q is typically absorbed at a unique point $\infty \in B$ satisfying $\Lambda(\infty, \cdot) = \pi$, the unique invariant probability measure of the kernel P. Instead we are in the case mentioned in Remark 8 of the appendix.

3 Probabilistic intertwinings

We extend here the conditioned algebraic intertwining relation of Theorem 1 into a conditioned probabilistic intertwining relation.

Let the kernels P, Q and Λ and the sets $A \subset V$, $B, \bar{B} \subset W$ be as in Section 2.

Our main assumption there was the existence of non-negative kernel G from $\bar{B}$ to B such that

$$\forall\ y \in \bar{B},\ \forall\ x \in A, \qquad \Lambda(y, x) = \sum_{y'\in B} G(y, y')\Lambda(y', x) \tag{17}$$

enabling us to define the Markov kernel $\widetilde{Q}$ via

$$\forall\ y, y' \in B, \qquad \widetilde{Q}(y, y') \ \coloneqq\ \frac{1}{\theta \Lambda[h](y)} \sum_{z \in \bar{B}} Q(y, z) G(z, y') \Lambda[h](y') \tag{18}$$

The two other important definitions were:

$$\begin{aligned} \forall\ x, x' \in A, \qquad \widetilde{P}(x, x') &\coloneqq \frac{P(x, x')h(x')}{\theta h(x)} \\ \forall\ y \in B,\ \forall\ x \in A, \qquad \widetilde{\Lambda}(y, x) &\coloneqq \frac{\Lambda(y, x)h(x)}{\Lambda[h](y)} \end{aligned}$$

We would like to construct a probabilistic intertwining relation from $\widetilde{Q}$ to $\widetilde{P}$. Namely, for any $x, x' \in A$ with $\widetilde{P}(x, x') > 0$ we want to find a Markov kernel from B_x to $B_{x'}$ $\widetilde{K}_{x,x'} \coloneqq (\widetilde{K}_{x,x'}(y, y'))_{y \in B_x,\, y' \in B_{x'}}$, such that defining

$$\forall\ (x, y), (x', y') \in A \times B, \qquad \widetilde{K}((x, y), (x', y')) \ \coloneqq\ \widetilde{P}(x, x') \widetilde{K}_{x,x'}(y, y') \tag{19}$$

we have

$$\forall\ y, y' \in B,\ \forall\ x' \in A, \qquad \sum_{x \in A} \widetilde{\Lambda}(y, x) \widetilde{K}((x, y), (x', y')) \ =\ \widetilde{Q}(y, y') \widetilde{\Lambda}(y', x') \tag{20}$$

Following (4), we used the notation

$$\forall\ x \in A, \qquad B_x \ \coloneqq\ \{y \in B\ :\ (x, y) \in A \ltimes B\}$$

where

$$\begin{aligned} A \ltimes B &\coloneqq \{(x, y) \in A \times B\ :\ \widetilde{\Lambda}(y, x) > 0\} \\ &= \{(x, y) \in A \times B\ :\ \Lambda(y, x) > 0\} \end{aligned}$$

since h is positive on A.

Here is our main result:

Theorem 2 *The probabilistic $\widetilde{\Lambda}$ intertwining relation from $\widetilde{Q}$ to $\widetilde{P}$ is satisfied if we take for any $x, x' \in A$ with $\widetilde{P}(x, x') > 0$,*

$$\forall\ y \in B_x,\ y' \in B_{x'}, \qquad \widetilde{K}_{x,x'}(y, y') \ \coloneqq\ \sum_{z \in \bar{B} \cap W_{x'}} K_{x,x'}(y, z) H(z, y', x')$$

where $W_{x'}$ is defined in (4), and with

$$\forall\ x' \in A,\ \forall\ y' \in B_{x'},\ \forall\ z \in \bar{B} \cap W_{x'}, \qquad H(z, y', x') \ \coloneqq\ \frac{G(z, y')\Lambda(y', x')}{\Lambda(z, x')}$$

(note that $\Lambda(z, x') > 0$ for $z \in \bar{B} \cap W_{x'}$).

We do have that $(\widetilde{K}_{x,x'}(y, y'))_{y,y' \in B}$ is a Markov kernel from B_x to $B_{x'}$.

Proof

To check (20), we start with its r.h.s., fixing $y, y' \in B$ and $x' \in A$. We compute

$$\begin{aligned}\sum_{x\in A}\widetilde{\Lambda}(y,x)\widetilde{K}((x,y),(x',y')) &= \sum_{x\in A}\frac{\Lambda(y,x)}{\Lambda[h](y)}P(x,x')\frac{h(x')}{\theta}\sum_{z\in\bar{B}\cap W_{x'}}K_{x,x'}(y,z)H(z,y',x')\\ &= \sum_{z\in\bar{B}\cap W_{x'}}\frac{h(x')H(z,y',x')}{\theta\Lambda[h](y)}\sum_{x\in A}\Lambda(y,x)P(x,x')K_{x,x'}(y,z)\\ &= \sum_{z\in\bar{B}\cap W_{x'}}\frac{h(x')H(z,y',x')}{\theta\Lambda[h](y)}Q(y,z)\Lambda(z,x')\\ &= \sum_{z\in\bar{B}}\frac{h(x')}{\theta\Lambda[h](y)}Q(y,z)G(z,y')\Lambda(y',x')\\ &= \frac{h(x')}{\Lambda[h](y')}\widetilde{Q}(y,y')\Lambda(y',x')\\ &= \widetilde{Q}(y,y')\widetilde{\Lambda}(y',x')\end{aligned}$$

where for the third equality we were able to replace $x \in A$ by $x \in V$, since $y \in B$, for the fourth equality we used that the relation $H(z,y',x')\Lambda(z,x') = G(z,y')\Lambda(y',x')$ is always satisfied, for the fifth equality we took (18) into account, and for the last equality used the definition of $\widetilde{\Lambda}$.

To check the last assertion of the above theorem, we have to show that for any $x, x' \in A$ with $\widetilde{P}(x,x') > 0$,

$$\forall\ y \in B_x, \qquad \sum_{y'\in B_{x'}}\widetilde{K}_{x,x'}(y,y') = 1$$

By definition, the above l.h.s. is equal to

$$\sum_{z\in\bar{B}\cap W_{x'}}K_{x,x'}(y,z)\sum_{y'\in B_{x'}}H(z,y',x') = \sum_{z\in\bar{B}\cap W_{x'}}K_{x,x'}(y,z) \tag{21}$$

due to (17).

From (3), we get that for $y \in B_x$, we have $K_{x,x'}(y,z) = 0$ when $Q(y,z) = 0$, in particular if $z \notin \bar{B}$. It follows that in the r.h.s. of (21), we can remove the restriction $z \in \bar{B}$ and we get

$$\sum_{z\in W_{x'}}K_{x,x'}(y,z) = 1$$

since $K_{x,x'}$ is a Markov kernel from W_x to $W_{x'}$. ■

4 Discrete Pitman intertwining

We consider here the example of the discrete (probabilistic) intertwining of Pitman [12], which served as a preliminary version of his famous (probabilistic) intertwining from the Bessel-3 process to the Brownian motion (see also Rogers and Pitman [13]).

The primal Markov chain is the simple random walk $X \coloneqq (X(n))_{n\in\mathbb{Z}_+}$ on $\mathbb{Z}$, jumping with probability 1/2 to the two nearest neighbors and starting from 0. Introduce the dual process $Y \coloneqq (Y(n))_{n\in\mathbb{Z}_+}$ taking values in $\mathbb{Z}_+$ and defined by

$$\forall\ n \in \mathbb{Z}_+, \qquad Y(n) \coloneqq 2\max_{m\in[\![0,n]\!]}X(m) - X(n) \tag{22}$$

Consider the Markov kernel Λ from $\mathbb{Z}_+$ to $\mathbb{Z}$ associating to any $y \in \mathbb{Z}_+$ the uniform distribution on $\{-y, -y+2, ..., y-2, y\}$.

Pitman [12] proved the probabilistic Λ-intertwining from Y to X:

$$\forall\ n \in \mathbb{Z}_+, \qquad \mathcal{L}(X(n)|Y(\llbracket 0,n\rrbracket)) \;=\; \Lambda(Y(n),\cdot) \tag{23}$$

which is an instance of (6).

