

Duality and Intertwining of Markov Chains

J.M. Swart

ALEA in Europe School, October 2013,
CIRM, Luminy (Marseille).

Contents

1	Markov chains	5
1.1	Discrete time	5
1.2	Continuous time	10
2	Duality	15
2.1	Basic facts	15
2.2	Pathwise duality	17
2.3	Monotone systems duality	19
2.4	Additive systems duality	25
2.5	Linear systems duality	32
2.6	Cancellative systems duality	34
2.7	Lloyd-Sudbury theory	37
2.8	Concluding remarks	46
2.9	Exercises	47
3	Intertwining	53
3.1	Markov functionals	53
3.2	Intertwining of Markov processes	56
3.3	Thinning	59
3.4	Concluding remarks	62
3.5	Exercises	63

Chapter 1

Markov chains

1.1 Discrete time

Let S be a finite set. By definition, a *probability kernel* on S is a function $P : S \times S \rightarrow [0, 1]$ such that $\sum_y P(x, y) = 1$ for all $x \in S$. Let \mathbb{R}^S denote the space of all functions $f : S \rightarrow \mathbb{R}$. We write

$$Pf(x) := \sum_y P(x, y)f(y) \quad (x \in S),$$

i.e., we associate a probability kernel P with the linear map $P : \mathbb{R}^S \rightarrow \mathbb{R}^S$ such that $(P(x, y))_{x, y \in S}$ is its matrix.

By definition, a *Markov chain* with *transition kernel* P is an S -valued stochastic process $X = (X_k)_{k \geq 0}$ such that

$$\mathbb{E}[f(X_{k+1}) \mid (X_0, \dots, X_k)] = Pf(X_k) \quad \text{a.s.} \quad (k \geq 0, f \in \mathbb{R}^S). \quad (1.1)$$

Lemma 1.1 (Markov chain) *An S -valued stochastic process $X = (X_k)_{k \geq 0}$ is a Markov chain with transition kernel P if and only if*

$$\mathbb{P}[(X_0, \dots, X_n) = (x_0, \dots, x_n)] = \mathbb{P}[X_0 = x_0]P(x_0, x_1) \cdots P(x_{n-1}, x_n) \quad (1.2)$$

for all $n \geq 0$, $x_0, \dots, x_n \in S$.

Proof We first prove by induction that (1.1) implies (1.2). The statement is

certainly true for $n = 0$ so assume that it holds for n . Then (1.1) implies that

$$\begin{aligned}
& \mathbb{P}[(X_0, \dots, X_{n+1}) = (x_0, \dots, x_{n+1})] \\
&= \mathbb{E}[\mathbb{E}[1_{\{(X_0, \dots, X_n) = (x_0, \dots, x_n)\}} 1_{\{X_{n+1} = x_{n+1}\}} \mid (X_0, \dots, X_n)]] \\
&= \mathbb{E}[1_{\{(X_0, \dots, X_n) = (x_0, \dots, x_n)\}} \mathbb{E}[1_{\{X_{n+1} = x_{n+1}\}} \mid (X_0, \dots, X_n)]] \\
&= \mathbb{E}[1_{\{(X_0, \dots, X_n) = (x_0, \dots, x_n)\}} P(X_n, x_{n+1})] \\
&= \mathbb{P}[X_0 = x_0] P(x_0, x_1) \cdots P(x_{n-1}, x_n),
\end{aligned}$$

where in the last step we have used the induction hypothesis.

To check that conversely (1.2) implies (1.1), by the definition of the conditional expectation, we must check that $Pf(X_k)$ is measurable with respect to the σ -field $\sigma(X_0, \dots, X_k)$ generated by the random variables X_0, \dots, X_k , and moreover, for any $A \in \sigma(X_0, \dots, X_k)$,

$$\mathbb{E}[1_A Pf(X_k)] = \mathbb{E}[1_A f(X_{k+1})].$$

By linearity, it suffices to check this for events of the form $A = \{(X_0, \dots, X_k) = (x_0, \dots, x_k)\}$ and functions $f(x) = 1_{\{X_{k+1} = x_{k+1}\}}$, which we leave to the reader. ■

Lemma 1.1 shows that the law of a Markov chain $X = (X_k)_{k \geq 0}$ is uniquely determined by its *initial law* $\mu(x) := \mathbb{P}[X_0 = x]$ and its transition kernel. We write $\mathbb{P}^\mu, \mathbb{E}^\mu$ to indicate that we are considering the Markov chain with initial law μ and more specifically $\mathbb{P}^x := \mathbb{P}^{\delta_x}, \mathbb{E}^x := \mathbb{E}^{\delta_x}$ for the Markov chain with initial state $X_0 = x$. (Here $\delta_x(y) := 1_{\{x=y\}}$ is the Dirac measure at x .) We see from (1.2) that

$$\mathbb{P}^x[X_k = y] = \sum_{x_1} \cdots \sum_{x_{k-1}} P(x, x_1) \cdots P(x_{k-1}, y) = P^k(x, y), \quad (1.3)$$

where P^k denotes the k -th power of P . For any function $f \in \mathbb{R}^S$ and probability law μ on S , let us set

$$\mu f := \sum_x \mu(x) f(x),$$

i.e., we associate μ with a linear form $\mu : \mathbb{R}^S \rightarrow \mathbb{R}$. Then μP^k is the concatenation of a linear map $P^k : \mathbb{R}^S \rightarrow \mathbb{R}^S$ and a linear form $\mu : \mathbb{R}^S \rightarrow \mathbb{R}$, which yields another linear form $(\mu P^k) : \mathbb{R}^S \rightarrow \mathbb{R}$, which by (1.3) corresponds to

$$(\mu P^k)(y) = \mathbb{P}^\mu[X_k = y],$$

i.e., if the initial law is μ , then μP^k is the law of the Markov chain at time k . Similarly

$$P^k f(x) = \mathbb{E}^x[f(X_k)]$$

gives the expectation of f at time k as a function of the initial state.

Let $\mu_k(x) := \mathbb{P}^\mu[X_k = x]$ be the law at time k of the Markov chain with initial law μ . Then $(\mu_k)_{k \geq 0}$ solves the *forward equations*

$$\mu_{k+1} = \mu_k P = P^\dagger \mu_k,$$

where we multiply the vector¹ μ from the right with the matrix P , or equivalently from the left with the *hermitian conjugate*

$$P^\dagger(x, y) := P(y, x) \quad (x, y \in S)$$

of P . Likewise, if $f_k(x) := \mathbb{E}^x[f(X_k)]$ denotes the expectation of f at time k as a function of the initial state, then $(f_k)_{k \geq 0}$ solves the *backward equations*

$$f_{k+1} = P f_k,$$

where this time the multiplication is from the left.

A fixed point of the forward equations, i.e., a probability measure π such that

$$\pi = \pi P \tag{1.4}$$

is called an *invariant law*. If $(X_k)_{k \geq 0}$ is started in an invariant law π , then $\mathbb{P}^\pi[X_k = x] = \pi(x)$ for all k and in fact $(X_k)_{k \geq 0}$ is a stationary process. A fixed point of the backward equations, i.e., a function h such that

$$h = P h \tag{1.5}$$

is called a *harmonic function*. Harmonic functions are characterized by the property that $(h(X_k))_{k \geq 0}$ is a *martingale*. A typical example of a harmonic function is the following. Let $S' \subset S$ be a set such that $P(x, y) = 0$ for all $x \in S'$ and $y \in S \setminus S'$, i.e., once the process enters S' , no escape is possible. Then the *trapping probability*

$$h(x) := \mathbb{P}^x[\exists k \geq 0 \text{ s.t. } X_k \in S'] \quad (x \in S)$$

is a harmonic function. In our present set-up, where S is finite, it can even be proved that all harmonic functions are linear combinations of functions of this form.

Every probability kernel P on a finite set has at least one invariant law and at least one harmonic function; the last statement is trivial, since the constant function 1

¹Throughout these notes, we do not distinguish between row and column vectors.

is always harmonic. It can be shown that one has uniqueness of the invariant law provided P is *irreducible*, which means that

$$\forall x, y \in S \exists k \geq 0 \text{ s.t. } P^k(x, y) > 0. \quad (1.6)$$

In this case moreover $\pi(x) > 0$ for all $x \in S$. Irreducibility also implies that all harmonic functions are constant. If P is irreducible, then the greatest common divisor of $\{k \geq 1 : P^k(x, x) > 0\}$ does not depend on the choice of the point $x \in S$; P is called *aperiodic* if this number is one. If P is irreducible and aperiodic, then it is *ergodic* in the sense that for any initial law μ , one has

$$\mu P^k \xrightarrow[k \rightarrow \infty]{} \pi,$$

where π is the unique invariant law. From this, it is easy to see that also

$$P^k f \xrightarrow[k \rightarrow \infty]{} (\pi f)1,$$

i.e., $P^k f$ converges to a constant multiple of the constant function 1.

A probability law π is *reversible* if it satisfies the *detailed balance* equations

$$\pi(x)P(x, y) = \pi(y)P(y, x) \quad (x, y \in S). \quad (1.7)$$

One can show that this is equivalent to the statement that π is an invariant law and moreover

$$\mathbb{P}^\pi[(X_0, \dots, X_n) = (x_0, \dots, x_n)] = \mathbb{P}^\pi[(X_0, \dots, X_n) = (x_n, \dots, x_0)],$$

i.e., the stationary process with one-dimensional law π is symmetric with respect to *time reversal*. Reversibility is a ‘nice’ property; for an irreducible probability kernel, it will ‘typically’ not be the case that the unique invariant law is reversible. We may define an inner product (or pseudo-inner product, if π is not everywhere positive) on \mathbb{R}^S by

$$\langle f|g \rangle_\pi := \sum_x \pi(x) f(x) g(x). \quad (1.8)$$

Then reversibility implies that P is self-adjoint with respect to this inner product:

$$\langle f|Pg \rangle_\pi = \sum_{x,y} \pi(x) f(x) P(x, y) g(y) = \sum_{x,y} \pi(y) P(y, x) f(x) g(y) = \langle Pf|g \rangle_\pi. \quad (1.9)$$

In particular, if π is the uniform distribution, then this says that $P = P^\dagger$ but in general we need to distinguish the hermitian conjugate of a matrix from its

adjoint with respect to the (pseudo) inner product $\langle \cdot | \cdot \rangle_\pi$. By inserting $f = 1_{\{x'\}}$ and $g = 1_{\{y'\}}$ into this equation we see that (1.7) is in fact equivalent to (1.9).² Note that from (1.9), it is immediately clear that if π is reversible for P , then it is also reversible for any power P^n of P . For reversible Markov chains, there is a symmetry between the forward and backward equations. Let π be a reversible law and let $f : S \rightarrow [0, \infty)$ be a function such that $\pi f = 1$. Let f_k be a solution of the backward equations with $f_0 = f$, i.e., $f_k = P^k f$. Define measures μ_k by

$$\mu_k(x) := f_k(x)\pi(x) \quad (k \geq 0),$$

i.e., μ_k is given by the density f_k with respect to π . Then we claim that the μ_k are in fact probability measures and that they solve the forward equations, i.e.,

$$\mu_k = \mu_0 P^k.$$

Indeed, this follows by writing

$$\mu_0 P^k(x) = \sum_y f(y)\pi(y)P^k(y, x) = \sum_y f(y)\pi(x)P^k(x, y) = f_k(x)\pi(x),$$

where we have used that P^k satisfies (1.7).

Markov chains describe systems with a random dynamics, where the state at time $k + 1$ depends in a random way on the state at time k , using ‘fresh’ randomness in each time step. In practise, this often takes the following form. Imagine that one is given an i.i.d. collection of random variables $(Z_k)_{k \geq 1}$ taking values in some measurable space (E, \mathcal{E}) , with common law μ , and that $\phi : S \times E \rightarrow S$ is a measurable function such that

$$P(x, y) = \mathbb{P}[\phi(x, Z_1) = y] \quad (x, y \in S). \quad (1.10)$$

Let X_0 be an S -valued random variable, independent of the $(Z_k)_{k \geq 1}$. Then the inductive formula

$$X_k := \phi(X_{k-1}, Z_k) \quad (k \geq 1)$$

defines a Markov chain $X = (X_k)_{k \geq 0}$ with transition kernel P . A probability space (E, \mathcal{E}, μ) together with a map $\phi : S \times E \rightarrow S$ such that (1.10) holds is called a *random mapping representation* of the probability kernel P . Each probability kernel on a finite set S has a random mapping representation, which is far from

²If $\pi(x) > 0$ for all x , then the adjoint of P w.r.t. the inner product $\langle \cdot | \cdot \rangle_\pi$ is given by $P^*(x, y) = \pi(y)P(y, x)\pi(x)^{-1}$. If π is not reversible, then $P^* \neq P$ is the transition kernel of the reversed chain.

unique. Note that simulating a Markov chain on a computer usually involves a random mapping representation, e.g., in code such as

```

if rand < 0.3
  X = X + 1
else
  X = X - 1
end

```

which applies the function $\phi(x, z) := x + 1_{\{z < 0.3\}} - 1_{\{z \geq 0.3\}}$ to uniformly distributed random variables $(Z_k)_{k \geq 1}$ to generate a Markov chain X that jumps from x to $x+1$ with probability 0.3 and to $x-1$ with the remaining probability. Random mapping representations also provide an important way of *coupling* Markov processes with different initial states.

1.2 Continuous time

Let S be a finite set. By definition, a *Markov semigroup* (of a Markov process in S) is a collection $(P_t)_{t \geq 0}$ of probability kernels on S such that

$$P_s P_t = P_{s+t} \quad (s, t \geq 0) \quad \text{and} \quad \lim_{t \downarrow 0} P_t f = P_0 f = f \quad (f \in \mathbb{R}^S).$$

One can show that each such Markov semigroup is of the form

$$P_t = e^{tG} = \sum_{n=0}^{\infty} \frac{1}{n!} t^n G^n, \quad (1.11)$$

where the *generator* G satisfies

$$G(x, y) \geq 0 \quad \forall x \neq y \quad \text{and} \quad \sum_y G(x, y) = 0.$$

Conversely, each matrix G of this form defines a Markov semigroup. By definition, a *Markov process* with semigroup $(P_t)_{t \geq 0}$ is an S -valued stochastic process $X = (X_t)_{t \geq 0}$ with cadlag (i.e., right continuous with left limits) sample paths such that (compare (1.1))

$$\mathbb{E}[f(X_u) \mid (X_s)_{0 \leq s \leq t}] = P_{u-t} f(X_t) \quad \text{a.s.} \quad (0 \leq t \leq u, f \in \mathbb{R}^S). \quad (1.12)$$

One can show that such a Markov process exists, and is unique in distribution, for each Markov semigroup $(P_t)_{t \geq 0}$ and initial law $\mathbb{P}[X_0 \in \cdot]$. If X is such a Markov

process, then for each $\varepsilon > 0$, the random variables $(X_{\varepsilon k})_{k \geq 0}$ form a Markov chain with transition kernel P_ε , which by (1.11) satisfies

$$P_\varepsilon(x, y) = 1_{\{x=y\}} + \varepsilon G(x, y) + O(\varepsilon^2) \quad \text{as } \varepsilon \downarrow 0. \quad (1.13)$$

Note that the condition $\sum_y G(x, y) = 0$ ensures that $\sum_y P_\varepsilon(x, y) = 1 + O(\varepsilon^2)$ as it should be for a probability kernel. Because of this condition, to specify the generator (and hence the semigroup), it suffices to specify the nonnegative numbers

$$(G(x, y))_{x \neq y}.$$

Then (1.13) says that for each $x \neq y$, the Markov chain $(X_{\varepsilon k})_{k \geq 0}$ jumps from x to y with probability $\varepsilon G(x, y) + O(\varepsilon^2)$ and stays at x with the remaining probability $1 - \varepsilon \sum_{y \neq x} G(x, y)$. It can be shown that if P_ε is any probability kernel on S which depends on a parameter ε in such a way that (1.13) holds, then the associated Markov chain approximates the continuous-time Markov process $(X_t)_{t \geq 0}$ after a rescaling of time by a factor ε . Usually, this gives the right intuition for thinking about continuous-time Markov processes with finite state space S . For $x \neq y$, we say that the process $(X_t)_{t \geq 0}$ jumps from x to y with *rate* $G(x, y)$. Because of the similarity with discrete time, continuous-time Markov processes with finite state space are often called *continuous-time Markov chains*.

Much of the theory of Markov chains $(X_k)_{k \geq 0}$ with finite state space now generalizes in a straightforward manner to continuous-time Markov processes $(X_t)_{t \geq 0}$ with finite state space. If μ is a probability measure on S , then $\mu_t := \mu P_t$ is the law of the process at time t , which satisfies the *forward equations*

$$\frac{\partial}{\partial t} \mu_t = \mu_t G = G^\dagger \mu_t.$$

Likewise, for any $f \in \mathbb{R}^S$, the function $f_t := P_t f$ satisfies the *backward equations*

$$\frac{\partial}{\partial t} f_t = G f_t.$$

Both forward and backward equations can be summarized in the formula

$$\frac{\partial}{\partial t} P_t = G P_t = P_t G. \quad (1.14)$$

An invariant law is a fixed point of the forward equations, i.e., a probability measure π such that $\pi = \pi P_t$ ($t \geq 0$) or equivalently $\pi G = 0$, and a harmonic function is a fixed point of the backward equations, i.e., a function h such that $Gh = 0$. The probability kernels P_t are either irreducible for no $t > 0$ or for all $t > 0$; the latter happens if and only if for each $x, y \in S$ there exist $x = x_0, \dots, x_n = y$ such that

$G(x_{k-1}, x_k) > 0$ for all $k = 1, \dots, n$. In this case we also say that G is irreducible or, informally, that the Markov process $(X_t)_{t \geq 0}$ is irreducible. If a continuous-time Markov chain is irreducible, then the P_t are always aperiodic for all $t > 0$ and in fact $P_t(x, y) > 0$ for all $t > 0$ and $x, y \in S$. A probability law π is *reversible* if the *detailed balance* equations (1.7) hold for P_t for all $t > 0$, or equivalently, if

$$\pi(x)G(x, y) = \pi(y)G(y, x) \quad (x, y \in S, x \neq y), \quad (1.15)$$

which is equivalent to saying that G is self-adjoint w.r.t. the inner product $\langle \cdot | \cdot \rangle_\pi$.

