On times to quasi-stationarity for birth and death processes

Persi Diaconis and Laurent Miclo

Department of Mathematics and Statistics
Stanford University, USA
and CNRS, Université de Nice-Sophia Antipolis, France

Laboratoire d'Analyse, Topologie, Probabilités, U.M.R. 6632
Université de Provence and C.N.R.S.
France

Abstract

The purpose of this paper is to present a probabilistic proof of the well-known result stating that the time needed by a continuous-time finite birth and death process for going from the left end to the right end of its state space is a sum of independent exponential variables whose parameters are the negatives of the eigenvalues of the underlying generator when the right end is treated as an absorbing state. The exponential variables appear as fastest strong quasi-stationary times for successive dual processes associated to the original absorbed process. As an aftermath, we get an interesting probabilistic representation of the time marginal laws of the process in terms of "local equilibria".

Keywords: birth and death processes, absorption times, sums of independent exponential variables, Dirichlet eigenvalues, fastest strong stationary times, strong dual processes, strong quasi-stationary times, local equilibria.

MSC2000: first: 60J27, secondary: 60J80, 60J35, 37A30, 15A18.

1 Introduction

The goal of this paper is to give a probabilistic derivation of the law of the time needed by an irreducible continuous time birth and death process on a finite path to go from the left end to the right end. It is known that this distribution is that of a sum of independent exponential variables of parameters the negatives of the eigenvalues of the underlying generator with a Dirichlet condition imposed at the right end. While this simple statement seems of probabilistic nature, its proof (until now) is indirect, via Laplace transforms. The main drawback of the latter method is that it prevents any probabilistic interpretation for the Dirichlet eigenvalues (except for the first one, which corresponds to the asymptotic rate to attain the right end), which is really the motivation for the following study.

More precisely, on the state space $[\![0,N]\!]$, with $N\in\mathbb{N}^*$, consider a birth and death process $X\coloneqq(X_t)_{t\geq0}$ starting from 0 and absorbed at N (this assumption is not restrictive, since we will not be concerned by what happens after the process has reached this point). The simplest way to specify its evolution is through a generator L acting on $\mathcal{D}$, the space of functions defined on $[\![0,N]\!]$ and vanishing at N. Thus

$$\forall\, f\in\mathcal{D},\,\forall\, x\in[\![0,N-1]\!],\qquad L[f](x)\;=\;b_x(f(x+1)-f(x))+d_x(f(x-1)-f(x))$$

where $(b_x)_{0\leq x<N}$ and $(d_x)_{0\leq x<N}$ are respectively the birth and death rates. Necessarily $d_0=0$ and we assume that all the other rates are positive (by a continuity argument, next result can afterward be extended to the case of positive birth rates and nonnegative death rates). To see L as an operator on $\mathcal{D}$, we take by convention $L[f](N)=0$, which constitutes the "Dirichlet condition" in this discrete setting. It is then well-known that $-L$ is diagonalizable with positive, distinct eigenvalues $0<\lambda_1<\lambda_2<\cdots<\lambda_N$.
Coming back to the process X, we are interested in its absorption time

$$\tau\;\coloneqq\;\inf\{t\geq0\,:\,X_t=N\}$$

The next point is the start of our study (see for instance the book of Keilson [15], we give the problem's history at the end of the introduction).

Proposition 1 *Assume that X starts from 0, i.e. $X_0=0$. Then the law of τ coincides with that of $T_1+\cdots+T_N$, where the T_i, for $1\leq i\leq N$, are independent and respectively distributed as exponential laws with parameters λ_i (i.e. with mean $1/\lambda_i$).*

The corresponding result for discrete time (where exponential laws are replaced by geometric laws for birth and death chains which are monotone and whose associated eigenvalues belong to $[0,1]$) is used to build and interpret various stopping times for some irreducible birth and death chains on $[\![0,N]\!]$ in Diaconis and Fill [7] and as the basic tool for proving a conjecture of Peres on the cut-off phenomena in Diaconis and Saloff-Coste [8].
As announced, our purpose is to give a probabilistic proof of the identity in law contained in Proposition 1, which can serve as a probabilistic interpretation of the Dirichlet eigenvalues λ_i, for $1\leq i\leq N$.

To proceed, let L' be the generator on $[\![0,N]\!]$ whose birth rates are given by

$$(b'_x)_{0\leq x<N}\;\coloneqq\;(\lambda_{N-x})_{0\leq x<N}$$

and whose death rates all vanish. Of course, if τ' is the absorption time at N for a corresponding Markov process $(X'_t)_{t\geq0}$ starting from 0, then the law of τ' is equal to that of $T_1+\cdots+T_N$ as in Proposition 1.
Most of our efforts will consist in constructing a coupling $(X'_t,X_t)_{t\geq0}$ of $(X'_t)_{t\geq0}$ and $(X_t)_{t\geq0}$ such that $(X'_t)_{t\geq0}$ and $(X_t)_{t\geq0}$ are intertwined in the following sense: there exists a Markovian kernel Λ

from $[\![0, N]\!]$ to $[\![0, N]\!]$, satisfying $\Lambda(x, [\![0, x]\!]) = 1$ for all $0 \leq x \leq N$ (namely it is lower triangular as a matrix) such that for any $t \geq 0$, we have a.s.,

$$\mathcal{L}(X_t | X'_s,\, 0 \leq s \leq t) \;=\; \Lambda(X'_t, \cdot) \tag{1}$$

where the l.h.s. stands for the conditional law of X_t knowing the σ-field generated by $(X'_s)_{0 \leq s \leq t}$. Traditional notations for kernels will be used, for instance for any $x \in [\![0, N]\!]$ and any function f on $[\![0, N]\!]$, $\Lambda(x, f) \coloneqq \sum_{y \in [\![0,N]\!]} \Lambda(x, y) f(y)$.par Let us check rapidly that the above property implies that the absorption time τ' of $(X'_t)_{t \geq 0}$ at N is a.s. equal to the absorption time τ of $(X_t)_{t \geq 0}$ at N. The equality in law $\tau = \tau'$ then follows, so Proposition 1 is proved.
Consider for any function f defined on $[\![0, N]\!]$,

$$\mathbb{E}[f(X_t)] \;=\; \mathbb{E}[\Lambda(X'_t, f)]$$

so letting t go to infinity, we get that $f(N) = \Lambda(N, f)$ (because both X'_t and X_t are a.s. convergent to N), which means that $\Lambda(N, \cdot) = \delta_N$. Then using for any $t \geq 0$ the relation

$$\begin{aligned}
\mathbb{P}[\tau' \leq t,\, \tau \leq t] &= \mathbb{E}[\mathbb{1}_{\tau' \leq t} \delta_N(X_t)] \\
&= \mathbb{E}[\mathbb{1}_{\tau' \leq t} \mathbb{E}[\delta_N(X_t) | X'_s,\, 0 \leq s \leq t]] \\
&= \mathbb{E}[\mathbb{1}_{\tau' \leq t} \Lambda(X'_t, N)] \\
&= \mathbb{E}[\mathbb{1}_{\tau' \leq t} \Lambda(N, N)] \\
&= \mathbb{P}[\tau' \leq t]
\end{aligned}$$

it must be that $\tau \leq \tau'$ a.s. Next for any $t \geq 0$ we also have

$$\begin{aligned}
\mathbb{P}[\tau' > t,\, \tau \leq t] &= \mathbb{E}[\mathbb{1}_{\tau' > t} \Lambda(X'_t, N)] \\
&= 0
\end{aligned}$$

since by our assumption on Λ, we have $\Lambda(x, N) = 0$ for all $0 \leq x < N$. It follows that $\tau = \tau'$ a.s.

The construction of the coupling is quite involved. We will need $N + 1$ intermediate processes $(X^{(i)}_t)_{t \geq 0}$, for $0 \leq i \leq N$, which are intertwined. More precisely, for any $i \in [\![0, N-1]\!]$, we will construct a lower triangular link $\Lambda^{(i)}$ such that (1) is true with X replaced by $X^{(i)}$, X' by $X^{(i+1)}$ and Λ by $\Lambda^{(i)}$. Then $\Lambda \coloneqq \Lambda^{(N-1)} \Lambda^{(N-2)} \cdots \Lambda^{(0)}$ will provide the link needed to intertwine $X^{(0)} = X$ and $X^{(N)} = X'$ (where equality here is in law). The probabilistic interpretation of Proposition 1 is encapsulated in the construction of these processes and their intertwining links. In particular the exponential times will appear as exit times under some "initial" quasi-stationary distributions (and the eigenvalues λ_i, for $1 \leq i \leq N$, will be interpreted as first Dirichlet eigenvalues, or exit rates, of the corresponding subdomains). Heuristically the picture is the following: starting from 0, X spends a time T_N (distributed as an exponential of parameter λ_N, as above) before reaching a "local equilibrium". Next it needs a time T_{N-1} to go to another local equilibrium etc. Finally it takes a time T_1 to escape from the $(N-1)^{\text{th}}$ local equilibrium to be absorbed in N.

In two recent papers [10, 11], Fill gives another proof of Proposition 1, but it relies on a nontrivial linear algebra result of Micchelli and Willoughby [20]. Indeed our arguments can be seen as a probabilistic approach to this result in the particular case we consider here.

This point of view can be extended to times to stationarity. Indeed, let Y be an irreducible birth and death process on $[\![0, N]\!]$ and denote by π its reversible probability. A strong stationary time S is a randomized stopping time such the law of Y_S is π and such that S and Y_S are independent. Among all such times, some are stochastically smaller than or equal to all the other ones and they are called *fastest times to stationarity*. Their distribution is directly related to the separation distance between the time-marginal laws and the equilibrium law. For more information on this subject, we refer to the articles of Diaconis and Fill [7] and Fill [9] and to the bibliography contained therein. To describe these times in the above continuous time setting when Y starts from 0, Fill [9] introduced a dual birth and death process Y^* also starting from 0, whose absorption time in N is a time to stationarity for Y (once these processes have been coupled through an appropriate intertwining relation). This procedure followed similar ideas developed in Diaconis and Fill [7] for discrete time chains. Since the eigenvalues of the generator of Y (except for the trivial eigenvalue 0) are the same as the eigenvalues of the generator of Y^* with a Dirichlet condition at N, Fill [9] deduced via Proposition 1 that the law of S is the convolution of exponential distributions of parameters the negatives of the non-zero eigenvalues of the generator of Y. We will revisit this result, which enters in the previous heuristic picture, except that in the last step, Y goes from the $(N-1)^{\text{th}}$ local equilibrium to the global equilibrium π within time T_1.

The case of birth and death processes starting from 0 is quite restrictive, but we see it as a step in the direction of a better probabilistic understanding of the relation between the eigenvalues of the generator and the convergence to equilibrium. Matthews [19] also provides a contribution in this direction. Indeed, note that even the analogous situation in discrete time remains puzzling when the transition matrix admits negative eigenvalues (some mysterious Bernoulli distributions appear, see for instance formula (4.23) in Diaconis and Fill [7]). We still hope that a probabilistic explanation can be found.

The notion of intertwining appears in an article of Rogers and Pitman [25]. For other examples of intertwined Markov semi-groups, see the article of Carmona, Petit and Yor [5]. For an analogue of Proposition 1 for one-dimensional diffusions, see Kent [16].

The outline of this paper is as follows: in the next section we will recover the construction of the dual process Y^* of Fill [9] (but whose setting is more general), by adopting a continuous space inspired formalism for birth and death processes. It is particularly well-adapted to deal with one-dimensional diffusions, but we will not develop the corresponding theory here. In section 3 we will extend these considerations from times to stationarity to times to quasi-stationarity, which will enable us to construct a first dual process $X^{(1)}$ of $X^{(0)} = X$. The main result will be that for birth and death processes starting at 0 and first absorbed at N, we can stop the process with a strong time in such a way that the distribution of the position at this stopping time is the quasi-stationary distribution. In section 4, the iteration of this procedure will lead to the whole familly $X^{(i)}$, for $1 \leq i \leq N$, and especially to $X' = X^{(N)}$. We will discuss the notion of local equilibrium, which is the key to our proof of Proposition 1. It also leads to a probabilistic representation of the time marginal laws of X. The last section will deal with two illustrative examples.