Remark 2 While the above Pitman relation corresponds to the historical birth of intertwinings, this example can be disturbing for the reader, since up to now we only considered finite state spaces. Nevertheless, it is possible to come back to the situation of finite state spaces by modifying the above chains. Let $M \in \mathbb{N}$ be fixed. We consider the Markov chain $X_M := (X_M(n))_{n\in\mathbb{Z}_+}$ on $\llbracket -M, M\rrbracket$, starting at 0 and jumping, in $\llbracket -M+1, M-1\rrbracket$, with probability $1/2$ to the two nearest neighbors, but once at M (respectively $-M$), it jumps at $M-1$ (resp. $-M+1$) with probability $1/2$ and choose uniformly a point in $\{-M+1, -M+3, ..., M-1\}$ otherwise. The dual process Y_M is defined as in (22), up to the time τ it reaches M. At time τ, Y_M jumps to $M-1$. Afterward the transitions of Y_M are coupled with those of X_M via the kernel given in (25) (restricted to $\llbracket -M+1, M-1\rrbracket$), until Y_M reaches M again. Then we start again the procedure as from τ above. From Pitman [12], we get that $X_M(\tau)$ is uniformly distributed on $\{-M, -M+2, ..., M\}$, even conditionally to $Y_M(\llbracket 0,\tau\rrbracket)$ and it is immediate to check that $X_M(\tau+1)$ is uniformly distributed on $\{-M+1, -M+3, ..., M-1\}$, conditionally to $Y_M(\llbracket 0,\tau+1\rrbracket)$ (whose information coincides with that of $Y_M(\llbracket 0,\tau\rrbracket)$). Next the transitions of (X_M, Y_M) coincides with those of (X, Y), until Y_M hits M again. It can be shown by iteration that the intertwining relation (23) holds for X_M and Y_M, where Λ is the restriction of the above link to $\llbracket -M, M\rrbracket$. This observation enables us to come back to the finite state spaces $\llbracket -M, M\rrbracket$ and $\llbracket 0, M\rrbracket$ for X_M and Y_M respectively. What follows can indeed be transcribed to this finite setting, where X is conditioned to stay in $\llbracket -N+1, N-1\rrbracket$ (the conditioned chain does not correspond to X_N), as soon as $N \leqslant M$, since the conditioned chain is indifferent to what goes on outside $\llbracket -N+1, N-1\rrbracket$. □

The above remark enables us to justify an extension of the considerations and notations of the previous sections to the present countable state space situation, where we take $V := \mathbb{Z}$ endowed with the Markov kernel P given by

$$\forall\ x, x' \in V, \qquad P(x,x') \;:=\; \begin{cases} 1/2 & \text{, if } |x'-x| = 1 \\ 0 & \text{, otherwise} \end{cases}$$

We condition the associated Markov chain X starting from 0 to stay in $A := \llbracket 1-N, N-1\rrbracket$, for any given $N \geqslant 4$. It is well-known and easy to check that $\theta = \cos\left(\frac{\pi}{2N}\right)$ and that a corresponding function h is given by

$$h\,:\, V \ni x \;\mapsto\; \cos\left(\frac{\pi x}{2N}\right)\mathbb{1}_{\llbracket 1-N,N-1\rrbracket}(x) \tag{24}$$

Consider on $W := \mathbb{Z}_+$ the Markov kernel given by

$$\forall\ y, y' \in W, \qquad Q(y,y') \;:=\; \begin{cases} \frac{y+2}{2(y+1)} & \text{, if } y \geqslant 1 \text{ and } y' = y+1 \\ \frac{y}{2(y+1)} & \text{, if } y' = y-1 \\ 0 & \text{, otherwise} \end{cases}$$

as well as the Markov kernel Λ from W to V given by

$$\forall\ y \in W,\ x \in V \qquad \Lambda(y,x) \;:=\; \begin{cases} \frac{1}{y+1} & \text{, if } x \in \{-y, -y+2, ..., y-2, y\} \\ 0 & \text{, otherwise} \end{cases}$$

The probabilistic Λ-intertwining of Pitman [12] from Q to P recalled above corresponds to the Markov kernels $K_{x,x'} := (K_{x,x'}(y,y'))_{y,y'\in W}$, for $x, x' \in V$ with $|x-x'| = 1$, given by

$$\forall\ y, y' \in W, \qquad K_{x,x'}(y,y') \;:=\; \begin{cases} 1 & \text{, if } x = y \text{ and } y' = y+1 \\ 1 & \text{, if } x \in \{-y, -y+2, ..., y-2\} \text{ and } y' = y-(x'-x) \\ 0 & \text{, otherwise} \end{cases} \tag{25}$$

As a consequence, we also get an algebraic Λ-intertwining from Q to P.

Let us specify in this situation, the objects previously introduced to get algebraic as well as probabilistic $\widetilde{\Lambda}$-intertwinings from $\widetilde{Q}$ to $\widetilde{P}$.

The conditioned Markov kernel $\widetilde{P}$ is given, for any $x, x' \in [\![1-N, N-1]\!]$, by

$$\widetilde{P}(x,x') \quad \coloneqq \quad \begin{cases} \frac{\cos(\pi x'/(2N))}{2\cos\left(\frac{\pi}{2N}\right)\cos(\pi x/(2N))} & \text{, if } |x'-x| = 1 \\ 0 & \text{, otherwise} \end{cases}$$

By identifying an element $y \in W$ with the subset $S_y \coloneqq \{-y, -y+2, ..., y-2, y\} \subset V$, we are in the situation of Example 1 of Section 2 (with π the counting measure on $\mathbb{Z}$), so B, $\bar{B}$, φ and g are given by

$$\begin{aligned} B &= [\![0, N-1]\!] \\ \bar{B} &= [\![0, N]\!] \\ \forall\, y \in \bar{B}, \qquad \varphi(y) &= \begin{cases} y & \text{, if } y \in [\![0, N-1]\!] \\ N-2 & \text{, if } y = N \end{cases} \\ \forall\, y \in \bar{B}, \qquad g(y) &= \begin{cases} 1 & \text{, if } y \in [\![0, N-1]\!] \\ \frac{N-1}{N+1} & \text{, if } y = N \end{cases} \end{aligned}$$

We compute

$$\begin{aligned} \forall\, y \in B, \qquad \Lambda[h](y) &= \frac{1}{y+1}\sum_{x\in S_y}\cos\left(\frac{\pi x}{2N}\right) \\ &= \frac{1}{y+1}\sum_{z\in[\![0,y]\!]}\cos\left(\frac{\pi(2z-y)}{2N}\right) \\ &= \frac{1}{2(y+1)}\sum_{z\in[\![0,y]\!]}\exp\left(\mathrm{i}\frac{\pi(2z-y)}{2N}\right) + \exp\left(-\mathrm{i}\frac{\pi(2z-y)}{2N}\right). \end{aligned}$$

The transformation $z \mapsto y-z$ shows that the two sums in the right are equal. So

$$\begin{aligned} \Lambda[h](y) &= \frac{\exp\left(-\mathrm{i}\frac{\pi y}{2N}\right)}{(y+1)}\sum_{z\in[\![0,y]\!]}\exp\left(\mathrm{i}\frac{\pi z}{N}\right) \\ &= \frac{\exp\left(-\mathrm{i}\frac{\pi y}{2N}\right)}{(y+1)}\frac{\exp\left(\mathrm{i}\frac{\pi(y+1)}{N}\right)-1}{\exp\left(\mathrm{i}\frac{\pi}{N}\right)-1} \\ &= \frac{\exp\left(-\mathrm{i}\frac{\pi y}{2N}\right)}{(y+1)}\frac{\exp\left(\mathrm{i}\frac{\pi(y+1)}{2N}\right)\left(\exp\left(\mathrm{i}\frac{\pi(y+1)}{2N}\right)-\exp\left(-\mathrm{i}\frac{\pi(y+1)}{2N}\right)\right)}{\exp\left(\mathrm{i}\frac{\pi}{2N}\right)\left(\exp\left(\mathrm{i}\frac{\pi}{2N}\right)-\exp\left(-\mathrm{i}\frac{\pi}{2N}\right)\right)} \\ &= \frac{1}{y+1}\frac{\sin\left(\frac{\pi(y+1)}{2N}\right)}{\sin\left(\frac{\pi}{2N}\right)} \;>\; 0. \end{aligned}$$

As a consequence, we get

$$\begin{aligned} \forall\, y \in B,\, \forall\, x \in A, \qquad \widetilde{\Lambda}(y,x) &\coloneqq \frac{\Lambda(y,x)h(x)}{\Lambda[h](y)} \\ &= \frac{\sin\left(\frac{\pi}{2N}\right)\cos\left(\frac{\pi x}{2N}\right)}{\sin\left(\frac{\pi(y+1)}{2N}\right)}\mathbb{1}_{S_y}(x) \end{aligned}$$

We also deduce that

$$\forall\, y,y'\in B,\qquad \widetilde{Q}(y,y') \;=\; \frac{1}{\theta\Lambda[h](y)}\sum_{z\in\bar{B}\,:\,\varphi(z)=y'} Q(y,z)g(z)\Lambda[h](y')$$

$$\;=\; \frac{1}{\cos\left(\frac{\pi}{2N}\right)}\sum_{z\in\bar{B}\,:\,\varphi(z)=y'} Q(y,z)g(z)\frac{(y+1)\sin\left(\frac{\pi(y'+1)}{2N}\right)}{(y'+1)\sin\left(\frac{\pi(y+1)}{2N}\right)}$$

It leads us to consider two cases:

- When $y'<N-2$, then $\{z\in\bar{B}\,:\,\varphi(z)=y'\}=\{y'\}$ and thus

$$\widetilde{Q}(y,y') \;=\; \frac{1}{\cos\left(\frac{\pi}{2N}\right)}\frac{(y+1)\sin\left(\frac{\pi(y'+1)}{2N}\right)}{(y'+1)\sin\left(\frac{\pi(y+1)}{2N}\right)}Q(y,y')$$