There is also a sort of equivalent of the random mapping representation for continuous-time Markov chains. Let \mathcal{M} be a finite set whose elements are maps $m : S \rightarrow S$, and let $(r_m)_{m \in \mathcal{M}}$ be nonnegative constants. Let Δ be a Poisson point subset of $\mathcal{M} \times \mathbb{R} = \{(m, t) : m \in \mathcal{M}, t \in \mathbb{R}\}$ with intensity $r_m dt$, where dt denotes Lebesgue measure, and for $s \leq t$, set $\Delta_{s,t} := \Delta \cap (\mathcal{M} \times (s, t])$. Define random maps $\Phi_{s,t} : S \rightarrow S$ ($s \leq t$) by

$$\begin{aligned} \Phi_{s,t}(x) &:= m_n \circ \dots \circ m_1(x) \\ \text{where } \Delta_{s,t} &:= \{(m_1, t_1), \dots, (m_n, t_n)\}, \quad t_1 < \dots < t_n, \end{aligned} \quad (1.16)$$

with the convention that $\Phi_{s,t}(x) = x$ if $\Delta_{s,t} = \emptyset$.

Proposition 1.2 (Poisson construction of Markov process) *Let X_0 be an S -valued random variable, independent of Δ . Then*

$$X_t := \Phi_{0,t}(X_0) \quad (t \geq 0) \quad (1.17)$$

defines a Markov process $(X_t)_{t \geq 0}$ with generator

$$Gf(x) = \sum_{m \in \mathcal{M}} r_m (f(m(x)) - f(x)) \quad (x \in S, f \in \mathbb{R}^S). \quad (1.18)$$

We call (1.18) a *random mapping representation* for the generator G . It is not hard to see that each Markov generator can be written (usually in many different ways) in the form (1.18). We may order the elements of $\Delta_{0,\infty} := \Delta \cap (\mathcal{M} \times (0, \infty))$ as

$$\Delta_{0,\infty} = \{(m_k, \tau_k) : k \geq 1\} \quad \text{with } 0 < \tau_1 < \tau_2 < \dots$$

Then $\{\tau_1, \tau_2, \dots\}$ is a Poisson point set on $[0, \infty)$ with intensity $R := \sum_{m \in \mathcal{M}} r_m$ and hence $(\tau_k - \tau_{k-1})_{k \geq 1}$ (with $\tau_0 := 0$) are i.i.d. exponentially distributed random variables with mean R^{-1} . Conditional on the times τ_1, τ_2, \dots , the random variables $(m_k)_{k \geq 1}$ are i.i.d. with common law

$$\mathbb{P}[m_k = m] = R^{-1} r_m \quad (m \in \mathcal{M}).$$

Thus, the evolution of X may be described as follows: wait an exponential time with mean R^{-1} , and then apply the map m with probability $R^{-1}r_m$. Note that τ_1, τ_2, \dots are not necessarily the times when the process jumps, i.e., it may happen that $X_{\tau_k} = m_k(X_{\tau_k-}) = X_{\tau_k-}$ for some k .

Proof of Proposition 1.2 It is easy to see from the definition that X has cadlag sample paths. We set

$$P_t(x, y) := \mathbb{P}[\Phi_{0,t}(x) = y] \quad (x, y \in S, t \geq 0).$$

Let \mathcal{G}_t be the σ -field generated by the random variables X_0 and $\Delta_{0,t}$. Fix $0 \leq s \leq t$. Since X_0 is independent of Δ and since Δ is a Poisson point process, we see that $X_0, \Delta_{0,s}$ and $\Delta_{s,t}$ are independent. Since $\Delta_{s,t}$ is up to a time shift equally distributed with $\Delta_{0,t-s}$, it follows that

$$\mathbb{P}[X_t \in \cdot \mid \mathcal{G}_s] = \mathbb{P}[\Phi_{s,t}(X_s) \in \cdot \mid X_0, \Delta_{0,s}] = P_{t-s}(X_s, \cdot).$$

Since $(X_s)_{0 \leq s \leq t}$ is a function of X_0 and $\Delta_{0,t}$, we have $\mathcal{F}_t \subset \mathcal{G}_t$, so it follows that for any $f \in \mathbb{R}^S$,

$$\mathbb{E}[f(X_t) \mid \mathcal{F}_s] = \mathbb{E}[\mathbb{E}[f(X_t) \mid \mathcal{G}_s] \mid \mathcal{F}_s] = \mathbb{E}[P_{t-s}f(X_s) \mid \mathcal{F}_s] = P_{t-s}f(X_s).$$

To finish the proof, we must show that $(P_t)_{t \geq 0}$ is a Markov semigroup with generator G given by (1.18). The fact that $\lim_{t \downarrow 0} P_t(x, y) = P_0(x, y) = \delta_x(y)$ follows from the fact that $\mathbb{P}[\Delta_{0,t} = \emptyset] \rightarrow 1$ as $t \downarrow 0$. To see that $P_s P_t = P_{s+t}$, let X^x be the process started in $X_0 = x$. By what we have already proved,

$$\begin{aligned} P_{s+t}f(x) &= \mathbb{E}[f(X_{s+t}^x)] = \mathbb{E}[\mathbb{E}[f(X_{s+t}^x) \mid \mathcal{F}_s]] \\ &= \mathbb{E}[P_t f(X_s^x)] = P_s P_t f(X_0^x) = P_s P_t f(x). \end{aligned}$$

To see that the generator G of $(P_t)_{t \geq 0}$ is given by (1.18), we observe that

$$P_t f(x) = \mathbb{E}[f(X_t^x)] = f(x) + t \sum_{m \in \mathcal{M}} r_m (f(m(x)) - f(x)) + O(t^2) \quad \text{as } t \downarrow 0,$$

which follows from the fact that $\mathbb{P}[|\Delta_{0,t}| \geq 2] = O(t^2)$ while

$$\mathbb{P}[\Delta_{0,t} = \{(m, s)\} \text{ for some } s \in (0, t)] = tr_m + O(t^2).$$

■

Chapter 2

Duality

2.1 Basic facts

Duality is a widely used, and in general not precisely defined concept in mathematics that usually involves two objects plus a relation between them that is symmetric in the sense that one can get from one object to the other ‘in the same way’ as from the other object back to the first one. Examples are dual linear spaces (which in the infinite-dimensional case are not always so nicely symmetric, though) and planar graph duality. We cite the general definition of Markov process duality that we are about to give from Liggett [Lig85, Def. II.3.1]. He may indeed have been the first to formulate the concept in this generality, though the topic is much older. Up to the present day, the term duality (of Markov processes) is sometimes used in meanings that do not fit the definition below. (For example for intertwining relations, which will be discussed in the next chapter.) On the other hand, some authors also use the term ‘dual Markov process’ in a more restricted meaning than we will (usually concerning a specific duality function) and use terms such as ‘quasi-dual’ for the more general concept.

Let $X = (X_t)_{t \geq 0}$ be a continuous-time Markov chain with finite state space S , generator G , and semigroup $(P_t)_{t \geq 0}$. Likewise, let $Y = (Y_t)_{t \geq 0}$ be a continuous-time Markov chain with finite state space R , generator H , and semigroup $(Q_t)_{t \geq 0}$. Finally, let $\psi : S \times R \rightarrow \mathbb{R}$ be a function. By definition, X and Y are *dual* to each other with *duality function* ψ if

$$\mathbb{E}^x[\psi(X_t, y)] = \mathbb{E}^y[\psi(x, Y_t)] \quad (x \in S, y \in R, t \geq 0). \quad (2.1)$$

More generally, we obtain from this that if X and Y are independent with (possibly) random initial states, then

$$\mathbb{E}[\psi(X_s, Y_{t-s})] \quad \text{does not depend on } s \in [0, t]. \quad (2.2)$$

To see this,¹ note that in terms of semigroups, formula (2.1) says that

$$\sum_{x' \in S} P_t(x, x') \psi(x', y) = \sum_{y' \in R} Q_t(y, y') \psi(x, y'). \quad (2.3)$$

Using this, we see that for any $0 \leq s \leq s' \leq t$,

$$\begin{aligned} \mathbb{E}[\psi(X_s, Y_{t-s})] &= \sum_{x, x', y, y'} \mathbb{P}[X_0 = x] P_s(x, x') \mathbb{P}[Y_0 = y] Q_{t-s}(y, y') \psi(x', y') \\ &= \sum_{x, x', y, y'} \mathbb{P}[X_0 = x] P_s(x, x') \mathbb{P}[Y_0 = y] Q_{t-s'}(y, y') \sum_{y''} Q_{s'-s}(y', y'') \psi(x', y'') \\ &= \sum_{x, x', y, y'} \mathbb{P}[X_0 = x] P_s(x, x') \mathbb{P}[Y_0 = y] Q_{t-s'}(y, y') \sum_{x''} P_{s'-s}(x', x'') \psi(x'', y'), \end{aligned}$$

which by the same reasoning backwards equals $\mathbb{E}[\psi(X_{s'}, Y_{t-s'})]$.

Lemma 2.1 (Generator characterization) *Two Markov processes X, Y with finite state spaces S, R and generators G, H are dual with duality function ψ if and only if*

$$G\psi(\cdot, y)(x) = H\psi(x, \cdot)(y) \quad (x \in S, y \in R). \quad (2.4)$$

Proof We may define a linear operator $\psi : \mathbb{R}^R \rightarrow \mathbb{R}^S$ by

$$\psi f(x) := \sum_{y \in R} \psi(x, y) f(y) \quad (x \in S). \quad (2.5)$$

Then (2.3) can more succinctly be written as

$$P_t \psi = \psi Q_t^\dagger \quad (t \geq 0). \quad (2.6)$$

Since $P_t = 1 + tG + O(t^2)$ and likewise $Q_t = 1 + tH + O(t^2)$ as $t \downarrow 0$, a necessary condition for this is that

$$G\psi = \psi H^\dagger, \quad (2.7)$$

which is just another way of writing (2.4). To see that this condition is also sufficient, we calculate, using (1.14),

$$\begin{aligned} \frac{\partial}{\partial s} (P_s \psi Q_{t-s}^\dagger) &= \left(\frac{\partial}{\partial s} P_s \right) \psi Q_{t-s}^\dagger + P_s \psi \left(\frac{\partial}{\partial s} Q_{t-s} \right)^\dagger \\ &= P_s G \psi Q_{t-s}^\dagger + P_s \psi (-Q_{t-s} H)^\dagger = P_s (G\psi - \psi H^\dagger) Q_{t-s}^\dagger = 0, \end{aligned}$$

¹It is in fact sufficient if in the expression $\mathbb{E}[\psi(X_s, Y_{t-s})]$, the random variables X_s and Y_{t-s} are independent, which is weaker than the statement that $X = (X_t)_{t \geq 0}$ and $Y = (Y_t)_{t \geq 0}$ are independent as processes.

which shows that $P_s \psi Q_{t-s}^\dagger$ does not depend on $s \in [0, t]$. ■

In particular, if $\psi : \mathbb{R}^R \rightarrow \mathbb{R}^S$ is invertible, then (2.6) says that

$$P_t = \psi Q_t^\dagger \psi^{-1} \quad (t \geq 0), \quad (2.8)$$

so a duality relates the *backward* evolution of one Markov process to the *forward* evolution of the other process (and vice versa, as can be seen by taking adjoints of this equation). In particular, we have that

$$\pi \text{ invariant law for } Y \quad \text{implies} \quad \psi \pi \text{ harmonic function for } X, \quad (2.9)$$

as can be seen by writing

$$P_t \psi \pi = \psi Q_t^\dagger \pi = \psi \pi \quad (t \geq 0).$$

(Note this argument does not need ψ to be invertible.) Similarly, if h is a harmonic function for the Markov process Y , then $\psi^\dagger h$ is a fixed point of P_t^\dagger , as can be seen by writing

$$P_t^\dagger \psi^\dagger h = (\psi P_t)^\dagger h = (Q_t^\dagger \psi)^\dagger h = \psi^\dagger Q_t h = \psi^\dagger h.$$

In general, $\psi^\dagger h$ does not need to be nonnegative, however, so it may not correspond to an invariant measure. (If $\psi^\dagger h$ is nonnegative, then by the finiteness of the state space, it can of course be normalized to a probability measure.)

2.2 Pathwise duality

Although a priori, Markov process duality is only a statement about expectations, in practise, many dualities can be turned into an almost sure relation. Let S and R be finite sets and let $\psi : S \times R \rightarrow \mathbb{R}$ be a function. Let $m : S \rightarrow S$ and $\hat{m} : R \rightarrow R$ be maps. Then we say that \hat{m} is *dual* to m with respect to the duality function ψ if

$$\psi(m(x), y) = \psi(x, \hat{m}(y)) \quad (x \in S, y \in R). \quad (2.10)$$

Now let X be a continuous-time Markov chain with finite state space S , generator G , and semigroup $(P_t)_{t \geq 0}$. Let us assume that we are given a random mapping representation of G as in (1.18), i.e.,

$$Gf(x) = \sum_{m \in \mathcal{M}} r_m (f(m(x)) - f(x)) \quad (x \in S, f \in \mathbb{R}^S),$$

where \mathcal{M} is a finite collection of maps $m : S \rightarrow S$ and $(r_m)_{m \in \mathcal{M}}$ is a collection of nonnegative rates. Assume that each map $m \in \mathcal{M}$ has a dual map $\hat{m} : R \rightarrow R$

with respect to the duality function ψ . Let $\hat{\mathcal{M}} := \{\hat{m} : m \in \mathcal{M}\}$ and define a Markov generator H by

$$Hf(y) := \sum_{\hat{m} \in \hat{\mathcal{M}}} r_{\hat{m}}(f(\hat{m}(y)) - f(y)) \quad (y \in S, f \in \mathbb{R}^R),$$

with $r_{\hat{m}} := r_m$ ($m \in \mathcal{M}$). Then we have the following elementary result.

Lemma 2.2 (Pathwise dual) *The processes X and Y are dual to each other with duality function ψ . Moreover, for each $t \geq 0$, the processes X and Y (with arbitrary initial laws) can be coupled such that for each $s \in [0, t]$, $(X_u)_{0 \leq u \leq s}$ is independent of $(Y_u)_{0 \leq u \leq t-s}$, and moreover, a.s.,*

$$\psi(X_{s-}, Y_{t-s}) \quad \text{does not depend on } s \in [0, t]. \quad (2.11)$$

Proof We will only prove (2.11). The statement about X and Y being dual then follows by taking expectations and setting $s = 0, t$. Let Δ be a Poisson point subset of $\mathcal{M} \times \mathbb{R}$ with local intensity $r_m dt$ and define random maps $\Phi_{s,t}$ with $s \leq t$ as in (1.16). Set $\Delta_{s-,t-} := \Delta \cap (\mathcal{M} \times [s, t))$ and in analogy with (1.16), define dual maps $\hat{\Phi}_{s-,t-}$ ($s \leq t$) by

$$\begin{aligned} \hat{\Phi}_{s-,t-}(x) &:= \hat{m}_1 \circ \cdots \circ \hat{m}_n(x) \\ \text{where } \Delta_{s-,t-} &:= \{(m_1, t_1), \dots, (m_n, t_n)\}, \quad t_1 < \cdots < t_n, \end{aligned} \quad (2.12)$$

with the convention that $\hat{\Phi}_{s-,t-}(x) = x$ if $\Delta_{s-,t-} = \emptyset$. Fix $t \geq 0$ and let X_0, Y_0 be random variables with values in S and R , respectively, independent of each other and of Δ . Then by Proposition 1.2 applied to Δ and the same turned upside down and shifted by t , we see that

$$X_s := \Phi_{0,s}(X_0) \quad \text{and} \quad Y_s := \hat{\Phi}_{(t-s)-,t-}(Y_0) \quad (s \geq 0)$$

are Markov processes with generator G and H and initial states X_0, Y_0 , respectively. For deterministic $s \in [0, t]$, one has $\Delta_{s-,t-} = \Delta_{s,t}$ a.s. and the latter is independent of $\Delta_{0,s}$ by the properties of Poisson point sets. Thus, since $(X_u)_{0 \leq u \leq s}$ and $(Y_u)_{0 \leq u \leq t-s}$ depend only on X_0 and $\Delta_{0,s}$ resp. Y_0 and $\Delta_{s-,t-}$, we see that these processes are independent.

Write

$$\Delta_{0,t} = \{(m_1, t_1), \dots, (m_n, t_n)\} \quad \text{with} \quad t_1 < \cdots < t_n.$$

and let s, s' be such that

$$t_{k-1} < s \leq t_k < s' \leq t_{k+1}.$$

Then, by (2.10),

$$\begin{aligned}\psi(X_s, Y_{t-s}) &= \psi(m_{k-1} \circ \cdots \circ m_1(X_0), \hat{m}_k \circ \cdots \circ \hat{m}_n(Y_0)) \\ &= \psi(m_k \circ \cdots \circ m_1(X_0), \hat{m}_{k+1} \circ \cdots \circ \hat{m}_n(Y_0)) = \psi(X_{s'}, Y_{t-s'}).\end{aligned}$$

Applying this repeatedly, we see that the expression in (2.11) does not depend on $s \in [0, t]$. \blacksquare

The same proof shows that apart from the expression in (2.11), for the same coupling, also $\psi(X_s, Y_{(t-s)-})$ does not depend on $s \in [0, t]$. Note that for either X or Y , we have to choose a version with caglad (continuous from the left, limits from the right) sample paths, contrary to our usual habit, since otherwise there would be exceptional times where the expression takes a different value.

