On the hypergroup property

Laurent Miclo

On the hypergroup property

Laurent Miclo*

Institut de Mathématiques de Toulouse, UMR 5219
Université de Toulouse and CNRS, France

Abstract

The hypergroup property satisfied by certain reversible Markov chains can be seen as a generalization of the convolution related features of class random walks on groups. Carlen, Geronimo and Loss [3] developed a method for checking this property in the context of Jacobi eigen-polynomials. A probabilistic extension of their approach is proposed here, enabling to recover the discrete example of the biased Ehrenfest model due to Eagleson [8]. Next a spectral characterization is provided for finite birth and death chains enjoying the hypergroup property with respect to one of the boundary points.

Keywords: hypergroup property, finite reversible Markov processes, biaised Ehrenfest model, finite birth and death processes, Neumann eigenvectors, Dirichlet (minor) eigenvalues.

MSC2010: primary: 20N20, secondary: 60J05, 60J10, 15A18, 15A42, 60J80, 65F15.

*Partially supported by ANR-12-BS01-0019

1 A theoretical result

There are several definitions of the hypergroup property for a reversible Markov kernel P. One of them, recalled in (4) below, is the non-negativity of certain sums of products of quantities related to the eigenfunctions associated to P. In the context of Jacobi polynomials, Carlen, Geronimo and Loss [3] developed a method in order to check this property. Here we begin by extending it, giving it a general probabilistic flavor by replacing in their criterion some mappings by Markov kernels. Next we will see how the resulting abstract condition can be applied to recover the first instance of the hypergroup property, namely the example of the biased Ehrenfest model due to Eagleson [8]. We will investigate further the hypergroup property for birth and death Markov chains, by providing a spectral criterion in the last section. For general motivations relative to the notion of the hypergroup property, see for instance Diaconis and Griffiths [4] or Bakry and Huet [2].

Let $(\bar{S}, \bar{\mathcal{S}}, \bar{\mu}, \bar{P})$ be a reversible Markov framework: $(\bar{S}, \bar{\mathcal{S}})$ is a measurable space endowed with a probability measure $\bar{\mu}$ and $\bar{P}$ is a self-adjoint Markovian operator on $\mathbb{L}^2(\bar{\mu})$. Recall that the Markov property consists in two assumptions: on one hand, for any non-negative function $f \in \mathbb{L}^2(\bar{\mu})$, $\bar{P}[f]$ is non-negative, and on the other hand, $\bar{P}[\mathbb{1}] = \mathbb{1}$, where $\mathbb{1} \in \mathbb{L}^2(\bar{\mu})$ is the constant function taking the value 1 ($\bar{\mu}$-a.s.).

Consider another measurable space $(S, \mathcal{S})$, as well as a Markov kernel Q from $(\bar{S}, \bar{\mathcal{S}})$ to $(S, \mathcal{S})$: it is a mapping from $\bar{S} \times \mathcal{S}$ to $[0, 1]$ such that for any $\bar{x} \in \bar{S}$, $Q(\bar{x}, \cdot)$ is a probability distribution and for any $A \in \mathcal{S}$, $Q(\cdot, A)$ is a measurable mapping (for our purpose, the requirements with respect to the first variable only need to be satisfied $\bar{\mu}$-a.s.). The kernel Q can be seen as a Markov operator from $\mathcal{B}(S)$, the space of bounded measurable functions defined on $(S, \mathcal{S})$, to $\mathcal{B}(\bar{S})$, via the formula

$$\forall\ f \in \mathcal{B}(S),\ \forall\ \bar{x} \in \bar{S}, \qquad Q[f](\bar{x}) := \int f(x)\, Q(\bar{x}, dx)$$

Denote by μ the image of $\bar{\mu}$ by Q:

$$\forall\ A \in \mathcal{S}, \qquad \mu(A) \ := \ \int Q(\bar{x}, A)\, \bar{\mu}(d\bar{x})$$

Then we have

$$\forall\ f \in \mathcal{B}(S), \qquad \mu[f] \ = \ \bar{\mu}[Q[f]]$$

and since by Cauchy-Schwarz inequality, $(Q[f])^2 \leqslant Q[f^2]$, it appears that Q can be extended into an operator of norm 1 from $\mathbb{L}^2(\mu)$ to $\mathbb{L}^2(\bar{\mu})$. Denote by Q^* the adjoint operator of Q, it is in particular an operator of norm 1 from $\mathbb{L}^2(\bar{\mu})$ to $\mathbb{L}^2(\mu)$. In fact we have a little better:

Lemma 1 *The operator Q^* is Markovian.*

Proof

To check the preservation of non-negativeness, it is sufficient to see that for any non-negative $f \in \mathbb{L}^2(\bar{\mu})$ and $g \in \mathbb{L}^2(\mu)$, $\langle Q^*[f], g\rangle_\mu \geqslant 0$, where $\langle \cdot, \cdot\rangle_\mu$ stands for the scalar product in $\mathbb{L}^2(\mu)$. This property is an immediate consequence of

$$\begin{aligned} \langle Q^*[f], g\rangle_\mu &= \langle f, Q[g]\rangle_{\bar{\mu}} \\ &\geqslant 0 \end{aligned}$$

For the computation of $Q^*[\mathbb{1}]$, note that for any $f \in \mathbb{L}^2(\mu)$,

$$\begin{aligned} \langle Q^*[\mathbb{1}], f\rangle_\mu &= \langle \mathbb{1}, Q[f]\rangle_{\bar{\mu}} \\ &= \bar{\mu}[Q[f]] \\ &= \mu[f] \\ &= \langle \mathbb{1}, f\rangle_\mu \end{aligned}$$

Since this is valid for all $f \in \mathbb{L}^2(\mu)$, we conclude that $Q^*[\mathbb{1}] = \mathbb{1}$.