- When $y'=N-2$, we have $\{z\in\bar{B}\,:\,\varphi(z)=N-2\}=\{N-2,N\}$. Thus we get for $y\in B$,

$$\widetilde{Q}(y,N-2)$$

$$= \frac{1}{\cos\left(\frac{\pi}{2N}\right)}\frac{(y+1)\sin\left(\frac{\pi(N-1)}{2N}\right)}{(N-1)\sin\left(\frac{\pi(y+1)}{2N}\right)}Q(y,N-2)+\frac{1}{\cos\left(\frac{\pi}{2N}\right)}\frac{(y+1)\sin\left(\frac{\pi(N-1)}{2N}\right)}{(N-1)\sin\left(\frac{\pi(y+1)}{2N}\right)}\frac{N-1}{N+1}Q(y,N)$$

$$= \frac{y+1}{\sin\left(\frac{\pi(y+1)}{2N}\right)}\left(\frac{1}{N-1}Q(y,N-2)+\frac{1}{N+1}Q(y,N)\right)$$

and this expression vanishes, except, first when $y=N-3$, in which case we get

$$\widetilde{Q}(N-3,N-2) \;=\; \frac{N-2}{(N-1)\sin\left(\frac{\pi(N-2)}{2N}\right)}Q(N-3,N-2)$$

$$\;=\; \frac{1}{2\cos\left(\frac{\pi}{N}\right)}$$

(where we took into account that $N\geqslant 4$), and second when $y=N-1$, in which case we get

$$\widetilde{Q}(N-1,N-2) \;=\; \frac{N}{\sin\left(\frac{\pi N}{2N}\right)}\left(\frac{1}{N-1}Q(N-1,N-2)+\frac{1}{N+1}Q(N-1,N)\right)$$

$$\;=\; N\left(\frac{1}{N-1}\frac{N-1}{2N}+\frac{1}{N+1}\frac{N+1}{2N}\right)$$

$$\;=\; 1$$

- When $y'=N-1$, we have $\{z\in\bar{B}\,:\,\varphi(z)=N-1\}=\{N-1\}$. Thus we get for $y\in B$,

$$\widetilde{Q}(y,N-1) \;=\; \frac{1}{\cos\left(\frac{\pi}{2N}\right)}\frac{(y+1)\sin\left(\frac{\pi N}{2N}\right)}{N\sin\left(\frac{\pi(y+1)}{2N}\right)}Q(y,N-1)$$

$$\;=\; \frac{y+1}{N\cos\left(\frac{\pi}{2N}\right)\sin\left(\frac{\pi(y+1)}{2N}\right)}Q(y,N-1)$$

and this expression vanishes, except when $y=N-2$, in which case we get

$$\widetilde{Q}(N-2,N-1) \;=\; \frac{N-1}{N\cos\left(\frac{\pi}{2N}\right)\sin\left(\frac{\pi(N-1)}{2N}\right)}Q(N-2,N-1)$$

$$\;=\; \frac{1}{2\cos^2\left(\frac{\pi}{2N}\right)}$$

It appears that a Markov chain $\widetilde{Y} := (\widetilde{Y}(n))_{n\in\mathbb{Z}_+}$ with transition kernel $\widetilde{Q}$ is not absorbed at $N-1$ (this is related to the fact that the conditioned process $\widetilde{X}$ does not converge in law for large times, due to periodicity). In fact the set $[\![1, N-1]\!]$ is a recurrence class for $\widetilde{Y}$, which in particular always returns to the point 1. As we are to see below, this leads to a strange phenomenon. For a better understanding, let us consider the probabilistic $\widetilde{\Lambda}$-intertwining.

By definition, we have, for any $x, x' \in A$ with $|x' - x| = 1$,

$$\forall\, y, y' \in B, \qquad \widetilde{K}_{x,x'}(y, y') \;:=\; \sum_{z\in\bar{B}\,:\,\varphi(z)=y'} K_{x,x'}(y, z)$$

thus again we are led to consider two cases.

- When $y' < N-2$ or $y' = N-2$, we get

$$\begin{aligned}\forall\, y \in B, \qquad \widetilde{K}_{x,x'}(y, y') &= K_{x,x'}(y, y') \\ &= \mathbb{1}_{y=x,\, y'=y+1} + \mathbb{1}_{y>x,\, y'=y-(x'-x)}\end{aligned}$$

- When $y' = N-2$, we get

$$\forall\, y \in B, \qquad \widetilde{K}_{x,x'}(y, N-2) \;=\; K_{x,x'}(y, N-2) + K_{x,x'}(y, N)$$

expression which vanishes, except, first for $y = N-3$,

$$\begin{aligned}\widetilde{K}_{x,x'}(N-3, N-2) &= K_{x,x'}(N-3, N-2) \\ &= \mathbb{1}_{x=N-3} + \mathbb{1}_{x<N-3,\, x'=x-1}\end{aligned}$$

and second for $y = N-1$,

$$\begin{aligned}\widetilde{K}_{x,x'}(N-1, N-2) &= K_{x,x'}(N-1, N-2) + K_{x,x'}(N-1, N) \\ &= \mathbb{1}_{x<N-1,\, x'=x+1} + \mathbb{1}_{x=N-1} + \mathbb{1}_{x<N-1,\, x'=x-1} \\ &= 1\end{aligned}$$

Thus given a transition from x to x' for $\widetilde{X}$, $\widetilde{Y}$ changes similarly to Y when X is making a transition from x to x', except when $\widetilde{Y}$ is equal to $N-1$, then the next position is necessary $N-2$. Of course this last fact is sufficient to prevent (22) to hold for $\widetilde{X}$ and $\widetilde{Y}$, but it does hold until $\widetilde{Y}$ hits $N-1$ for the first time. In the traditional theory of Diaconis and Fill [4], the dual chain indicates how the primal chain is progressing towards equilibrium, until the dual chain is absorbed, then the primal chain is at equilibrium in the sense of strong stationary times. Here the situation is different: at some times as large as we want, we can deduce concentration for the primal chain $\widetilde{X}$ from the observation of the trajectory of the dual chain $\widetilde{Y}$, since at the stopping times τ satisfying $\widetilde{Y}(\tau) = 1$, we know from the probabilistic intertwining relation that $\widetilde{X}(\tau) = 1$ or $\widetilde{X}(\tau) = -1$, each event occurring with probability 1/2.

Remark 3 The periodicity can be removed by considering P^2, which can be decomposed into its two irreducible parts, one on $2\mathbb{Z}$ and the other one on $1 + 2\mathbb{Z}$. The Markov kernel P^2 is intertwined with Q^2 with the same link Λ. The irreducible parts of Q^2 are $2 + 2\mathbb{Z}_+$ and $1 + 2\mathbb{Z}_+$, the remaining singleton $\{0\}$ being transient (it is left in one transition). Consider for instance $\widehat{P}$ the restriction of P^2 to $\widehat{V} := 2\mathbb{Z}$, $\widehat{Q}$ the restriction of Q^2 to $\widehat{W} := 2\mathbb{Z}$ and $\widehat{\Lambda}$ the restriction of Λ from $\widehat{W}$ to $\widehat{V}$. We have the algebraic $\widehat{\Lambda}$-intertwining relation

$$\widehat{Q}\widehat{\Lambda} \;=\; \widehat{\Lambda}\widehat{P}$$

and the probabilistic intertwining is given with the restriction to $\widehat{V} \times \widehat{W}$ of $\widehat{K} = K^2$. In particular we have for any $x, x' \in \widehat{V}$ with $\widehat{P}(x, x') > 0$,

$$\forall\, y, y' \in \widehat{W}, \qquad \widehat{K}_{x,x'}(y, y') \;=\; \sum_{x''\in\widehat{V}} \frac{P(x, x'')P(x'', x')}{P(x, x')} (K_{x,x''}K_{x'',x'})(y, y')$$

Assume we want to condition the underlying Markov chain to stay in $\widehat{A} \coloneqq [\![2-N, N-2]\!]$ with $N \geqslant 6$ and even. We have $\widehat{P}\widehat{h} = \widehat{\theta}\widehat{h}$, with $\widehat{h}$ the restriction to $\widehat{A}$ of the function h given in (24), and with $\widehat{\theta} = \theta^2$. Denote $\breve{P}$ the obtained conditioned transition kernel. It appears that $\breve{P} = \widetilde{P}^2$. It is $\breve{\Lambda}$-intertwined with $\breve{Q} = \widetilde{Q}^2$ restricted to $\widehat{B} = [\![0, 2, ..., N-2]\!]$ and with $\breve{\Lambda}$ the restriction of $\widetilde{\Lambda}$ from $\breve{B}$ to $\breve{A}$. The Markov kernel $\breve{Q}$ is irreducible and aperiodic on $[\![2, 4, ..., N-2]\!]$ and thus it is no more absorbed there than $\widetilde{Q}$, so it is not helpful to construct a strong stationary time for $\breve{P}$, in the spirit of Diaconis and Fill [4]. It is nevertheless possible to construct a probabilistic $\breve{\Lambda}$-intertwining from $\breve{Q}$ to $\breve{P}$, described as above through a family of kernels $\breve{K}_{x,x'}$ from $\breve{B}$ to $\breve{B}$, for any $x, x' \in \breve{A}$ with $|x'-x| \leqslant 2$. □

5 Top-to-random shuffle intertwining

We consider here the classical top-to-random shuffle (probabilistic) intertwining of Aldous and Diaconis [1]. We condition it so that the last initial card ends up staying in the first half of the deck. Our interest is more on an illustration of the main assumption (10) than on this particular conditioning in itself.

For $N \in \mathbb{N}\backslash\{1\}$, let V be the symmetric group $\mathcal{S}_N$, seen as the set of decks of N cards, whose values are the elements of $[\![N]\!]$. The identity corresponds to the ordering where 1 is at the top, 2 at the second place, etc., N being the card at the bottom of the deck. Introduce the Markov chain $X \coloneqq (X(n))_{n\in\mathbb{Z}_+}$ on $\mathcal{S}_N$ starting from the identity and whose transitions are described by: the top card is removed and put back uniformly at random in the deck (independently from the past).