The term ‘pathwise duality’ was (to the best of my knowledge) first coined (in a more general meaning) in a paper by Sabine Jansen and Noemi Kurt [JK12], although the subject is much older.

2.3 Monotone systems duality

For any finite set S , we let $\mathcal{P}(S)$ denote the set of all subsets of S , which is of course still finite, although, with $2^{|S|}$ elements, it is usually much larger than S . Consider the function $\psi : S \times \mathcal{P}(S) \rightarrow \mathbb{R}$ given by

$$\psi(x, A) := 1_{\{x \in A\}} \quad (x \in S, A \in \mathcal{P}(S)). \quad (2.13)$$

For any map $m : S \rightarrow S$, let $m^{-1} : \mathcal{P}(S) \rightarrow \mathcal{P}(S)$ denote the *inverse image* map, i.e., $m^{-1}(A) := \{x \in S : m(x) \in A\}$. Then m and m^{-1} are dual in the sense of (2.10), i.e.,

$$\psi(m(x), A) = 1_{\{m(x) \in A\}} = 1_{\{x \in m^{-1}(A)\}} = \psi(x, m^{-1}(A)).$$

If X is any Markov process in S with generator G , and we are given any random mapping representation of G as in (1.18), i.e.,

$$Gf(x) = \sum_{m \in \mathcal{M}} r_m(f(m(x)) - f(x)) \quad (x \in S, f \in \mathbb{R}^S),$$

then we can *always* find a pathwise dual to X with the duality function ψ in (2.13), which is the $\mathcal{P}(S)$ -valued Markov process $\mathcal{X} = (\mathcal{X}_t)_{t \geq 0}$ with generator

$$\mathcal{G}f(A) = \sum_{m \in \mathcal{M}} r_m(f(m^{-1}(A)) - f(A)) \quad (A \in \mathcal{P}(S), f \in \mathbb{R}^{\mathcal{P}(S)}). \quad (2.14)$$

In practise, this duality is of little use since the space $\mathcal{P}(S)$ is so large. More interesting pathwise duals can sometimes be found, however, if $\mathcal{P}(S)$ contains interesting subspaces that are a.s. preserved under the dynamics of $(\mathcal{X}_t)_{t \geq 0}$. For monotone systems, one such subspace is formed by the set of all monotone subsets of S , as we explain now.

By definition, a *partial order* over S is a binary relation \leq that satisfies, for any $x, y, z \in S$

- (i) $x \leq x$,
- (ii) $x \leq y$ and $y \leq x$ implies $x = y$,
- (iii) $x \leq y \leq z$ implies $x \leq z$.

A partial order is called a *total order* if

$$x \leq y \text{ or } y \leq x \quad \text{for all } x, y \in S, x \neq y.$$

Finite (nonempty) totally ordered sets are always isomorphic to a finite discrete interval of the form $\{1, \dots, n\}$ with $n \geq 1$. If S, S' are partially ordered sets, then the *product order* on $S \times S'$ is defined by

$$(x, x') \leq (y, y') \quad \text{iff} \quad x \leq y \text{ and } x' \leq y'.$$

A similar definition applies to the cartesian product of more than two partially ordered sets. Even if the original sets are totally ordered, the product order is only a partial order (trivial cases excluded). For example, $(0, 1) \not\leq (1, 0)$ and $(0, 1) \not\geq (1, 0)$.

If S is a partially ordered set and $A \subset S$, then a *minimal element* of A is an element $x \in A$ such that there is no $y \in A$ with $x \neq y$ and $y \leq x$. A subset of a totally ordered set can have at most one minimal element, but the same is not true for subsets of partially ordered sets. Finite subsets always have at least one minimal element.

A map $m : S \rightarrow S'$ from one partially ordered set to another is called *monotone* if

$$x \leq y \quad \text{implies} \quad m(x) \leq m(y).$$

We will say that a subset $A \subset S$ is *increasing* if its indicator function $1_A : S \rightarrow \{0, 1\}$ is monotone, i.e., if $A \ni x \leq y$ implies $y \in A$. It is not hard to see² that a

²OK, here is a proof: Imagine $A \subset S'$ is increasing. Then, for any $m^{-1}(A) \ni x \leq y$ we have $A \ni m(x) \leq m(y)$ and hence $m(y) \in A$, proving that $y \in m^{-1}(A)$ so $m^{-1}(A)$ is increasing. Conversely, if (2.15) holds, then using notation as in (2.17) we have that $m^{-1}(\{m(x)\}^\uparrow) = \{z : m(z) \geq m(x)\}$ is a increasing set for each $x \in S$. In particular, since x lies in this set, any $y \geq x$ must also lie in this set, which implies that $m(y) \geq m(x)$.

map m from one partially ordered set into another is monotone if and only if

$$A \text{ is increasing} \implies m^{-1}(A) \text{ is increasing.} \quad (2.15)$$

Finite increasing sets are uniquely characterized by their minimal elements. Indeed, if A is finite and increasing, then

$$A = (A_{\min})^\uparrow, \quad (2.16)$$

where A_{\min} denotes the set of minimal elements of A and for any set $B \subset S$, let us call³

$$B^\uparrow := \{y \in S : y \geq x \text{ for some } x \in B\} \quad (2.17)$$

the *episet* of B .

Let X be a Markov process with finite state space S as at the start of this section, with a given random mapping representation of its generator G in terms of random maps $m \in \mathcal{M}$ with rates r_m . Assume that S is equipped with a partial order and that the maps $m \in \mathcal{M}$ are all monotone. In this case, the maps $\Phi_{s,t}$ ($s \leq t$) defined in (1.16) are all monotone. From this, we see that the Markov process with generator G has the following property: If X, X' are processes started in initial states $X_0 = x$ and $X'_0 = x'$ that are ordered as $x \leq x'$, then we can couple X and X' such that $X_t \leq X'_t$ for all $t \geq 0$. Markov processes with this property are called *monotone*. Let $\mathcal{P}_{\text{inc}}(S)$ denote the space of all increasing subsets of S . Then by the monotonicity of the maps $m \in \mathcal{M}$ and (2.15), the pathwise dual $\mathcal{X} = (\mathcal{X}_t)_{t \geq 0}$ of X with generator \mathcal{G} as in (2.14) has the property that $\mathcal{X}_0 \in \mathcal{P}_{\text{inc}}(S)$ implies that a.s. $\mathcal{X}_t \in \mathcal{P}_{\text{inc}}(S)$ for all $t \geq 0$.

To find a really useful dual, one usually needs a property that is somewhat stronger than monotonicity. For any finite partially ordered set S , let us define

$$\mathcal{P}_{\text{linc}}(S) := \{A \in \mathcal{P}_{\text{inc}} : A \text{ has a unique minimal element}\}.$$

If the ordering on S is a total order, then $\mathcal{P}_{\text{linc}}(S) = \mathcal{P}_{\text{inc}}(S) \setminus \{\emptyset\}$, while for partially ordered sets, $\mathcal{P}_{\text{linc}}(S)$ is usually much smaller than $\mathcal{P}_{\text{inc}}(S)$. Instead of just assuming (2.15), let us assume that each $m \in \mathcal{M}$ has the stronger⁴ property that

$$A \in \mathcal{P}_{\text{linc}}(S) \implies m^{-1}(A) \in \mathcal{P}_{\text{linc}}(S). \quad (2.18)$$

³This is not standard terminology. I would be grateful if someone could tell me the established lattice theoretic names for this and the hyposets introduced below.

⁴To see that indeed (2.18) implies (2.15), use (2.16) to see that each set in $A \in \mathcal{P}_{\text{linc}}(S)$ is the union of finitely many sets in $\mathcal{P}_{\text{linc}}(S)$, as $A = \bigcup_{x \in A_{\min}} \{x\}^\uparrow$, and use this to conclude from (2.18) that $m^{-1}(A) = \bigcup_{x \in A_{\min}} m^{-1}(\{x\}^\uparrow) \in \mathcal{P}_{\text{inc}}(S)$.

There is a natural bijection between S and $\mathcal{P}_{\text{inc}}(S)$ given by the map

$$S \ni x \mapsto \{x\}^\uparrow \in \mathcal{P}_{\text{inc}}(S).$$

In view of this, for each map $m : S \rightarrow S$ satisfying (2.18), we can define a map $\hat{m} : S \rightarrow S$ by

$$m^{-1}(\{x\}^\uparrow) =: \{\hat{m}(x)\}^\uparrow. \quad (2.19)$$

We define the set $\mathcal{P}_{\text{dec}}(S)$ of *decreasing* sets, a *maximal element* of a set, and the *hyposet* A^\downarrow of a set $A \subset S$ in analogy with $\mathcal{P}_{\text{inc}}(S)$, minimal elements, and A^\uparrow , but with the order \leq replaced by its converse \geq . We also write $\mathcal{P}_{\text{!dec}}(S)$ for the set of all decreasing subsets of S that have a unique maximal element.

Proposition 2.3 (Monotone systems duality) *Let S be a finite set equipped with a partial order and let $m : S \rightarrow S$ satisfy (2.18). Then the map $\hat{m} : S \rightarrow S$ defined in (2.19) satisfies*

$$A \in \mathcal{P}_{\text{!dec}}(S) \quad \text{implies} \quad \hat{m}^{-1}(A) \in \mathcal{P}_{\text{!dec}}(S), \quad (2.20)$$

and \hat{m} is dual to m with respect to the duality function

$$\psi(x, y) := 1_{\{x \geq y\}} \quad (x, y \in S).$$

Moreover, the linear operator $\psi : \mathbb{R}^S \rightarrow \mathbb{R}^S$ associated with the matrix ψ is invertible.

Proof Formula (2.19) says that $\{y : m(y) \geq x\} = \{y : y \geq \hat{m}(x)\}$, i.e.,

$$x \leq m(y) \quad \text{iff} \quad \hat{m}(x) \leq y \quad (x, y \in S). \quad (2.21)$$

It follows that for any $x \in S$

$$\hat{m}^{-1}(\{x\}^\downarrow) = \{y : \hat{m}(y) \leq x\} = \{y : y \leq m(x)\} = \{m(x)\}^\downarrow,$$

which proves that \hat{m}^{-1} maps the space $\mathcal{P}_{\text{!dec}}(S)$ into itself. We also see from (2.21) that

$$\psi(m(x), y) = 1_{\{m(x) \geq y\}} = 1_{\{x \geq \hat{m}(y)\}} = \psi(x, \hat{m}(y)),$$

i.e., \hat{m} is dual to m with respect to the duality function ψ .

To see that $\psi : \mathbb{R}^S \rightarrow \mathbb{R}^S$ is invertible, we must show that the functions $\{\psi(x, \cdot) : x \in S\}$ are linearly independent. Let $(a_x)_{x \in S}$ be real numbers such that

$$\sum_{x \in S} a_x \psi(x, \cdot) = 0.$$

Since S is finite, it contains a maximal element, x_1 , say. The functions $\psi(x, \cdot)$ with $x \neq x_1$ are all zero at x_1 so we must have $a_{x_1} = 0$. But now $S \setminus \{x_1\}$ also has a maximal element, x_2 , say, so continuing the process we see that $a_x = 0$ for all $x \in S$, i.e., the $\{\psi(x, \cdot) : x \in S\}$ are linearly independent. ■

By applying Proposition 2.3 to the converse order, we see that also to each map m with the property that m^{-1} maps $\mathcal{P}_{\text{!dec}}(S)$ into itself there is a dual map, \tilde{m} , say, such that

$$\psi(x, m(y)) = \psi(\tilde{m}(x), y) \quad (x, y \in S).$$

Clearly,

$$\tilde{\tilde{m}} = m \quad \text{and} \quad \hat{\hat{m}} = m,$$

provided m^{-1} maps the space $\mathcal{P}_{\text{!inc}}(S)$ resp. $\mathcal{P}_{\text{!dec}}(S)$ into itself. We warn the reader, however, that since the duality function ψ is not symmetric (i.e., $\psi(x, y) \neq \psi(y, x)$), we need to distinguish \tilde{m} from \hat{m} . In particular, if a map m has the property that m^{-1} maps both $\mathcal{P}_{\text{!inc}}(S)$ and $\mathcal{P}_{\text{!dec}}(S)$ into themselves, then it may happen that $\hat{m} \neq \tilde{m}$.

Birth-and-death processes

As a simple application of Proposition 2.3, let us consider a Markov process X with state space $S = \{0, \dots, n\}$ and generator of the form

$$Gf(x) = b_{x+1}(f(x+1) - f(x)) + d_x(f(x-1) - f(x)) \quad (x = 0, \dots, n), \quad (2.22)$$

where b_1, \dots, b_n and d_1, \dots, d_n are nonnegative rates, and for notational convention we set $b_{n+1} = 0 = d_0$, i.e., the first (resp. second) term in (2.22) are absent if $x = n$ (resp. $x = 0$). Processes with generators as in (2.22) are called *birth-and-death processes*. Let us define maps $\text{birth}_z : S \rightarrow S$ and $\text{death}_z : S \rightarrow S$ by

$$\text{birth}_z(x) := \begin{cases} x+1 & \text{if } x+1 = z, \\ x & \text{otherwise,} \end{cases} \quad \text{death}_z(x) := \begin{cases} x-1 & \text{if } x = z, \\ x & \text{otherwise.} \end{cases}$$

Then a random mapping representation for the operator G in (2.22) is

$$Gf(x) = \sum_{z=1}^n b_z(f(\text{birth}_z(x)) - f(x)) + \sum_{z=1}^n d_z(f(\text{death}_z(x)) - f(x)).$$

The set $S = \{0, \dots, n\}$ is totally ordered and the maps birth_z and death_z are monotone. With the exception of death_n , their inverse images map $\mathcal{P}_{\text{!inc}}(S)$ into

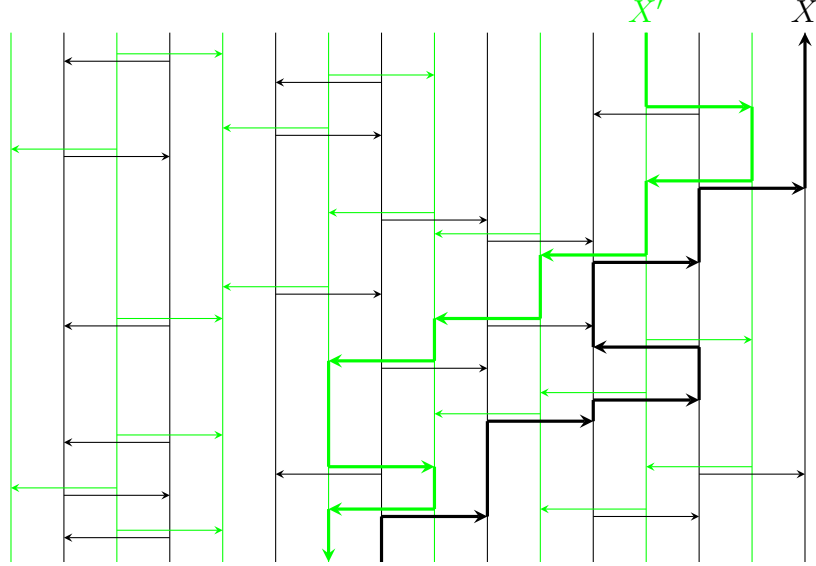


Figure 2.1: Noncrossing duality between two birth-and-death processes. Arrows to the right and left indicate birth and death events, respectively. The dual (downward) process is drawn in green and for readability has been shifted a distance $1/2$ to the left.

itself, so by Proposition 2.3 there exist duals of these maps with respect to the duality function $\psi(x, y) = 1_{\{x \geq y\}}$. Indeed,

$$\text{birth}_z(x) \geq y \Leftrightarrow x \geq y \text{ or } x + 1 = y = z \Leftrightarrow x \geq \text{death}_z(y),$$

and

$$\text{death}_z(x) \geq y \Leftrightarrow x \geq y \text{ and not } x = y = z \Leftrightarrow x \geq \text{birth}_{z+1}(y),$$

so

$$\widehat{\text{birth}}_z = \text{death}_z \quad \text{and} \quad \widehat{\text{death}}_z = \text{birth}_{z+1}.$$

Now let X, X' be two birth-and-death processes with generators as in (2.22) and rates $b_1, \dots, b_n, d_1, \dots, d_n$ resp. $b'_1, \dots, b'_n, d'_1, \dots, d'_n$. Assume that

$$d_n = 0 \quad \text{and} \quad b'_1 = 0,$$

and that

$$d'_z = b_z \quad (z = 1, \dots, n) \quad \text{and} \quad b'_{z+1} = d_z \quad (z = 1, \dots, n-1).$$

Then Lemma 2.2 and Proposition 2.3 together imply that for each $t \geq 0$, the processes X and X' can be coupled such that

$$1_{\{X_s \geq X'_{t-s}\}} \quad \text{does not depend on } s \in [0, t].$$

See Figure 2.1 for a graphical demonstration of this duality. This form of duality has been known for a long time and is nowadays usually associated with the name of Siegmund [Sie76]; see also [KM57, CR83]. One-dimensional diffusion processes (including Brownian motion) with the right boundary conditions satisfy similar relations. A very readable discussion of this (from which I have taken these references) can be found in [Lig85, Sect. II.3]. Note that the assumption that the process only moves up and down in steps of one can be relaxed, as long as the generator has a random mapping representation in terms of monotone maps whose inverse images map $\mathcal{P}_{\text{inc}}(S)$ into itself.

2.4 Additive systems duality

In the previous section we saw one example of monotone systems duality when we considered birth-and-death processes, which take values in the totally ordered set $\{0, \dots, n\}$. In the present section, we will look at monotone Markov processes taking values in very different partially ordered sets, namely, the set of subsets of some finite set Λ .