We have indicated by an empty box $\square$ the end of the remarks, which should all be skipped at a first reading.

Historical Note:

The earliest appearance of Proposition 1 that we know is in Karlin and McGregor [13], Equation 45 (thanks to Laurent Saloff-Coste for this reference). Their proof is via the orthogonal polynomials associated to the birth and death process. They give an expression for the Laplace transform of the first absorption time which is equivalent to the probabilistic formulation of Proposition 1. The result was used by Keilson [14] who gives an independent proof using complex variables. The topic is developed further in chapter 5 of Keilson [15]. In all these proofs, the exponential variables appear through analysis, without probabilistic motivation.

A different proof of Proposition 1 follows from Kent [17]. Briefly, Kent considers the first hitting time of N for an irreducible birth and death process started at 0. Let S_i, for $i \in [\![0, N-1]\!]$, be the time spent in state i before N is reached. Kent shows that the vector $S = (S_0, ..., S_{N-1})$ has the law of $Y + Z$, with Y and Z independent vectors, distributed as coordinate-wise squares of independent Gaussian vectors V and W, each having mean 0 and covariance matrix Σ, with $\Sigma^{-1} := 2Q$. Here Q is the upper $N \times N$ block of the matrix associated to $-L$ symmetrized by the stationary distribution. In particular, for nonnegative s_i, $i \in [\![0, N-1]\!]$,

$$\mathbb{E}\left[\exp\left(-\sum_{i\in[\![0,N-1]\!]} s_i S_i\right)\right] = \frac{\det(Q)}{\det(Q+\widetilde{S})}$$

with $\widetilde{S}$ a diagonal matrix with s_i in position (i,i). Notice that if all the s_i are equal, the right side is invariant under the conjugacy mapping $Q \mapsto A^{-1}QA$, for any invertible matrix A, so this right side can be taken with Q replaced by a diagonal matrix containing the eigenvalues of the original Q. Since $S_0 + \cdots + S_{N-1}$ is the time until N is reached, we have another proof of Proposition 1. Kent [17] passes to the limit and uses the result to give a proof of the Ray-Knight theorem expressing the local time of Brownian motion as the sum of two independent square Bessel processes.

2 Times to stationarity for birth and death processes

We will revisit here the reduction to absorption times of times to stationarity for birth and death processes starting from 0. The main point is to introduce a differential formalism which is different from the approach used by Fill [9] to construct dual processes: he rather built on the elementarily probabilistic treatment of Diaconis and Fill [7] for discrete-time chains to get corresponding results for continuous-time chains by simple passages to limits.

We still consider $V := [\![0, N]\!]$ as state space, but it is also convenient to introduce $V^- := [\![-1, N]\!]$, $V^+ := [\![0, N+1]\!]$ and $\bar{V} := [\![-1, N+1]\!]$. The spaces $\mathcal{F}$, $\mathcal{F}^-$, $\mathcal{F}^+$ and $\bar{\mathcal{F}}$ respectively stand for the collections of real valued functions defined on the previous sets. We denote by ∂^+ the operator from $\bar{\mathcal{F}}$ to $\mathcal{F}^-$ given by

$$\forall\, f \in \bar{\mathcal{F}}, \forall\, x \in V^-, \qquad \partial^+ f(x) \;:=\; f(x+1) - f(x)$$

By restriction of $\partial^+ f$ to V, this operator can also be seen to go from $\mathcal{F}^+$ to $\mathcal{F}$. In a symmetrical way, we consider $\partial^- : \bar{\mathcal{F}} \to \mathcal{F}^+$ (or from $\mathcal{F}^-$ to $\mathcal{F}$),

$$\forall\, f \in \bar{\mathcal{F}}, \forall\, x \in V^+, \qquad \partial^- f(x) \;:=\; f(x-1) - f(x)$$

Next let L be an irreducible birth and death generator on V. We denote by $(b_x)_{0\le x\le N}$ and $(d_x)_{0\le x\le N}$ respectively its birth and death rates, which are positive, except for $d_0 = b_N = 0$. Throughout, let

$$\forall\, x \in V, \qquad \pi(x) \;:=\; Z^{-1} \prod_{1\le y\le x} \frac{b_{y-1}}{d_y}$$

(where Z is the normalizing constant) be the stationary distribution for L. Let $u \in \mathcal{F}$ and $v \in \mathcal{F}^-$ be the functions defined by

$$\begin{aligned}
\forall\, x \in V, \qquad u(x) \;&:=\; \frac{1}{\pi(x)} \\
\forall\, x \in V^-, \qquad v(x) \;&:=\; \pi(x)L(x, x+1)
\end{aligned}$$

(in particular $v(-1) = v(N) = 0$). Then the generator L can be rewritten in the form

$$L \;=\; -u\partial^- v \partial^+ \tag{2}$$

In formulas such as (2), the functions u, v act by multiplication, so $uf(x) = u(x)f(x)$ for any x in the underlying set. More rigorously, the operator on the right side of (2) goes from $\bar{\mathcal{F}}$ to $\mathcal{F}$, but it happens that for $f \in \bar{\mathcal{F}}$, $-u\partial^- v\partial^+ f$ does not depend on the values $f(-1)$ and $f(N+1)$, so there is no ambiguity in interpreting $-u\partial^- v\partial^+$ as an operator from $\mathcal{F}$ to itself.

One feature of the formulation (2) is that it makes it easy to find a "first order" difference operator D with a "0-order" term (namely an operator of the form $f \mapsto a\partial^+ f + bf$ or of the form $f \mapsto a\partial^- f + bf$, where a, b are functions and where b is seen as a "0-order" operator acting by multiplication) and a birth and death generator L^* absorbed in N such that

$$LD = DL^* \tag{3}$$

which is our next goal:

Lemma 2 *Define $D := -\frac{1}{\pi}\partial^- H$ and $L^* := -\frac{v}{H}\partial^+ \frac{1}{\pi}\partial^- H$ (here and in the whole paper, the traditional convention $0 \cdot \infty$ is assumed to hold), where the probability π has been extended to $\bar{V}$ by $\pi(-1) = 0 = \pi(N+1)$ and where H is the cumulative function of π:*

$$\forall\, x \in \bar{V}, \qquad H(x) := \sum_{0 \le y \le x} \pi(y)$$

A priori, $D : \mathcal{F}^- \to \mathcal{F}$ and $L^* : \bar{\mathcal{F}} \to \mathcal{F}$, but as before these operators can be naturally interpreted as going from $\mathcal{F}$ to itself and the algebraic duality relation (3) is satisfied.

Proof

For $f \in \mathcal{F}^-$, we compute that

$$\begin{aligned} Df(0) &= -\frac{1}{\pi(0)}\partial^- Hf(0) \\ &= -\frac{1}{\pi(0)}(H(-1)f(-1) - H(0)f(0)) \\ &= f(0) \end{aligned}$$

doesn't depend on $f(-1)$, nor does any other value of Df, so Df depends only on the restriction of f to V. Similarly, for $f \in \bar{\mathcal{F}}$, we have

$$\begin{aligned} L^* f(0) &= -\frac{v(0)}{H(0)}\left(\frac{\partial^- Hf(1)}{\pi(1)} - \frac{\partial^- Hf(0)}{\pi(0)}\right) \\ &= -L(0,1)\left(\frac{H(0)f(0) - H(1)f(1)}{\pi(1)} - \frac{H(-1)f(-1) - H(0)f(0)}{\pi(0)}\right) \\ &= L(0,1)\left(\frac{\pi(0)}{\pi(1)} + 1\right)(f(1) - f(0)) \end{aligned}$$

doesn't depend on $f(-1)$, nor does any other value of $L^* f$, and $L^* f(N) = 0$, because $v(N) = 0$, and so we find that $L^* f$ doesn't depend on $f(-1)$ or $f(N)$. Thus L^* can equally be seen as an operator from $\mathcal{F}$ to itself. Furthermore, since $L^* f(N)$ always vanishes, any Markov process generated by L^* (for instance in the sense of the corresponding martingale problem) is absorbed at N. Indeed, we check that L^* is a birth and death generator on V with rates given for any $x \in V$ by

$$\begin{aligned} b^*(x) &= \frac{v(x)H(x+1)}{H(x)\pi(x+1)} \\ &= d(x+1)\frac{H(x+1)}{H(x)} \\ d^*(x) &= \frac{v(x)H(x-1)}{H(x)\pi(x)} \\ &= b(x)\frac{H(x-1)}{H(x)} \end{aligned}$$

Next the formal verification of (3) is immediate:

$$\begin{aligned} LD &= \frac{1}{\pi}\partial^- v\partial^+ \frac{1}{\pi}\partial^- H \\ DL^* &= \frac{1}{\pi}\partial^- H\frac{v}{H}\partial^+ \frac{1}{\pi}\partial^- H \\ &= \frac{1}{\pi}\partial^- v\partial^+ \frac{1}{\pi}\partial^- H \end{aligned}$$

■

The operator $D : \mathcal{F} \to \mathcal{F}$ is in fact one-to-one. To see this, let us compute its inverse Λ. Let $g \in \mathcal{F}$ be given, we want to find $f \in \mathcal{F}$ such that $Df = g$, namely

$$\begin{aligned} \forall\, x \in V, \quad Df(x) \;=\; g(x) &\iff \forall\, x \in V, \quad \partial^- Hf(x) \;=\; -\pi(x)g(x) \\ &\iff f(x) \;=\; \frac{1}{H(x)}\sum_{0\leq y\leq x}\pi(y)g(y) \end{aligned}$$

We recover the link Λ considered by Fill [9] in the above case of a birth and death process starting from 0 (or starting from a distribution m_0 such that m_0/π is nonincreasing, see below):

$$\forall\, x \in V, \qquad \Lambda[g](x) \;:=\; \frac{1}{H(x)}\sum_{0\leq y\leq x}\pi(y)g(y)$$

It is clear from this expression that Λ can be interpreted as a Markov kernel going from V to V and satisfying the lower triangularity mentioned in the introduction. Furthermore, we deduce from Lemma 2 that

$$\Lambda L \;=\; L^*\Lambda \tag{4}$$

The same relation was deduced by Fill [9], using the approach of Diaconis and Fill [7].
Let us denote by $(P_t)_{t\geq 0}$ and $(P_t^*)_{t\geq 0}$ the semigroups associated to L and L^*, i.e.

$$\forall\, t \geq 0, \qquad \begin{cases} P_t &:= \exp(tL) \\ P_t^* &:= \exp(tL^*) \end{cases}$$

From (4), they also satisfy the intertwining relation

$$\forall\, t \geq 0, \qquad \Lambda P_t \;=\; P_t^*\Lambda \tag{5}$$

Let m_0 (respectively m_0^*) be a probability on V. There exists a Markov process $(X_t)_{t\geq 0}$ (resp. $(X_t^*)_{t\geq 0}$) with cdlg (right continuous with left hand limits) trajectories, whose initial law $\mathcal{L}(X_0)$ is m_0 (resp. $\mathcal{L}(X_0^*)$ is m_0^*) and whose generator is L (resp. L^*). Furthermore uniqueness of these processes holds in law. Assume that

$$m_0 \;=\; m_0^*\Lambda \tag{6}$$

then (5) implies that for any $t \geq 0$, we have $\mathcal{L}(X_t) = \mathcal{L}(X_t^*)\Lambda$. But one can go further, since Fill showed in Theorem 2 of [9] that under the assumptions (4) and (6), there exists a Markovian coupling of $(X_t)_{t\geq 0}$ and $(X_t^*)_{t\geq 0}$, still denoted by $(X_t, X_t^*)_{t\geq 0}$, such that

$$\mathcal{L}(X_t|\mathcal{X}_t^*) \;=\; \Lambda(X_t^*, \cdot) \tag{7}$$

where $\mathcal{X}_t^*$ stands for the σ-field generated by $(X_s^*)_{0\leq s\leq t}$.
Next we define

$$\tau^* \;:=\; \inf\{t \geq 0 \,:\, X_t^* = N\}$$

Under the hypotheses (4) and (6), Fill [9] showed that τ^* is a strong stationary time for X. Let us verify this assertion by using only (7). We begin by extending this relation to any a.s. finite stopping time T^* relative to the filtration $(\mathcal{X}_t^*)_{t\geq 0}$, namely we have

$$\mathcal{L}(X_{T^*}|\mathcal{X}^*_{T^*}) \quad = \quad \Lambda(X^*_{T^*}, \cdot) \tag{8}$$