Consider the $Y \coloneqq (Y_1(n), Y_2(n))_{n\in\mathbb{Z}_+}$ where $Y_1(n)$ is the position of the card N in the deck $X(n)$, up to the time where this position reaches the top position 1, then the value of $Y_1(n)$ is 0 for all subsequent times, and where $Y_2(n)$ is the $Y_1(n)$-tuple of the values of the cards at positions 1, 2, ..., $Y_1(n)$, with $Y_2(n) = \varnothing$ if $Y_1(n) = 0$. In particular the last value of $Y_2(n)$ is N at position $Y_1(n)$, as long as $Y_1(n) \geqslant 1$. The chain Y is Markovian and absorbed at $(0, \varnothing)$. Denote by W its state space. The corresponding Markov kernel Q is described by

$$\forall\, y \coloneqq (y_1, y_2) \neq (0, \varnothing),\ y' \coloneqq (y'_1, y'_2) \in W, \quad Q(y, y') \quad = \quad \begin{cases} \frac{1}{N} & \text{, if } y'_1 = y_1 > 1 \text{ and } y'_2 \in \mathcal{Y}(y_2) \\ 1 - \frac{y_1-1}{N} & \text{, if } y'_1 = y_1 - 1 \text{ and } y'_2 = y''_2 \\ 0 & \text{, otherwise} \end{cases}$$

where y''_2 is obtained from y_2 by removing its first value (with the convention that $y_2 = \varnothing$ if $y_1 = 1$) and where $\mathcal{Y}(y_2)$ is the set obtained from y_2 by inserting the first value of y_2 somewhere in y''_2 except in last position. Furthermore, $Q((0, \varnothing), \cdot)$ is the Dirac mass at $(0, \varnothing)$.

There is an intertwining relation from Y to X with the link Λ given by the requirement that for any $y \coloneqq (y_1, y_2) \in W$, $\Lambda(y, \cdot)$ is the uniform distribution over the $x \in \mathcal{S}_N$ such that the card N is at position y_1 and the values of x up to this position are given by y_2. If $y = (0, \varnothing)$, $\Lambda(y, \cdot)$ is just the uniform distribution over $\mathcal{S}_N$. Thus Y can be seen as a subset-valued dual process, by identifying $y \coloneqq (y_1, y_2) \in W$, with the subset of $x \in \mathcal{S}_N$ such that the card N is at position y_1 and the values of x up to this position are given by y_2. In particular $(0, \varnothing)$ is identified with $\mathcal{S}_N$.

For simplicity, assume furthermore that N is even and consider the set $A \subset V$ consisting of permutations $x \in \mathcal{S}_N$ such that the card N is in the first half of the deck, i.e. its position belongs to $[\![N/2]\!]$. Initially this condition is not satisfied, since the card N is at position N. So introduce the stopping time

$$T \quad \coloneqq \quad \inf\{n \in \mathbb{Z}_+ \,:\, X(n) \in A\} \tag{26}$$

and let m_0 be the law of $X(T)$. We shift the origin of times to T and thus rather consider the Markov chain X starting with the initial distribution m_0.

To compute a Doob function h corresponding to the subset A, let us remark that $Z \coloneqq (Z(n))_{n\in\mathbb{Z}_+}$ is a Markov chain on $[\![N]\!]$, where for any $n \in \mathbb{Z}_+$, $Z(n)$ stands for the position of card N in $X(n)$ (in particular Z coincides with the Markov chain Y_1 up to the hitting time of 1 by Y_1. We have $Z(0) = N/2$ and the transition kernel R of Z is given by

$$\forall\ z, z' \in [\![N]\!], \qquad R(z, z') \quad = \quad \begin{cases} (z-1)/N & \text{, if } z' = z \text{ and } z \geqslant 2 \\ 1-(z-1)/N & \text{, if } z' = z-1 \text{ and } z \geqslant 2 \\ 1/N & \text{, if } z = 1 \\ 0 & \text{, otherwise} \end{cases}$$

Thus conditioning X to stay in A amounts to condition Z to stay in $\underline{A} \coloneqq [\![N/2]\!]$, and denoting $\underline{\theta}$ and $\underline{h}$ corresponding largest Dirichlet eigenvalue and Doob function, we have

$$\theta \quad = \quad \underline{\theta} \tag{27}$$

$$\forall\ x \in V, \qquad h(x) \quad = \quad \underline{h}(z(x)) \tag{28}$$

where $z(x)$ is the position of the card N in x.

Since the dual chain Y is subset-valued, we could think we are in situation of Example 1. This is not completely true, because W is only a subset of the set of all non-empty subsets of $\mathcal{S}_N$.

From our general definitions, it appears that

$$\begin{aligned} B &\quad=\quad \{y \coloneqq (y_1, y_2) \in W\ :\ y_1 \in [\![N/2]\!]\} \\ \forall\ y \coloneqq (y_1, y_2) \in B,\ \forall\ x \in A, \qquad \widetilde{\Lambda}(y, x) &\quad=\quad \frac{\Lambda(y, x)h(x)}{\Lambda[h](y)} \\ &\quad=\quad \Lambda(y, x) \\ &\quad=\quad \frac{1}{(N-y_1)!}\mathbb{1}_{x\in y} \end{aligned}$$

where we used that h is constant on y, which is the support of $\Lambda(y, \cdot)$ (with value $\underline{h}(y_1)$).

Since $\bar{B} = \{(0, \varnothing)\}$, our main assumption (10) consists in the existence of a family $(g(y))_{y\in B}$ of non-negative numbers such that

$$\forall\ x \in A, \qquad \Lambda((0, \varnothing), x) \quad = \quad \sum_{y\in B} g(y)\Lambda(y, x) \tag{29}$$

Note that for any $x \in A$, there is only one $y \in B$ such that $\Lambda(y, x) > 0$, call it $y(x)$. This $y(x) \coloneqq (y_1(x), y_2(x))$ is given by

$$y_1(x) \quad = \quad z(x) \qquad \text{and} \qquad y_2(x) = (x(1), x(2), ..., x(y_1))$$

It follows that (29) amounts to

$$\forall\ x \in A, \qquad \frac{1}{N!} \quad = \quad g(y(x))\frac{1}{(N-y_1(x))!}$$

relation which is satisfied by taking

$$\begin{aligned} \forall\ y \coloneqq (y_1, y_2) \in B, \qquad g(y) &\quad\coloneqq\quad \frac{(N-y_1)!}{N!} \\ &\quad=\quad \frac{1}{N(N-1)\cdots(N-y_1+1)} \end{aligned}$$

Taking Remark 1 into account, we get from (13), for any $y := (y_1, y_2) \in B$ and $y' := (y'_1, y'_2) \in B$,

$$\widetilde{Q}(y,y') \quad = \quad \begin{cases} Q(y,y') & \text{, if } y_1 \geqslant 2 \\ Q(y,(0,\emptyset))g(y')\frac{\Lambda[h](y')}{\theta\Lambda[h](y)} & \text{, if } y_1 = 1 \end{cases}$$

The last term can be simplified: on one hand $Q(y,(0,\emptyset)) = 1$ when $y_1 = 1$, and on the other hand, note that for any $y := (y_1, y_2)$ with $y_1 \geqslant 1$, the support of $\Lambda(y,\cdot)$ is included into the set of $x \in \mathcal{S}_N$ whose N card is at position y_1, so that $\Lambda[h](y) = \underline{h}(y_1)$. It follows that for any $y := (y_1, y_2) \in B$ and $y' := (y'_1, y'_2) \in B$,

$$y_1 \;=\; 1 \quad\Rightarrow\quad \widetilde{Q}(y,y') \;=\; \frac{1}{N(N-1)\cdots(N-y_1+1)}\frac{\underline{h}(y'_1)}{\theta\underline{h}(y_1)}$$

As in the previous section the Markov kernel $\widetilde{Q}$ is not absorbed but ergodic, so the associated dual Markov chain $\widetilde{Y}$ cannot be used to construct a strong stopping time.

Nevertheless, to finish this section, let us go a little further in the simplification of the situation, since it shows how it may sometimes be interesting to first extend the initial intertwining relation. Coming back to the subset interpretation of W, we replace it by

$$\bar{W} \quad := \quad (W\backslash\{(0,\emptyset)\}) \cup \{\hat{1}, \hat{2}, ..., \hat{N}\}$$

where for any $l \in [\![N]\!]$, $\hat{l}$ stands for the set of permutations from $\mathcal{S}_N$ which are such that the card N is at position l. In particular we have $\hat{1} = (1, N)$ with the notations of the beginning of this section (and this is the unique point belonging both to $W\backslash\{(0,\emptyset)\}$ and $\{\hat{1}, \hat{2}, ..., \hat{N}\}$). Inspired by (12) (where π is the uniform distribution on $\mathcal{S}_N$), we get a Markov kernel $\bar{\Lambda}$ from $\bar{W}$ to $\mathcal{S}_N$. Replace Q by the Markov kernel $\bar{Q}$ on $\bar{W}$, defined by

$$\forall\, y, y' \in \bar{W}, \qquad \bar{Q}(y,y') \;:=\; \begin{cases} Q(y,y') & \text{, if } y, y' \in W \\ \frac{1}{N} & \text{, if } y = \hat{1} \text{ and } y' \in \{\hat{1}, \hat{2}, ..., \hat{N}\} \\ \frac{l-1}{N} & \text{, if } y = y' = \hat{l} \text{ for some } l \in [\![2, N]\!] \\ \frac{N-l+1}{N} & \text{, if } y = \hat{l} \text{ for some } l \in [\![2, N]\!] \text{ and } y' = \widehat{l-1} \\ 0 & \text{, otherwise} \end{cases}$$

The algebraic Λ-intertwining relation from Q to P can be extended into a algebraic $\bar{\Lambda}$-intertwining relation from $\bar{Q}$ to P.