Let Λ be a finite set. We will be interested in Markov processes X taking values in the set $\mathcal{P}(\Lambda)$ of all subsets of Λ . The set $\mathcal{P}(\Lambda)$ is of course partially ordered by inclusion \subset . Let us assume that we have a random mapping representation for the generator G of X , i.e.,

$$Gf(x) = \sum_{m \in \mathcal{M}} r_m (f(m(x)) - f(x)) \quad (x \in \mathcal{P}(\Lambda), f \in \mathbb{R}^{\mathcal{P}(\Lambda)}),$$

where the maps $m \in \mathcal{M}$ are all monotone, i.e., $x \subset y$ implies $m(x) \subset m(y)$. Let $x^c := \Lambda \setminus x$ denote the complement of a set $x \subset \Lambda$. Then $X^c = (X_t^c)_{t \geq 0}$ is of course also a Markov process, whose generator (informally denoted by G^c here) has the random mapping representation

$$G^c f(x) = \sum_{m \in \mathcal{M}} r_m (f(m(x^c)^c) - f(x)) \quad (x \in \mathcal{P}(\Lambda), f \in \mathbb{R}^{\mathcal{P}(\Lambda)}).$$

We observe that if $m : \mathcal{P}(\Lambda) \rightarrow \mathcal{P}(\Lambda)$ is monotone, then the map

$$x \mapsto m(x^c)^c$$

is also monotone. Indeed:

$$x \subset y \Rightarrow x^c \supset y^c \Rightarrow m(x^c) \supset m(y^c) \Rightarrow m(x^c)^c \subset m(y^c)^c.$$

Thus, X^c is also a monotone Markov process. If X and Y are pathwise dual with the duality function $1_{\{x \geq y\}}$, then X^c and Y are pathwise dual with duality function $1_{\{x^c \geq y\}} = 1_{\{x \cap y = \emptyset\}}$, or equivalently, with the duality function $1_{\{x \cap y \neq \emptyset\}}$. It turns out that this formulation of the duality is more convenient to work with, so for the remainder of this section, we shift our attention to this duality function.

By definition, a map $m : \mathcal{P}(\Lambda) \rightarrow \mathcal{P}(\Lambda)$ is *additive* if

$$m(\emptyset) = \emptyset \quad \text{and} \quad m(x \cup y) = m(x) \cup m(y) \quad (x, y \in \mathcal{P}(\Lambda)).$$

Proposition 2.4 (Additive systems duality) *Let $m : \mathcal{P}(\Lambda) \rightarrow \mathcal{P}(\Lambda)$. Then the following statements are equivalent.*

- (i) $m^{-1}(A) \in \mathcal{P}_{\text{dec}}(\mathcal{P}(\Lambda))$ for all $A \in \mathcal{P}_{\text{dec}}(\mathcal{P}(\Lambda))$.
- (ii) There exists an $m^\dagger : \mathcal{P}(\Lambda) \rightarrow \mathcal{P}(\Lambda)$ that is dual to m with respect to the duality function $\psi(x, y) = 1_{\{x \cap y \neq \emptyset\}}$ ($x, y \in \mathcal{P}(\Lambda)$).
- (iii) m is additive.

Proof (i) \Rightarrow (ii): By Proposition 2.3, (i) implies the existence of a dual map $\hat{m} : \mathcal{P}(\Lambda) \rightarrow \mathcal{P}(\Lambda)$ such that

$$1_{\{m(x) \subset y\}} = 1_{\{x \subset \hat{m}(y)\}} \quad (x, y \in \mathcal{P}(\Lambda)).$$

Setting $m^\dagger(x) := \hat{m}(x^c)^c$, we see that

$$1_{\{m(x) \cap y = \emptyset\}} = 1_{\{m(x) \subset y^c\}} = 1_{\{x \subset \hat{m}(y^c)\}} = 1_{\{x \cap \hat{m}(y^c)^c = \emptyset\}},$$

so

$$m(x) \cap y \neq \emptyset \quad \text{if and only if} \quad x \cap m^\dagger(y) \neq \emptyset \quad (x, y \in \mathcal{P}(\Lambda)).$$

(ii) \Rightarrow (iii): We observe that

$$m(\emptyset) = \{i \in \Lambda : \{i\} \cap m(\emptyset) \neq \emptyset\} = \{i \in \Lambda : m^\dagger(\{i\}) \cap \emptyset \neq \emptyset\} = \emptyset,$$

and

$$\begin{aligned} m(x \cup x') &= \{i \in \Lambda : \{i\} \cap m(x \cup x') \neq \emptyset\} = \{i \in \Lambda : m^\dagger(\{i\}) \cap (x \cup x') \neq \emptyset\} \\ &= \{i \in \Lambda : m^\dagger(\{i\}) \cap x \neq \emptyset\} \cup \{i \in \Lambda : m^\dagger(\{i\}) \cap x' \neq \emptyset\} = m(x) \cup m(x'), \end{aligned}$$

which shows that m is additive.

(iii) \Rightarrow (i): Set

$$\hat{m}(y) := \{i \in \Lambda : m(\{i\}) \subset y\}.$$

Then, using additivity,

$$m^{-1}(\{y\}^\downarrow) = \{x : m(x) \subset y\} = \{x : \bigcup_{i \in x} m(\{i\}) \subset y\} = \{\hat{m}(y)\}^\downarrow,$$

proving that m^{-1} maps $\mathcal{P}_{\text{dec}}(\mathcal{P}(\Lambda))$ into itself. ■

For any additive map $m : \mathcal{P}(\Lambda) \rightarrow \mathcal{P}(\Lambda)$, let us write

$$m(i, j) := 1_{\{j \in m(\{i\})\}} \quad (i, j \in \Lambda).$$

Since

$$m(x) = \bigcup_i m(\{i\}) = \{j : m(i, j) = 1 \text{ for some } i \in x\},$$

the additive map m is uniquely characterized by the matrix $m(i, j)$.

Lemma 2.5 (Dual map) *For each additive map $m : \mathcal{P}(\Lambda) \rightarrow \mathcal{P}(\Lambda)$, there exists a unique additive map $m^\dagger : \mathcal{P}(\Lambda) \rightarrow \mathcal{P}(\Lambda)$ such that*

$$1_{\{m(x) \cap y \neq \emptyset\}} = 1_{\{x \cap m^\dagger(y) \neq \emptyset\}} \quad (x, y \in \mathcal{P}(\Lambda)),$$

and the matrix of m^\dagger is given by $m^\dagger(i, j) = m(j, i)$.

Proof Existence of m^\dagger has been proved in Proposition 2.4. Since

$$m^\dagger(i, j) = 1_{\{\{j\} \cap m^\dagger(\{i\}) \neq \emptyset\}} = 1_{\{m(\{j\}) \cap \{i\} \neq \emptyset\}} = m(j, i),$$

we see that m^\dagger is unique and its matrix is the adjoint of the matrix of m . ■

By definition, a Markov process X with state space $\mathcal{P}(\Lambda)$ is *additive* if its generator G has a random mapping representation involving only additive maps. We construct such a system with the help of random maps $\Phi_{s,t}$ as in Proposition 1.2, which in turn are defined in terms of the Poisson point set Δ . For additive systems, there is a nice way of visualizing Δ , yielding a *graphical representation* of X . We draw Λ horizontally, time vertically, and for each $(m, t) \in \Delta$, we draw:

- an arrow from (i, t) to (j, t) for each $i, j \in \Lambda$, $i \neq j$ such that $m(i, j) = 1$,
- a blocking symbol \blacksquare at (i, t) for each $i \in \Lambda$ such that $m(i, i) = 0$.

We will be interested in paths that walk upwards in time, that may follow arrows, but must avoid blocking symbols. More precisely, for any $s \leq u$, we say that a cadlag function $\gamma : [s, t] \rightarrow \Lambda$ is an *open path* if for each $t \in (s, u]$, it satisfies the following requirements:

- (i) If $(m, t) \in \Delta$ for some $m \in \Delta$, then $m(\gamma_{t-}, \gamma_t) = 1$.
- (ii) If there is no $m \in \mathcal{M}$ such that $(m, t) \in \Delta$, then $\gamma_t = \gamma_{t-}$.

(Note that by the properties of a Poisson point set, there a.s. are no times $t \in \mathbb{R}$ such that $(t, m) \in \Delta$ and $(t, m') \in \Delta$ for *two* maps $m, m' \in \mathcal{M}$.) We write $(i, s) \rightsquigarrow (j, u)$ if there exists an open path $\gamma : [s, u] \rightarrow \Lambda$ such that $\gamma_s = i$ and $\gamma_u = j$.

Proposition 2.6 (Graphical representation) *Let X_0, Y_0 be $\mathcal{P}(\Lambda)$ -valued random variables, independent of each other and of the Poisson set Δ , and let $t \geq 0$. Define processes X and Y with cadlag sample paths by*

$$\begin{aligned} X_s &:= \{j \in \Lambda : \exists i \in X_0 \text{ s.t. } (i, 0) \rightsquigarrow (j, s)\}, \\ Y_{s-} &:= \{i \in \Lambda : \exists j \in Y_0 \text{ s.t. } (i, t-s) \rightsquigarrow (j, t)\}. \end{aligned}$$

Then X and Y are Markov processes with generators G, H given by

$$\begin{aligned} Gf(x) &:= \sum_{m \in \mathcal{M}} r_m (f(m(x)) - f(x)), \\ Hf(y) &:= \sum_{m \in \mathcal{M}} r_m (f(m^\dagger(y)) - f(y)), \end{aligned}$$

and X and Y are pathwise duals of each other, in the sense that the event

$$\{X_s \cap Y_{(t-s)-} \neq \emptyset\} = \{\exists i \in X_0, j \in Y_0 \text{ s.t. } (i, 0) \rightsquigarrow (j, t)\} \quad (2.23)$$

a.s. does not depend on $s \in [0, t]$.

Proof Open paths are defined in such a way that X changes only at times of the Poisson set Δ . If $(m, s) \in \Delta$, then, again by the way open paths are defined

$$X_s = \{j : \exists i \in X_{s-} \text{ s.t. } m(i, j) = 1\} = m(X_{s-}),$$

so by Proposition 1.2 we see that $X_s = \Phi_{0,s}(X_0)$ is the Markov process with generator G . In the same way, using also Lemma 2.5, we see that Y is the Markov process with generator H . Now (2.23) is obvious from our construction, and also follows more abstractly from Lemma 2.2 and the fact that the maps m and m^\dagger are dual with duality function $\psi(x, y) = 1_{\{x \cap y \neq \emptyset\}}$. \blacksquare

We note that the construction of the dual process in terms of the arrows and blocking signs of the graphical representation is the same as for the original, ‘forward’ process, except that we follow open paths downwards in time and hence traverse the arrows in the opposite direction. In particular, the dual m^\dagger of an additive map m is obtained by reversing the arrows of m and keeping the blocking symbols in place (see Figures 2.2 and 2.3 below).

The voter model

Let Λ be a finite set as before, and for any $i, j \in \Lambda$, $i \neq j$, let us define a ‘voter model map’ $\text{vot}_{i,j} : \mathcal{P}(\Lambda) \rightarrow \mathcal{P}(\Lambda)$ by

$$\text{vot}_{i,j}(x) := \begin{cases} x \cup \{j\} & \text{if } i \in x, \\ x \setminus \{j\} & \text{if } i \notin x. \end{cases} \quad (2.24)$$

Let us introduce the notation

$$x(i) := 1_{\{i \in x\}} \quad (x \in \mathcal{P}(\Lambda), i \in \Lambda),$$

which basically says that we identify a set $x \in \mathcal{P}(\Lambda)$ with its indicator function, which is an element of $\{0, 1\}^\Lambda$. In the voter model context, we interpret $x(i)$ as the *type* of the *site* i , i.e., we think of x as describing a population of individuals occupying sites in a lattice, with one individual per site, where each individual can be of two types, labeled 0 and 1. Then (2.24) says that applying the map $\text{vot}_{i,j}$ has the effect that the individual at site j adopts the type of the individual at the site i , regardless of what type previously occupied j . In biology, types are often interpreted as genetic types and $\text{vot}_{i,j}$ describes the event that the individual at j dies and is replaced by a offspring of the individual at site i .

For any collection of nonnegative constants $p(i, j)$ ($i, j \in \Lambda$, $i \neq j$) we may consider a Markov process X with values in $\mathcal{P}(\Lambda)$, or equivalently $\{0, 1\}^\Lambda$, with generator given by

$$G_{\text{vot}} f(x) := \sum_{i \neq j} p(i, j) (f(\text{vot}_{i,j}(x)) - f(x)). \quad (2.25)$$

Then X is a *voter model* where the site j adopts the type of site i with rate $p(i, j)$. It is straightforward to check that the map $\text{vot}_{i,j}$ is additive. In graphical representations, $\text{vot}_{i,j}$ is represented by an arrow from i to j and a blocking symbol at j . In Figure 2.2 on the left, we have drawn an example of a graphical representation for a one-dimensional nearest-neighbor voter model on a set of the form $\Lambda = \{0, \dots, n\}$ where $p(i, j) > 0$ if and only if $|i - j| = 1$.

Let $\text{rw}_{j,i} := \text{vot}_{i,j}^\dagger$ be the dual map of $\text{vot}_{i,j}$. In graphical representations, $\text{rw}_{j,i}$ is represented by an arrow from j to i and a blocking symbol at j . From this, we see that

$$\text{rw}_{j,i}(x) := \begin{cases} (x \setminus \{j\}) \cup \{i\} & \text{if } j \in x, \\ x & \text{if } j \notin x. \end{cases} \quad (2.26)$$

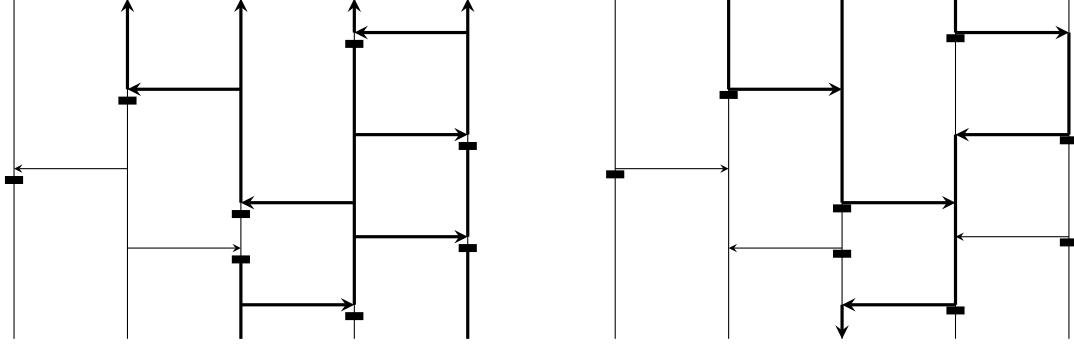


Figure 2.2: The voter model (on the left, with time running upwards) and its dual system of coalescing random walks (on the right, with time running downwards).

The voter model X with generator G_{vot} as in (2.25) is dual to the Markov process Y with generator given by

$$G_{\text{rw}}f(y) := \sum_{i \neq j} p(i, j) (f(\text{rw}_{j,i}(y)) - f(y)). \quad (2.27)$$

An example of a graphical representation of Y is drawn in Figure 2.2 on the right. If $i \in Y_s$, then let us say that the site i is at time s occupied by a particle. Then these particles form a system of *coalescing random walks* which independently jump from i to j with rate $p^\dagger(i, j) = p(j, i)$ and which coalesce (i.e., two particles become one) as soon as one particle jumps on top of another one.

The contact process

With Λ a finite set as before, we define the following maps on $\mathcal{P}(\Lambda)$:

$$\begin{aligned} \text{rec}_i(x) &:= x \setminus \{i\} & (i \in \Lambda), \\ \text{inf}_{i,j}(x) &:= \begin{cases} x \cup \{j\} & \text{if } i \in \Lambda \\ x & \text{otherwise,} \end{cases} & (i, j \in \Lambda, i \neq j). \end{aligned} \quad (2.28)$$

Both maps are additive. In graphical representations, rec_i is represented by a single blocking symbol at i while $\text{inf}_{i,j}$ corresponds to a single arrow from i to j . Given rates $\delta \geq 0$ and $\lambda(i, j) \geq 0$ ($i, j \in \Lambda, i \neq j$), we call the Markov process X with generator

$$G_{\text{cont}}f(x) := \delta \sum_i (f(\text{rec}_i(x)) - f(x)) + \sum_{i \neq j} \lambda(i, j) (f(\text{inf}_{i,j}(x)) - f(x)) \quad (2.29)$$

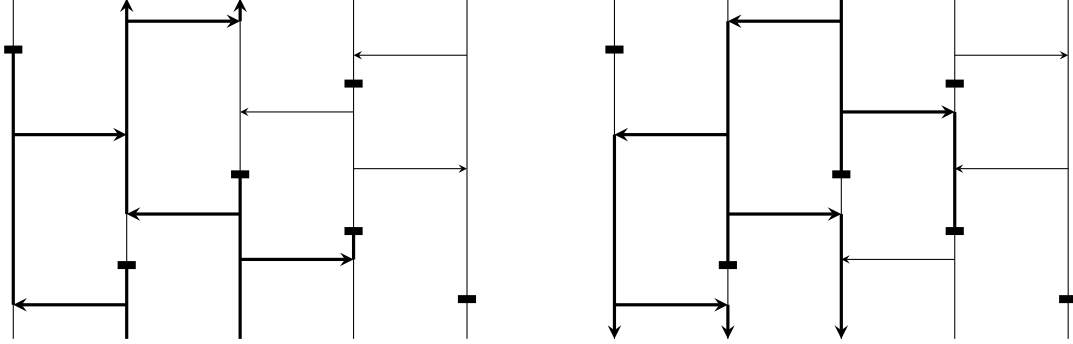


Figure 2.3: The contact process, which is self-dual, with time running upwards and downwards on the left and right, respectively.

the *contact process* with *recovery rate* δ and *infection rate* $\lambda(i, j)$. An example of a graphical representation is drawn in Figure 2.2. If $\lambda(i, j) = \lambda(j, i)$, then the contact process is *self-dual*. In general, the contact process with rates $\lambda(i, j)$ is dual to the contact process with rates $\lambda^\dagger(i, j) = \lambda(j, i)$.