■

Define

$$P \quad \coloneqq \quad Q^* \bar{P} Q \tag{1}$$

By composition, P is a Markov operator from $\mathbb{L}^2(\mu)$ to $\mathbb{L}^2(\mu)$, which is clearly self-adjoint, by self-adjointness of $\bar{P}$. To get a more interesting property of P, we need to introduce the following notion. A Markov operator G from $\mathbb{L}^2(\bar{\mu})$ to itself is said to be Q-compatible if we have

$$QQ^*GQ \quad = \quad GQ \tag{2}$$

Lemma 2 *If $\bar{P}$ is Q-compatible, then the operators $\bar{P}$ and P are intertwined through Q:*

$$QP = \bar{P}Q$$

Proof

By definition, we have

$$\begin{aligned} QP &= QQ^*\bar{P}Q \\ &= \bar{P}Q \end{aligned}$$

by Q-compatibility.

■

From now on, $\bar{P}$ is assumed to be Q-compatible. It seems that an important tool to investigate the Markov operator P intertwined with $\bar{P}$ is the set $\mathcal{G}$ of Markov operators G from $\mathbb{L}^2(\bar{\mu})$ to itself which commute with $\bar{P}$, $G\bar{P} = \bar{P}G$, and which are Q-compatible. This set $\mathcal{G}$ has the structure of a semi-group: for all $G, G' \in \mathcal{G}$, $GG' \in \mathcal{G}$. Indeed, GG' clearly commutes with $\bar{P}$ if both G and G' commute with $\bar{P}$. If (2) is satisfied by G and G', then the same is true for GG', because

$$\begin{aligned} QQ^*GG'Q &= QQ^*GQQ^*G'Q \\ &= GQQ^*G'Q \\ &= GG'Q \end{aligned}$$

In particular $\mathcal{G}$ contains $\{\bar{P}^n : n \in \mathbb{Z}_+\}$, the semi-group generated by $\bar{P}$, but as it can be observed on the example of the next section, $\mathcal{G}$ can be larger than a temporal evolution semi-group. Under the above setting, to each $G \in \mathcal{G}$, we can associate a Markov operator K_G on $\mathbb{L}^2(\mu)$, via

$$K_G \quad \coloneqq \quad Q^*GQ$$

Proposition 3 *For all $G \in \mathcal{G}$, K_G and P commute.*

Proof

The argument is similar to that used in the proof of Lemma 2: using Lemma 5 and Assumption (2), it appears that

$$\begin{aligned} K_GP &= Q^*GQQ^*\bar{P}Q \\ &= Q^*G\bar{P}Q \\ &= Q^*\bar{P}GQ \\ &= Q^*\bar{P}QQ^*GQ \\ &= PK_G \end{aligned}$$

■

It is time to come to the main application of the above considerations. Assume that S is a finite set of cardinal $N \in \mathbb{N}$ (then up to lumping some of its elements together, there is no loss of generality in taking for $\mathcal{S}$ the σ-algebra consisting of all the subsets of S and up to removing the μ-negligible points from S, we furthermore assume that μ gives a positive weight to all the points of S). By symmetry, P is diagonalizable, let $(\varphi_l)_{l\in[\![N]\!]}$ be an orthonormal basis of $\mathbb{L}^2(\mu)$ consisting of its eigenvectors. We make the hypothesis that all the eigenvalues of P are of multiplicity 1 and that there exists $x_0 \in S$ such that for all $x_1 \in S$, there exists $G \in \mathcal{G}$ such that

$$K_G(x_0, \cdot) \;=\; \delta_{x_1} \tag{3}$$

Theorem 4 *Under the above conditions, P satisfies the hypergroup property with respect to x_0, namely we have $\varphi_l(x_0) \neq 0$ for all $l \in [\![N]\!]$ and*

$$\forall\ x, y, z \in S, \qquad \sum_{l\in[\![N]\!]} \frac{\varphi_l(x)\varphi_l(y)\varphi_l(z)}{\varphi_l(x_0)} \;\geqslant\; 0 \tag{4}$$

Proof

Fix $l \in [\![N]\!]$ and denote θ_l the eigenvalue associated to the eigenvector φ_l. From $P[\varphi_l] = \theta_l\varphi_l$ and Proposition 3, we deduce that for any $G \in \mathcal{G}$,

$$\begin{aligned} P[K_G[\varphi_l]] &= K_G[P[\varphi_l]] \\ &= \theta_l K_G[\varphi] \end{aligned}$$

namely, either $K_G[\varphi_l] = 0$ or $K_G[\varphi_l]$ is an eigenvector of P associated to the eigenvalue θ_l. Due the multiplicity 1 of this eigenvalue, we deduce that $K_G[\varphi_l]$ is proportional to φ_l (this being also true if $K_G[\varphi_l] = 0$), say $K_G[\varphi_l] = \lambda(G,l)\varphi_l$. Since this is true for all $l \in [\![N]\!]$, the spectral decomposition of K_G is given by $(\lambda(G,l), \varphi_l)_{l\in[\![N]\!]}$ and we have

$$\forall\ y, z \in S, \qquad K_G(y,z) \;=\; \sum_{l\in[\![N]\!]} \lambda(G,l)\varphi_l(y)\varphi_l(z)\mu(z) \tag{5}$$

Fix $x_1 \in S$ and let $G \in \mathcal{G}$ be as in (3). We get from this equation that

$$\begin{aligned} \lambda(G,l)\varphi(x_0) &= K_G[\varphi_l](x_0) \\ &= \varphi(x_1) \end{aligned}$$

If $\varphi_l(x_0)$ was to vanish, the same would be true of $\varphi_l(x_1)$, for all $x_1 \in S$, contradicting that φ is a vector of norm 1. Thus $\varphi_l(x_0) \neq 0$ and $\lambda(G,l) = \varphi(x_1)/\varphi(x_0)$. It follows from (5) that for all $y, z \in S$,

$$\begin{aligned} \sum_{l\in[\![N]\!]} \frac{\varphi_l(x_1)\varphi_l(y)\varphi_l(z)}{\varphi_l(x_0)} &= \sum_{l\in[\![N]\!]} \lambda(G,l)\varphi_l(y)\varphi_l(z) \\ &= \frac{K_G(y,z)}{\mu(z)} \\ &\geqslant 0 \end{aligned}$$

because K_G is a Markovian matrix. This is the wanted hypergroup property, because x_1 was chosen arbitrarily.