Starting again X from the distribution m_0 defined after (26) and conditioning it by its card N staying in the first half of the deck, we are led to introduce the state space

$$\widehat{B} \;:=\; B \cup \left\{\hat{1}, ..., \frac{\hat{N}}{2}\right\}$$

as well as the Markov kernel $\widehat{Q}$ defined on $\widehat{B}$ by

$$\forall\, y, y' \in \widehat{B}, \qquad \widehat{Q}(y,y') \;:=\; \begin{cases} \widetilde{Q}(y,y') & \text{, if } y, y' \in W \\ \frac{\underline{h}(l)}{N\theta\underline{h}(1)} & \text{, if } y = \hat{1} \text{ and } y' = \hat{l} \text{ with } l \in [\![N/2]\!] \\ \frac{l-1}{N\theta} & \text{, if } y = y' = \hat{l} \text{ for some } l \in [\![2, N/2]\!] \\ 1 - \frac{l-1}{N\theta} & \text{, if } y = \hat{l} \text{ for some } l \in [\![2, N/2]\!] \text{ and } y' = \widehat{l-1} \\ 0 & \text{, otherwise} \end{cases}$$

(taking into account Remark 1, we have $G(y,\cdot) = \delta_y(\cdot)$ for any $y \in \{\hat{1}, \hat{2}, ..., \widehat{N/2}\}$ and $G(y,\cdot) = 0$ for $y \in \{\widehat{N/2+1}, ..., \hat{N}\}$).

From Theorem 1, we get an algebraic $\widehat{\Lambda}$-intertwining from $\widehat{Q}$ to $\widetilde{P}$, where the Markov kernel $\widehat{\Lambda}$ from $\widehat{B}$ to A is given by

$$\forall\ y \in \widehat{B},\ \forall\ x \in A, \qquad \widehat{\Lambda}(y,x) \quad = \quad \begin{cases} \widetilde{\Lambda}(y,x) & \text{, if } y \in W \\ \frac{1}{(N-1)!} & \text{, if } y = \hat{l} \text{ with } l \in [\![N/2]\!] \end{cases}$$

(as previously, we used that h is equal to $\underline{h}(l)$ on $\widehat{l}$ for any $l \in [\![N/2]\!]$).

Consider the associated dual process $\widehat{Y} := (\widehat{Y}(n))_{n\in\mathbb{Z}_+}$. It coincides (in law) with the previous dual chain $\widetilde{Y}$ until the hitting time $\tau_{\widehat{1}}$ of $\widehat{1} = (1, N)$. After this time, $\widehat{Y}$ is a Markov chain on $[\![\widehat{N/2}]\!] := \{\hat{1}, \hat{2}, ..., \widehat{N/2}\}$ whose transition kernel $\widehat{R}$ is given by

$$\forall\ \hat{k},\ \hat{l} \in [\![\widehat{N/2}]\!], \qquad \widehat{R}(\hat{k}, \hat{l}) \quad := \quad \begin{cases} \frac{\underline{h}(l)}{N\theta \underline{h}(1)} & \text{, if } \hat{k} = \hat{1} \text{ and } \hat{l} \in [\![\widehat{N/2}]\!] \\ \frac{l-1}{N\theta} & \text{, if } \hat{k} = \hat{l} \neq \hat{1} \\ 1 - \frac{l-1}{N\theta} & \text{, if } \hat{k} \neq 1 \text{ and } \hat{l} = \widehat{k-1} \\ 0 & \text{, otherwise} \end{cases}$$

This Markov chain is relatively simple and if we find a corresponding strong stationary time $\widehat{\tau}$ (when $\widehat{Y}$ starts from $\widehat{1}$), then $\tau_{\widehat{1}} + \widehat{\tau}$ (where $\tau_{\widehat{1}}$ and $\widehat{\tau}$ are independent) has the same law as a strong stationary time for $\widetilde{X}$ (starting from the distribution m_0). Furthermore, $\tau_{\widehat{1}}$ has the same law as the hitting time of $\hat{1}$ when $\widehat{Y}$ is starting from $\widehat{N/2}$, thus to compute the law of $\tau_{\widehat{1}} + \widehat{\tau}$, we just need to work with the Markov chain $\widehat{Y}$ restricted to $[\![\widehat{N/2}]\!]$, i.e. we just need to consider the Markov kernel $\widehat{R}$. This will not be done here, since this question is no longer related to conditioning of Markov chains. An investigation of some features of Markov kernels such as $\widehat{R}$ is provided in [10], but it does not enable to get an estimate on corresponding strong stationary times.

6 Birth-and-death intertwinings

Here we consider a situation where the primal Markov chain is absorbed and intertwined with a simpler dual Markov chain, which enables to deduce the law of the absorption time. We will see that we get a new intertwining relation by conditioning the primal Markov chain not to be absorbed. Nevertheless if we condition the primal Markov chain to stay in a subset strictly smaller that the set of non-absorbing points, our main assumption (10) may not be satisfied.

Let P be a birth-and-death Markov kernel on $[\![0, N]\!]$, with $N \in \mathbb{N}$, $N \geqslant 2$, whose transition probabilities between two neighbors are positive, except that $P(N, N-1) = 0$, namely N is an absorbing point. Let $\theta_0 < \theta_1 < \theta_2 < \cdots < \theta_{N-1}$ be the eigenvalues of $P_{[\![0,N-1]\!]}$, which is diagonalizable as a birth-and-death sub-Makrov kernel on $[\![0, N-1]\!]$. A priori these eigenvalues belong to $(-1, 1)$ and are all distinct (as a consequence of irreducibility and of the birth-and-death feature). The Dirichlet eigenvalue θ appearing in (8) is here equal to θ_{N-1}. Let us assume that $\theta_0 \geqslant 0$, so that all the eigenvalues of $P_{[\![0,N-1]\!]}$ are non-negative.

Consider $X := (X(n))_{n\in\mathbb{Z}_+}$ a corresponding Markov chain starting from 0. It ends up being absorbed at N in finite time.

Fill [6] (see also Fill [7] and Diaconis and Miclo [5]) has shown how to construct a probabilistic intertwining dual birth-and-death Markov chain $Y := (Y(n))_{n\in\mathbb{Z}_+}$ on $[\![0, N]\!]$, starting from 0, non-decreasing, ending being absorbed at N and such that, denoting Λ the corresponding link, we have

- for any $y \in [\![0, N-1]\!]$, the support of $\Lambda(y, \cdot)$ is included in $[\![0, y]\!]$ and contains y,
- $\Lambda(N-1, \cdot)$ is the quasi-stationary distribution associated to P,
- $\Lambda(N, \cdot)$ is the Dirac mass at N.

This (algebraic) intertwining is not of form of the subset-valued duals presented in Example 1, it rather corresponds to a measure-valued dual. Despite the primal random walk of Section 4 is also birth-and-death (but not absorbed and on $\mathbb{Z}$), the discrete Pitman intertwining is not of this type.

Let $M \in [\![0, N-1]\!]$ be given, we want to condition X to stay in $A \coloneqq [\![0, M]\!]$. Our purpose here is to investigate whether or not the considerations of Section 2 and Section 3 can be applied to this situation. To start with, note that the kernel $P_{[\![0,M]\!]}$ is irreducible. The set B is equal to $[\![0, M]\!]$ and we have $\bar{B} = [\![0, M+1]\!]$. Taking into account Remark 1 of Section 2, we just have to check there exists a family of non-negative numbers $(G(M+1, y))_{y\in[\![0,M]\!]}$ such that

$$\forall\, x \in [\![0, M]\!], \qquad \Lambda(M+1, x) \;=\; \sum_{y\in[\![0,M]\!]} G(M+1, y)\Lambda(y, x) \tag{30}$$

By backward iteration, taking into account that the support of the $\Lambda(y, \cdot)$ are the $[\![0, y]\!]$ for $y \in [\![0, N-1]\!]$, (30) determines $G(M+1, M)$, $G(M+1, M-1)$, ..., $G(M+1, 0)$, so the family $(G(M+1, y))_{y\in[\![0,M]\!]}$ exists and is unique. Our goal is to check that it consists of non-negative numbers.

The cases $M = N-1$ and $M < N-1$ lead to different results.

Proposition 1 *When $M = N-1$, we have*

$$\forall\, y \in [\![0, M]\!], \qquad G(N, y) \;=\; 0$$

The Markov chain $\widetilde{Y}$ associated to the kernel $\widetilde{Q}$ ends up being absorbed at $N-1$ and thus can be used to construct a strong stationary time for the conditioned chain $\widetilde{X}$.

Remark 4 Recall from Fill [6] that Q is given by

$$Q \;=\; \begin{pmatrix} \theta_0 & 1-\theta_0 & 0 & \cdots & 0 \\ 0 & \theta_1 & 1-\theta_1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ddots & 0 & \theta_{N-1} & 1-\theta_{N-1} \\ 0 & \cdots & \ddots & 0 & 1 \end{pmatrix} \tag{31}$$

(the fact Q is bi-diagonal comes from the fact that Y is birth-and-death and non-decreasing).