In the traditional interpretation, sites in Λ are thought of as being occupied by an organism (e.g., a tree) that can be infected by some bug (e.g., an insect). Then X_t is the set of infected organisms at time t , $\lambda(i, j)$ is the rate at which infected organisms infect healthy organisms, and δ is the rate at which infected organisms recover (after which they can be infected again).

Alternatively, sites $i \in X_t$ can be interpreted as being occupied by an organism of any kind (not necessarily a parasite). Then infection and recovery may simply be interpreted as the production of offspring and death.

Notes and generalizations

Let S be a partially ordered set. It may happen that the inverse image m^{-1} of a monotone map m does not map the space $\mathcal{P}_{\text{inc}}(S)$ into itself, but that it does map the slightly larger set $\mathcal{P}_{\text{inc}}(S) \cup \{\emptyset\}$ into itself. In this case, one can still define a dual by a recipe similar to (2.19) where we need one extra point in the state space of the dual process to represent the set \emptyset . In this way, one can, for example, find duals to variations of the voter model, coalescing random walks, or contact process, that also have ‘spontaneous births’ corresponding to the map

$$\text{birth}_i(x) := x \cup \{i\}. \quad (2.30)$$

More generally, one can allow maps m that satisfy $m(x \cup y) = m(x) \cup m(y)$ but not $m(\emptyset) = \emptyset$.

We have so far restricted our attention to two very special partially ordered sets: $S = \{0, \dots, n\}$ and $S = \mathcal{P}(\Lambda)$, where the latter is naturally isomorphic to $\{0, 1\}^\Lambda$, equipped with the product order. I do not know what happens when one replaces the latter by, for example, $\{0, 1, 2\}^\Lambda$, equipped with the product order. A natural question is whether there exists a generalization of Proposition 2.4 in this context, i.e., can one classify all monotone maps on $S = \{0, 1, 2\}^\Lambda$ whose inverse image maps $\mathcal{P}_{\text{linc}}(S)$ into itself? There might be interesting dualities waiting to be discovered.

2.5 Linear systems duality

In this section, we will temporarily drop the assumption that the state space be finite, and look at linear systems, which take values in a linear space. We hope the reader will trust us that -at least for the processes we will consider- the technicalities concerning the construction of these processes and their associated semigroups from their generators and from a Poisson set of events are basically the same as in the case of finite state space. Obviously, in continuous space, more complicated Markov processes like Brownian motion can be constructed, but the processes that we will be interested in will have piecewise constant sample paths and are from the viewpoint of their construction very similar to processes with finite state spaces.

Let Λ be a finite set and let \mathbb{R}^Λ be the real vectorspace consisting of all functions $x : \Lambda \rightarrow \mathbb{R}$. We equip \mathbb{R}^Λ with the inner product

$$\langle x, y \rangle := \sum_{i \in \Lambda} x(i)y(i) \quad (x, y \in \mathbb{R}^\Lambda).$$

Let \mathcal{M} be a finite set whose elements are *linear* maps $m : \mathbb{R}^\Lambda \rightarrow \mathbb{R}^\Lambda$ and let $(r_m)_{m \in \mathcal{M}}$ be a collection of nonnegative rates. Then the Markov process X with generator

$$Gf(x) = \sum_{m \in \mathcal{M}} r_m (f(m(x)) - f(x))$$

can be constructed from a Poisson subset $\Delta \subset \mathcal{M} \times \mathbb{R}$ just as we did for processes with finite state spaces. We will be interested in pathwise duals of X .

For each linear map $m : \mathbb{R}^\Lambda \rightarrow \mathbb{R}^\Lambda$, let $m(i, j)$ denote the matrix of m , so that $m(x)(i) = \sum_j m(i, j)x(j)$. Let $m^\dagger(i, j) := m(j, i)$ denote its adjoint w.r.t. the inner product $\langle \cdot, \cdot \rangle$. Then

$$\langle m(x), y \rangle = \langle x, m^\dagger(y) \rangle \quad (x, y \in \mathbb{R}^\Lambda).$$

In the language of (2.10), this says that the maps m and m^\dagger are dual with respect to the duality function $\psi(x, y) = \langle x, y \rangle$. Thus, by (a straightforward generalization of) Lemma 2.2, we see that the process Y with generator

$$G_\dagger f(y) = \sum_{m \in \mathcal{M}} r_m (f(m^\dagger(y)) - f(y))$$

is a pathwise dual of X , i.e., for each $t \geq 0$, we can couple X and Y in such a way that for each $s \in [0, t]$, $(X_u)_{0 \leq u \leq s}$ is independent of $(Y_u)_{0 \leq u \leq t-s}$, and moreover, a.s.,

$$\langle X_{s-}, Y_{t-s} \rangle \quad \text{does not depend on } s \in [0, t]. \quad (2.31)$$

There is a way of interpreting this duality in terms of open paths, similar to what we saw for additive systems. The only complication is that we have to give each path a weight. Let Δ be the Poisson set used for constructing X . Draw Λ horizontally, time vertically, and for each $(m, t) \in \Delta$, draw:

- an arrow with *weight* $m(i, j)$ from (i, t) to (j, t)
- for each $i, j \in \Lambda$ with $i \neq j$ such that $m(i, j) \neq 0$,
- a symbol \blacksquare with *weight* $m(i, i)$ at (i, t)
- for each $i \in \Lambda$ such that $m(i, i) \neq 1$.

We say that a cadlag function $\gamma : [s, t] \rightarrow \Lambda$ is an *open path* if for each $t \in (s, u]$, it satisfies the following requirements:

- (i) If $(m, t) \in \Delta$ for some $m \in \Delta$, then $m(\gamma_{t-}, \gamma_t) \neq 0$.
- (ii) If there is no $m \in \mathcal{M}$ such that $(m, t) \in \Delta$, then $\gamma_t = \gamma_{t-}$.

We write $(i, s) \rightsquigarrow (j, u)$ if γ is an open path with $\gamma_s = i$ and $\gamma_u = j$. We give each open path γ a *weight*

$$w(\gamma) := \text{the product of the weights of all arrows and symbols } \blacksquare \text{ on } \gamma.$$

The following result is very similar to Proposition 2.6.

Proposition 2.7 (Graphical representation) *Let X_0, Y_0 be \mathbb{R}^Λ -valued random variables, independent of each other and of the Poisson set Δ , and let $t \geq 0$. Define processes X and Y with cadlag sample paths by*

$$\begin{aligned} X_s(j) &:= \sum_{i \in \Lambda} \sum_{(i,0) \rightsquigarrow (j,s)} X_0(i) w(\gamma), \\ Y_{s-}(i) &:= \sum_{j \in \Lambda} \sum_{(i,t-s) \rightsquigarrow (j,t)} w(\gamma) Y_0(j). \end{aligned}$$

Then X and Y are Markov processes with generators G, H given by

$$\begin{aligned} Gf(x) &:= \sum_{m \in \mathcal{M}} r_m (f(m(x)) - f(x)), \\ Hf(y) &:= \sum_{m \in \mathcal{M}} r_m (f(m^\dagger(y)) - f(y)), \end{aligned}$$

and X and Y are pathwise duals of each other, in the sense that the function

$$\langle X_s, Y_{(t-s)-} \rangle = \sum_i \sum_j \sum_{(i,0) \rightsquigarrow (j,t)} X_0(i) w(\gamma) Y_0(j) \quad (2.32)$$

a.s. does not depend on $s \in [0, t]$.

As a simple example, consider the graphical representation of the contact process, where we give each arrow weight 1 and each blocking symbol weight zero. If X_0, Y_0 take values in \mathbb{N}^Λ , then the processes X and Y constructed in Proposition 2.7 take values in \mathbb{N}^Λ at all times; up to a deterministic rescaling of time, this is known as the *binary contact path process* in the literature. Now the duality formula (2.32) just says that

$$\langle X_s, Y_{(t-s)-} \rangle = \left| \{ \gamma : (i,0) \rightsquigarrow (j,t) \text{ for some } i, j \text{ s.t. } X_0(i) = 1 = Y_0(j) \} \right|$$

a.s. does not depend on $s \in [0, t]$.

2.6 Cancellative systems duality

We recall that a *field* is a set F equipped with two operations $(x, y) \mapsto x + y$ (addition) and $(x, y) \mapsto xy$ (multiplication) such that:

- (i) Addition and multiplication are associative.
- (ii) Addition and multiplication are commutative.
- (iii) Addition and multiplication satisfy the distributive property.
- (iv) Addition and multiplication have identity elements 0 and 1, respectively, with $0 \neq 1$.
- (v) Each element x has inverses $-x$ and x^{-1} with respect to addition and multiplication, respectively.

Linear spaces can be defined over any field. In particular, the concepts of linear independence, a basis, and the matrix of a linear map with respect to a given basis of the source and target space work just in the same way as for the well-known fields \mathbb{R} and \mathbb{C} .

There exist many other fields than \mathbb{R} and \mathbb{C} . In particular, there exist *finite fields*. The simplest example is $\{0, 1\}$ equipped with addition *modulo 2* and the usual product. Let us (somewhat nonstandardly) write \oplus for addition modulo 2 to remind ourselves that this is not the usual addition. Then, for any finite set Λ , the space $\{0, 1\}^\Lambda$ equipped with \oplus is a linear space over the finite field $\{0, 1\}$. We can even equip $\{0, 1\}^\Lambda$ with something similar to an inner product by setting

$$\langle x, y \rangle := \bigoplus_{i \in \Lambda} x(i)y(i).$$

This is not positive definite but it is true that

$$\langle x, y \rangle = 0 \quad \forall y \in \{0, 1\}^\Lambda \quad \text{implies} \quad x = 0.$$

Each linear map $m : \{0, 1\}^\Lambda \rightarrow \{0, 1\}^\Lambda$ (where linearity should be interpreted over the finite field $\{0, 1\}$) is uniquely characterized by its matrix $m(i, j)$, where

$$m(x)(i) = \bigoplus_j m(i, j)x(j),$$

and the entries $m(i, j)$ of m take of course values in $\{0, 1\}$. Setting $m^\dagger(i, j) := m(j, i)$, we have that

$$\langle m(x), y \rangle = \langle x, m^\dagger(y) \rangle,$$

which says that m and m^\dagger are dual with respect to the duality function $\psi(x, y) = \langle x, y \rangle$.

In view of this, all that we have said in the previous section about linear systems duality applies without a change to Markov processes X with a generator of the form

$$Gf(x) = \sum_{m \in \mathcal{M}} r_m (f(m(x)) - f(x))$$

where the maps $m \in \mathcal{M}$ are linear maps $m : \{0, 1\}^\Lambda \rightarrow \{0, 1\}^\Lambda$ with respect to the finite field $\{0, 1\}$. Such Markov processes are called *cancellative systems*, as defined⁵ in the classical monograph [Gri79]. Let Y be the dual process with generator

$$Hf(y) = \sum_{m \in \mathcal{M}} r_m (f(m^\dagger(y)) - f(y)).$$

⁵Actually, Griffeath's definition of cancellative systems is somewhat more general, in the sense that he also allows for spontaneous births. (Compare Exercise 2.8.)

For given $t \geq 0$, we may construct X and Y as in Proposition 2.7. Since the weight of each arrow must be a number in the finite field $\{0, 1\}$, different from zero, and the weight of each symbol \blacksquare must be a number in the finite field $\{0, 1\}$ different from one, we see that each arrow has weight one and each symbol \blacksquare is, indeed, a blocking symbol. Now every open path (that must avoid blocking symbols) has weight 1. Let

$$\Gamma := \{\gamma : (i, 0) \rightsquigarrow (j, t) \text{ for some } i, j \text{ s.t. } X_0(i) = 1 = Y_0(j)\}.$$

Then the duality formula (2.32) just says that

$$\langle X_s, Y_{(t-s)-} \rangle = \begin{cases} 0 & \text{if } |\Gamma| \text{ is even} \\ 1 & \text{if } |\Gamma| \text{ is odd,} \end{cases}$$

does not depend on $s \in [0, t]$.

The voter model revisited

Identifying sets with their indicator functions and $\mathcal{P}(\Lambda) \cong \{0, 1\}^\Lambda$, we may write the voter model map $\text{vot}_{i,j}$ defined in (2.24) as

$$\text{vot}_{i,j}(x)(k) = \begin{cases} x(i) & \text{if } k = j, \\ x(k) & \text{otherwise,} \end{cases}$$

which is clearly linear modulo 2. Thus, the voter model is also a cancellative system. In fact, as a cancellative system, it can be constructed from the same arrows and blocking symbols that are used in its construction as an additive system. Its dual with respect to the cancellative duality function

$$\psi(x, y) = \langle x, y \rangle = \bigoplus_i x(i)y(i) = 1_{\{|x \cap y| \text{ is odd}\}} \quad (2.33)$$

is different, however, from its dual with respect to the additive duality function

$$1_{\{x \cap y \neq \emptyset\}}.$$

Indeed,

$$\langle \text{vot}_{i,j}(x), y \rangle = \langle x, \text{ann}_{j,i}(y) \rangle,$$

where $\text{ann}_{j,i}$ is the map

$$\text{ann}_{i,j}(y)(k) = \begin{cases} 0 & \text{if } k = i, \\ y(i) \oplus y(j) & \text{if } k = j, \\ y(k) & \text{otherwise,} \end{cases} \quad (2.34)$$

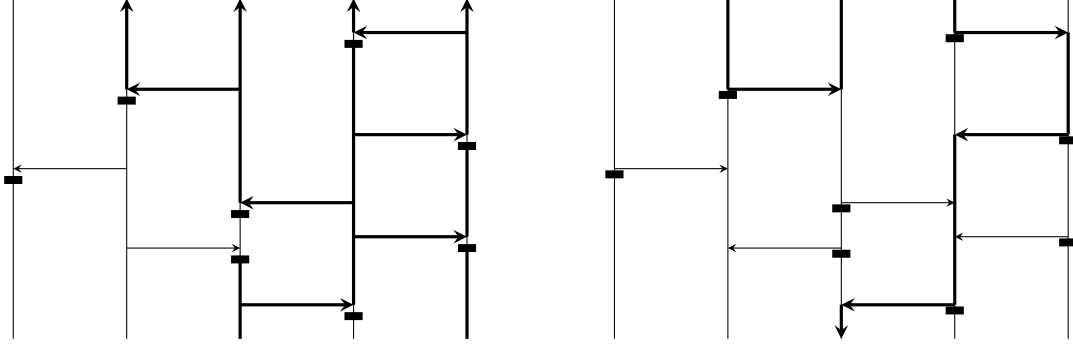


Figure 2.4: The voter model (on the left, with time running upwards) and its dual system of annihilating random walks (on the right, with time running downwards).

which corresponds to a transition of *annihilating random walks*, where a particle at i jumps to j with the rule that if a particle lands on an occupied site, then both particles annihilate each other, i.e., both particles disappear; see Figure 2.4 and compare Figure 2.2.

2.7 Lloyd-Sudbury theory

In a series of papers⁶ [LS95, LS97, Sud00], Lloyd and Sudbury systematically look for dualities of Markov processes with state space of the form $\{0, 1\}^{\mathbb{Z}}$. They do not require that their duals be pathwise, and in fact, it seems that almost all of the ‘new’ duals they find (compared to those that we have already seen in the previous sections) cannot be constructed in a pathwise way.

Recall that $\mathbb{R}^{\{0,1\}^{\Lambda}}$ is the space of all functions $f : \Lambda \rightarrow \{0, 1\}$. If Λ_1, Λ_2 are disjoint finite sets, and $f_1 \in \mathbb{R}^{\{0,1\}^{\Lambda_1}}$, $f_2 \in \mathbb{R}^{\{0,1\}^{\Lambda_2}}$ are real functions on $\{0, 1\}^{\Lambda_1}$ and $\{0, 1\}^{\Lambda_2}$, respectively, then we may define a function $f_1 \otimes f_2$ on $\{0, 1\}^{\Lambda_1 \cup \Lambda_2} \cong \{0, 1\}^{\Lambda_1} \times \{0, 1\}^{\Lambda_2}$ by

$$f_1 \otimes f_2(x_1, x_2) := f_1(x_1)f_2(x_2) \quad (x_1 \in \{0, 1\}^{\Lambda_1}, x_2 \in \{0, 1\}^{\Lambda_2}).$$

⁶The names of these papers are slightly confusing. The papers [LS95, LS97] are called *Quantum operators in classical probability theory. II* and *IV*, respectively. The authors also have a paper with the same name and serial number I, but this is on a somewhat different topic than the other two and certainly no required reading before one can understand parts II and IV. A part III has apparently also been planned but as far as I have been able to find out, no such paper has ever appeared.

In this way, we may identify $\mathbb{R}^{\{0,1\}^\Lambda}$ with the tensor product

$$\mathbb{R}^{\{0,1\}^\Lambda} \cong (\mathbb{R}^{\{0,1\}})^{\otimes \Lambda}$$

of $|\Lambda|$ copies of the two-dimensional vector space $\mathbb{R}^{\{0,1\}}$. For any $x \in \{0,1\}^\Lambda$, define $|x\rangle \in \mathbb{R}^{\{0,1\}^\Lambda}$ by

$$|x\rangle(y) := 1_{\{x=y\}} \quad (y \in \{0,1\}^\Lambda).$$

The functions $|x\rangle$ with $x \in \{0,1\}^\Lambda$ obviously form a basis for $\mathbb{R}^{\{0,1\}^\Lambda}$ and in fact

$$|x\rangle(y) = \prod_{i \in \Lambda} 1_{\{x(i)=y(i)\}} = \left(\bigotimes_{i \in \Lambda} |x(i)\rangle \right)(y),$$

where $\{|0\rangle, |1\rangle\}$ is the obvious basis of $\mathbb{R}^{\{0,1\}}$.