■

To put more flesh on the notion of Q-compatibility, let us present a traditional instance of the above setting. Instead of a Markov kernel Q, assume that we are given a measurable mapping q

from $(\bar{S}, \bar{\mathcal{S}})$ to $(S, \mathcal{S})$. It can also be seen as a "deterministic" Markov kernel from $(\bar{S}, \bar{\mathcal{S}})$ to $(S, \mathcal{S})$, via

$$\forall\, x \in \bar{S}, \qquad Q(\bar{x}, \cdot) \;\coloneqq\; \delta_{q(\bar{x})}(\cdot) \tag{6}$$

so that the above development applies.

Let $\mathcal{T}$ be the σ-field generated by q. In this context, the operator Q is an isometry from $\mathbb{L}^2(\mu)$ to $\mathbb{L}^2(\bar{\mu})$, because for all $f \in \mathcal{B}(S)$, $(Q[f])^2 = Q[f^2]$ (this identity is in fact a characterization of the $\bar{\mu}$-a.s. determinism of Q). A convenient property of Q^* is:

Lemma 5 *The Markov operator QQ^* corresponds to the conditional expectation with respect to $\mathcal{T}$.*

Proof

By composition, QQ^* is a Markov operator from $\mathbb{L}^2(\bar{\mu})$ to itself. To show that it corresponds to the conditional expectation with respect to $\mathcal{T}$, we must prove that

$$\forall\, f, g \in \mathbb{L}^2(\bar{\mu}), \qquad \langle QQ^*[f], Q[g]\rangle_{\bar{\mu}} \;=\; \langle f, Q[g]\rangle_{\bar{\mu}}$$

This comes from the fact that Q is an isometry, which implies that the l.h.s. is equal to

$$\begin{aligned}\bar{\mu}[Q[Q^*[f]g]] &= \mu[Q^*[f]g] \\ &= \langle Q^*[f], g\rangle_\mu \\ &= \langle f, Q[g]\rangle_{\bar{\mu}}\end{aligned}$$

■

The notion of Q-compatibility (2) of a Markov kernel G is then equivalent to

$$GQ \text{ is a Markov operator from } \mathbb{L}^2(\mu) \text{ to } \mathbb{L}^2(\bar{\mu}, \mathcal{T}) \tag{7}$$

where $\mathbb{L}^2(\bar{\mu}, \mathcal{T})$ is the subspace of $\mathbb{L}^2(\mu)$ consisting of functions measurable with respect to $\mathcal{T}$.

Thus in the context of a deterministic Q, Lemma 2 amounts to the famous criterion of Dynkin [7] insuring that a function of a Markov chain is itself a Markov chain.

2 An example

Eagleson [8] proved that the biased Ehrenfest model satisfies the hypergroup property, let us show how Theorem 4 enables to recover this result.

We begin by recalling the underlying birth and death Markov transition kernel P on $S \coloneqq [\![0, N]\!]$, with $N \in \mathbb{N}^*$ (so there is a slight modification of the notations of Theorem 4: the cardinal of S is now $N + 1$), parametrized by $p \in (0, 1)$:

$$\forall\, x, y \in [\![0, N]\!], \qquad P(x, y) \;\coloneqq\; \begin{cases} \frac{N-x}{N} p & \text{, if } y = x + 1 \\ \frac{x}{N}(1 - p) & \text{, if } y = x - 1 \\ 1 - p - (1 - 2p)\frac{x}{N} & \text{, if } y = x \\ 0 & \text{, otherwise} \end{cases} \tag{8}$$

This birth and death kernel is irreducible and its unique reversible probability measure μ is given by

$$\forall\, x \in [\![0, N]\!], \qquad \mu(x) \;=\; \binom{N}{x} p^x (1 - p)^{N-x} \tag{9}$$

This can be computed directly or deduced, as in the previous section, from the existence of a simple reversible Markov framework $(\bar{S}, \bar{\mathcal{S}}, \bar{\mu}, \bar{P})$ "above" P. Indeed, take $\bar{S} = \{0,1\}^N$, endowed with the σ-field $\bar{\mathcal{S}}$ of all its subsets, and consider the mapping q going from $\bar{S}$ to S defined by

$$\forall\ \bar{x} \coloneqq (\bar{x}_l)_{l\in[\![0,N]\!]} \in \bar{S}, \qquad q(\bar{x}) \ \coloneqq \ \sum_{l\in[\![N]\!]} \bar{x}_l$$

The probability measure $\bar{\mu}$ is given by

$$\forall\ \bar{x} \in \bar{S}, \qquad \bar{\mu}(\bar{x}) \ \coloneqq \ p^{q(\bar{x})}(1-p)^{q(\bar{x})}$$

For any $l \in [\![0,N]\!]$, consider the Markov transition matrix $\bar{P}_l$ defined by

$$\forall\ x \coloneqq (\bar{x}_k)_{k\in[\![0,N]\!]}, y \coloneqq (\bar{y}_k)_{k\in[\![0,N]\!]} \in \bar{S}, \quad \bar{P}_l(\bar{x},\bar{y}) \ \coloneqq \ \begin{cases} p & \text{, if } \bar{y}_l = 1 \text{ and } \bar{y}_k = \bar{x}_k, \text{ for } k \neq l \\ 1-p & \text{, if } \bar{y}_l = 0 \text{ and } \bar{y}_k = \bar{x}_k, \text{ for } k \neq l \\ 0 & \text{, otherwise} \end{cases}$$

The measure $\bar{\mu}$ is clearly reversible for $\bar{P}_l$, as well as for

$$\bar{P} \ \coloneqq \ \frac{1}{N}\sum_{l\in[\![N]\!]} \bar{P}_l$$

As in the end of last section, we reinterpret the mapping q as the Markov kernel from $(\bar{S}, \bar{\mathcal{S}})$ to $(S, \mathcal{S})$ given in (6). The associated σ-field $\mathcal{T} \subset \mathcal{S}$ consists of the events which are left invariant by all the permutations of the indices. In particular, Assumption (2) is satisfied, $\bar{P}Q$ being clearly a Markov operator from $\mathbb{L}^2(\mu)$ to $\mathbb{L}^2(\bar{\mu}, \mathcal{T})$. Furthermore the Markovian matrix P defined in (1) is given by (8) and the image of $\bar{\mu}$ by q coincides with μ described in (9). Thus μ is necessarily reversible with respect to P.