Note that $\widetilde{Q}$ has a similar bi-diagonal structure (but of size $N \times N$ instead of $(N+1) \times (N+1)$). From Proposition 1, we have $\widetilde{Q}(N-1, N-1) = 1$ and from (13) applied with $y' = y$, we get

$$\begin{aligned} \forall\, y \in [\![0, N-2]\!], \qquad \widetilde{Q}(y, y) \;&=\; \frac{1}{Z(y)} Q(y, y)\Lambda[h](y) \\ &=\; \frac{1}{\theta_{N-1}\Lambda[h](y)} \theta_y \Lambda[h](y) \\ &=\; \frac{\theta_y}{\theta_{N-1}} \end{aligned}$$

(thus a posteriori it appears that this equality is also satisfied for $y = N-1$).

Due to the algebraic $\widetilde{\Lambda}$-intertwining from $\widetilde{Q}$ to $\widetilde{P}$ and to the fact that $\widetilde{\Lambda}$ is invertible (since it is triangular with a non-degenerate diagonal, similarly to Λ), we get that $\widetilde{Q}$ and $\widetilde{P}$ have the same eigenvalues. Thus if we denote $\widetilde{\theta}_0 < \widetilde{\theta}_1 < \widetilde{\theta}_2 < \cdots < \widetilde{\theta}_{N-1} = 1$ the eigenvalues of $\widetilde{P}$, we get

$$\forall\, y \in [\![0, N-1]\!], \qquad \widetilde{\theta}_y \;=\; \frac{\theta_y}{\theta_{N-1}}$$

Of course, this result can also be obtained directly from (9), where $\widetilde{P}$ appears to be similar to $P_{[\![0,N-1]\!]}$, up to the factor $1/\theta_{N-1}$. □

Let us now come to the second result of this section: Proposition 1 is no longer necessarily true for $M \leqslant N-2$. Introduce the transition kernel P_0 of the usual random walk on $[\![0,N]\!]$, reflected at 0 and absorbed at N, given by

$$\forall\, x,x' \in [\![0,N]\!], \qquad P_0(x,x') \;=\; \begin{cases} 1/2 & \text{, if } |x-x'|=1 \text{, except if } x=0 \text{ or } x'=N \\ 1 & \text{, if } x=0 \text{ and } x'=1 \\ 0 & \text{, otherwise} \end{cases}$$

Some of the eigenvalues of this kernel are negative, so rather consider

$$P_{1/2} \;\coloneqq\; \frac{I+P_0}{2}$$

(where I is the identity kernel on $[\![0,N]\!]$), whose eigenvalues are non-negative.

Proposition 2 *For $N \geqslant 2$ and $P = P_{1/2}$, the family $(G(2,y))_{y\in[\![0,1]\!]}$ is non-negative if and only if $N=2$. Thus the considerations of Sections 2 and 3 cannot be applied for $N \geqslant 3$.*

Both propositions will be consequences of a general computation of the coefficients $(G(M+1,y))_{y\in[\![0,1]\!]}$, based on the theory of divided differences see e.g. de Boor [3] or [14].

Denote $\widehat{\theta}_0 < \widehat{\theta}_1 < \widehat{\theta}_2 < \cdots < \widehat{\theta}_M$ the eigenvalues of $P_{[\![0,M]\!]}$. Taking into account the variational formulation of these eigenvalues, we get

$$\forall\, k \in [\![0,M]\!], \qquad \theta_k \;\leqslant\; \widehat{\theta}_k \;\leqslant\; \theta_{N-1-M+k} \tag{32}$$

(strict inequalities even hold as soon as $M < N-1$). In particular, the eigenvalues of $P_{[\![0,M]\!]}$ are all non-negative.

Denote

$$\widehat{R}_{M+1}(X) \;\coloneqq\; (X-\widehat{\theta}_0)(X-\widehat{\theta}_1)\cdots(X-\widehat{\theta}_M)$$

Lemma 1 *With the bracket notation of divided difference, we have*

$$\forall\, y \in [\![0,M]\!], \qquad G(M+1,y) \;=\; -\left(\prod_{k\in[\![y,M]\!]}(1-\theta_k)\right)^{-1} \widehat{R}_{M+1}[\theta_0,\theta_1,\ldots,\theta_y]$$

Proof

Recall that from Fill [6], we have

$$\forall\, y \in [\![0,N]\!], \qquad \Lambda(y,\cdot) \;=\; \delta_0 \prod_{k=0}^{y-1} \frac{P-\theta_k I}{1-\theta_k} \tag{33}$$

with the convention that the product is the identity operator I if $y=0$.

For $l \in [\![0,N-1]\!]$, consider the polynomial of degree l,

$$R_l(X) \;\coloneqq\; (X-\theta_0)(X-\theta_1)\cdots(X-\theta_{l-1})$$

(again for $l=0$, the convention is that $R_0=1$).

From (30) and (33), we have

$$\forall\, x \in [\![0,M]\!], \qquad R_{M+1}(P)(0,x) \;=\; \sum_{y\in[\![0,M]\!]} H(M+1,y) R_y(P)(0,x) \tag{34}$$

with

$$\forall\ y\in[\![0,M]\!],\qquad H(M+1,y) \;:=\; G(M+1,y)\frac{\prod_{k\in[\![0,M]\!]}(1-\theta_k)}{\prod_{k'\in[\![0,y-1]\!]}(1-\theta_{k'})} \;=\; G(M+1,y)\prod_{k\in[\![y,M]\!]}(1-\theta_k)$$

Thus to prove the above lemma, it is sufficient to show that

$$\forall\ y\in[\![0,M]\!],\qquad H(M+1,y) \;=\; -\widehat{R}_{M+1}[\theta_0,\theta_1,...,\theta_y]$$

Introduce the polynomial of degree M,

$$S_M(X) \;:=\; \sum_{y\in[\![0,M]\!]} H(M+1,y)R_y(X) \tag{35}$$

We deduce from the theory of divided difference that

$$\forall\ y\in[\![0,M]\!],\qquad H(M+1,y) \;=\; S_M[\theta_0,\theta_1,...,\theta_y] \tag{36}$$

Our next goal is to prove that $S_M(X)$ is the polynomial of degree M coinciding with R_{M+1} on $\{\widehat{\theta}_0,\widehat{\theta}_1,...,\widehat{\theta}_{y-1}\}$.

For $k\in[\![0,M]\!]$, let $\widehat{\psi}_k$ be an eigenvector associated to the eigenvalue $\widehat{\theta}_k$ of $P_{[\![0,M]\!]}$. Extend $\widehat{\psi}_k$ into a function defined on $[\![0,N]\!]$ by imposing that $\widehat{\psi}_k$ vanishes on $[\![M+1,N]\!]$. We deduce from (34) that

$$R_{M+1}(P)[\widehat{\psi}_k](0) \;=\; \sum_{y\in[\![0,M]\!]} H(M+1,y)R_y(P)[\widehat{\psi}_k](0)$$

which is equivalent (since $\widehat{\psi}_k$ vanishes on $[\![M+1,N]\!]$) to

$$R_{M+1}(P_{[\![0,M]\!]})[\widehat{\psi}_k](0) \;=\; \sum_{y\in[\![0,M]\!]} H(M+1,y)R_y(P_{[\![0,M]\!]})[\widehat{\psi}_k](0)$$

and thus to

$$R_{M+1}(\widehat{\theta}_k)\widehat{\psi}_k(0) \;=\; \left(\sum_{y\in[\![0,M]\!]} H(M+1,y)R_y(\widehat{\theta}_k)\right)\widehat{\psi}_k(0) \tag{37}$$

Note that $\widehat{\psi}_k(0)\neq 0$, otherwise we would conclude by iteration over $\widehat{\psi}_k(x)$ for $x\in[\![0,M]\!]$, from $P_{[\![0,M]\!]}\widehat{\psi}_k=\widehat{\theta}_k\widehat{\psi}_k$, that $\widehat{\psi}_k=0$.

It follows that for any $k\in[\![0,M]\!]$, we have $R_{M+1}(\widehat{\theta}_k)=S_M(\widehat{\theta}_k)$ and we get the desired characterization of S_M.