(recall that $\mathcal{X}^*_{T^*}$ is the σ-field generated by $(X^*_{T^*\wedge t})_{t\geq 0}$). We first consider the usual approximation $T^*_n := \lceil nT^* \rceil / n$ of T^*, for $n \in \mathbb{N} \setminus \{0\}$, where $\lceil t \rceil$ is the smallest integer larger (or equal) than t. Its advantages are, on one hand that it is also a stopping time with respect to $(\mathcal{X}_t^*)_{t\geq 0}$, and on the other hand that it takes only a countable number of values, the values m/n for $m \in \mathbb{N}$. Let $f \in \mathcal{F}$ and G be a bounded mesurable functional on the set of cdlg trajectories from $\mathbb{R}_+$ to V. This path space is endowed with the σ-field generated by the coordinates, coinciding with the Borel σ-field associated to the Skorokhod topology, which is Polish, so we are ensured of the existence of regular conditional probabilities. We compute that

$$\begin{aligned}
\mathbb{E}[f(X_{T^*_n})G((X^*_{T^*_n\wedge t})_{t\geq 0})] &= \sum_{m\in\mathbb{N}} \mathbb{E}[f(X_{T^*_n})G((X^*_{T^*_n\wedge t})_{t\geq 0})\mathbb{1}_{T^*_n=m/n}] \\
&= \sum_{m\in\mathbb{N}} \mathbb{E}[\mathbb{E}[f(X_{m/n})|\mathcal{X}^*_{m/n}]G((X^*_{(m/n)\wedge t})_{t\geq 0})\mathbb{1}_{T^*_n=m/n}] \\
&= \sum_{m\in\mathbb{N}} \mathbb{E}[\Lambda[f](X^*_{m/n})G((X^*_{(m/n)\wedge t})_{t\geq 0})\mathbb{1}_{T^*_n=m/n}] \\
&= \mathbb{E}[\Lambda[f](X^*_{T^*_n})G((X^*_{T^*_n\wedge t})_{t\geq 0})]
\end{aligned}$$

(the third equality comes from (7)). The validity of these relations for any f and G as above is equivalent to (8), where T^* is replaced by T^*_n. But we remark that $X_{T^*_n}$, $X^*_{T^*_n}$ and $(X^*_{T^*_n\wedge t})_{t\geq 0}$ are a.s. convergent to X_{T^*}, $X^*_{T^*}$ and $(X^*_{T^*\wedge t})_{t\geq 0}$, when n goes to infinity, so for a continuous function G, we get

$$\mathbb{E}[f(X_{T^*})G((X^*_{T^*\wedge t})_{t\geq 0})] \quad = \quad \mathbb{E}[\Lambda[f](X^*_{T^*})G((X^*_{T^*\wedge t})_{t\geq 0})]$$

and this is sufficient to be able to conclude (8).
We can now check that τ^* and X_{τ^*} are independent and that X_{τ^*} is distributed as π. Given $f \in \mathcal{F}$ and g a bounded mesurable mapping from $\mathbb{R}_+$ to $\mathbb{R}$, since τ^* is an a.s. finite stopping time which is measurable with respect to $\mathcal{X}^*_{\tau^*}$, we compute that

$$\begin{aligned}
\mathbb{E}[f(X_{\tau^*})g(\tau^*)] &= \mathbb{E}[\mathbb{E}[f(X_{\tau^*})|\mathcal{X}^*_{\tau^*}]g(\tau^*)] \\
&= \mathbb{E}[\Lambda(f)(X^*_{\tau^*})g(\tau^*)] \\
&= \Lambda(f)(N)\mathbb{E}[g(\tau^*)] \\
&= \pi(f)\mathbb{E}[g(\tau^*)]
\end{aligned}$$

(the third equality is an application of (8)), which is the announced result (the identity $\Lambda(N, \cdot) = \pi$ comes from the definition of Λ but as in the introduction, it could be deduced from (7) by letting t going to infinity).
It remains to check that τ^* is a randomized stopping time for $(X_t)_{t\geq 0}$, namely that it is a stopping time with respect to a filtration of the kind $(\sigma(U, X_s \, : \, 0 \leq s \leq t))_{t\geq 0}$, where U is "random noise" independent from $X := (X_t)_{t\geq 0}$. This is equivalent to

$$\mathcal{L}(\tau^*|X^{\tau^*}) \quad = \quad \mathcal{L}(\tau^*|X)$$

(where $X^{\tau^*} := (X_{t\wedge\tau^*})_{t\geq 0}$), since this equality means that the time τ^* depends on the trajectory X only through its positions up to time τ^*. This can be rewritten under the following form: for any bounded mesurable function $g : \mathbb{R}_+ \to \mathbb{R}$ and any bounded mesurable functional G on the set of càdlàg trajectories from $\mathbb{R}_+$ to V, we have

$$\mathbb{E}[\mathbb{E}[g(\tau^*)|X^{\tau^*}]G(X)] = \mathbb{E}[g(\tau^*)G(X)] \quad (9)$$

Via the monotone class theorem, we can restrict to functions G of the form

$$G(X) = G_1(X^{\tau^*})G_2(X^{\tau^*,+}) \quad (10)$$

where $X^{\tau^*,+} := (X_{\tau^*+t})_{t\geq 0}$. Using the strong Markov property of $(X_t, X_t^*)_{t\geq 0}$, we compute that

$$\begin{aligned}
\mathbb{E}[g(\tau^*)G_1(X^{\tau^*})G_2(X^{\tau^*,+})] &= \mathbb{E}[g(\tau^*)G_1(X^{\tau^*})\mathbb{E}[G_2(X^{\tau^*,+})|(X_t, X_t^*)_{0\leq t\leq \tau^*}]] \\
&= \mathbb{E}[g(\tau^*)G_1(X^{\tau^*})\mathbb{E}[G_2(X^{\tau^*,+})|(X_{\tau^*}, X_{\tau^*}^*)]] \\
&= \mathbb{E}[g(\tau^*)G_1(X^{\tau^*})\mathbb{E}[G_2(X^{\tau^*,+})|(X_{\tau^*}, N)]] \\
&= \mathbb{E}[g(\tau^*)G_1(X^{\tau^*})\mathbb{E}[G_2(X^{\tau^*,+})|X_{\tau^*}]]
\end{aligned}$$

We note that the random variable $G_1(X^{\tau^*})\mathbb{E}[G_2(X^{\tau^*,+})|X_{\tau^*}]$ is measurable with respect to X^{τ^*}, so the last expectation can be rewritten as

$$\begin{aligned}
&\mathbb{E}[\mathbb{E}[g(\tau^*)|X^{\tau^*}]G_1(X^{\tau^*})\mathbb{E}[G_2(X^{\tau^*,+})|X_{\tau^*}]] \\
&= \mathbb{E}[\mathbb{E}[g(\tau^*)|X^{\tau^*}]G_1(X^{\tau^*})\mathbb{E}[G_2(X^{\tau^*,+})|(X_t, X_t^*)_{0\leq t\leq \tau^*}]]
\end{aligned}$$

by the same computations as in the previous display. Since $\mathbb{E}[g(\tau^*)|X^{\tau^*}]G_1(X^{\tau^*})$ is measurable with respect to $(X_t, X_t^*)_{0\leq t\leq \tau^*}$, the last quantity is equal to

$$\mathbb{E}[\mathbb{E}[g(\tau^*)|X^{\tau^*}]G_1(X^{\tau^*})G_2(X^{\tau^*,+})]$$

which is just the left hand side of (9) when G is given by (10). It follows a posteriori that in the above expressions we could have replaced $\mathbb{E}[G_2(X^{\tau^*,+})|X_{\tau^*}]$ by $\mathbb{E}[G_2(X^{\tau^*,+})|X^{\tau^*}]$, since the strong Markov property is satisfied by X with respect to randomized stopping times.

Remark 3 Fill [9] (see his equation (2.12)) noticed another property of his coupling, which can be rewritten (again through the monotone class theorem) as

$$\forall\ t \geq 0, \qquad \mathcal{L}((X_{t+s})_{s\geq 0}|X_t, X_t^*) = \mathcal{L}((X_{t+s})_{s\geq 0}|X_t) \quad (11)$$

(note that the l.h.s also coincides with $\mathcal{L}((X_{t+s})_{s\geq 0}|X_u, X_u^* : 0 \leq u \leq t)$). Similarly to what we have done before, this identity can be extended to stopping times with respect to the filtration $(\mathcal{X}_t^*)_{t\geq 0}$. Then the above computation shows that all stopping times with respect to $(\mathcal{X}_t^*)_{t\geq 0}$ are indeed randomized stopping times for X. We did not need (11) to get this property for τ^*, because it has a particular feature: $X_{\tau^*}^*$ is deterministic. We remark that among all stopping times for X^*, τ^* is the smallest one such that X_{τ^*} is distributed as π. Indeed, consider such an a.s. finite stopping time T^*, since we have $\mathcal{L}(X_{T^*}) = \mathcal{L}(X_{T^*}^*)\Lambda$ and that $\Lambda(x, N) = 0$ for $x \in [\![0, N-1]\!]$, it appears that $\mathcal{L}(X_{T^*}) = \pi$ implies $\mathcal{L}(X_{T^*}^*) = \delta_N$.
Of course the fact that τ^* is a fastest time to stationarity is a priori asking for more: namely that τ^* is stochastically smaller than all other strong stationary times. We will not prove this and refer again to Fill [9], because we don't want to enter here into the relationship between strong stationary times and separation distance.

□

3 Times to quasi-stationarity

After having introduced the notion of strong quasi-stationary times, we will extend the dual construction of the previous section to them. To do so, our first step will work in the opposite direction: we shall extract a recurrent Markov process from an absorbing one.

For simplicity, we will be working in the setting of birth and death processes starting from 0 and absorbed at N. From now on, L will be a generator as in the introduction, going from $\mathcal{D}(L) \coloneqq \{f \in \mathcal{F} \,:\, f(N) = 0\}$ into itself. Again $X \coloneqq (X_t)_{t\geq 0}$ designates any Markov process on V admitting L as generator. The distribution of X is determined by its initial law $\mathcal{L}(X_0)$. We are mainly concerned by the case $\mathcal{L}(X_0) = \delta_0$, but there is another interesting initial law, the *quasi-stationary law* ρ. To define it, let $\widehat{L}$ be the adjoint operator of L with respect to ℓ, the counting measure on $[\![0, N-1]\!]$:

$$\forall\, f, g \in \mathcal{D}(L), \qquad \ell(fLg) \;=\; \ell(g\widehat{L}f)$$

The operator $-\widehat{L}$ has the same spectrum $\lambda_1 < \lambda_2 < \cdots < \lambda_N$ as $-L$. So let ψ be an eigenfunction of $-\widehat{L}$ associated to λ_1. A standard application of Perron-Frobenius theorem shows that ψ has a constant (strict) sign on $[\![0, N-1]\!]$, so $\rho \coloneqq \psi/\ell(\psi)$ is a probability which does not vanish on $[\![0, N-1]\!]$.

The next result is classical (see for instance the book [3] of Aldous and Fill), but it will be very important for us, since all of our exponential variables will be created from it, so we include a proof.