But the interest of the above construction is that it enables to recover the hypergoup property of P via Theorem 4. From now on, assume that $p \in (0, 1/2]$ (if $p \in (1/2, 1)$, reverse the order of the segment $[\![0,N]\!]$ to come back to the situation where $p \in (0, 1/2)$). For $l \in [\![N]\!]$, consider the Markov transition matrix H_l defined by

$$\forall\ x \coloneqq (\bar{x}_k)_{k\in[\![N]\!]}, y \coloneqq (\bar{y}_k)_{k\in[\![N]\!]} \in \bar{S},$$

$$H_l(\bar{x},\bar{y}) \ \coloneqq \ \begin{cases} 1 & \text{, if } \bar{x}_l = 1,\ \bar{y}_l = 0 \text{ and } \bar{y}_k = \bar{x}_k, \text{ for } k \neq l \\ p/(1-p) & \text{, if } \bar{x}_l = 0,\ \bar{y}_l = 1 \text{ and } \bar{y}_k = \bar{x}_k, \text{ for } k \neq l \\ (1-2p)/(1-p) & \text{, if } \bar{x}_l = 0,\ \bar{y}_l = 0 \text{ and } \bar{y}_k = \bar{x}_k, \text{ for } k \neq l \\ 0 & \text{, otherwise} \end{cases}$$

It is immediate to check that

$$\bar{P}_l \ = \ pI + (1-p)H_l$$

where I is the identity matrix (seen as the motionless Markov kernel). In particular, H_l commutes with $\bar{P}_l$ and with $\bar{P}$. More generally, for $A \subset [\![N]\!]$, let H_A be given by

$$H_A \ \coloneqq \ \prod_{l\in A} H_l$$

(in r.h.s. the order of the compositions of the Markov kernels does not matter, because they commute among themselves). Again, H_A is a Markov kernel commuting with $\bar{P}$. Nevertheless, it lacks symmetry to belong to $\mathcal{G}$. So for any $l \in [\![N]\!]$, consider

$$G_l \ \coloneqq \ \frac{1}{\binom{N}{l}} \sum_{A\subset[\![N]\!]\,:\,\mathrm{card}(A)=l} H_A$$

which is easily seen to belong to $\mathcal{G}$.

This leads us to consider the Markov kernel K_{G_l} on $[\![0,N]\!]$. It appears without difficulty that

$$\forall\ l \in [\![N]\!], \qquad K_{G_l}(N,\cdot) \;=\; \delta_{N-l}(\cdot)$$

This observation enables to apply Theorem 4 to get that P satisfies the hypergroup property with respect to the point N (if $p \geqslant 1/2$, P satisfies the hypergroup property with respect to the point 0).

To investigate the extent of the applicability of the approach of the previous section, it would be interesting to study the multidimensional Krawtchouk polynomials, which are a multidimensional extension of the above example, cf. Diaconis and Griffiths [4, 5]. Nevertheless, to generalize the result of this section, staying in the one-dimensional setting of finite birth and death chains, already presents surprising challenges, as we are now going to see.

3 On birth and death chains

Instead of working with a "covering Markov framework" $(\bar{S}, \bar{\mathcal{S}}, \bar{\mu}, \bar{P})$, where hidden symmetries in the initial model $(S, \mathcal{S}, \mu, P)$ are more obvious, one can also try to find directly the commuting Markov kernels. We investigate here the situation of finite birth and death chains, by providing a spectral characterization of the hypergroup property with respect to the left boundary point. This enables to construct a practical algorithm for checking this property. Next we conjecture two seemingly natural discrete versions of Achour-Trimèche's theorem [1] (see also Bakry and Huet [2]), asserting the hypergroup property under certain log-concavity of the reversible measure. Using numerical implementations of the proposed algorithm, it appears they are both wrong.

We begin by recalling the framework of finite birth and death chains. For some $N \in \mathbb{N}$, we take $S := [\![0,N]\!]$ endowed with its total σ-field $\mathcal{S}$ and an irreducible birth and death Markov kernel P, i.e. whose permitted transitions are those to the nearest neighbors, S being given its usual discrete line graph structure (with self-connecting loop at each vertex, to allow non-zero diagonal entries for P). Then there exists a unique invariant probability measure μ for P and it is reversible. Our purpose is to investigate the set of Markov kernels commuting with P, namely the set

$$\mathcal{K} \;:=\; \{K \in \mathcal{M}\ :\ KP = PK\}$$

where $\mathcal{M}$ is the set Markov kernels on S. Note that the elements of $\mathcal{K}$ admits μ as invariant probability. Indeed, we have

$$\begin{aligned}\mu KP &= \mu PK\\ &= \mu K\end{aligned}$$

This shows that μK is invariant by P, so that $\mu K = \mu$.