Denote $\mathcal{F}$ the linear mapping associating to any polynomial H of degree at most $M+1$ the polynomial of degree M coinciding with H on $\{\widehat{\theta}_0,\widehat{\theta}_1,...,\widehat{\theta}_M\}$. In particular we have $S_M(X)=\mathcal{F}[R_{M+1}](X)$ by the above characterization. Note that $\mathcal{F}[\widehat{R}_{M+1}](X)=0$, so that we also get

$$\begin{aligned} S_M(X) &= \mathcal{F}[R_{M+1}-\widehat{R}_{M+1}](X) \\ &= R_{M+1}(X)-\widehat{R}_{M+1}(X) \end{aligned} \tag{38}$$

since $R_{M+1}-\widehat{R}_{M+1}$ is of degree at most M. It follows from (36) that

$$\forall\ y\in[\![0,M]\!],\qquad H(M+1,y) \;=\; R_{M+1}[\theta_0,\theta_1,...,\theta_y]-\widehat{R}_{M+1}[\theta_0,\theta_1,...,\theta_y]$$

and the desired result follows from the fact that

$$\forall\ y \in [\![0, M]\!], \qquad R_{M+1}[\theta_0, \theta_1, ..., \theta_y] \ = \ 0$$

■

Remark 5 The above proof is valid even if the eigenvalues of $P_{[\![0,N-1]\!]}$ are not assumed to be non-negative. In fact the l.h.s. of (33) is not modified if P is replaced by the affine combination $aI+(1-a)P$, with $a \in [0,1)$, since then θ_k has also to be replaced by $a+(1-a)\theta_k$, for all $k \in [\![0, y-1]\!]$. Thus even if we had only proven Lemma 1 under the assumption of non-negativity of the eigenvalues, we could extend it to the general case. The only problem with negative eigenvalues comes from the matrix Q from (31), which is non longer Markovian. □

We can now come to the

Proof of Proposition 1

When $M = N-1$, we have

$$\forall\ k \in [\![0, N-1]\!], \qquad \widehat{\theta}_k \ = \ \theta_k$$

so that $\widehat{R}_{M+1} = R_{M+1}$,

$$\begin{aligned}\forall\ y \in [\![0, N-1]\!], \qquad \widehat{R}_{M+1}[\theta_0, \theta_1, ..., \theta_y] \ &= \ R_{M+1}[\theta_0, \theta_1, ..., \theta_y] \\ &= \ 0\end{aligned}$$

and we deduce from Lemma 1 that

$$\forall\ y \in [\![0, N-1]\!], \qquad G(N, y) \ = \ 0$$

In particular the $(G(N,y))_{y\in[\![0,M]\!]}$ is non-negative and we can apply the considerations of Section 2 and Section 3. It follows from (11) of Section 2 with $y = N-1$, that

$$\begin{aligned}\forall\ y' \in [\![0, N-2]\!], \qquad \widetilde{Q}(N-1, y') \ &\coloneqq \ \frac{1}{Z(N-1)} \sum_{z\in[\![N-1,N]\!]} Q(N-1, z)G(z, y')\Lambda[h](y') \\ &= \ \frac{1}{Z(y)} Q(N-1, N-1)G(N-1, y')\Lambda[h](y') \\ &= \ 0\end{aligned}$$

where we used, for the second equality, that $G(N, y') = 0$ since $(G(N,y))_{y\in[\![0,N-1]\!]} = 0$, and for the third equality that $G(N-1, y') = 0$ for $y' < N-1$, from Remark 3 of Section 2.

It follows $\widetilde{Q}(N-1, N-1) = 1$, namely $\widetilde{Q}$ is absorbed at $N-1$, as announced in Proposition 1. ■

A more direct proof of Proposition 1 consists in noting that since $\Lambda(N, \cdot) = \delta_N$, the coefficients $(G(N,y))_{y\in[\![0,N-1]\!]}$ in the r.h.s. of (30) necessarily vanish.

Finally, we come to the

Proof of Proposition 2

Due to Remark 5, it is sufficient to prove Proposition 2 with $P_{1/2}$ replaced by P_0. With $N \geqslant 2$, $M = 1$ and $P = P_0$, classical computations give us, on one hand,

$$\forall\ k \in [\![0, N-1]\!], \qquad \theta_k \ = \ \cos\left(\frac{\pi(2N-1-k)}{2N}\right)$$

and in particular

$$\theta_0 \ = \ \cos\left(\frac{2N-1}{2N}\pi\right) \quad \text{and} \quad \theta_1 \ = \ \cos\left(\frac{2N-3}{2N}\pi\right)$$

and on the other hand,

$$\widehat{\theta}_0 = \cos\left(\frac{3\pi}{4}\right) \quad \text{and} \quad \widehat{\theta}_1 = \cos\left(\frac{\pi}{4}\right)$$

i.e.

$$\widehat{\theta}_0 = -\widehat{\theta}_1 = -\frac{1}{\sqrt{2}}$$

Coming back to (38), we get

$$\begin{aligned} S_2(X) &= (X-\theta_0)(X-\theta_1)-(X-\widehat{\theta}_0)(X-\widehat{\theta}_1) \\ &= (X-\theta_0)(X-\theta_1)-(X^2-\widehat{\theta}_0^2) \\ &= \theta_0\theta_1+\widehat{\theta}_0^2-(\theta_0+\theta_1)X \\ &= \widehat{\theta}_0^2-\theta_0^2-(\theta_0+\theta_1)(X-\theta_0) \end{aligned}$$

and thus

$$\begin{aligned} G(2,0) &= \widehat{\theta}_0^2-\theta_0^2 \\ G(2,1) &= -(\theta_0+\theta_1) \end{aligned}$$

So the desired non-negativity amounts to

$$\theta_0^2 \geqslant \widehat{\theta}_0^2 \tag{39}$$

$$\theta_0+\theta_1 \leqslant 0 \tag{40}$$

Condition (40) is satisfied, because the mapping

$$\mathbb{N}\setminus\{1\} \ni N \mapsto \cos\left(\frac{2N-1}{2N}\pi\right)+\cos\left(\frac{2N-3}{2N}\pi\right)$$

is decreasing and vanishes at $N=2$.

Condition (39) is trivially satisfied for $N=2$ but it is not satisfied for $N \geqslant 3$, because the mapping

$$\mathbb{N}\setminus\{1\} \ni N \mapsto \cos\left(\frac{2N-1}{2N}\pi\right)$$

is non-positive and decreasing. ■

We conjecture that Proposition 2 is true more generally: whatever the birth-and-death kernel P as above, we can apply the considerations of Section 2 and Section 3 if and only if $M=N-1$. Despite (32) and interesting related results of Micchelli and Willoughby [8], in particular their Lemma 2.2, we did not succeed in proving or disproving this generalization of Proposition 2.

Remark 6

a) Proposition 1 can be extended to (absorbed or irreducible) skip-free Markov kernels P, under the assumptions that its eigenvalues are non-negative (they can now be complex) and that its spectral polynomials are non-negative, see Fill [7]. The arguments are the same, since Λ satisfies the same properties as those mentioned above.

b) It would be interesting to investigate the situation of more general absorbed Markov chains, conditioned not to be absorbed. See the last section of Fill [7] and Miclo [9] for corresponding intertwinings. □

A Conditioning of finite Markov chains

Consider the setting described at the beginning of Section 2: P is a Markov kernel on the finite set V, A is a proper subset of V not reduced to a singleton, P_A is the restriction of P to $A \times A$ and $\widetilde{P}$ is defined in (9) under the assumption that P_A is irreducible. The goal of this appendix is to recall that $\widetilde{P}$ is the transition kernel of the Markov chain conditioned to stay in A. This result is classical and the proof is given here for the convenience of the reader, since we did not find a suitable reference (for the continuous-time setting, see Section 3.2 of Collet, Martínez and San Martín [2]).

More precisely, let $X \coloneqq (X(n))_{n\in\mathbb{Z}_+}$ be a Markov chain whose transition probabilities are given by P. Consider its exit time from A:

$$\tau_A \quad \coloneqq \quad \inf\{n \in \mathbb{Z}_+ \,:\, X(n) \notin A\}$$

For any $x \in V$, denote $\mathbb{P}_x$ an underlying probability when $X(0) = x$ a.s. For any $n \in \mathbb{Z}_+$, let $\mathcal{B}(n)$ be the sigma-field generated by $X([\![0,n]\!]) \coloneqq (X(m))_{m\in[\![0,n]\!]}$. Before treating the general situation, we deal with the simpler case when P_A is primitive, namely irreducible and aperiodic.

Lemma 2 *Assume that P_A is primitive. For any $x \in A$ and any $B \in \mathcal{B}(n)$ with fixed $n \in \mathbb{Z}_+$, we have*

$$\lim_{N\to+\infty} \mathbb{P}_x[B|\tau_A > N] \quad = \quad \widetilde{\mathbb{P}}_x[B]$$

where $\widetilde{\mathbb{P}}_x$ is the law of a Markov chain on A, starting from x and whose transition probabilities are given by $\widetilde{P}$.

Proof

It is sufficient to consider the case where

$$B \quad = \quad \{X(0) = x_0, X(1) = x_1, ..., X(n) = x_n\} \tag{41}$$

with $(x_l)_{l\in[\![0,n]\!]}$ a sequence of elements from A with $x_0 = x$.

Then we have for $N \geqslant n$,

$$\begin{aligned}
\mathbb{P}_x[B|\tau_A > N] \quad &= \quad \frac{\mathbb{P}_x[X(0) = x_0, X(1) = x_1, ..., X(n) = x_n, \tau_A > N]}{\mathbb{P}_x[\tau_A > N]} \\
&= \quad \frac{\mathbb{P}_{x_n}[\tau_A > N-n]}{\mathbb{P}_x[\tau_A > N]} \prod_{l\in[\![n]\!]} P(x_{l-1}, x_l) \\
&= \quad \frac{\mathbb{P}_{x_n}[\tau_A > N-n]}{\mathbb{P}_x[\tau_A > N]} \frac{h(x)}{h(x_n)} \theta^n \prod_{l\in[\![n]\!]} \widetilde{P}(x_{l-1}, x_l)
\end{aligned}$$

Similarly, we compute that for any $x \in A$ and $N \in \mathbb{Z}_+$,

$$\mathbb{P}_x[\tau_A > N] \quad = \quad \theta^N h(x) \widetilde{\mathbb{E}}_x\left[\frac{1}{h}(\widetilde{X}(N))\right]$$

where $\widetilde{X} \coloneqq (\widetilde{X}(n))_{n\in\mathbb{Z}_+}$ be a Markov chain on A whose transition probabilities are given by P_A and which is starting from $x \in A$.