Lemma 4 *Assume that $\mathcal{L}(X_0) = \rho$, then*

$$\tau \;\coloneqq\; \inf\{t \geq 0 \,:\, X_t = N\}$$

is distributed as an exponential variable of parameter λ_1.

Proof

Let $(P_t)_{t\geq 0} \coloneqq (\exp(tL))_{t\geq 0}$ (respectively $(\widehat{P}_t)_{t\geq 0} \coloneqq (\exp(t\widehat{L}))_{t\geq 0}$) be the semigroup associated to L (resp. $\widehat{L}$). For any $f \in \mathcal{D}(L)$, we have

$$\forall\, t \geq 0, \forall\, x \in V, \qquad P_t[f](x) \;=\; \mathbb{E}_x[f(X_t)]$$

where the subscript x indicates that X starts from x. So, if $m_0 = \mathcal{L}(X_0)$ is such that $m_0(N) = 0$, we get for any $t \geq 0$,

$$\begin{aligned}
\mathbb{E}[f(X_t)] &= \sum_{x\in[\![0,N-1]\!]} m_0(x)\mathbb{E}_x[f(X_t)] \\
&= \ell(m_0 P_t[f]) \\
&= \ell(\widehat{P}_t[m_0]f)
\end{aligned}$$

But if $m_0 = \rho$, we have by definition $\widehat{P}_t[\rho] = \exp(-\lambda_1 t)\rho$, so

$$\begin{aligned}
\mathbb{E}[f(X_t)] &= \exp(-\lambda_1 t)\ell(\rho f) \\
&= \exp(-\lambda_1 t)\rho(f) \qquad (12)
\end{aligned}$$

In particular for $f = \mathbb{1}_{[\![0,N-1]\!]}$ and any $t \geq 0$,

$$\begin{aligned}
\mathbb{P}[\tau > t] &= \mathbb{E}[\mathbb{1}_{[\![0,N-1]\!]}(X_t)] \\
&= \exp(-\lambda_1 t)\rho([\![0, N-1]\!]) \\
&= \exp(-\lambda_1 t)
\end{aligned}$$

Since this is true for any $t \geq 0$, it follows that τ is distributed as an exponential variable of parameter λ_1.

■

From (12), we deduce that for any $t \geq 0$,

$$\mathcal{L}(X_t) \quad = \quad \exp(-\lambda_1 t)\rho + (1 - \exp(-\lambda_1 t))\delta_N$$

if $\mathcal{L}(X_0) = \rho$. This justifies the name of quasi-stationary distribution for ρ. Seneta [26] is a useful reference for quasi-stationarity.

Coming back to the case where X starts from 0, we say that an a.s. finite randomized stopping time S for X is a *strong quasi-stationary time* if S and X_S are independent and X_S is distributed as ρ (note in particular that we must have $S < \tau$). It is furthermore called a *fastest time to quasi-stationarity* if it is stochastically smaller than any other strong quasi-stationary time.

Our main goal is to get the following result:

Proposition 5 *For a birth and death process starting at 0 and first absorbed at N, there exists a strong quasi-stationary time S for X.*

Before coming to the proof ot this proposition, let us explain how it will be used to prove Proposition 1. By Lemma 4, the time needed after S to be absorbed is an exponential variable of parameter λ_1 which is furthermore independent of S (since it only depends on the past up to time S through X_S). But by construction, S will have the same law as that of the absorbtion time of a birth and death process starting from 0, absorbed in $N-1$ and whose negatives of the Dirichlet eigenvalues are exactly $\lambda_2 < \lambda_3 < \cdots < \lambda_N$. So Proposition 1 will follow by iteration.

Indeed, to prove Proposition 5, we will construct a dual process X^* on $[\![0, N-1]\!]$, whose absorption time in $N-1$ is a strong quasi-stationary time (and even a fastest time to quasi-stationarity as it will appear later on, see Remark 13) for X, once X and X^* are appropriately intertwined.
To continue, we first need to derive from L an irreducible generator $\widetilde{L}$ on $[\![0, N-1]\!]$.

Lemma 6 *Let φ_1 be an eigenfunction of L associated to $-\lambda_1$ and define the operator*

$$\begin{array}{rcl} \widetilde{L} : \mathcal{F}([\![0, N-1]\!]) & \to & \mathcal{F}([\![0, N-1]\!]) \\ f & \mapsto & L[\varphi_1 f] + \lambda_1 \varphi_1 f \end{array}$$

where $\mathcal{F}([\![0, N-1]\!])$ is the vector space of real functions defined on $[\![0, N-1]\!]$ (in the above formula, we also naturally identify it with $\mathcal{D}(L)$, extending functions by 0 at N, or in the reverse way, taking restriction to $[\![0, N-1]\!]$). Then $\widetilde{L}$ is an irreducible birth and death generator whose reversible probability is ρ.

Proof
If we identify $\widetilde{L}$ with its matrix $(\widetilde{L}(x,y))_{x,y\in[\![0,N-1]\!]}$, it follows from our assumption on the birth and death rates of L that

$$\forall\, x, y \in [\![0, N-1]\!], \qquad \left\{ \begin{array}{lcl} |x-y| = 1 & \Rightarrow & \widetilde{L}(x,y) > 0 \\ |x-y| > 1 & \Rightarrow & \widetilde{L}(x,y) = 0 \end{array} \right.$$

So to check that $\widetilde{L}$ is an irreducible birth and death generator $\widetilde{L}$ on $[\![0, N-1]\!]$, it is sufficient to verify that $\widetilde{L}[\mathbb{1}_{[\![0,N-1]\!]}] = 0$ on $[\![0, N-1]\!]$, but this is a direct consequence of the definition of φ_1.

The probability ρ will be invariant for $\widetilde{L}$, if and only if for any $f \in \mathcal{F}(\llbracket 0, N-1 \rrbracket)$, we have $\rho(\widetilde{L}[f]) = 0$, and we compute

$$\begin{aligned}\rho(\widetilde{L}[f]) &= \rho(L[\varphi_1 f]) + \lambda_1 \rho(\varphi_1 f) \\ &= -\lambda_1 \rho(\varphi_1 f) + \lambda_1 \rho(\varphi_1 f) \\ &= 0\end{aligned}$$

The second equality comes from the definition of ρ: for any $g \in \mathcal{D}(L)$, $\rho(L[g]) = \ell(\psi L[g])/\ell(\psi) = \ell(gL[\psi])/\ell(\psi) = -\lambda_1 \ell(g\psi)/\ell(\psi) = -\lambda_1 \rho(g)$.
The probability ρ is indeed reversible, as it is always the case for an invariant measure associated to a finite birth and death generator.

■

As a consequence, we obtain the following expression for L.

Lemma 7 *There exists a function $v \in \mathcal{F}(\llbracket -1, N-1 \rrbracket$ such that seen as an operator on $\mathcal{F}(\llbracket 0, N-1 \rrbracket)$, L can be rewritten*

$$L = -\lambda_1 - \frac{1}{\rho} \partial^- v \partial^+ \frac{1}{\varphi_1}$$

where $1/\rho$ is seen as an element of $\mathcal{F}(\llbracket 0, N-1 \rrbracket)$ and $1/\varphi_1$ has been extended as a function in $\mathcal{F}(\llbracket -1, N \rrbracket)$ by making it vanish on $\{-1, N\}$. The latter convention is not really necessary, because we must have $v(-1) = v(N-1) = 0$. Furthermore v is positive on $\llbracket 0, N-2 \rrbracket$.

Proof

It follows from (2) applied to the irreducible birth and death generator $\widetilde{L}$ and the preceding lemma that there exists a function $v \in \mathcal{F}(\llbracket -1, N-1 \rrbracket)$ such that

$$\widetilde{L} = -\frac{1}{\rho} \partial^- v \partial^+$$

and v satisfies the properties stated above. Translating this expression to L, we get the claimed result.

■

It is not difficult to check that that the function v is indeed unique. The advantage of this formulation is that it makes it easy to find a "first order" difference operator D and a "second order" difference operator L^* such that an algebraic duality relation is satisfied:

Lemma 8 *Define $D := -\frac{1}{\rho} \partial^- R$ and $L^* := -\lambda_1 - \frac{v}{R} \partial^+ \frac{1}{\varphi_1 \rho} \partial^- R$, where R is the cumulative function of ρ:*

$$\forall\, x \in \llbracket -1, N \rrbracket, \qquad R(x) := \sum_{0 \le y \le x} \rho(y)$$

Since $v(N-1) = 0$, the function $\frac{1}{\varphi_1 \rho}$ need not be defined at N and the operators $D : \mathcal{F}(\llbracket -1, N-1 \rrbracket) \to \mathcal{F}(\llbracket 0, N-1 \rrbracket)$ and $L^* : \mathcal{F}(\llbracket -1, N \rrbracket) \to \mathcal{F}(\llbracket 0, N-1 \rrbracket)$, can be naturally interpreted as going from $\mathcal{F}(\llbracket 0, N-1 \rrbracket)$ to itself. Then we have

$$LD = DL^* \tag{13}$$

Furthermore, the restriction of L^* to $\mathcal{D}_{N-1} := \{f \in \mathcal{F}(\llbracket 0, N-1 \rrbracket) : f(N-1) = 0\}$ is a Markov generator absorbed at $N-1$.

Proof
The first assertions are immediate to check. Concerning (13), it is equivalent to verify that

$$\left(\frac{1}{\rho}\partial^- v\partial^+\frac{1}{\varphi_1}\right)\circ D \quad = \quad D\circ\left(\frac{v}{R}\partial^+\frac{1}{\varphi_1\rho}\partial^- R\right)$$

This is true, because both sides are equal to

$$-\frac{1}{\rho}\partial^- v\partial^+\frac{1}{\varphi_1\rho}\partial^- R$$

So the only point which needs some care is the last sentence of Lemma 8. First the image of $\mathcal{D}_{N-1}$ by L^* is included in $\mathcal{D}_{N-1}$: let $f\in\mathcal{D}_{N-1}$, we compute that

$$\begin{aligned} L^*[f](N-1) &= -\lambda_1 f(N-1)-\frac{v(N-1)}{R(N-1)}\partial^+\frac{1}{\varphi_1\rho}\partial^- Rf(N-1)\\ &= 0\end{aligned}$$

because $v(N-1)=0$. Next by definition of L^* and the fact that R, φ_1 and ρ (respectively v) are positive on $[\![0,N-1]\!]$ (resp. $[\![0,N-2]\!]$) we already see that

$$\forall\, x,y\in[\![0,N-1]\!],\qquad \begin{cases}|x-y|=1 &\Rightarrow\quad L^*(x,y)>0\\ |x-y|>1 &\Rightarrow\quad L^*(x,y)=0\end{cases}$$

so to conclude that the restriction of L^* to $\mathcal{D}_{N-1}$ is a Markov generator, it remains to check that $L^*[\mathbb{1}_{[\![0,N-1]\!]}]=0$ on $[\![0,N-2]\!]$. But using that $\partial^- R=-\rho$ on $[\![0,N-1]\!]$, this property can be rewritten

$$\begin{aligned} &\forall\, x\in[\![0,N-2]\!],\quad \left(\frac{v}{R}\partial^+\frac{1}{\varphi_1\rho}\partial^- R\right)(x) \;=\; -\lambda_1\\ \iff\quad &\forall\, x\in[\![0,N-2]\!],\quad \left(\frac{v}{R}\partial^+\frac{1}{\varphi_1}\right)(x)\;=\;\lambda_1\\ \iff\quad &\forall\, x\in[\![0,N-2]\!],\quad v(x)\left(\partial^+\frac{1}{\varphi_1}\right)(x)\;=\;\lambda_1 R(x)\end{aligned}$$

We note that the last equality is satisfied if we consider it at $x=-1$, both sides being equal to zero. So taking differences with respect to ∂^-, we get

$$\begin{aligned} &\forall\, x\in[\![0,N-2]\!],\quad L^*[\mathbb{1}_{[\![0,N-1]\!]}](x)\;=\;0\\ \iff\quad &\forall\, x\in[\![0,N-2]\!],\quad \left(\partial^- v\partial^+\frac{1}{\varphi_1}\right)(x)\;=\;\lambda_1\partial^- R(x)\\ \iff\quad &\forall\, x\in[\![0,N-2]\!],\quad \left(\partial^- v\partial^+\frac{1}{\varphi_1}\right)(x)\;=\;-\lambda_1\rho(x)\\ \iff\quad &\forall\, x\in[\![0,N-2]\!],\quad \left(\frac{1}{\rho}\partial^- v\partial^+\frac{1}{\varphi_1}\right)(x)\;=\;-\lambda_1\\ \iff\quad &\forall\, x\in[\![0,N-2]\!],\quad L[\mathbb{1}_{[\![0,N-1]\!]}](x)\;=\;0\end{aligned}$$

(by Lemma 7) which is satisfied, since the restrictions of $L[\mathbb{1}_{[\![0,N-1]\!]}]$ and $L[\mathbb{1}_{[\![0,N]\!]}]$ to $[\![0,N-2]\!]$ coincide and $L[\mathbb{1}_{[\![0,N]\!]}]$ vanishes on $[\![0,N]\!]$.