We are looking for conditions on P which ensure that for any probability distribution μ_0, there exists a Markov kernel $K_{\mu_0} \in \mathcal{K}$ such that $K_{\mu_0}(0,\cdot) = \mu_0$, namely we are trying to check the hypergroup property with respect to 0. By convexity of $\mathcal{K}$, this amounts to find, for any given $x_1 \in S$, $K_{x_1} \in \mathcal{K}$ such that $K_{x_1}(0,\cdot) = \delta_{x_1}(\cdot)$, since we can next take for any probability distribution μ_0,

$$K_{\mu_0} \;=\; \sum_{x_1 \in S} \mu_0(x_1) K_{x_1} \tag{10}$$

Remark 6 The commutation relation $KP = PK$ can be seen as a discrete wave equation in K, by interpreting the first (respectively, second) variable in the matrix K as a time (resp., space) variable. More precisely, denote k the density kernel associated to K:

$$\forall\ t, x \in [\![0,N]\!], \qquad k(t,x) \;:=\; \frac{K(t,x)}{\mu(x)}$$

Using that for all $x, y \in [\![0, N]\!]$, $\mu(x)P(x, y) = \mu(y)P(y, x)$, we can transform the equality

$$\forall\ t, x \in [\![0, N]\!], \qquad \sum_{y \in S} K(t, y)P(y, x) \quad = \quad \sum_{y \in S} P(t, y)K(y, x)$$

into

$$\forall\ t, x \in [\![0, N]\!], \qquad \mu(x) \sum_{y \in S} P(x, y)k(t, y) \quad = \quad \mu(x) \sum_{y \in S} P(t, y)k(y, x)$$

Dividing by $\mu(x)$ and considering the generator matrix $L = P - I$, we get

$$\forall\ t, x \in [\![0, N]\!], \qquad L^{(1)}[k](t, x) \quad = \quad L^{(2)}[k](t, x)$$

where for $i \in \{1, 2\}$, $L^{(i)}$ stands for the generator acting on the i-th variable as L. A least formally, one recognizes a wave equation. Thus our objective is to see when a wave equation starting from a non-negative initial condition remains non-negative.

□

The biased Ehrenfest birth and death processes of the previous section with $p \in [1/2, 1)$ provide examples of the Markov kernels we want to characterize. We will denote by M_p the Markov matrix defined in (8) with $p \in [1/2, 1)$. Here is a simpler example where the discrete wave interpretation is particularly obvious:

Example 7 Consider the birth and death random walk on $[\![0, N]\!]$: its Markov kernel M_0 (not to be confused with the notation M_p, for $p \in [1/2, 1)$, defined above) is given by

$$\forall\ x, y \in [\![0, N]\!], \qquad M_0(x, y) \quad = \quad \begin{cases} 1 & \text{, if } (x, y) = (0, 1) \text{ or } (x, y) = (N, N-1) \\ 1/2 & \text{, if } |x - y| = 1 \text{ and } x \in \{0, N\} \\ 0 & \text{, otherwise} \end{cases}$$

For any $x_0 \in [\![0, N]\!]$ and $\varepsilon \in \{-1, 1\}$, let $(\psi_{x_0,\varepsilon}(x))_{x \in \mathbb{Z}_+}$ the deterministic and discrete time evolution in $[\![0, N]\!]$ constructed in the following way: $\psi_{x_0,\varepsilon}(0) = x_0$ and if $x_0 \in [\![1, N-1]\!]$, then we take $\psi_{x_0,\varepsilon}(1)$ to be $x_0 + \varepsilon$. If $x_0 = 0$ (respectively $x_0 = N$), we take $\psi_{x_0,\varepsilon}(1) = 1$ (resp. $\psi_{x_0,\varepsilon}(1) = N - 1$). Next for $x \in \mathbb{N}$, if $\psi_{x_0,\varepsilon}(x - 1)$ and $\psi_{x_0,\varepsilon}(x)$ have been constructed with $d_{x_0,\varepsilon}(x) \coloneqq \psi_{x_0,\varepsilon}(x) - \psi_{x_0,\varepsilon}(x - 1) \in \{-1, 1\}$, then we take $\psi_{x_0,\varepsilon}(x + 1) = \psi_{x_0,\varepsilon}(x) + d_{x_0,\varepsilon}(x)$, except if it is not possible (i.e. $\psi_{x_0,\varepsilon}(x) \in \{0, N\}$), in which case we consider $\psi_{x_0,\varepsilon}(x + 1) = \psi_{x_0,\varepsilon}(x) - d_{x_0,\varepsilon}(x)$. Visually, it corresponds to a trajectory of a particle issued from x_0, starting to go to $x_0 + \varepsilon$ and keeping in the same direction until it is reflected on one of the "walls at -1 and $N + 1$".

We leave to the reader as an exercise to check that for any $x_0 \in [\![0, N]\!]$, the Markov kernel K_{x_0} defined by

$$\forall\ x, y \in [\![0, N]\!], \qquad K_{x_0}(x, y) \quad = \quad \frac{1}{2}\left(\delta_{\psi_{x_0,1}(y)} + \delta_{\psi_{x_0,-1}(y)}\right)$$

does commute with M_0.

□

We come back to the situation of a general irreducible birth and death Markov kernel P. To describe the main theoretical result of this section, we need some further notations. For $n \in [\![0, N-1]\!]$, denote by $-1 < \theta_{n,0} < \theta_{n,1} < \cdots < \theta_{n,n-1} < 1$ the n eigenvalues of the minor of P corresponding to the rows and columns indexed by $[\![0, n-1]\!]$.

Proposition 8 *The Markov kernel P satisfies the hypergroup property with respect to 0 if and only if for all $n \in [\![0, N-1]\!]$, the matrix*

$$(P - \theta_{n,0})(P - \theta_{n,1}) \cdots (P - \theta_{n,n-1})$$

has non-negative entries.

We excluded the case $n = N$, because the corresponding product matrix vanishes by the Hamilton-Cayley theorem.

Remark 9 Markov kernels usually refer to discrete time processes. Continuous time processes rather use Markov generators. A matrix L is called a Markov generator if its off-diagonal entries are non-negative and if the raw-sums all vanish. It is equivalent to the fact that we can find a positive number $l > 0$ and a Markov kernel P such that $L = l(P - I)$, where I is the corresponding identity matrix. A technical advantage of Markov generators over Markov kernels is that it is straightforward to perturb them (in addition to the fact that continuous time is often easier to manage than discrete time). It is convenient for them to rewrite the above result under the following form.