It follows that

$$\mathbb{P}_x[B|\tau_A > N] \quad = \quad \frac{\widetilde{\mathbb{E}}_{x_n}\left[\frac{1}{h}(\widetilde{X}(N-n))\right]}{\widetilde{\mathbb{E}}_x\left[\frac{1}{h}(\widetilde{X}(N))\right]} \widetilde{\mathbb{P}}_x[B] \tag{42}$$

Not that primitivity is a property only depending on the underlying graph (whose edge set corresponds to positive probability transitions), so that $\widetilde{P}$ is also primitive. As a consequence the Markov chain $\widetilde{X}$ is ergodic and the law of $\widetilde{X}(n)$ converges for large $n \in \mathbb{Z}_+$ toward the invariant measure $\widetilde{\pi}$ of $\widetilde{P}$, whatever the initial distribution of $\widetilde{X}(0)$. We deduce that

$$\begin{aligned}\lim_{N\to+\infty} \mathbb{P}_x[B|\tau_A > N] &= \frac{\widetilde{\pi}[1/h]}{\widetilde{\pi}[1/h]}\widetilde{\mathbb{P}}_x[B] \\ &= \widetilde{\mathbb{P}}_x[B]\end{aligned}$$

as desired. ■

Let us now come to the general situation, only assuming P_A irreducible, and consider how the previous proof should be modified. Denote $d \in \mathbb{N}$ its period. The aperiodicity (and thus primitivity of A) corresponds to the case $d = 1$. Let C_k, for $k \in \mathbb{Z}_d := \mathbb{Z}/(d\mathbb{Z})$, be the underlying periodic classes, indexed so that

$$\forall\ k \in \mathbb{Z}_d,\ \forall\ x \in C_k, \qquad P_A[\mathbb{1}_{C_{k+1}}] = 1 \tag{43}$$

Note that $\widetilde{P}$ has the same period and the same periodicity classes as P_A and that (43) also holds with P_A replaced by $\widetilde{P}$. The ergodicity of $\widetilde{X}$ has to be replaced by the following convergence, for any test function f defined on A,

$$\forall\ k, l \in \mathbb{Z}_d,\ \forall\ x \in C_k, \qquad \lim_{N\to\infty} \widetilde{\mathbb{E}}_x[f(\widetilde{X}(l + dN))] = \frac{\widetilde{\pi}[f\mathbb{1}_{C_{l+k}}]}{\widetilde{\pi}[\mathbb{1}_{C_{l+k}}]}$$

where $\widetilde{\pi}$ is still the unique invariant probability of $\widetilde{P}$ (and where elements of $\mathbb{Z}_d$ have been identified with their representative elements in $[\![0, d-1]\!]$, as it will also be the case in the sequel).

Let us revisit the above proof in view of this change. First in (41), we can restrict ourselves to the situation where $x_0 = x \in C_k$, for some $k \in \mathbb{Z}_d$, and $x_l \in C_{k+l}$, for $l \in [\![n]\!]$. Indeed, otherwise, we have

$$\begin{aligned}\forall\ N \in \mathbb{Z}_+, \qquad \mathbb{P}_x[B|\tau_A > N] &= 0 \\ \widetilde{\mathbb{P}}_x[B] &= 0\end{aligned}$$

so Lemma 2 holds for such events.

We can thus concentrate on events B of the form described above. From (42), which is also valid in the periodic case, we get, for any $l \in \mathbb{Z}_d$,

$$\mathbb{P}_x[B|\tau_A > l + dN] = \frac{\widetilde{\mathbb{E}}_{x_n}\left[\frac{1}{h}(\widetilde{X}(l + dN - n))\right]}{\widetilde{\mathbb{E}}_x\left[\frac{1}{h}(\widetilde{X}(l + dN))\right]}\widetilde{\mathbb{P}}_x[B]$$

and for large N the r.h.s. is converging toward

$$\frac{\widetilde{\pi}[\mathbb{1}_{C_{l-n+n+k}}/h]}{\widetilde{\pi}[\mathbb{1}_{C_{l-n+n+k}}]}\frac{\widetilde{\pi}[\mathbb{1}_{C_{l+k}}]}{\widetilde{\pi}[\mathbb{1}_{C_{l+k}}/h]}\widetilde{\mathbb{P}}_x[B] = \widetilde{\mathbb{P}}_x[B]$$

Since the limit $\widetilde{\mathbb{P}}_x[B]$ for large N of the quantity $\mathbb{P}_x[B|\tau_A > l+dN]$ does not depend on $l \in [\![0, d-1]\!]$, we deduce that Lemma 2 holds.

Remark 7 Even if it was not required by the above computations, let us mention that it is well-known that $\widetilde{\pi}$ is the probability on A admitting a density proportional to h with respect to ν, the quasi-invariant probability, i.e. the unique probability on A satisfying

$$\nu P_A = \theta\nu \tag{44}$$

(as for h, only the irreducibility of P_A is needed for the existence and uniqueness of ν).

Indeed, for any test function f defined on A, we have

$$\begin{aligned} \nu[h\widetilde{P}[f]] &= \frac{\nu[P_A[hf]]}{\theta} \\ &= \frac{\theta\nu[hf]}{\theta} \\ &= \nu[hf] \end{aligned}$$

□

Remark 8 The irreducibility of P_A was convenient to deduce the existence of h and of the quasi-stationary measure ν.

Nevertheless this irreducibility condition is not necessary for Lemma 2 to hold. In fact, looking at the above arguments, it appears only the existence of h satisfying (8) (with $h > 0$ on A and $h = 0$ on $V\backslash A$) and the uniqueness of the invariant probability $\widetilde{\pi}$ for $\widetilde{P}$ are needed. The uniqueness of $\widetilde{\pi}$ is equivalent to the uniqueness of a minimal recurrent class for P_A, namely a subset $\emptyset \neq C \subset A$ such that $P_{C\times C}$ is a Markov transition and such that $P_{C'\times C'}$ is not a Markov transition for any proper subset $C' \subset C$.

This happens in the following situation: assume the existence of h as above and that P is absorbed at a point $\infty \in A$ (i.e. $P(\infty, \infty) = 1$). Furthermore suppose that for any $x \in A$ there is a P-path going from x to ∞ and staying in A. Then it is not difficult to check that δ_∞ is the unique invariant measure of $\widetilde{P}$. Note that in this case a Markov chain associated to $\widetilde{P}$ ends up being absorbed at ∞. Let us give a very simple example of this situation. We take $V = \{0, 1, 2\}$, $A = \{0, 1\}$ and P such that

$$P_A = \begin{pmatrix} 1 & 0 \\ a & b \end{pmatrix}$$

with $a, b > 0$ such that $a + b < 1$. Then 0 is absorbing for P_A and (8) is satisfied with $\theta = 1$ and

$$h = \begin{pmatrix} 1-b \\ a \end{pmatrix}$$

We deduce that

$$\widetilde{P} = \begin{pmatrix} 1 & 0 \\ 1-b & b \end{pmatrix}$$

□

References

[1] David Aldous and Persi Diaconis. Shuffling cards and stopping times. *Amer. Math. Monthly*, 93(5):333–348, 1986.

[2] Pierre Collet, Servet Martínez, and Jaime San Martín. *Quasi-stationary distributions*. Probability and its Applications (New York). Springer, Heidelberg, 2013. Markov chains, diffusions and dynamical systems.

[3] Carl de Boor. Divided differences. *Surv. Approx. Theory*, 1:46–69, 2005.

[4] Persi Diaconis and James Allen Fill. Strong stationary times via a new form of duality. *Ann. Probab.*, 18(4):1483–1522, 1990.

[5] Persi Diaconis and Laurent Miclo. On times to quasi-stationarity for birth and death processes. *J. Theoret. Probab.*, 22(3):558–586, 2009.

[6] James Allen Fill. On hitting times and fastest strong stationary times for skip-free and more general chains. *J. Theoret. Probab.*, 22(3):587–600, 2009.

[7] James Allen Fill. The passage time distribution for a birth-and-death chain: strong stationary duality gives a first stochastic proof. *J. Theoret. Probab.*, 22(3):543–557, 2009.

[8] Charles A. Micchelli and R. A. Willoughby. On functions which preserve the class of Stieltjes matrices. *Linear Algebra Appl.*, 23:141–156, 1979.

[9] Laurent Miclo. On absorption times and Dirichlet eigenvalues. *ESAIM Probab. Stat.*, 14:117–150, 2010.

[10] Laurent Miclo and Chi Zhang. On a family of isospectral pure-birth processes. *ALEA, Lat. Am. J. Probab. Math. Stat.*, 18(2):1759–1771, 2021.

[11] Soumik Pal and Mykhaylo Shkolnikov. Intertwining diffusions and wave equations. *ArXiv e-prints*, June 2013.

[12] Jim W. Pitman. One-dimensional Brownian motion and the three-dimensional Bessel process. *Advances in Appl. Probability*, 7(3):511–526, 1975.

[13] L. Chris G. Rogers and Jim W. Pitman. Markov functions. *Ann. Probab.*, 9(4):573–582, 1981.

[14] Wikipedia contributors. Divided differences — Wikipedia, the free encyclopedia. `https://en.wikipedia.org/w/index.php?title=Divided_differences&oldid=1127366892`, 2022. [Online; accessed 30-December-2022].

Institut de Mathématiques de Bordeaux
Univ. Bordeaux, CNRS, Bordeaux INP, IMB
F-33400 Talence, France

Institut Élie Cartan de Lorraine
Université de Lorraine

Toulouse School of Economics
Institut de Mathématiques de Toulouse
1, Esplanade de l'université
31080 Toulouse cedex 06, France