■

To follow the development presented in the previous section, we will have to be more careful with the domains of the operators.

First, $D : \mathcal{F}(\llbracket 0, N-1 \rrbracket) \to \mathcal{F}(\llbracket 0, N-1 \rrbracket)$ is one-to-one and its inverse Λ is Markovian and given by

$$\forall\, f \in \mathcal{F}(\llbracket 0, N-1 \rrbracket),\, \forall\, x \in \llbracket 0, N-1 \rrbracket, \qquad \Lambda[f](x) \;\coloneqq\; \frac{1}{R(x)} \sum_{0 \le y \le x} \rho(y) f(y)$$

So we deduce from (13) that we have on $\mathcal{F}(\llbracket 0, N-1 \rrbracket)$,

$$\Lambda L \;=\; L^* \Lambda \tag{14}$$

and as a consequence,

$$\forall\, t \ge 0, \qquad \Lambda \exp(tL) \;=\; \exp(tL^*) \Lambda$$

But the semigroup $(\exp(tL^*))_{t \ge 0}$ is Markovian only if we restrict it to $\mathcal{D}_{N-1}$. So for the previous formula to be useful for intertwining, we need to slightly change the point of view on the Markov processes associated to L and L^*. More precisely, let

$$\mathcal{S}_{N-1} \coloneqq \{f \in \mathcal{F}(\llbracket 0, N-1 \rrbracket) \,:\, \rho(f) = 0\}$$

so that the image of $\mathcal{S}_{N-1}$ by Λ is $\mathcal{D}_{N-1}$. We now see the operators $\Lambda : \mathcal{S}_{N-1} \to \mathcal{D}_{N-1}$ and $D : \mathcal{D}_{N-1} \to \mathcal{S}_{N-1}$ as inverses of each other. Let $\check{L}$ be the irreducible (but non-reversible) Markov generator on $\llbracket 0, N-1 \rrbracket$ whose jump rates are given by

$$\forall\, x, y \in \llbracket 0, N-1 \rrbracket, \text{ with } x \neq y,$$
$$\check{L}(x,y) \;=\; \begin{cases} L(x,y) & , \text{if } x \neq N-1 \\ L(N-1,N)\rho(y) & , \text{if } x = N-1 \text{ and } y \neq N-2 \\ L(N-1,N-2) + L(N-1,N)\rho(N-2) & , \text{if } x = N-1 \text{ and } y = N-2 \end{cases}$$

Then ρ is the invariant probability associated to $\check{L}$. This can be computed directly, but it is clearer from a probabilistic point of view: the Markov process corresponding to $\check{L}$, instead of jumping from $N-1$ to N (as the Markov process associated to L), redistributes itself according to the quasi-stationary distribution. In particular $\check{L}$ can be seen as an operator from $\mathcal{S}_{N-1}$ to $\mathcal{S}_{N-1}$. We also observe that on $\mathcal{S}_{N-1}$, L and $\check{L}$ coincide, so we deduce from (14) the following commutative diagram

$$\begin{array}{ccc} \mathcal{S}_{N-1} & \xrightarrow{\check{L}} & \mathcal{S}_{N-1} \\ \Lambda\big\downarrow & & \Lambda\big\downarrow \\ \mathcal{D}_{N-1} & \xrightarrow{L^*} & \mathcal{D}_{N-1} \end{array} \tag{15}$$

By definition, $\check{L}$ is a Markov generator on $\llbracket 0, N-1 \rrbracket$, so in particular we have $\check{L}[\mathbb{1}_{\llbracket 0,N-1 \rrbracket}] = 0$. Let us consider $\check{L}^*$ the operator which coincides with L^* on $\mathcal{D}_{N-1}$ and which satisfies $\check{L}^*[\mathbb{1}_{\llbracket 0,N-1 \rrbracket}] = 0$ (in view of the proof of Lemma 8, this amounts to just replacing the entry $L^*(N-1,N-1) = -\lambda_1$ by $\check{L}^*(N-1,N-1) = 0$). It appears that $\check{L}^* : \mathcal{F}(\llbracket 0, N-1 \rrbracket) \to \mathcal{F}(\llbracket 0, N-1 \rrbracket)$ is a Markovian generator on $\llbracket 0, N-1 \rrbracket$ absorbed at $N-1$. Since $\Lambda(\mathbb{1}_{\llbracket 0,N-1 \rrbracket}) = \mathbb{1}_{\llbracket 0,N-1 \rrbracket}$, we get that $\Lambda\check{L}[\mathbb{1}_{\llbracket 0,N-1 \rrbracket}] = 0 = L^*\Lambda[\mathbb{1}_{\llbracket 0,N-1 \rrbracket}]$, so Diagram (15) can be extended to

$$\begin{array}{ccc} \mathcal{F}(\llbracket 0, N-1 \rrbracket) & \xrightarrow{\check{L}} & \mathcal{F}(\llbracket 0, N-1 \rrbracket) \\ \Lambda\big\downarrow & & \Lambda\big\downarrow \\ \mathcal{F}(\llbracket 0, N-1 \rrbracket) & \xrightarrow{\check{L}^*} & \mathcal{F}(\llbracket 0, N-1 \rrbracket) \end{array}$$

and it follows that for any $t \geq 0$,

$$\begin{array}{ccc} \mathcal{F}([\![0,N-1]\!]) & \xrightarrow{\exp(t\check{L})} & \mathcal{F}([\![0,N-1]\!]) \\ \Lambda \big\downarrow & & \Lambda \big\downarrow \\ \mathcal{F}([\![0,N-1]\!]) & \xrightarrow{\exp(t\check{L}^*)} & \mathcal{F}([\![0,N-1]\!]) \end{array}$$

which is an intertwining relation between "true" Markovian semigroups.

Next let $(\check{X}_t)_{t\geq 0}$ (respectively $(\check{X}^*_t)_{t\geq 0}$) be a Markov process starting from 0 with generator $\check{L}$ (resp. $\check{L}^*$). Since their initial conditions $\check{m}_0 = \delta_0 = \check{m}^*_0$ satisfy $\check{m}^*_0\Lambda = \check{m}_0$, again we can use Theorem 2 of Fill [9] to get a strong Markovian coupling of $(\check{X}_t)_{t\geq 0}$ and $(\check{X}^*_t)_{t\geq 0}$, still denoted by $(\check{X}_t, \check{X}^*_t)_{t\geq 0}$, such that for any $t \geq 0$, a.s.,

$$\mathcal{L}(\check{X}_t | \check{X}^*_t,\, 0 \leq s \leq t) \;=\; \Lambda(\check{X}^*_t, \cdot) \tag{16}$$

Now the situation has been reduced to that of the previous section. So

$$\check{\tau}^* \;\coloneqq\; \inf\{t \geq 0 \,:\, \check{X}^*_t = N-1\}$$

is a strong stationary time for $\check{X}$. Furthermore, since Λ satisfies

$$\forall\, x \in [\![0, N-1]\!], \qquad \Lambda(x, [\![0, x]\!]) \;=\; 1$$

it follows from (16) that we have a.s.

$$\forall\, t \geq 0, \qquad \check{X}_t \;\leq\; \check{X}^*_t$$

thus $\check{\tau}^* \leq \check{\tau} \coloneqq \inf\{t \geq 0 \,:\, \check{X}_t = N-1\}$. But up to time $\check{\tau}$, $\check{X}$ and X (recall that it is a Markov process starting from 0 and whose generator is L) have the same law, so we have proven Proposition 5.

Remark 9 The above arguments can be extended to the case where m_0 the initial distribution of X_0 satisfies that m_0/ρ is non-increasing on V (implying in particular that $m_0(N) = 0$).
If we assume that m_0/ρ is increasing, there is no quasi-stationary time for X. Indeed, observe that the mapping $[\![0, N-1]\!] \ni x \mapsto \mathbb{E}_x[\tau]$ is decreasing, so we get that $\mathbb{E}_{m_0}[\tau] < \mathbb{E}_\rho[\tau] = 1/\lambda_1$. But if there exists a strong quasi-stationary times S, then τ is stochastically larger (or equal) than an exponential variable of parameter λ_1 (see the argument below this remark), in contradiction with the above bound.
We wonder about a necessary and sufficient condition in terms of m_0 for the existence of a quasi-stationary time for X

□

Let us define

$$T_1 \;\coloneqq\; \inf\{t \geq 0 \,:\, X_{S+t} = N\}$$

By the strong Markov property applied to the randomized stopping time S, T_1 depends on $(X_{S\wedge t})_{t\geq 0}$ only through X_S and is thus independent of S, which is a strong quasi-stationary time for X. Furthermore, the strong Markov property and Lemma 4 imply that T_1 is distributed as an exponential variable of parameter λ_1. So, writing

$$\tau \;=\; S + T_1$$

is the first step in the iterative proof of Proposition 1. Indeed, by the above characterization of S, it is the absorption time $\check{\tau}^*$ of the birth and death process $\check{X}^*$ starting from 0, so Proposition 1 is proven if we verify that the eigenvalues of $-L^* : \mathcal{D}_{N-1} \to \mathcal{D}_{N-1}$ are exactly $\lambda_2 < \lambda_3 < \cdots < \lambda_N$. By (15), the eigenvalues of $L^* : \mathcal{D}_{N-1} \to \mathcal{D}_{N-1}$ are exactly those of $\check{L} : \mathcal{S}_{N-1} \to \mathcal{S}_{N-1}$, which also coincide with those of $L : \mathcal{S}_{N-1} \to \mathcal{S}_{N-1}$. But the vector spaces $\mathrm{Vect}(\varphi_1)$ and $\mathcal{S}_{N-1}$ are both stable by $L : \mathcal{D}_N \to \mathcal{D}_N$ (with the obvious notation $\mathcal{D}_N = \{f \in \mathcal{F} : f(N) = 0\}$) and $\mathrm{Vect}(\varphi_1)$ is the eigenspace associated to the eigenvalue $-\lambda_1$, so necessarily the eigenvalues of $-L : \mathcal{S}_{N-1} \to \mathcal{S}_{N-1}$ are the $\lambda_2 < \lambda_3 < \cdots < \lambda_N$.

Nevertheless, the particular intertwining relation mentioned in the introduction contains more information, that is why we will construct it in the next section.

4 Intertwining processes

We now slightly modify the intertwining described in last section, so that it can be iterated.