If Λ_1, Λ_2 are again disjoint finite sets, and A_1, A_2 are linear operators on $\mathbb{R}^{\{0,1\}^{\Lambda_1}}$ and $\mathbb{R}^{\{0,1\}^{\Lambda_2}}$, respectively, then we may define an operator $A_1 \otimes A_2$ on

$$\mathbb{R}^{\{0,1\}^{\Lambda_1 \cup \Lambda_2}} \cong \mathbb{R}^{\{0,1\}^{\Lambda_1}} \otimes \mathbb{R}^{\{0,1\}^{\Lambda_2}}$$

by

$$(A_1 \otimes A_2)(f_1 \otimes f_2) := (A_1 f_1) \otimes (A_2 f_2).$$

In coordinates, this says that the matrix of $A_1 \otimes A_2$ has the structure

$$(A_1 \otimes A_2)((x_1, x_2), (x_1, x_2)) = A_1(x_1, y_1) A_2(x_2, y_2)$$

($x_1, y_1 \in \{0,1\}^{\Lambda_1}$, $x_2, y_2 \in \{0,1\}^{\Lambda_2}$). It is easy to see that $(A_1 \otimes A_2)(B_1 \otimes B_2) = A_1 B_1 \otimes A_2 B_2$. In particular, two operators of the form $A \otimes 1$ and $1 \otimes B$ always commute. We may informally describe an operator of the form $A \otimes 1$ as ‘let A act on the coordinates in Λ_1 and do nothing with the coordinates in Λ_2 ’.

Lloyd and Sudbury look for dualities between Markov processes X and Y that both have the state space $\{0,1\}^\Lambda$ (with the same Λ). Their starting point is the algebraic formulation of duality in (2.7). Let G and H be the generators of X and Y , respectively, which are linear operators on $\mathbb{R}^{\{0,1\}^\Lambda}$, and let ψ be the duality function, which we also associate with a linear operator on $\mathbb{R}^{\{0,1\}^\Lambda}$ as in (2.5). Lloyd and Sudbury restrict themselves to processes with two-point interactions, in a sense that the generator G , and likewise H , can be written as

$$G = \sum_{\substack{\{i,j\} \subset \Lambda \\ i \neq j}} G_{ij} \quad \text{and} \quad H = \sum_{\substack{\{i,j\} \subset \Lambda \\ i \neq j}} H_{ij}, \quad (2.35)$$

where the sums run over all subsets of Λ with exactly two elements, and G_{ij} and H_{ij} are operators that act only on the coordinates i, j in the sense described above, i.e., there are operators \tilde{G}_{ij} and \tilde{H}_{ij} acting on $\mathbb{R}^{\{0,1\}^{\{i,j\}}}$ such that the matrices of G_{ij} and H_{ij} have the special form

$$\begin{aligned} G_{ij}(x, y) &= \tilde{G}_{ij}((x(i), x(j)), (y(i), y(j))), \\ H_{ij}(x, y) &= \tilde{H}_{ij}((x(i), x(j)), (y(i), y(j))). \end{aligned}$$

The main idea of Lloyd and Sudbury is now to try duality functions whose corresponding operator ψ is a product of operators acting on a single coordinate only. In fact, they choose the same operator $\tilde{\psi}$ for each coordinate, which yields a duality function of the form

$$\psi(x, y) = \prod_{i \in \Lambda} \tilde{\psi}(x(i), y(i)). \quad (2.36)$$

To satisfy (2.7), it suffices to match up the terms G_{ij} and H_{ij} separately, i.e.,

$$G_{ij}\psi = \psi H_{ij}^\dagger \quad (\{i, j\} \subset \Lambda, i \neq j). \quad (2.37)$$

Identifying $\{0, 1\}^\Lambda \cong \{0, 1\}^{\{i,j\}} \times \{0, 1\}^{\Lambda \setminus \{i,j\}}$ we may write $x \in \{0, 1\}^\Lambda$ as $x = (x_1, x_2)$ with $x_1 \in \{0, 1\}^{\{i,j\}}$ and $x_2 \in \{0, 1\}^{\Lambda \setminus \{i,j\}}$. Then

$$\begin{aligned} G_{ij}|x\rangle &= G_{ij}(|x_1\rangle \otimes |x_2\rangle) = (\tilde{G}_{ij}|x_1\rangle) \otimes |x_2\rangle, \\ H_{ij}^\dagger|x\rangle &= (\tilde{H}_{ij}^\dagger|x_1\rangle) \otimes |x_2\rangle. \end{aligned}$$

Because of the special form of ψ ,

$$\begin{aligned} \psi|x\rangle &= \psi\left(\bigotimes_{i \in \Lambda} |x(i)\rangle\right) = \bigotimes_{i \in \Lambda} \tilde{\psi}|x(i)\rangle = \psi_1|x_1\rangle \otimes \psi_2|x_2\rangle, \\ \text{where } \psi_1 &:= \tilde{\psi} \otimes \tilde{\psi} \quad \text{and} \quad \psi_2 := \bigotimes_{k \neq i, j} \tilde{\psi}. \end{aligned}$$

Now

$$\begin{aligned} G_{ij}\psi|x\rangle &= G_{ij}[\psi_1|x_1\rangle \otimes \psi_2|x_2\rangle] = \tilde{G}_{ij}\psi_1|x_1\rangle \otimes \psi_2|x_2\rangle, \\ \psi H_{ij}^\dagger|x\rangle &= \psi[\tilde{H}_{ij}^\dagger|x_1\rangle \otimes |x_2\rangle] = \psi_1 \tilde{H}_{ij}^\dagger|x_1\rangle \otimes \psi_2|x_2\rangle, \end{aligned}$$

so, recalling that $\psi_1 = \tilde{\psi} \otimes \tilde{\psi}$, we see that (2.37) is equivalent to

$$\tilde{G}_{ij}(\tilde{\psi} \otimes \tilde{\psi}) = (\tilde{\psi} \otimes \tilde{\psi}) \tilde{H}_{ij}^\dagger \quad (i, j \in \Lambda, i \neq j), \quad (2.38)$$

which is just a statement about 4×4 matrices.

Lloyd and Sudbury [LS95, LS97] give several examples where (2.38) is satisfied and hence two processes X and Y are dual. Their most elaborate example concerns two-point interactions described by a matrix $\tilde{G}_{ij} = G$ of the form

$$\begin{pmatrix} G(00,00) & G(00,01) & G(00,10) & G(00,11) \\ G(01,00) & G(01,01) & G(01,10) & G(01,11) \\ G(10,00) & G(10,01) & G(10,10) & G(10,11) \\ G(11,00) & G(11,01) & G(11,10) & G(11,11) \end{pmatrix} = \begin{pmatrix} \cdot & 0 & 0 & 0 \\ d & \cdot & e & b \\ d & e & \cdot & b \\ a & c & c & \cdot \end{pmatrix}. \quad (2.39)$$

For clarity, we have not written the diagonal entries, which are determined by the requirement that $\sum_y G(x, y) = 0$. (For example, $G(01, 01) = -d - e - b$.) The generator in (2.39) corresponds to a Markov process in $\{0, 1\}^2$ that makes the following transitions with the following rates:

$$\begin{aligned} \text{annihilation} \quad 11 &\mapsto 00 && \text{with rate } a, \\ \text{branching} \quad 01 &\mapsto 11 && \text{with rate } b, \\ \text{coalescence} \quad 11 &\mapsto 01 && \text{with rate } c, \\ \text{death} \quad 01 &\mapsto 00 && \text{with rate } d, \\ \text{exclusion} \quad 01 &\mapsto 10 && \text{with rate } e. \end{aligned}$$

Here, we have not written down transitions that are mirror images of other transitions, i.e., it is understood that the transition $01 \mapsto 11$ happens at the same rate as $10 \mapsto 11$, etc. Note that within the class of two-point interactions that are symmetric in this sense and for which 00 is trap, this is as general as one can get. For generators of the form (2.39), it turns out that the most useful duality functions, that give the richest class of dualities, are those of the form

$$\begin{pmatrix} \tilde{\psi}(0,0) & \tilde{\psi}(0,1) \\ \tilde{\psi}(1,0) & \tilde{\psi}(1,1) \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & q \end{pmatrix}, \quad (2.40)$$

where $q \in \mathbb{R} \setminus \{1\}$ is a constant. (This is not the complete picture, however. Lloyd and Sudbury also find some dualities with a more complicated duality function.) The following theorem is proved in [LS95], but stated more clearly in [Sud00]. In the definition of the generator G below, we have added a factor $\frac{1}{2}$ in front of the rate a to avoid double counting. (Note that a is the only rate that occurs only once in the matrix in (2.39).)

Theorem 2.8 (Lloyd-Sudbury duals) *Let Λ be a finite set. Let $p : \Lambda \times \Lambda \rightarrow [0, \infty)$ satisfy $p(i, j) = p(j, i)$. Let $a, b, c, d, e \geq 0$ be constants. Let X be the*

Markov process with state space $\mathcal{P}(\Lambda)$ and generator

$$Gf(x) = \sum_{\substack{i,j \in \Lambda \\ i \neq j}} p(i,j) \left[1_{\{i \in x, j \in x\}} \left\{ \frac{1}{2}a(f(x \setminus \{i,j\}) - f(x)) + c(f(x \setminus \{i\}) - f(x)) \right\} \right. \\ \left. + 1_{\{i \notin x, j \in x\}} \left\{ b(f(x \cup \{i\}) - f(x)) + d(f(x \setminus \{j\}) - f(x)) \right. \right. \\ \left. \left. + e(f((x \setminus \{j\}) \cup \{i\}) - f(x)) \right\} \right].$$

Let $q \in \mathbb{R} \setminus \{1\}$ and let X' be independent of X with similar dynamics but with rates $a', b', c', d', e' \geq 0$ satisfying

$$a' = a + 2q\gamma, \quad b' = b + \gamma, \quad c' = c - (1+q)\gamma, \quad d' = d + \gamma, \quad e' = e - \gamma, \quad (2.41)$$

where $\gamma := (a + c - d + qb)/(1 - q)$. Then

$$\mathbb{E}[q^{|X_t \cap X'_0|}] = \mathbb{E}[q^{|X_0 \cap X'_t|}] \quad (t \geq 0). \quad (2.42)$$

Remark The symmetry assumption that $p(i, j) = p(j, i)$ can be dropped and one can even allow the rates $a, b, \dots = a(i, j), b(i, j), \dots$ to depend individually on i and j , at the cost of replacing (2.41) by a somewhat more complicated set of conditions; see [Swa06, Appendix A in the version on the ArXiv].

Proof We observe that, identifying sets with their indicator functions

$$q^{|x \cap y|} = \prod_{i \in \Lambda} q^{x(i)y(i)} = \prod_{i \in \Lambda} \tilde{\psi}(x(i), y(i)) = \psi(x, y),$$

with $\tilde{\psi}$ as in (2.40). In view of this, it suffices to check (2.37). Dividing out a factor $p(i, j) + p(j, i)$ on each site, this amounts to checking that

$$G(\tilde{\psi} \otimes \tilde{\psi}) = (\tilde{\psi} \otimes \tilde{\psi})H^\dagger,$$

where by a slight abuse of notation G now denotes the generator in (2.39) and H is the same but with the rates a, b, c, d, e replaced by a', b', c', d', e' . Here

$$\tilde{\psi} \otimes \tilde{\psi} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & q & 1 & q \\ 1 & 1 & q & q \\ 1 & q & q & q^2 \end{pmatrix},$$

so, setting $f := b + d + e$ and $g := a + 2c$ and similarly for the primed parameters, we need to check that

$$\begin{aligned} & \begin{pmatrix} 0 & 0 & 0 & 0 \\ d & -f & e & b \\ d & e & -f & b \\ a & c & c & -g \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & q & 1 & q \\ 1 & 1 & q & q \\ 1 & q & q & q^2 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & q & 1 & q \\ 1 & 1 & q & q \\ 1 & q & q & q^2 \end{pmatrix} \begin{pmatrix} 0 & d' & d' & a' \\ 0 & -f' & e' & c' \\ 0 & e' & -f' & c' \\ 0 & b' & b' & -g' \end{pmatrix}. \end{aligned} \quad (2.43)$$

We can simplify these equations a bit by noting that by the fact that $\sum_y G(x, y) = 0$, and similarly for H ,

$$G \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix} = 0 = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix} H^\dagger.$$

Thus, subtracting the matrix with all entries equal to one and dividing out a factor $q - 1$, we may equivalently solve (2.43) with

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & q & 1 & q \\ 1 & 1 & q & q \\ 1 & q & q & q^2 \end{pmatrix} \text{ replaced by } \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & q+1 \end{pmatrix}.$$

This gives (note the transpose sign on the right-hand side!)

$$\begin{aligned} & \left(\begin{array}{c|c|c|c} 0 & 0 & 0 & 0 \\ 0 & b-f & b+e & (q+1)b+e-f \\ 0 & b+e & b-f & (q+1)b+e-f \\ 0 & c-g & c-g & 2c-(q+1)g \end{array} \right) \\ &= \left(\begin{array}{c|c|c|c} 0 & 0 & 0 & 0 \\ 0 & b'-f' & b'+e' & (q+1)b'+e'-f' \\ 0 & b'+e' & b'-f' & (q+1)b'+e'-f' \\ 0 & c'-g' & c'-g' & 2c'-(q+1)g' \end{array} \right)^\dagger. \end{aligned}$$

This yields the equations

$$\begin{aligned}
\text{(i)} \quad & b - f = b' - f', \\
\text{(ii)} \quad & b + e = b + e', \\
\text{(iii)} \quad & (q + 1)b + e - f = c' - g', \\
\text{(iv)} \quad & c - g = (q + 1)b' + e' - f', \\
\text{(v)} \quad & 2c - (q + 1)g = 2c' - (q + 1)g'.
\end{aligned}$$

Filling in the definitions of f and g yields

$$\begin{aligned}
\text{(i)} \quad & d + e = d' + e', \\
\text{(ii)} \quad & b + e = b' + e', \\
\text{(iii)} \quad & d - qb = a' + c', \\
\text{(iv)} \quad & a + c = d' - qb', \\
\text{(v)} \quad & (q + 1)a + 2qc = (q + 1)a' + 2qc'.
\end{aligned}$$

Given a, b, c, d, e , these are five equations for five variables a', b', c', d', e' . The first two equations give $b' = b + (e - e')$ and $d' = d + (e - e')$, so setting $\gamma := e - e'$ we have that

$$b' = b + \gamma, \quad d' = d + \gamma, \quad \text{and} \quad e' = e - \gamma.$$

Equation (iv) now says that

$$a + c = (d + \gamma) - q(b + \gamma) = d - qb + (1 - q)\gamma,$$

which forces us to choose

$$\gamma = \frac{a + c - d + qb}{1 - q}.$$

We are left with the equations

$$\begin{aligned}
\text{(iii)} \quad & d - qb = a' + c', \\
\text{(v)} \quad & (q + 1)a + 2qc = (q + 1)a' + 2qc'.
\end{aligned}$$

Multiplying the first equation by $2q$ and subtracting this from the second equation yields

$$(1 - q)a + 2q(a + c - d' + qb) = (1 - q)a'.$$

Recalling the definition of γ , we may rewrite this as

$$a' = a + 2q\gamma.$$

Inserting this into equation (v), finally, we obtain that

$$c' = c + \frac{q+1}{2q}(a - a') = c - (q+1)\gamma.$$

■

We have, in fact, already seen two special cases of the duality function

$$\psi(x, y) = q^{|x \cap y|} \quad (x, y \in \mathcal{P}(\Lambda))$$

that occurs in Theorem 2.8. First, when $q = 0$, this reduces to

$$\psi(x, y) = 0^{|x \cap y|} = 1_{\{x \cap y = \emptyset\}},$$

which is the duality function familiar to us from additive systems duality. Secondly, taking $q = -1$ yields (letting \oplus denote addition modulo 2)

$$\psi(x, y) = (-1)^{|x \cap y|} = 1 - 2 \bigoplus_{i \in \Lambda} x(i)y(i),$$

which is equivalent to the duality function that we used for cancellative systems.

In general, if X and Y are $\mathcal{P}(\Lambda)$ -valued Markov processes that are dual with respect to the duality function

$$\psi_q(x, y) = q^{|x \cap y|} \quad (x, y \in \mathcal{P}(\Lambda)) \quad (2.44)$$

for some $q \in \mathbb{R} \setminus \{1\}$, then we will say that Y is a q -dual of X . Some of the most useful dualities covered by Theorem 2.8 have $-1 < q < 0$, which interpolates between additive and cancellative systems duality. We conclude this section with some examples.

Duals of the voter model

Let Λ be a finite set and let $p : \Lambda \times \Lambda \rightarrow [0, \infty)$ be a function. We recall from (2.25) that the voter model X with these rates is the $\mathcal{P}(\Lambda)$ -valued Markov process with generator

$$\begin{aligned} G_{\text{vot}} = \sum_{i \neq j} p(i, j) \{ & 1_{\{i \in x(i), j \notin x(j)\}} (f(x \cup \{j\}) - f(x)) \\ & + 1_{\{i \notin x(i), j \in x(j)\}} (f(x \setminus \{j\}) - f(x)) \}. \end{aligned}$$

Assuming moreover that $p(i, j) = p(j, i)$, we can write this generator in the form of Theorem 2.8 by choosing

$$a = 0, \quad b = 1, \quad c = 0, \quad d = 1, \quad e = 0.$$

It follows that the parameter γ from Theorem 2.8 equals

$$\gamma = \frac{a + c - d + qb}{1 - q} = -1,$$

and for a given value of $q \in \mathbb{R} \setminus \{0\}$, we find that X is q -dual to the process X' with rates

$$a' = -2q, \quad b' = 0, \quad c' = 1 + q, \quad d' = 0, \quad e' = 1,$$

provided these rates are all nonnegative, which requires us to choose $-1 \leq q \leq 0$. Set $\alpha := -q$, which satisfies $0 \leq \alpha \leq 1$. Using the maps rw_{ij} and ann_{ij} defined in (2.26) and (2.34), we may write the generator G' of X' in the form

$$G'f(x) = \sum_{i \neq j} p(i, j) \{ \alpha (f(\text{ann}_{i,j}(x)) - f(x)) + (1 - \alpha) (f(\text{rw}_{i,j}(x)) - f(x)) \}. \quad (2.45)$$

We may interpret X' as a system of particles performing independent random walks, jumping from i to j with rate $p(i, j)$, with the rule that when one particle lands on an already occupied site, it *annihilates* (resulting in an empty site) with probability α and *coalesces* (resulting in an occupied site) with probability $1 - \alpha$. In particular, for $\alpha = 0, 1$, respectively, we find back the duals of the voter model in the sense of additive systems duality and cancellative systems duality, respectively.