Consider an irreducible birth and death Markov L generator on $[\![0, N]\!]$. For $n \in [\![0, N-1]\!]$, denote by $-1 < \lambda_{n,0} < \lambda_{n,1} < \cdots < \lambda_{n,n-1} < 1$ the eigenvalues of the minor of L corresponding to the rows and columns indexed by $[\![0, n]\!]$. The Markov generator L satisfies the hypergroup property (4) with respect to $x_0 = 0$ (and N replaced by $N + 1$) if and only if for all $n \in [\![0, N-1]\!]$, the matrix

$$(L - \lambda_{n,0})(L - \lambda_{n,1}) \cdots (L - \lambda_{n,n-1})$$

has non-negative entries.

□

At the end of this section, we will explain how Proposition 8 enables to construct a relatively efficient algorithm to check the hypergroup property. First we prove Proposition 8 through a sequence of intermediate results.

We begin by showing that it is always possible to solve the commutation equation $K_{\mu_0} P = P K_{\mu_0}$ explicitly in terms of P and μ_0. It will remain to see if the obtained solution is non-negative, but the condition $K_{\mu_0} \mathbb{1} = \mathbb{1}$ will be automatically satisfied. Indeed, for any matrix K commuting with P, we have

$$K\mathbb{1} \;=\; KP\mathbb{1} \;=\; PK\mathbb{1}$$

so that $K\mathbb{1}$ is an eigenfunction associated to the eigenvalue 1 and thus must be constant by irreducibility of P. The first component of $K_{\mu_0}\mathbb{1}$ is equal to $\delta_0 K_{\mu_0} \mathbb{1} = \mu_0(\mathbb{1}) = 1$, so that we get $K_{\mu_0}\mathbb{1} = \mathbb{1}$.

Define

$$\forall\, n, m \in [\![0, N]\!], \qquad a(n, m) \;:=\; P^n(0, m)$$

(note that $a(n, n) > 0$ for $n \in [\![0, N]\!]$) and for $n \in [\![0, N]\!]$, the polynomial $R_n(X)$ given by

$$\begin{aligned} R_n(X) \;:=\; & \frac{1}{a(n,n)} X^n - \frac{1}{a(n,n)} \sum_{n_1 \in [\![0,n-1]\!]} a(n, n_1) \frac{1}{a(n_1, n_1)} X^{n_1} \\ & + \frac{1}{a(n,n)} \sum_{n_1 \in [\![0,n-1]\!]} \sum_{n_2 \in [\![0,n_1-1]\!]} a(n, n_1) \frac{1}{a(n_1, n_1)} a(n_1, n_2) \frac{1}{a(n_2, n_2)} X^{n_2} + \cdots \\ & + (-1)^n \frac{1}{a(n,n)} \sum_{n_1 \in [\![0,n-1]\!]} \sum_{n_2 \in [\![0,n_1-1]\!]} \cdots \sum_{n_n \in [\![0,n_{n-1}-1]\!]} a(n, n_1) \frac{1}{a(n_1, n_1)} a(n_1, n_2) \frac{1}{a(n_2, n_2)} \\ & \qquad \cdots a(n_{n-1}, n_n) \frac{1}{a(n_n, n_n)} X^{n_n} \end{aligned}$$

This polynomial has degree n and the last sum over n_n is empty except if $n_{n-1} = 1$, because for any $l \in [\![0, n]\!]$, $n_l \leqslant n - l$. The interest of R_n comes from

Lemma 10 *For any probability distribution μ_0, there exists a unique matrix K_{μ_0} commuting with P and whose first line coincides with μ_0. It is given by*

$$\forall\ n \in [\![0, N]\!], \qquad K_{\mu_0}(n, \cdot) \;=\; \mu_0 R_n(P)(\cdot) \tag{11}$$

Proof

We begin by showing that a solution K satisfying the two requirements of this proposition is necessarily given by the above formula. To simplify the notations, we consider the case where $\mu_0 = \delta_{x_1}$, with $x_1 \in [\![0, N]\!]$ given. Fix some $n \in [\![1, N]\!]$. From the commutation relation, we get that $P^n K = K P^n$. The first line of this matrix identity reads

$$\sum_{m \in [\![0,n]\!]} P^n(0, m) K(m, \cdot) \;=\; P^n(x_1, \cdot)$$

because $P^n(0, m) = 0$ for $m \in [\![n + 1, N]\!]$. It follows that

$$K(n, \cdot) \;=\; \frac{1}{a(n, n)} \left(P^n(x_1, \cdot) - \sum_{m \in [\![0,n-1]\!]} a(n, m) K(m, \cdot) \right)$$

which provides an iteration formula for the computations of $K(n, \cdot)$, starting from $K(0, \cdot) = \delta_{x_1}$. It leads without difficulty to the announced expression, $K(n, \cdot) = \delta_{x_1} R_n(P)(\cdot)$. These arguments extend to the situation of a general probability measure μ_0. Conversely, the matrix defined by (11) satisfies on one hand, $K_{\mu_0}(0, \cdot) = \mu_0(\cdot)$ and on the other hand, for all $n \in [\![0, N]\!]$,

$$K_{\mu_0}(n, \cdot) \;=\; \frac{1}{a(n, n)} \left(\mu_0 P^n(\cdot) - \sum_{m \in [\![0,n-1]\!]} a(n, m) K_{\mu_0}(m, \cdot) \right)$$

namely

$$\sum_{m \in [\![0,n]\!]} P^n(0, m) K_{\mu_0}(m, \cdot) \;=\; \mu_0 P^n(\cdot)$$

or equivalently, we have the equality of the first line of $P^n K_{\mu_0}$ and $K_{\mu_0} P^n$:

$$\delta_0 P^n K_{\mu_0} \;=\; \delta_0 K_{\mu_0} P^n$$

Since this is true for all $n \in [\![0, N]\!]$, we deduce that

$$\forall\ n \in [\![1, N]\!], \qquad \delta_0 P^{n-1} P K_{\mu_0} \;=\; \delta_0 P^{n-1} K_{\mu_0} P$$

Note that the support of the measure $\delta_0 P^{n-1}$ is exactly $[\![0, n-1]\!]$, thus by iteration, it follows that all the lines of $P K_{\mu_0}$ coincide with the corresponding ones of $K_{\mu_0} P$, i.e. K_{μ_0} commutes with P.