More precisely, we begin by extending the operator L^* defined in Lemma 8 on $\mathcal{F}([\![0, N-1]\!])$, identified with $\mathcal{D}_N$, into an operator, still denoted L^*, on $\mathcal{F}$, by imposing that $L^*[\mathbb{1}_V] = 0$. Recall that $V := [\![0, N]\!]$ and that $\mathcal{F} := \mathcal{F}([\![0, N]\!])$. From a matrix point of view, this operation amounts to adding to $(L^*(x, y))_{x,y \in [\![0,N-1]\!]}$ a row N of zeroes and a column N of zeroes, except for the entry $(N-1, N)$ which is equal to λ_1. Of course L^* is now a Markov generator on V which is absorbed at N.
Next we extend Λ into a Markov kernel on V, by taking

$$\Lambda(N, \cdot) \quad := \quad \delta_N(\cdot)$$

It is not difficult to check that this is in fact the only possible choice if we want this kernel to coincide on $[\![0, N-1]\!]$ with the previous one and so that we have an algebraic duality relation on $\mathcal{F}$,

$$\Lambda L \quad = \quad L^* \Lambda \tag{17}$$

With the above interpretation, this is an intertwining relation between true Markov generators, contrary to the one (14) considered in last section. So by Theorem 2 of Fill [9], if m_0 and m_0^* are two probabilities on V satisfying $m_0 = m_0^* \Lambda$, we can construct a Markov process $(X_t, X_t^*)_{t \geq 0}$ such that:
- the process $(X_t)_{t \geq 0}$ is Markovian with generator L and initial distribution m_0
- the process $(X_t^*)_{t \geq 0}$ is Markovian with generator L^* and initial distribution m_0^*
- for any $t \geq 0$, we have a.s., $\mathcal{L}(X_t | X_s^*, \, 0 \leq s \leq t) \; = \; \Lambda(X_t^*, \cdot)$.

By definition, the Markov kernel Λ satisfies

$$\begin{aligned} \forall\, x \in [\![0, N-1]\!], \qquad \Lambda(x, [\![0, x]\!]) &= 1 \\ \Lambda(N, \{N\}) &= 1 \end{aligned}$$

Thus, by the arguments given in the introduction, we get that a.s. $\tau = \tau^*$, where as usual,

$$\begin{aligned} \tau &= \inf\{t \geq 0 : X_t = N\} \\ \tau^* &= \inf\{t \geq 0 : X_t^* = N\} \end{aligned}$$

This property allows other extensions of (17). Assume for instance that we are given $M \in \mathbb{N}$ and $\widetilde{L}$ a Markov generator on $[\![0, N+M]\!]$ such that

$$\forall\, x, y \in [\![0, N+M]\!], \qquad \widetilde{L}(x, y) \quad = \quad \begin{cases} L(x, y) & , \text{ if } x \in [\![0, N-1]\!] \text{ and } y \in [\![0, N]\!] \\ 0 & , \text{ if } x \in [\![N, N+M]\!] \text{ and } y \in [\![0, N-1]\!] \end{cases}$$

Of course, we must have $\widetilde{L}(x,y) = 0$ for $x \in [\![0, N-1]\!]$ and $y \in [\![N+1, N+M]\!]$, but the entries of $(\widetilde{L}(x,y))_{(x,y)\in[\![N,N+M]\!]^2}$ are free, as long as $\widetilde{L}$ remains a Markov generator.
Next, let $\widetilde{L}^*$ be the Markov generator defined on $[\![0, N+M]\!]$ by

$$\forall\, x,y \in [\![0,N+M]\!], \qquad \widetilde{L}^*(x,y) \;=\; \begin{cases} L^*(x,y) & \text{, if } x \in [\![0,N-1]\!] \text{ and } y \in [\![0,N]\!] \\ \widetilde{L}(x,y) & \text{, if } x \in [\![N,N+M]\!] \text{ and } y \in [\![N,N+M]\!] \\ 0 & \text{, otherwise} \end{cases}$$

and let Λ be the Markov transition matrix defined on $[\![0, N+M]\!]$ which coincides with the previous one on $[\![0, N]\!]$ and which satisfies $\Lambda(x,\cdot) \coloneqq \delta_x$ for $x \in [\![N, N+M]\!]$.
Then it is immediate to check that we still have $\widetilde{\Lambda}\widetilde{L} = \widetilde{L}^*\widetilde{\Lambda}$ and then

$$\Lambda L \;=\; L^*\Lambda \tag{18}$$

So if m_0 and m_0^* are two probabilities on $[\![0, N+M]\!]$ such that $m_0 = m_0^*\Lambda$, then we can find a Markov process $(X_t, X_t^*)_{t\geq 0}$ such that, as before, the process $(X_t)_{t\geq 0}$ is Markovian with generator L and initial distribution m_0, the process $(X_t^*)_{t\geq 0}$ is Markovian with generator L^* and initial distribution m_0^* and for any $t \geq 0$, we have a.s.,

$$\mathcal{L}(X_t | X_s^*,\, 0 \leq s \leq t) \;=\; \Lambda(X_t^*, \cdot) \tag{19}$$

Indeed, this can be deduced from the previous construction (corresponding to $M = 0$): assume for instance that $m_0^*([\![0, N-1]\!]) = 1$, then we use the previous coupling up to the time $\tau = \tau^*$ and after this time, X and X^* stick together. Similarly, if $X_0^* \geq N$, we take $X = X^*$. This direct construction can also be used as an alternative to the matrix verification of (18): first consider $x \in [\![0, N-1]\!]$ and let $m_0^* = \delta_x$ and $m_0 = \Lambda(x,\cdot)$. Taking into account (19), we get for any $t \geq 0$ and any function $f \in \mathcal{F}([\![0, N+M]\!])$,

$$\sum_{y\in V} \Lambda(x,y)\mathbb{E}_y[f(X_t)] \;=\; \mathbb{E}_x[\Lambda(X_t^*, f)]$$

and thus by differentiation with respect to t at 0_+, we recover (18) on $[\![0, N-1]\!]$. To get it on $[\![N, N+M]\!]$, we use that for $x \in [\![N, N+M]\!]$, for any $t \geq 0$ and any function $f \in \mathcal{F}([\![0, N+M]\!])$,

$$\mathbb{E}_x[f(X_t)] \;=\; \mathbb{E}_x[f(X_t^*)]$$

With all these preliminaries, we can now construct iteratively the generators $L^{(i)}$ of the processes $X^{(i)}$ mentioned in the introduction, for $i \in [\![0, N]\!]$. We start with $L^{(0)} = L$. Next we assume that for some $i \in [\![0, N-1]\!]$, we have constructed a birth and death generator $L^{(i)}$ on V such that:
• The corresponding birth and death rates $(b_x^{(i)})_{0\leq x<N}$ and $(d_x^{(i)})_{0\leq x<N}$ satisfy $b_x^{(i)} = \lambda_{N-x}$ for $N-i \leq x < N$, $d_x^{(i)} = 0$ for $N-i < x \leq N$ and are positive otherwise.
• $\lambda_{i+1}, \lambda_{i+2}, ..., \lambda_N$ are the eigenvalues of the operator

$$\mathcal{F}([\![0, N-i-1]\!]) \ni f \;\;\mapsto\;\; (-L^{(i)}[\bar{f}](x))_{x\in[\![0,N-i]\!]} \in \mathcal{F}([\![0, N-i-1]\!])$$

where $\bar{f}$ is the function from $\mathcal{F}$ which coincides with f on $[\![0, N-i-1]\!]$ and which vanishes on $[\![N-i, N]\!]$ (so the above operator is just the restriction of $-L^{(i)}$ on $[\![0, N-i]\!]$ with a Dirichlet boundary condition on $N-i$).
• There is a Markov kernel $\bar{\Lambda}^{(i)}$ from V to V such that

$$\begin{aligned} \forall\, x \in [\![0, N-1]\!], \qquad \bar{\Lambda}^{(i)}(x, [\![0, x]\!]) \;&=\; 1 \\ \bar{\Lambda}^{(i)}(N, \{N\}) \;&=\; 1 \end{aligned}$$

(such a kernel will be called lower triangular in the sequel) and $\bar{\Lambda}^{(i)}$ serves as a link between $L^{(i)}$ and L on $\mathcal{F}$:

$$\bar{\Lambda}^{(i)}L = L^{(i)}\bar{\Lambda}^{(i)} \tag{20}$$

(for definiteness, we can take $\bar{\Lambda}^{(0)} = \mathrm{Id}$, the identity kernel).

We now construct $L^{(i+1)}$ and $\bar{\Lambda}^{(i+1)}$. We begin by considering the restriction of $L^{(i)}$ on $[\![0, N-i]\!]$ with a Dirichlet boundary condition on $N-i$. Applying the construction of the previous section, we get a dual operator $L^{(i)*}$ on $\mathcal{F}([\![0, N-i-1]\!])$ and a Markov kernel $\Lambda^{(i)}$ on $[\![0, N-i-1]\!]$ such that on $\mathcal{F}([\![0, N-i-1]\!])$,

$$\Lambda^{(i)}L^{(i)} = L^{(i)*}\Lambda^{(i)}$$

Note that $\Lambda^{(i)}$ satisfies

$$\forall\, x \in [\![0, N-i-1]\!], \qquad \Lambda^{(i)}(x, [\![0, x]\!]) = 1$$

and that the restriction of $-L^{(i)*}$ to $\{f \in \mathcal{F}([\![0, N-i-1]\!]) \,:\, f(N-i-1) = 0\}$ has $\lambda_{i+2}, \lambda_{i+3}, ..., \lambda_N$ as eigenvalues.

Next the considerations of the beginning of this section enable us to extend, on one hand, $L^{(i)*}$ into a birth and death generator $L^{(i+1)}$ on V and on the other hand, $\Lambda^{(i)}$ into a Markov kernel on V, again denoted $\Lambda^{(i)}$, such that we have on $\mathcal{F}$,

$$\Lambda^{(i)}L^{(i)} = L^{(i+1)}\Lambda^{(i)} \tag{21}$$

Furthermore $L^{(i+1)}$ has the required form and $\Lambda^{(i)}$ satisfies

$$\begin{aligned} \forall\, x \in [\![0, N-i-1]\!], &\qquad \Lambda^{(i)}(x, [\![0, x]\!]) = 1 \\ \forall\, x \in [\![N-i, N]\!], &\qquad \Lambda^{(i)}(x, \{x\}) = 1 \end{aligned}$$

So $\bar{\Lambda}^{(i+1)} = \Lambda^{(i)}\bar{\Lambda}^{(i)}$ is a lower triangular kernel and we get from (20) and (21) that

$$\begin{aligned} \bar{\Lambda}^{(i+1)}L &= \Lambda^{(i)}\bar{\Lambda}^{(i)}L \\ &= \Lambda^{(i)}L^{(i)}\bar{\Lambda}^{(i)} \\ &= L^{(i+1)}\Lambda^{(i)}\bar{\Lambda}^{(i)} \\ &= L^{(i+1)}\bar{\Lambda}^{(i+1)} \end{aligned}$$

Thus the iterative step is completed.

At the end of this procedure, we get the announced generator $L^{(N)}$, described by

$$\forall\, x, y \in V, \qquad L^{(N)}(x, y) = \begin{cases} -\lambda_{N-x} & \text{, if } x = y \\ \lambda_{N-x} & \text{, if } y = x+1 \\ 0 & \text{, otherwise} \end{cases}$$

which is intertwined with L by a lower triangular kernel $\bar{\Lambda}^{(N)}$. In particular we can find a coupling $(X^{(N)}, X)$ of $X^{(N)}$, a Markov process starting from 0 and generated by $L^{(N)}$ and X, a Markov process starting from 0 and generated by L, satisfying for all $t \geq 0$, a.s.

$$\mathcal{L}(X_t | X_s^{(N)}, 0 \leq s \leq t) = \bar{\Lambda}^{(N)}(X_t^{(N)}, \cdot)$$

As explained in the introduction, Proposition 1 follows at once from the existence of such a process.