Duals of the contact process

Let Λ be a finite set and let $p : \Lambda \times \Lambda \rightarrow [0, \infty)$ be a function satisfying $p(i, i) = 0$, $p(i, j) = p(j, i)$, and $\sum_j p(i, j) = 1$. Fix $\lambda \geq 0$ and let X be the contact process with generator (compare (2.29))

$$\begin{aligned} G_{\text{vot}} &= \sum_i (f(\text{rec}_i(x)) - f(x)) + \lambda \sum_{i \neq j} p(i, j) (f(\text{inf}_{i,j}(x)) - f(x)) \\ &= \sum_{i \neq j} p(i, j) \{ (f(\text{rec}_i(x)) - f(x)) + \lambda (f(\text{inf}_{i,j}(x)) - f(x)) \}. \end{aligned}$$

We observe that this generator is of the form of Theorem 2.8 if we choose

$$a = 0, \quad b = \lambda, \quad c = 1, \quad d = 1, \quad e = 0.$$

It follows that for a given value of the duality parameter $q \in \mathbb{R} \setminus \{1\}$, the parameter γ from Theorem 2.8 equals

$$\gamma = \frac{a + c - d + qb}{1 - q} = \frac{q\lambda}{1 - q}.$$

We find that X is q -dual to the process X' with rates

$$a' = \frac{q^2\lambda}{1 - q}, \quad b' = \frac{1}{1 - q}, \quad c' = 1 - q\frac{1 + q}{1 - q}\lambda, \quad d' = 1 + \frac{q\lambda}{1 - q}, \quad e' = -\frac{q\lambda}{1 - q},$$

provided these rates are all nonnegative. Here $b' \geq 0$ requires $1 - q > 0$ and hence $e' \geq 0$ requires $q \leq 0$. Now $a' \geq 0$ is automatically satisfied. Let us assume, for simplicity, that $-1 \leq q \leq 0$. Then $c' \geq 0$ is also automatically satisfied, while $d' \geq 0$ leads to the requirement $1 - q + q\lambda \geq 0$ or equivalently $(\lambda - 1)q \geq -1$, so sufficient conditions for the rates a', \dots, e' to be all nonnegative are that

$$q = 0 \quad \text{or} \quad 0 > q \geq \begin{cases} -1 & (\lambda \leq 2), \\ -(\lambda - 1)^{-1} & (\lambda > 2). \end{cases}$$

In particular, for $q = 0$, we find back the additive self-duality of the contact process.

2.8 Concluding remarks

There exists a lot of material on Markov process duality that we have been able to cover in the limited space of these notes. Several authors have tried to look for dualities in a systematic way. An interesting approach is that of [GKR07, CGGR13], who attempt to find dualities by writing Markov generators in terms of creation and annihilation operators⁷ look for symmetry groups, and link dualities to different representations of the same group.

We have restricted most of our analysis to Markov processes with state space of the form $\{0, 1\}^\Lambda$ with Λ a finite set. The extension to countably infinite Λ is in fact not hard, but skipped here for lack of space. Within this class, the approach by Lloyd and Sudbury, who look at duality functions that act ‘locally’ on each site as in (2.36), seems pretty compelling. Nevertheless, there exist useful dualities that do not fall into this class; an example can be found in [Swa13, Lemma 4].

⁷This is something that Lloyd and Sudbury [LS95, LS97] also do, though we have not made use of this in our brief exposition of their work.

Leaving the class of Markov processes with state space of the form $\{0, 1\}^\Lambda$, there exist many more dualities that we cannot cover here. Examples include processes with state space of the form \mathbb{N}^Λ and diffusion processes with state space of the form I^Λ , where $I \subset \mathbb{R}$ is some interval. As explained in [Swa06], many of these dualities can be found by starting with a Lloyd-Sudbury duality and then looking at suitable functions of processes with Λ tending to infinity, which in the right limit may converge to diffusion processes etc. Nevertheless, this approach cannot cover all cases. For example, it is difficult to see how the intricate self-duality of the mutual catalyst described in [Myt98] could be derived in such a way.

2.9 Exercises

Exercise 2.1 (Subduality) Let X and Y be Markov processes with finite state spaces S and R and generators G and H , respectively, and let $\psi : S \times R \rightarrow \mathbb{R}$ be a function. Let us say that Y is a *subdual* of X with duality function ψ if

$$\mathbb{E}^x[\psi(X_t, y)] \geq \mathbb{E}^y[\psi(x, Y_t)] \quad (x \in S, y \in R, t \geq 0). \quad (2.46)$$

Show that a necessary and sufficient condition for (2.46) is that

$$G\psi(\cdot, y)(x) \geq H\psi(x, \cdot)(y) \quad (x \in S, y \in R).$$

Let $\psi : \mathbb{R}^R \rightarrow \mathbb{R}^S$ be the operator defined in (2.5) and let π be an invariant law of Y . Show that $\psi\pi$ is a superharmonic function of X . *Note:* subduality has successfully been applied in, for example, [AS05, SS13].

Exercise 2.2 (Reversibility and duality) Let X be a Markov process with finite state space S and reversible invariant law π satisfying $\pi(x) > 0$ for all $x \in S$. Show that X is self-dual with respect to the duality function

$$\psi(x, y) = 1_{\{x=y\}}\pi(x)^{-1} \quad (x, y \in S).$$

Note Some authors, such as [GKR07], use this simple fact as a starting point to look for more interesting dualities.

Exercise 2.3 (Birth-and-death chains on \mathbb{N}) Let $(b_k)_{k \geq 1}$ and $(d_k)_{k \geq 1}$ be strictly positive constants. Let X be the continuous-time Markov process with state space \mathbb{N} ($= \{0, 1, \dots\}$) that jumps from $k-1$ to k with rate b_k and from k to $k-1$ with rate d_k . Let X' be another continuous-time Markov process with state space \mathbb{N} , that jumps from $k-1$ to k with rate $b'_k := d_{k-1}$ ($k \geq 2$) and from k to

$k - 1$ with rate $d'_k := b_k$ ($k \geq 1$). The process X' cannot jump from 0 to 1, i.e., 0 is a trap for X' . Assume that the trapping probability

$$h(k) := \mathbb{P}^k[\exists t \geq 0 \text{ s.t. } X'_t = 0]$$

satisfies $h(k) \rightarrow 0$ as $k \rightarrow \infty$. Prove that

$$\pi(k) := h(k) - h(k+1) \quad (k \geq 0)$$

is an invariant law for the process X .

Exercise 2.4 (Symmetric exclusion process) Let $\text{exc}_{i,j} : \{0,1\}^\Lambda \rightarrow \{0,1\}^\Lambda$ be the map defined by

$$\text{exc}_{i,j}(x)(k) := \begin{cases} x(i) & \text{if } k = j, \\ x(j) & \text{if } k = i, \\ x(k) & \text{otherwise.} \end{cases} \quad (2.47)$$

Show that $\text{exc}_{i,j}$ is both additive and cancellative and determine its dual maps with respect to the duality functions $\psi(x, y) = 1_{\{x \cap y \neq \emptyset\}}$ and $\psi(x, y) = \bigoplus_{i \in \Lambda} x(i)y(i)$, respectively.

Exercise 2.5 (Biased voter model (part 1)) Let Λ be a finite set and let $p : \Lambda \times \Lambda \rightarrow [0, 1]$ satisfy $\sum_j p(i, j) = 1$ and $p(i, j) = p(j, i)$. For any $x \in \{0, 1\}^\Lambda$, let

$$f_0(x, i) := \sum_{j \in \Lambda} p(i, j)(1 - x(j)) \quad \text{and} \quad f_1(x, i) := \sum_{j \in \Lambda} p(i, j)x(j)$$

denote the local frequency of zeros and ones around the site i , weighted with the kernel p . The *biased* voter model with bias $\beta \geq 0$ is the Markov process X with state space $\{0, 1\}^\Lambda$ such that in each jump of the process, only one coordinate $x(i)$ of the vector $x \in \{0, 1\}^\Lambda$ changes its value, and

$$x(i) \text{ jumps } \begin{cases} 0 \mapsto 1 & \text{with rate } (1 + \beta)f_1(x, i), \\ 1 \mapsto 0 & \text{with rate } f_0(x, i). \end{cases}$$

Show that X is dual, in the sense of additive systems duality, to a Markov process Y with the following description. Say that a site i is occupied by a particle at time t if $Y_t(i) = 1$. Then these particles form a system of branching and coalescing random walks, where a particle at i jumps with rate $p(i, j)$ to the site j , a particle at i gives with rate $\beta p(i, j)$ birth to a new particle at j , and particles that jump to or are created on an already occupied site immediately coalesce with the particle that is already present there.

Exercise 2.6 (Biased voter model (part 2)) Show that there exists some $q > 0$ (depending on $\beta > 0$) such that product measure with intensity q is a reversible law for the branching coalescing walk defined in the previous exercise.

Exercise 2.7 (Biased voter model (part 3)) The constant configurations 0 and 1 are traps for the biased voter model, and the trapping probability

$$h(x) := \mathbb{P}^x[\exists t \geq 0 \text{ s.t. } X_t = 1],$$

as a function of the initial state x , is a harmonic function for X . Give an expression for h in terms of the invariant law of the dual system of branching and coalescing random walks described in the previous exercise.

Exercise 2.8 (Contact process with spontaneous births (part 1)) Let Λ be a finite set and for each $i \in \Lambda$, define a map $\text{birth}_i : \mathcal{P}(\Lambda) \rightarrow \mathcal{P}(\Lambda)$ by

$$\text{birth}_i(x) := x \cup \{i\}.$$

Show that birth_i is not additive. Let $\Lambda_* := \Lambda \cup \{*\}$ be the set Λ with one extra adjoined element $*$. Define maps $\text{inf}_{i,*}$ from $\mathcal{P}(\Lambda_*)$ into itself in the obvious way, i.e.,

$$\text{inf}_{i,*}(x) = \begin{cases} x \cup \{*\} & \text{if } i \in x, \\ x & \text{otherwise.} \end{cases}$$

Let $\alpha, \delta \geq 0$, let $\lambda : \Lambda \times \Lambda \rightarrow [0, \infty)$ be a function and let X and Y be Markov processes with state spaces $\mathcal{P}(\Lambda)$ and $\mathcal{P}(\Lambda_*)$, respectively, and generators G and H given by

$$\begin{aligned} Gf(x) &:= \delta \sum_i (f(\text{rec}_i(x)) - f(x)) + \sum_{i \neq j} \lambda(i, j) (f(\text{inf}_{i,j}(x)) - f(x)) \\ &\quad + \alpha \sum_i (f(\text{birth}_i(x)) - f(x)), \\ Hf(y) &:= \delta \sum_i (f(\text{rec}_i(y)) - f(y)) + \sum_{i \neq j} \lambda(j, i) (f(\text{inf}_{i,j}(y)) - f(y)) \\ &\quad + \alpha \sum_i (f(\text{inf}_{i,*}(y)) - f(y)), \end{aligned}$$

where all sums run over $i, j \in \Lambda$ only (excluding $*$). Show that X and Y are pathwise dual with duality function

$$\psi(x, y) := 1_{\{x_* \cap y \neq \emptyset\}} \quad \text{where } x_* := x \cup \{*\}.$$

Note: By a similar trick, it is also possible to allow for spontaneous birth of particles in cancellative systems.

Exercise 2.9 (Contact process with spontaneous births (part 2)) Let X and Y be the contact process with spontaneous births and its dual from the previous exercise. Show that the process started in $X_0 = \Lambda$ satisfies

$$\mathbb{P}^\Lambda[X_t \in \cdot] \xrightarrow[t \rightarrow \infty]{} \mathbb{P}[X_\infty \in \cdot],$$

where X_∞ is a $\mathcal{P}(\Lambda)$ -valued random variable whose law is uniquely characterized by the relation

$$\begin{aligned} \mathbb{P}[X_\infty \cap y \neq \emptyset] &= \mathbb{P}^y[Y_t \neq \emptyset \forall t \geq 0] \\ &= \mathbb{P}^y[\exists t \geq 0 \text{ s.t. } * \in Y_t] \quad (y \in \mathcal{P}(\Lambda)). \end{aligned}$$

Note: The law of X_∞ is called the *upper invariant law*. For processes on infinite lattices, this can be nontrivial even if there are no spontaneous births.

Exercise 2.10 (Odd upper invariant law) Let X and Y be $\mathcal{P}(\Lambda)$ -valued Markov processes that are dual in the sense of cancellative systems duality. Identifying $\mathcal{P}(\Lambda) \cong \{0, 1\}^\Lambda$, let $\pi_{1/2}$ denote product measure on $\{0, 1\}^\Lambda$ with intensity $1/2$. Show that the process X started in the initial law $\pi_{1/2}$ satisfies

$$\mathbb{P}^{\pi_{1/2}}[X_t \in \cdot] \xrightarrow[t \rightarrow \infty]{} \mathbb{P}[X_\infty \in \cdot],$$

where X_∞ is a $\mathcal{P}(\Lambda)$ -valued random variable whose law is uniquely characterized by the relation

$$\mathbb{P}[|X_\infty \cap y| \text{ is odd}] = \frac{1}{2} \mathbb{P}^y[Y_t \neq \emptyset \forall t \geq 0] \quad (y \in \mathcal{P}(\Lambda)).$$

Hint Recall the form of the cancellative systems duality function from (2.33).

Exercise 2.11 (Moran model) Let Λ be a set with $|\Lambda| = N$ elements and let X be the voter model with generator (compare (2.25))

$$Gf(x) = \frac{1}{2} \sum_{i \neq j} (f(\text{vot}_{i,j}(x)) - f(x)).$$

Show that

$$M_t := |X_t| \quad (t \geq 0)$$

defines a Markov process $(M_t)_{t \geq 0}$ with state space $\{0, \dots, N\}$, and determine its generator. Show that M has a dual process $K = (K_t)_{t \geq 0}$ with state space $\{1, \dots, N\}$, with respect to the duality function

$$\psi(m, k) := \prod_{i=0}^{k-1} \frac{N - m - i}{N - i},$$

which corresponds to the probability that a sample of k individuals does not contain an individual of type one, when drawn (without replacement) from a population of N individuals of which m are of type one. Determine the generator of K . *Note* The process M is a continuous-time version of the *Moran model*, which is used in mathematical population dynamics.

Exercise 2.12 (Self-duality) In the set-up of Theorem 2.8, show that each model with $b > 0$ and $q := (d - a - c)/b \neq 1$ is self-dual with parameter q . Use this to show that the biased voter model with rates

$$a = 0, \quad b = 1 + \beta, \quad c = 0, \quad d = 1, \quad e = 0,$$

where $\beta > 0$, is self-dual. What happens if the bias β is zero?

Exercise 2.13 (Weak stochastic order) Let Λ be a finite set. The laws $\mu = \mathbb{P}[X \in \cdot]$ and $\mu' = \mathbb{P}[X' \in \cdot]$ of two $\mathcal{P}(\Lambda)$ -valued random variables X and X' are called *stochastically ordered*, denoted $\mu \leq \mu'$, if X and X' can be coupled such that $X \leq X'$. It is known that this is equivalent to the statement that $\mathbb{E}[f(X)] \leq \mathbb{E}[f(X')]$ for each monotone function $f : \{0, 1\}^\Lambda \rightarrow \mathbb{R}$ [Lig85, Thm II.2.4]. Let us say that μ and μ' are *weakly stochastically ordered*, denoted $\mu \prec \mu'$, if $\mathbb{P}[X \cap y \neq \emptyset] \leq \mathbb{P}[X' \cap y \neq \emptyset]$ for all $y \in \mathcal{P}(\Lambda)$. Prove the following statements.

- (i) $\mu \leq \mu'$ implies $\mu \prec \mu'$ but the converse statement does not hold.
- (ii) If $(P_t)_{t \geq 0}$ is the semigroup of a monotone Markov process, then $\mu \leq \mu'$ implies $\mu P_t \leq \mu' P_t$ for all $t \geq 0$.
- (iii) If $(P_t)_{t \geq 0}$ is the semigroup of an additive Markov process, then $\mu \prec \mu'$ implies $\mu P_t \prec \mu' P_t$ for all $t \geq 0$.

Chapter 3

Intertwining

3.1 Markov functionals

Let $X = (X_k)_{k \geq 0}$ be a Markov chain with finite state space S and transition kernel P , let $f : S \rightarrow R$ be a surjective function from S onto some other space R , and let $Y = (Y_k)_{k \geq 0}$ be the chain given by

$$Y_k := f(X_k) \quad (k \geq 0).$$

We ask ourselves the following question: Under what conditions is Y itself a Markov chain with some transition kernel Q ?

By definition, we say that $(Y_k)_{k \geq 0} = (f(X_k))_{k \geq 0}$ is *autonomous* if

$$f(x) = f(x') \quad \text{implies} \quad \mathbb{P}^x[f(X_1) = y] = \mathbb{P}^{x'}[f(X_1) = y] \quad (3.1)$$

for all $x, x' \in S$ and $y \in R$.

Lemma 3.1 (Autonomous Markov chain) *In the set-up above, if Y is autonomous, then Y is a Markov chain with transition kernel Q given by*

$$Q(y, y') := \mathbb{P}^x[f(X_1) = y'] = \sum_{x' \in S} 1_{\{f(x')=y'\}} P(x, x')$$

$(x \in S, y, y' \in R, f(x) = y)$.