■

It is natural to wonder if, for given $n \in [\![0, N]\!]$, the polynomial R_n is uniquely by the property (11). Indeed, assume that $\widetilde{R}_n$ is another polynomial of degree n satisfying the same equation. Since it must be true for all probability measure μ_0, we get that $R_n(P) = \widetilde{R}_n(P)$. Thus a priori, R_n is only determined up to an additional term belonging to the ideal generated by the unital minimal polynomial Q associated to the matrix P. Since P is an irreducible birth and death transition kernel, it is diagonalizable and all its eigenvalues are different.This implies that Q is of degree N. Thus if $n \in [\![0, N - 1]\!]$, R_n is uniquely determined, due to the fact that its degree is n. But this argument doesn't seem to work for $n = N$. There is a more convenient way to see that R_n is uniquely determined for all $n \in [\![0, N]\!]$, even under an apparently weaker requirement, as we are to see.

Note that if $\mu_0 = \delta_0$, the identity matrix I is a trivial solution to the problem corresponding to K_0. By the uniqueness statement of Lemma 10, we conclude that $K_0 = I$. In the wave equation interpretation, K_0 corresponds to a wave initially localized at 0 and which travels at speed 1 to the right, until it reaches N at time N. The polynomial R_n is in fact characterized by (11) with $\mu_0 = \delta_0$:

Lemma 11 *For all $n \in [\![0, N]\!]$, there is a unique polynomial R_n of degree n such that*

$$\delta_n(\cdot) \quad = \quad \delta_0 R_n(P)(\cdot) \tag{12}$$

Proof

Let us fix $n \in [\![0, N]\!]$ and write

$$R_n(X) \quad = \quad \sum_{p \in [\![0,n]\!]} r_p X^p$$

Since the support of the probability measure $\delta_0 P^p$ is $[\![0, p]\!]$, for all $p \in [\![0, n]\!]$, we deduce from (12) applied at n that $1 = r_n P^n(0, n)$, namely $r_n = 1/a(n, n)$. Next applying (12) at $n - 1$, we deduce that $0 = r_{n-1} P^{n-1}(0, n-1) + r_n P^n(0, n-1)$, i.e. $r_{n-1} = -a(n, n-1)/(a(n, n)a(n-1, n-1))$. It appears that we can deduce iteratively the values of r_{n-2}, r_{n-3}, ..., r_0.

■

The previous result enables an interesting interpretation of R_n, for fixed $n \in [\![0, N]\!]$, from which Proposition 8 follows at once. Consider the matrix $\widetilde{P}_n$, indexed by $[\![0, n]\!] \times [\![0, n]\!]$ and given by

$$\forall\, k, l \in [\![0, n]\!], \qquad \widetilde{P}_n(k, l) \quad := \quad \begin{cases} P(k, l) & , \text{if } k \in [\![0, n-1]\!] \\ \delta_n(l) & , \text{if } k = n \end{cases}$$

It is a Markov transition matrix absorbed at n. Its eigenvalue are $\theta_{n,n} := 1$ and the eigenvalues $-1 < \theta_{n,0} < \theta_{n,1} < \cdots < \theta_{n,n-1} < 1$ introduced before Proposition 8 and corresponding to eigenvectors vanishing at n.

Lemma 12 *For $n \in [\![0, N]\!]$ fixed as above, we have*

$$R_n(X) \quad = \quad \frac{1}{a(n, n)}(X - \theta_{n,0})(X - \theta_{n,1}) \cdots (X - \theta_{n,n-1})$$

Proof

Since P is a birth and death transition matrix, we have

$$\forall\, x \in [\![0, n]\!], \qquad \delta_0 R_n(P)(x) \quad = \quad \delta_0 R_n(\widetilde{P}_n)(x)$$

thus reinterpreting (12) on $[\![0, n]\!]$, we get

$$\delta_n(\cdot) \quad = \quad \delta_0 R_n(\widetilde{P}_n)(\cdot) \tag{13}$$

The same arguments as in the proof of Lemma 11 show that this equation determine the polynomial $R_n(X)$, in particular the coefficient of X^n is $1/a(n, n)$.

Consider the polynomial

$$Q(X) \quad := \quad (X - \theta_{n,0})(X - \theta_{n,1}) \cdots (X - \theta_{n,n-1})$$

Hamilton-Cayley theorem says that $Q(\widetilde{P}_n)(\widetilde{P}_n - I) = 0$ and in particular $\delta_0 Q(\widetilde{P}_n)(\widetilde{P}_n - I) = 0$, which means that $\delta_0 Q(\widetilde{P}_n)$ is an invariant measure for $\widetilde{P}_n$. Since the invariant measures of $\widetilde{P}_n$ are proportional to δ_n, we deduce that there exists a constant $c_n \in \mathbb{R}$ such that

$$c_n \delta_n(\cdot) \quad = \quad \delta_0 Q(\widetilde{P}_n)(\cdot)$$

Applying this inequality at n, we get that $c_n = P^n(0,n) = a(n,n)$ and the announced result is a consequence of the uniqueness statement of Lemma 11.

■

Proposition 8 suggests the following algorithm to check for the hypergroup property at 0 of a finite birth and death Markov kernel P: for all $n \in [\![0, N-1]\!]$, one computes the eigenvalues $\theta_{n,0}, \theta_{n,1}, \ldots, \theta_{n,n-1}$ and checks the non-negativity of its entries $(P-\theta_{n,0})(P-\theta_{n,1})\cdots(P-\theta_{n,n-1})$. From a theoretical point of view, it may seem simpler to compute the normalized eigenvectors $(\varphi_n)_{n\in[\![0,N]\!]}$ and to check directly the hypergroup property as it stated in (4) (with $x_0 = 0$ and the appropriate change of indices of the eigenvectors). But in practice it is more delicate to compute eigenvectors than eigenvalues and in the numerical experiments we made (using Scilab), first just to check the Markov kernels M_p for $p \in [1/2, 1)$ and M_0 (defined in Example 7) satisfy the hypergroup property, the algorithm based on Proposition 8 is more stable.