But one can construct more intertwined processes. For $0 \leq i \leq N$, let us denote by $X^{(i)}$ a Markov process starting from 0 and generated by $L^{(i)}$. Then from (21), for any $0 \leq i < j \leq N$, we can find a coupling $(X^{(j)}, X^{(i)})$ such that for all $t \geq 0$, a.s.

$$\mathcal{L}(X_t^{(i)} | X_s^{(j)}, 0 \leq s \leq t) = \bar{\Lambda}^{(j,i)}(X_t^{(j)}, \cdot)$$

where $\Lambda^{(j,i)} = \Lambda^{(j-1)}\Lambda^{(j-2)}\cdots\Lambda^{(i)}$. This Markov kernel satisfies

$$\begin{aligned}\forall\, x \in [\![0, N-i-1]\!], &\qquad \bar{\Lambda}^{(j,i)}(x, [\![0,x]\!]) &= 1\\ \forall\, x \in [\![N-i, N]\!], &\qquad \bar{\Lambda}^{(j,i)}(x, \{x\}) &= 1\end{aligned}$$

One can even go further and couple all the processes $X^{(i)}$, for $0 \le i \le N$, into a "big" Markov process:

Proposition 10 *There exists a Markov process $(X_t^{(N)}, X_t^{(N-1)}, \cdots, X_t^{(0)})_{t\ge 0}$, such that for all $i \in [\![0, N-1]\!]$ and all $t \ge 0$, we have a.s.*

$$\mathcal{L}(X_t^{(i)} | X_s^{(j)},\, j \in [\![i+1, N]\!],\, 0 \le s \le t) \quad = \quad \Lambda^{(i)}(X_t^{(i+1)}, \cdot)$$

Furthermore, in this formula, the path valued finite sequence $[\![0, N]\!] \ni n \mapsto X^{(N-n)}$ is in fact Markovian (and by consequence, $[\![0, N]\!] \ni n \mapsto X^{(n)}$ is equally a Markov chain).

Proof

We begin with the Markov process $(X^{(N)}, X^{(N-1)})$ constructed as above and call $L^{(N,N-1)}$ its generator. Then we consider the Markov kernel $\widetilde{\Lambda}^{(N,N-1)}$ from V^2 to V defined by

$$\forall\, (x_N, x_{N-1}) \in V^2,\, \forall\, x_{N-2} \in V, \qquad \widetilde{\Lambda}^{(N,N-1)}((x_N, x_{N-1}), x_{N-2}) \;\;:=\;\; \Lambda^{(N-2)}(x_{N-1}, x_{N-2})$$

Taking into account that for functions depending only on the x_{N-1} variable, $L^{(N,N-1)}$ coincides with $L^{(N-1)}$, (21) implies that

$$\widetilde{\Lambda}^{(N,N-1)} L^{(N-2)} \quad = \quad L^{(N,N-1)} \widetilde{\Lambda}^{(N,N-1)}$$

This algebraic duality relation enables us to construct $(X^{(N)}, X^{(N-1)}, X^{(N-2)})$, by resorting one more time to Theorem 2 of Fill [9]. This procedure can obviously be iterated, by considering for $2 \le i \le N-1$, the Markov kernel $\widetilde{\Lambda}^{(N,N-i)}$ from V^{i+1} to V defined by

$$\begin{aligned}&\forall\, (x_N, x_{N-1}, \cdots, x_{N-i}) \in V^{i+1},\, \forall\, x_{N-i-1} \in V,\\ &\qquad\qquad \widetilde{\Lambda}^{(N,N-i)}((x_N, x_{N-1}, \cdots, x_{N-i}), x_{N-i-1}) \;\;:=\;\; \Lambda^{(N-i-1)}(x_{N-i}, x_{N-i-1})\end{aligned}$$

$\blacksquare$

Nevertheless, we think the most interesting intertwined process remains $(X^{(N)}, X)$. Let us consider for $i \in [\![0, N]\!]$, the probability $\pi_i = \bar{\Lambda}^{(N)}(i, \cdot)$, in particular we have $\pi_0 = \delta_0$ and $\pi_N = \delta_N$. It can be shown that for $i \in [\![1, N-1]\!]$, the support of π_i is $[\![0, i]\!]$ and that π_i is decreasing on this discrete interval. Essentially, this comes from the fact that the quasi-stationary distribution ρ considered in section 3 is decreasing on $[\![0, N-1]\!]$ (see for instance Miclo [21]) and the iterative definitions of the Markov kernels used to intertwine the previous generators. On a picture, the evolution of π_i when i goes from 0 to $N-1$ looks like an avalanche going from the left to the right. Next define for $i \in [\![0, N]\!]$,

$$\tau_i^{(N)} \quad := \quad \min\{t \ge 0 \,:\, X_t^{(N)} = i\}$$

As in section 3, we can prove that all these variables are strong randomized stopping times for X (the adjective strong refer to the fact that the position reached at the randomized stopping time is independent of this time) and by definition we have that for any $i \in [\![0, N]\!]$, $X_{\tau_i^{(N)}}$ is distributed as π_i. In some sense, this distribution is kept for some random time, since we have for any $t \ge 0$,

$$\mathcal{L}(X_t | \tau_i^{(N)} \le t < \tau_{i+1}^{(N)}) \quad = \quad \pi_i \tag{22}$$

Indeed, this an immediate consequence of the equality $\{\tau_i^{(N)} \le t < \tau_{i+1}^{(N)}\} = \{X_t^{(N)} = i\}$. The property (22) leads us to call the π_i, for $i \in [\![0, N-1]\!]$, local equilibria. In the same spirit, we deduce the following probabilistic representation of the time-marginal of X.

Theorem 11 *For any $t \geq 0$, we get*

$$\mathcal{L}(X_t) \quad = \quad \sum_{i\in V} \mathbb{P}\left[\sum_{N-i+1\leq j\leq N} T_j \leq t < \sum_{N-i\leq j\leq N} T_j\right] \pi_i$$

where the $(T_i)_{i\in[\![1,N]\!]}$ are independent exponential variables of respective parameters the $(\lambda_i)_{i\in[\![1,N]\!]}$ and with the convention that $T_0 = +\infty$.

This formula can be rewritten in terms of the right eigendecomposition of L, even if the latter description is less meaningful from a probabilistic point of view. Let us recall that

Lemma 12 *For any $i \in [\![1, N]\!]$, we have*

$$\mathbb{P}[T_N + \cdots + T_{N-i+1} > t] \quad = \quad \sum_{j\in[\![N-i+1,N]\!]} \left(\prod_{k\in[\![N-i+1,N]\!]\setminus\{j\}} \left(1 - \frac{\lambda_j}{\lambda_k}\right)^{-1}\right) \exp(-\lambda_j t)$$

One way to deduce this result is through an iteration with respect to $i \in [\![1, N]\!]$: this equality is clear for $i = 1$ and next write that for $i \geq 1$,

$$\mathbb{P}[T_N + \cdots + T_{N-i} > t] \quad = \quad \mathbb{P}[T_{N-i} > t] + \int_0^t \mathbb{P}[T_N + \cdots + T_{N-i+1} > t - s] \exp(-\lambda_{N-i}s)\lambda_{N-i}\, ds$$

to deduce the result at stage $i + 1$ from the result at stage i, taking into account the rational identity

$$\sum_{j\in[\![N-i,N]\!]} \left(\prod_{k\in[\![N-i,N]\!]\setminus\{j\}} \left(1 - \frac{\lambda_j}{\lambda_k}\right)^{-1}\right) \quad = \quad 1$$

The latter can be proven considering the divided differences taken at the points $\lambda_N, ..., \lambda_{N-i}$ with the function $x \mapsto x^i$ and using the Peano form.

Thus we get for any $t \geq 0$,

$$\begin{aligned}
\mathcal{L}(X_t) \quad &= \quad \sum_{i\in V} (\mathbb{P}[T_N + \cdots + T_{N-i+1} \leq t] - \mathbb{P}[T_N + \cdots + T_{N-i} \leq t])\pi_i \\
&= \quad \sum_{i\in[\![1,N]\!]} \mathbb{P}[T_N + \cdots + T_{N-i+1} \leq t](\pi_i - \pi_{i-1}) + \pi_0 \\
&= \quad \sum_{i\in[\![1,N]\!]} (1 - \mathbb{P}[T_N + \cdots + T_{N-i+1} > t])(\pi_i - \pi_{i-1}) + \pi_0 \\
&= \quad \pi_N - \sum_{i\in[\![1,N]\!]} \mathbb{P}[T_N + \cdots + T_{N-i+1} > t](\pi_i - \pi_{i-1}) \\
&= \quad \delta_N - \sum_{i\in[\![1,N]\!]} \sum_{j\in[\![N-i+1,N]\!]} \prod_{k\in[\![N-i+1,N]\!]\setminus\{j\}} \left(1 - \frac{\lambda_j}{\lambda_k}\right)^{-1} \exp(-\lambda_j t)(\pi_i - \pi_{i-1}) \\
&= \quad \delta_N - \sum_{j\in[\![1,N]\!]} \left(\sum_{i\in[\![N-j+1,N]\!]} \prod_{k\in[\![N-i+1,N]\!]\setminus\{j\}} \left(1 - \frac{\lambda_j}{\lambda_k}\right)^{-1} (\pi_i - \pi_{i-1})\right) \exp(-\lambda_j t)
\end{aligned}$$

It follows that for any $j \in [\![1, N]\!]$, the signed measure

$$\mu_j \quad \coloneqq \quad \sum_{i\in[\![N-j+1,N]\!]} \prod_{k\in[\![N-i+1,N]\!]\setminus\{j\}} \left(1 - \frac{\lambda_j}{\lambda_k}\right)^{-1} (\pi_i - \pi_{i-1}) \tag{23}$$

is an eigenvector of L seen as an operator acting on the right (namely on measures). The normalisation of these vectors is such that we have

$$\delta_0 = \pi_N - \sum_{j\in[\![1,N]\!]} \mu_j \tag{24}$$

Conversely, (23) can be inverted and the $(\pi_i)_{i\in[\![1,N]\!]}$ can be expressed in terms of the eigenmeasures $(\mu_i)_{i\in[\![1,N]\!]}$ satisfying (24). One aftermath of these considerations is that the parameters of the independent exponential variables $(T_i)_{i\in[\![1,N]\!]}$ and the probabilities $(\pi_i)_{i\in[\![1,N]\!]}$ appearing in Theorem 11 are uniquely determined (in particular, the former are necessarily the negatives of the eigenvalues of the underlying generator with a Dirichlet condition at N).

Coming back to Theorem 11, we see that the time marginal laws of the process always belong to the convex hull generated by the $(\pi_i)_{i\in[\![1,N]\!]}$. Furthermore, if the quotients λ_{i+1}/λ_i, for $i \in [\![1, N-1]\!]$, are very large, the trajectory $\mathbb{R}_+ \ni t \mapsto \mathcal{L}(X_t)$ has a tendency to be close to π_i at time $1/\lambda_i$, where it stays for a period of the same order, before going directly in direction of π_{i+1}, etc. This is generically the case for the Metropolis algorithms at small temperature, at least for the eigenvalues which vanish exponentially fast (cf. Miclo [22]). Furthermore, for those eigenvalues, the time T_i and $\tau_i^{(N)}$ are equivalent and are supposedly also close to the exit times associated to certain cycles, which are known to be almost exponential variables with eigenvalues as parameters (see for instance Bovier, Eckhoff, Gayrard and Klein [4] or Miclo [23]). Thus it would seem that asymptotically at small temperature, the previous random stopping times become "true" stopping time and get a "spatial" interpretation.

These observations lead us to believe that some of the behaviors we have displayed for birth and death processes starting from 0 could be extended to more general situations and this could led to a better understanding of metastability.

Remark 13 Since $\tau_N^{(N)}$ is the absorption time for X at N, it is the fastest strong stopping time such that $X_{\tau_N^{(N)}}$ is distributed according to δ_N. One can deduce from this property that for any $i \in [\![1, N]\!]$, $\tau_i^{(N)}$ is a fastest strong stopping time such that $X_{\tau_i^{(N)}}$ is distributed according to π_i and even better: let τ' be another such strong stopping time and define

$$\tau'' := \inf\{t \geq 0 : X_{\tau'+t} = N\}$$

By the strong Markov property, we get that τ'' is independent from τ' and that it has the same law as $\tau_N^{(N)} - \tau_i^{(N)}$. But $\tau' + \tau''$ is also distributed as $\tau_N^{(N)}$, so it follows that τ' has the same law as $\tau_i^{(N)}$. In particular, there is only one possible law for the strong quasi-stationary time considered in section 3 (this is a difference with strong stationary times: for example if T is such a time, then $T + t$ is also a strong stationary time, for any fixed $t \geq 0$). This can be extended to the exponential times $\tau_{i+1}^{(N)} - \tau_i^{(N)}$, for $i \in [\![0, N-1]\!]$: if X_0 is distributed according to π_i, the law of a strong stopping time τ such that X_τ is distributed according to π_{i+1} is necessarily an exponential variable with parameter $N - i$.

□

5 Examples

This short section contains two illustrative examples. The first is the Ehrenfest urn, where independent exponential variables show up naturally inside a fastest strong stationary time. The second concerns the continuous time random walk on a segment, absorbed at the right end.

Example 14: The continuous time version of the Ehrenfest urn
This is the birth and death process on $V := [\![0, N]\!]$ whose generator is given by

$$\forall\, x, y \in V, \qquad L(x,y) \;:=\; \begin{cases} x & \text{, if } y = x-1 \\ N-x & \text{, if } y = x+1 \\ -N & \text{, if } y = x \\ 0 & \text{, otherwise} \end{cases}$$

There is a traditional probabilistic way to construct a corresponding Markov process X starting from 0. We start by defining a Markov process Y on the hypercube $\{0,1\}^V$. Given a configuration on this state space, we attach to each site of V an exponential clock of parameter 2 (each of them being independent from the others). When the first clock rings, say at site $i \in V$, we flip a fair coin and the i^{th} coordinate is changed or allowed to stay the same as the coin comes up heads or tails. The construction goes on in the same way, starting from the (new or not new with probability 1/2) configuration obtained and we end up with a $\{0,1\}^V$-valued Markov process Y. If Y starts from the configuration where all spins are 0, X can be obtained by counting the number of spins equal to 1 in Y.
Let τ be the first time all coordinates have seen their respective clocks ring at least once. This randomized stopping time τ can clearly be written as a sum of exponential variables $(T_i)_{i\in[\![1,N]\!]}$ of parameters $(2i)_{i\in[\![1,N]\!]}$. Indeed, the first time T_N at which any clock rings is a minimum of N independent exponential variables of parameter 2, so it is an exponential variable of parameter $2N$. Next, by the loss of memory property of exponential variables, we wait a new time T_{N-1} for a clock from the other $N-1$ sites to ring, so this is an exponential variable of parameter $2(N-1)$, which is independent from T_N. Etc., until the last site has finally had its clock ring. This takes time T_1 since the last-but-one site has seen its own clock ringing.
Using the same probabilistic arguments as in Example 4.38 of Diaconis and Fill [7] (see also their Example 3.2 and Remark 2.39), in continuous time instead of discrete time, it can be shown that τ is a fastest strong stationary time for X. But from the above considerations (in particular section 2), we know that τ is a sum of independent exponential variables whose parameters are $(\lambda_i)_{i\in[\![1,N]\!]}$, the positive eigenvalues of $-L$. As "there is only one way to write a sum of independent exponential variables as a sum of independent exponential variables", it follows that we necessarily have

$$\forall\, i \in [\![1, N]\!], \qquad \lambda_i \;=\; 2i$$

This example can be seen as an entirely probabilistic computation of eigenvalues. Of course there are more classical ways to deduce them (see for instance Kac [12] or Diaconis [6]).

There are many other examples where natural fastest strong times to stationarity have been constructed, see the original papers of Aldous and Diaconis [1, 2], Diaconis and Fill [7], Pak [24] or Lovasz and Winkler [18]. The theory of the present paper shows that at least in the case of birth and death processes starting from one end of their state space, these times are sums of exponential variables.

□

The next example goes in the reverse direction and takes advantage of a known eigen-decomposition to compute an absorption time.

Example 15: Continuous time nearest neighbor random walk
Consider the birth and death process X on V, starting from 0, absorbed at N with generator given by

$$\forall\, x, y \in V, \qquad L(x,y) \;:=\; \begin{cases} 2 & \text{, if } x = 0 \text{ and } y = 1 \\ 1 & \text{, if } x \in [\![1, N-1]\!] \text{ and } |y-x| = 1 \\ -\sum_{z\in V\setminus\{x\}} L(x,z) & \text{, if } y = x \\ 0 & \text{, otherwise} \end{cases}$$

Then the time needed to go from 0 to N is distributed as a sum of independent exponential variables with parameters $(2(1-\cos(2\pi(2n-1)/(4N))))_{n\in[\![1,N]\!]}$.
Indeed, it is sufficient to show that the negatives of the eigenvalues of the sub-Markovian generator $\widetilde{L} := (L(x,y))_{x,y\in[\![0,N-1]\!]}$ on $[\![0,N-1]\!]$ are the $\lambda_n := 2(1-\cos(2\pi(2n-1)/(4N)))$, for $n\in[\![1,N]\!]$ (namely that the latter are the Dirichlet eigenvalues of $-L$). Let us also check that the corresponding eigenfunctions are given by

$$\varphi_n \,:\, [\![0,N-1]\!] \ni x \mapsto \cos(2\pi(2n-1)x/(4N))$$

To do so, we consider the generator $\widehat{L}$ of the usual continuous time nearest neighbor random walk on $\mathbb{Z}/(4N\mathbb{Z})$ (with rates 1). One verifies at once that if $f\in\mathcal{F}(\mathbb{Z}/(4N\mathbb{Z}))$ is an even function such that $f(N)=0$, then $\widehat{L}[f]$ coincide with $\widetilde{L}[f]$ on $[\![0,N-1]\!]$, where in the last expression, f has been identified with its restriction to $[\![0,N-1]\!]$. But the eigenvalues of $\widehat{L}$ are the $2(1-\cos(2\pi k/(4N)))$ with associated (complex-valued) eigenfunction $\mathbb{Z}/(4N\mathbb{Z})\ni x\mapsto \exp(2\pi ikx/(4N))$, for $k\in[\![0,4N-1]\!]$. Since the eigenvalues associated to k and $N-k$ coincide, it appears that the eigenvalues of $\widehat{L}$ are the $2(1-\cos(2\pi k/(4N)))$, for $k\in[\![0,2N-1]\!]$, their multiplicity is 2 and the corresponding eigenspace is generated by the two mappings $\mathbb{Z}/(4N\mathbb{Z})\ni x\mapsto \sin(2\pi kx/(4N))$ and $\mathbb{Z}/(4N\mathbb{Z})\ni x\mapsto\cos(2\pi kx/(4N))$. The latter function is odd and for k odd, it vanishes at N. So as announced, its restriction to $[\![0,N-1]\!]$ is an eigenfunction for $\widetilde{L}$, and since we get $N-1$ of them in this way, we have in fact exhibited all of them.

We also remark that the quasi-stationary distribution ρ is proportional to the measure $m\varphi_1$ on $[\![0,N-1]\!]$, where $m := (m(x))_{x\in[\![0,N-1]\!]}$ is given by $m(0)=1/2$ and $m(x)=1$ for $x\in[\![1,N-1]\!]$. This follows from the fact that the matrix $M\widetilde{L}$ is symmetric, where $M=\mathrm{diag}(m)$.

□

Acknowledgments:

We thank Jim Fill for telling us about his earlier works [10, 11] on this problem. He has proved similar results in continuous and discrete time and he describes precisely the sample-path construction of dual processes which enables direct extraction of the exponential variables from X and extra randomness. We thank Susan Holmes and Laurent Saloff-Coste. We are grateful to the ANR's Chaire d'Excellence program for providing financial support and the Laboratoire J. A. Dieudonné (UMR 6621) in Nice, where this work was done, for its hospitality. Finally we are endebted to the referee for his/her very careful reading which helped us to improve this paper.

References

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

[2] David Aldous and Persi Diaconis. Strong uniform times and finite random walks. *Adv. in Appl. Math.*, 8(1):69–97, 1987.

[3] David Aldous and James Allen Fill. Reversible Markov chains and random walks on graphs. Monograph in preparation, available on the web site: `http://www.stat.berkeley.edu/~aldous/RWG/book.html`, 1994-2002.

[4] Bovier, Anton and Eckhoff, Michael and Gayrard, Véronique and Klein, Markus. Metastability and low lying spectra in reversible Markov chains. *Comm. Math. Phys.*, 2285(2):219–255, 2002.

[5] Philippe Carmona, Frédérique Petit, and Marc Yor. Beta-gamma random variables and intertwining relations between certain Markov processes. *Rev. Mat. Iberoamericana*, 14(2):311–367, 1998.

[6] Persi Diaconis. *Group representations in probability and statistics*. Institute of Mathematical Statistics Lecture Notes—Monograph Series, 11. Institute of Mathematical Statistics, Hayward, CA, 1988.

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

[8] Persi Diaconis and Laurent Saloff-Coste. Separation cut-offs for birth and death chains. *Ann. Appl. Probab.*, 16(4):2098–2122, 2006.

[9] James Allen Fill. Strong stationary duality for continuous-time Markov chains. I. Theory. *J. Theoret. Probab.*, 5(1):45–70, 1992.

[10] James Allen Fill. The passage time distribution for a birth-and-death chain: Strong stationary duality gives a first stochastic proof. preprint, 2007.

[11] James Allen Fill. On hitting times and fastest strong stationary times for skip-free processes. Preprint, 2007.

[12] Mark Kac. Random walk and the theory of Brownian motion. *Amer. Math. Monthly*, 54:369–391, 1947.

[13] Samuel Karlin and James McGregor. Coincidence properties of birth and death processes. *Pacific J. Math.*, 9:1109–1140, 1959.

[14] Julian Keilson. Log-concavity and log-convexity in passage time densities of diffusion and birth-death processes. *J. Appl. Probability*, 8:391–398, 1971.

[15] Julian Keilson. *Markov chain models—rarity and exponentiality*, volume 28 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1979.

[16] John T. Kent. The spectral decomposition of a diffusion hitting time. *Ann. Probab.*, 10(1):207–219, 1982.

[17] John T. Kent. The appearance of a multivariate exponential distribution in sojourn times for birth-death and diffusion processes. In *Probability, statistics and analysis*, volume 79 of *London Math. Soc. Lecture Note Ser.*, pages 161–179. Cambridge Univ. Press, Cambridge, 1983.

[18] László Lovász and Peter Winkler. Mixing times. In *Microsurveys in discrete probability (Princeton, NJ, 1997)*, volume 41 of *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.*, pages 85–133. Amer. Math. Soc., Providence, RI, 1998.

[19] Peter Matthews. Strong stationary times and eigenvalues. *J. Appl. Probab.*, 29(1):228–233, 1992.

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

[21] Laurent Miclo. On eigenfunctions of Markov processes on trees. Preprint, available on `http://hal.ccsd.cnrs.fr/ccsd-00147865`, 2007.

[22] Laurent Miclo. Comportement de spectres d'opérateurs de Schrödinger à basse température. *Bull. Sci. Math.*, 119(6):529–553, 1995.

[23] Laurent Miclo. Sur les problèmes de sortie discrets inhomogènes. *Ann. Appl. Probab.*, 6(4):1112–1156, 1996.

[24] Igor Pak. Random walks on groups: strong uniform time approach. Ph.D. thesis, Dept. of Mathematics, Harvard University, available on `http://math.mit.edu/ pak/research.html`, 1997.

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

[26] Eugene Seneta. Quasi-stationary distributions and time-reversion in genetics. (With discussion). *J. Roy. Statist. Soc. Ser. B*, 28:253–277, 1966.

miclo@latp.univ-mrs.fr
Laboratoire d'Analyse, Topologie, Probabilités
Centre de Mathématiques et Informatique
Université de Provence
39, rue Frédéric Joliot-Curie
13453 Marseille cedex 13, France