Thus we rather used the latter to proceed to the numerical experiments described below (the codes are available on request).

Let us recall the Achour-Trimèche's theorem [1]. Consider the differential operator $L = \partial^2 - U'\partial$ on $[0,1]$ with Neumann boundary conditions, where $U : [0,1] \to \mathbb{R}$ is a smooth convex potential, which is assumed to be either non-increasing or symmetric with respect to the point 1/2. Then L satisfies the hypergroup property with respect to 0 (the finite sum in (4) has to be naturally extended into a denumerable sum, see for instance Bakry and Huet [2]).

We would like to find an extension of this result to its discrete analogous setting of finite birth and death processes. Seeing $\partial^2 - U'\partial$ as a Metropolis modification of ∂^2 with respect to the probability measure admitting a density proportional to $\exp(-U)$ (for this point of view, cf. e.g. [6]), a first guess is as follows. Let U be a convex and non-increasing function on $[0, N]$ and consider the probability measure μ defined as

$$\forall\, x \in [\![0, N]\!], \qquad \mu(x) \quad := \quad \frac{1}{Z}\exp(-U(x))\,\pi_0(x) \tag{14}$$

where Z is the normalizing constant and π_0 is the invariant probability of M_0, namely

$$\forall\, x \in [\![0, N]\!], \qquad \pi_0(x) \quad = \quad \begin{cases} 1/N & , \text{ if } x \in [\![1, N-1]\!] \\ 1/(2N) & , \text{ if } x \in \{0, N\} \end{cases}$$

The usual choice for a Markov kernel admitting μ as reversible measure is the Metropolis perturbation of M_0 (initiated in Metropolis et al. [9]) given here by

$$\begin{aligned} \forall\, x \neq y \in [\![0, N]\!], \qquad P(x,y) \quad &:= \quad M_0(x,y)\exp(-(U(y) - U(x))_+) \\ &= \quad \begin{cases} M_0(x,y) & , \text{ if } x < y \\ M_0(x,y)\exp(-(U(y) - U(x))) & , \text{ if } x > y \end{cases} \end{aligned}$$

Numerical experiments based on Proposition 8 show that the conjecture that P satisfy the hypergroup property with respect to 0 is wrong. We checked this assertion by taking $N = 10$ and by sampling the convex function U according to the following procedure: let $(V(x))_{x\in[\![0,N-1]\!]}$ be independent exponential random variables of parameter $1/N$, we take

$$\forall\, x \in [\![1, N]\!], \qquad U(x-1) - U(x) \quad := \quad \sum_{n\in[\![0,N-x]\!]} V(n)$$

Finally, we replaced in the above considerations the exploration kernel M_0 by M_p, for $p \in [1/2, 1)$. This should reinforce the log-concavity of the probability measure μ defined as in (14), where π_0 is replaced by the invariant probability measure π_p of the Markov kernel M_p. Nevertheless the conjecture still seems to be wrong (but less so when p becomes closer to 1).

Of course these experiments suggest that the right notion of log-concavity of a measure (or rather of a Markov kernel) has yet to be found in the discrete setting.

Aknowledgments:
The hypergroup property, the Carlen, Geronimo and Loss [3] method and the result of Achour and Trimèche [1] were explained to me by Dominique Bakry. He also transmitted to me the question of Persi Diaconis of knowing whether Eagleson's example of the biased Ehrenfest model could be recovered by the approach of Carlen, Geronimo and Loss [3]. Despite our conversations, he categorically refused to co-sign this paper. Maybe to leave me wish him here a happy 60th birthday ;-). I'm also grateful to Sylvain Ervedoza for helping with some preliminary numerical simulations and to the ANR STAB (Stabilité du comportement asymptotique d'EDP, de processus stochastiques et de leurs discrétisations) for its support.

References

[1] Abdennebi Achour and Khalifa Trimeche. Opérateurs de translation généralisée associés à un opérateur différentiel singulier sur un intervalle borné. *C. R. Acad. Sci. Paris Sér. A-B*, 288(7):A399–A402, 1979.

[2] D. Bakry and N. Huet. The hypergroup property and representation of Markov kernels. In *Séminaire de probabilités XLI*, volume 1934 of *Lecture Notes in Math.*, pages 295–347. Springer, Berlin, 2008.

[3] Eric A. Carlen, Jeffrey S. Geronimo, and Michael Loss. On the Markov sequence problem for Jacobi polynomials. *Adv. Math.*, 226(4):3426–3466, 2011.

[4] Persi Diaconis and Robert Griffiths. Exchangeable pairs of Bernoulli random variables, Krawtchouck polynomials, and Ehrenfest urns. *Aust. N. Z. J. Stat.*, 54(1):81–101, 2012.

[5] Persi Diaconis and Robert Griffiths. An introduction to multivariate Krawtchouk polynomials and their applications. *J. Statist. Plann. Inference*, 154:39–53, 2014.

[6] Persi Diaconis and Laurent Miclo. On characterizations of Metropolis type algorithms in continuous time. *ALEA Lat. Am. J. Probab. Math. Stat.*, 6:199–238, 2009.

[7] E. B. Dynkin. *Markov processes. Vols. I, II*, volume 122 of *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.* Academic Press Inc., Publishers, New York, 1965.

[8] G. K. Eagleson. A characterization theorem for positive definite sequences on the Krawtchouk polynomials. *Austral. J. Statist.*, 11:29–38, 1969.

[9] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, 21(6):1087–1092, June 1953.

miclo@math.univ-toulouse.fr
Institut de Mathématiques de Toulouse
Université Paul Sabatier
118, route de Narbonne
31062 Toulouse Cedex 9, France