Proof We note that the fact that f is surjective says that for each $y \in R$ there exists an $x \in S$ with $f(x) = y$. Since by the definition of autonomy, $\mathbb{P}^x[f(X_1) = y']$ does not depend on the choice of such an x , this shows that Q is well-defined.

Now, by the Markov property of X and the definition of autonomy,

$$\begin{aligned}
& \mathbb{P}[Y_{k+1} = y \mid (Y_0, \dots, Y_k) = (y_0, \dots, y_k)] \\
&= \sum_{x \in S} 1_{\{f(x)=y_k\}} \mathbb{P}[Y_{k+1} = y \mid X_k = x, (Y_0, \dots, Y_k) = (y_0, \dots, y_k)] \\
&\quad \cdot \mathbb{P}[X_k = x \mid (Y_0, \dots, Y_k) = (y_0, \dots, y_k)] \\
&= \sum_{x \in S} 1_{\{f(x)=y_k\}} \mathbb{P}[Y_{k+1} = y \mid X_k = x] \cdot \mathbb{P}[X_k = x \mid (Y_0, \dots, Y_k) = (y_0, \dots, y_k)] \\
&= Q(y_k, y) \sum_{x \in S} 1_{\{f(x)=y_k\}} \cdot \mathbb{P}[X_k = x \mid (Y_0, \dots, Y_k) = (y_0, \dots, y_k)] \\
&= Q(y_k, y).
\end{aligned}$$

■

It seems that Lemma 3.1 is more or less optimal when the aim is to ensure the Markov property of Y by putting restrictions on the transition kernel of X alone. If, however, one is also prepared to put restrictions on the initial law of X , then it turns out that there are interesting cases where Y , on its own, is a Markov process, even though it is not autonomous.

Proposition 3.2 (Markov functionals) *Let S, R be finite spaces, let P, Q be probability kernels on S and R , respectively, let $f : S \rightarrow R$ be a surjective function, and let K be a probability kernel from R to S such that*

$$K(y, x) = 0 \quad \text{whenever} \quad f(x) \neq y \quad (x \in S, y \in R).$$

Let $X = (X_k)_{k \geq 0}$ be a Markov chain with transition kernel P and let $Y_k := f(X_k)$ ($k \geq 0$). Assume that

$$QK = KP. \tag{3.2}$$

Then, if X is started in an initial law such that

$$\mathbb{P}[X_0 = x \mid Y_0] = K(Y_0, x) \quad \text{a.s.} \quad (x \in S), \tag{3.3}$$

this implies that

$$\mathbb{P}[X_k = x \mid (Y_0, \dots, Y_k)] = K(Y_k, x) \quad \text{a.s.} \quad (x \in S), \tag{3.4}$$

and Y , on its own, is a Markov chain with transition kernel Q .

Proof We start by proving (3.4). For each $x \in S$ and $(y_0, \dots, y_k) \in R^k$ such that $\mathbb{P}[(Y_0, \dots, Y_k) = (y_0, \dots, y_k)] > 0$, let us define

$$\pi(x \mid y_0, \dots, y_k) := \mathbb{P}[X_k = x \mid (Y_0, \dots, Y_k) = (y_0, \dots, y_k)].$$

Let us also introduce the notation

$$P(x, x'; y) := 1_{\{f(x')=y\}} P(x, x') \quad (x, x' \in S, y \in R).$$

Then (3.2) and the fact that $K(y, x) = 0$ whenever $f(x) \neq y$ imply that

$$\begin{aligned} \sum_{x \in S} K(y, x) P(x, x'; y') &= 1_{\{f(x')=y'\}} (KP)(y, x') = 1_{\{f(x')=y'\}} (QK)(y, x') \\ &= \sum_{y'' \in R} Q(y, y'') K(y'', x') 1_{\{f(x')=y'\}} = Q(y, y') K(y', x') \end{aligned} \quad (3.5)$$

Filtering theory tells us that there is a systematic way of calculating the functions $\pi(x | y_0, \dots, y_k)$ for $k = 0, 1, \dots$. For $k = 0$, obviously, by (3.3),

$$\pi(x | y_0) = \mathbb{P}[X_0 = x | Y_0 = y_0] = K(y_0, x). \quad (3.6)$$

The *filtering equations* now tell us that for $k \geq 1$,

$$\pi(x | y_0, \dots, y_k) = \frac{\sum_{x' \in S} P(x', x; y_k) \pi(x' | y_0, \dots, y_{k-1})}{\sum_{x', x'' \in S} P(x', x''; y_k) \pi(x' | y_0, \dots, y_{k-1})}, \quad (3.7)$$

which can be proved by writing the left-hand side as

$$\frac{\mathbb{P}[X_k = x, Y_k = y_k | (Y_0, \dots, Y_{k-1}) = (y_0, \dots, y_{k-1})]}{\mathbb{P}[Y_k = y_k | (Y_0, \dots, Y_{k-1}) = (y_0, \dots, y_{k-1})]},$$

which obviously equals the right-hand side of (3.7).

To prove (3.4), we need to show that

$$\pi(x | y_0, \dots, y_k) = K(x, y_k) \quad (k \geq 0).$$

Formula (3.6) says that this is OK for $k = 0$. By induction, using the filtering equations (3.7) and (3.5),

$$\begin{aligned} \pi(x | y_0, \dots, y_{k+1}) &= \frac{\sum_{x' \in S} P(x', x; y_{k+1}) K(y_k, x')}{\sum_{x', x'' \in S} P(x', x''; y_{k+1}) K(y_k, x')} \\ &= \frac{Q(y_k, y_{k+1}) K(y_{k+1}, x)}{\sum_{x'' \in S} Q(y_k, y_{k+1}) K(y_{k+1}, x'')} = K(y_{k+1}, x), \end{aligned}$$

completing the induction step in the proof of (3.4).

Now, by the Markov property of X and formulas (3.4) and (3.5),

$$\begin{aligned}
& \mathbb{P}[Y_{k+1} = y \mid (Y_0, \dots, Y_k) = (y_0, \dots, y_k)] \\
&= \sum_{x \in S} \mathbb{P}[Y_{k+1} = y \mid X_k = x, (Y_0, \dots, Y_k) = (y_0, \dots, y_k)] \\
&\quad \cdot \mathbb{P}[X_k = x \mid (Y_0, \dots, Y_k) = (y_0, \dots, y_k)] \\
&= \sum_{x \in S} \mathbb{P}[Y_{k+1} = y \mid X_k = x] \pi(x \mid y_0, \dots, y_k) \\
&= \sum_{x, x' \in S} P(x, x'; y) K(y_k, x) = \sum_{x' \in S} Q(y_k, y) K(y_k, x') = Q(y_k, y),
\end{aligned} \tag{3.8}$$

proving that Y is a Markov chain with transition kernel Q . ■

3.2 Intertwining of Markov processes

In algebra, a relation between operators A, B, C of the form

$$AB = BC$$

is called an *intertwining relation*. In particular, if B is invertible, this implies that $A = BCB^{-1}$; the term ‘intertwining’, however, is especially used in situations where B is not (necessarily) invertible.

The relation (3.2) is thus an example of an intertwining relation as is the abstract formulation of duality in (2.6). We now make a more specific definition. Let X and Y be Markov chains with finite state spaces S and R and transition kernels P and Q , respectively, and let K be a probability kernel from R to S (in this order!). If

$$QK = KP, \tag{3.9}$$

then we say that X and Y are *intertwined*. Since the inverse of a probability kernel, if it exists, is usually not a probability kernel, intertwining of Markov processes (unlike duality) is not a symmetric relation. To distinguish the different roles of X and Y , we will say that X is an *intertwined chain on top of Y* .

By induction, (3.9) implies $Q^k K = KP^k$ ($k \geq 0$), so intertwining relations make a connection between the *forward* equations of X and the *forward* equations of Y . (Recall from (2.8) that a duality links the *forward* evolution of one process to the *backward* evolution of the other process.)

It seems the use of the word intertwining in this specific sense for Markov chains was first introduced by Marc Yor in the preprint [Yor88], and then quickly adopted by Diaconis and Fill [DF90] who proved Proposition 3.3 below. Since Marc Yor's work appeared only several years later and in a form quite different from the original preprint, the paper by Diaconis and Fill seems to be the first place in the literature where intertwining of Markov chains occurs as an abstract concept. Although examples of intertwining of Markov processes have been known and studied (without being called so) for a long time, the more systematic investigation of the subject is quite new. Much of it has been motivated by the study of mixing times (the book [LPW09] is a good general reference of this concept). Proposition 3.2 on Markov functionals can be traced back to Rogers and Pitman [RP81].

Note that in general (unlike the set-up in Proposition 3.2), we do not assume that there exists a function $f : S \rightarrow R$ such that $K(y, x) = 0$ unless $f(x) = y$. In this more general context, a result similar to Proposition 3.2 holds, but we have to formulate it somewhat differently since Y is in general no longer a function of X .

Proposition 3.3 (Intertwining of Markov chains) *Let S, R be finite spaces, let P, Q be probability kernels on S and R , respectively, let K be a probability kernel from R to S , and assume that (3.9) holds. Then there exists a Markov chain $(X, Y) = (X_k, Y_k)_{k \geq 0}$ with state space $\hat{S} := \{(x, y) \in S \times R : K(y, x) > 0\}$ such that*

- (i) X is autonomous with transition kernel P ,

and moreover, the condition

$$\mathbb{P}[X_0 = x \mid Y_0] = K(Y_0, x) \quad \text{a.s.} \quad (x \in S) \quad (3.10)$$

implies that

- (ii) Y , on its own, is a Markov chain with transition kernel Q ,
- (iii) $\mathbb{P}[X_k = x \mid (Y_0, \dots, Y_k)] = K(Y_k, x) \quad \text{a.s.} \quad (k \geq 0, x \in S).$

Remark Proposition 3.3 contains Proposition 3.2 as a special case. Indeed, if there exists a function $f : S \rightarrow R$ such that $K(y, x) = 0$ unless $f(x) = y$, then we can naturally identify the sets S and \hat{S} through the bijection $x \mapsto (x, f(x))$.

Proof The strategy will be to construct the transition kernel of (X, Y) ‘by hand’ and then apply Proposition 3.2 to the joint process (X, Y) . For each $x' \in S$ and $y \in R$ such that $QK(y, x') > 0$, we define a probability law $Q_{x'}(y, \cdot)$ on R by

$$Q_{x'}(y, y') := \frac{Q(y, y')K(y', x')}{QK(y, x')}.$$

For notational convenience, if $QK(y, x') = 0$, then we choose for $Q_{x'}(y, \cdot)$ some arbitrary probability on R . Next, we define a probability kernel \hat{P} on $S \times R$ by

$$\hat{P}(x, y; x', y') := P(x, x')Q_{x'}(y, y') \quad (x, x' \in S, y, y' \in R).$$

We claim that if $(x, y) \in \hat{S}$, then $\hat{P}(x, y; \cdot)$ is concentrated on \hat{S} and its definition does not depend on our arbitrary choice of $Q_{x'}(y, \cdot)$ for $QK(y, x') = 0$. Indeed, if $(x, y) \in \hat{S}$ and $\hat{P}(x, y; x', y') > 0$, then $P(x, x') > 0$ while $K(y, x) > 0$ by the definition of \hat{S} , so $K(y, x)P(x, x') > 0$ which by (3.9) implies that $QK(y, x') = KP(y, x') > 0$ and hence $Q_{x'}(y, \cdot)$ is unambiguous. Now $\hat{P}(x, y; x', y') > 0$ implies $Q_{x'}(y, y') > 0$ which by the (unambiguous) definition of the latter implies $K(y', x') > 0$ and hence $(x', y') \in \hat{S}$.

Let \hat{K} be the kernel from R to \hat{S} defined by

$$\hat{K}(y; x', y') := K(y, x')1_{\{y=y'\}},$$

and let $f : \hat{S} \rightarrow R$ be defined by $f(x, y) := y$. Then $\hat{K}(y; x', y') > 0$ implies $f(x', y') = y$. We claim that moreover

$$Q\hat{K} = \hat{K}\hat{P}.$$

Indeed, by (3.9),

$$\begin{aligned} \hat{K}\hat{P}(y; x', y') &= \sum_{(x'', y'') \in \hat{S}} K(y; x'', y'')P(x'', x')Q_{x'}(y'', y') \\ &= \sum_{x'' \in S} K(y, x'')P(x'', x') \frac{Q(y, y')K(y', x')}{QK(y, x')} = \sum_{x'' \in S} KP(y, x'') \frac{Q(y, y')K(y', x')}{KP(y, x')} \\ &= Q(y, y')K(y', x') = \sum_{y''} Q(y, y'')\hat{K}(y''; x', y') = Q\hat{K}(y; x', y'). \end{aligned}$$

This shows that all assumptions of Proposition 3.2 are satisfied. Claims (ii) and (iii) are now immediate from that proposition, while the fact that X is autonomous with kernel P is clear from our definition of \hat{P} . \blacksquare

We note that condition (3.10) does not put any restrictions on the initial law of Y . Proposition 3.3 can therefore be read in such a way that the intertwining condition (3.9) implies that the process Y , started in an arbitrary initial law, can be coupled to a process X such that (i) and (iii) hold. We may view X as some *added structure* that we have added on top of Y . Proposition 3.3 has a continuous-time analogue, that we cite here from [Fil92]. We note that an extension of this result, where X need not be autonomous (but in applications may be almost autonomous) is proved in [AS10].

Proposition 3.4 (Intertwining of Markov processes) *Let G, H be generators of Markov processes with finite state spaces S, R , let K be a probability kernel from R to S , and assume that*

$$HK = KG. \quad (3.11)$$

Then there exists a Markov process $(X, Y) = (X_t, Y_t)_{t \geq 0}$ with state space $\hat{S} := \{(x, y) \in S \times R : K(y, x) > 0\}$ such that

- (i) *X is autonomous with generator G ,*

and moreover, the condition

$$\mathbb{P}[X_0 = x \mid Y_0] = K(Y_0, x) \quad \text{a.s.} \quad (x \in S)$$

implies that

- (ii) *Y , on its own, is a Markov process with generator H ,*

- (iii) $\mathbb{P}[X_t = x \mid (Y_s)_{0 \leq s \leq t}] = K(Y_t, x) \quad \text{a.s.} \quad (t \geq 0, x \in S).$

Remark In particular, if there exists a function $f : S \rightarrow R$ such that $K(y, x) = 0$ unless $f(x) = y$, then by identifying the sets S and \hat{S} through the bijection $x \mapsto (x, f(x))$, one may derive a continuous-time analogue of Proposition 3.2.

3.3 Thinning

Let Λ be a finite set and let $(\chi(i))_{i \in \Lambda}$ be an i.i.d. collection of *Bernoulli* (i.e., $\{0, 1\}$ -valued) random variables with intensity $\mathbb{P}[\chi(i) = 1] = p \in [0, 1]$. Then $\chi = (\chi(i))_{i \in \Lambda}$ is a $\{0, 1\}^\Lambda$ -valued random variable. Alternatively, identifying sets with their indicator functions, we may view $\chi = \{i \in \Lambda : \chi(i) = 1\}$ as a random element of $\mathcal{P}(\Lambda)$, the space of all subsets of Λ . If x is another $\mathcal{P}(\Lambda)$ -valued random variable, independent of χ , then we call

$$\text{Thin}_p(x) := x \cap \chi$$

a *p-thinning* of x . Note that $x \cap \chi$ is obtained from x by independently throwing away elements of x (with probability $1 - p$) or keeping them (with probability p). We may define a probability kernel T_p on $\mathcal{P}(\Lambda)$ by

$$T_p(x, y) := \mathbb{P}[x \cap \chi = y] \quad (x, y \in \mathcal{P}(\Lambda)), \quad (3.12)$$

where in this formula, x is (of course) deterministic. It is not hard to see that the thinning of a thinning is again a thinning:

$$T_p T_{p'} = T_{pp'} \quad (p, p' \in [0, 1]). \quad (3.13)$$

Thinnings are closely related to the Lloyd-Sudbury duals of Theorem 2.8. Recall from (2.44) that $\mathcal{P}(\Lambda)$ -valued Markov processes X and Y are called q -dual for some $q \in \mathbb{R} \setminus \{1\}$ if they are dual with respect to the duality function

$$\psi_q(x, y) = q^{|x \cap y|} \quad (x, y \in \mathcal{P}(\Lambda)).$$

The following proposition is (more or less) [Sud00, Thm 2.1].

Proposition 3.5 (Thinnings and q -duality) *Let X, X' and Y be Markov processes with state space $\mathcal{P}(\Lambda)$ and generators G, G' and H , respectively. Assume that Y is a q -dual of X and a q' -dual of X' , for constants $q, q' \in \mathbb{R} \setminus \{1\}$ satisfying*

$$p := \frac{1 - q}{1 - q'} \in [0, 1].$$

Then the generators of X and X' satisfy the intertwining relation

$$GT_p = T_p G'.$$

In particular, the process X , started in an arbitrary initial law, can be coupled to a process X' such that

- (i) X' is an autonomous Markov process with generator G ,
- (ii) $\mathbb{P}[X'_t \in \cdot \mid (X_s)_{0 \leq s \leq t}] = T_p(X_t, \cdot) \quad \text{a.s.} \quad (t \geq 0).$

Proof We will give an algebraic proof. For a proof with a more probabilistic flavour, see [Swa06, Lemma 2 in the version on the ArXiv]. Let ψ_q denote the linear operator on $\mathbb{R}^{\mathcal{P}(\Lambda)}$ with matrix $\psi_q(x, y)$. We claim that ψ_q is invertible for each $q \neq 1$ and that

$$\psi_q \psi_{q'}^{-1} = T_p \quad \text{provided that} \quad p = \frac{1 - q}{1 - q'} \in [0, 1]. \quad (3.14)$$

We observe that both $\psi_{q'}$ and T_p are products of local operators that act on a single coordinate only. In view of this, it suffices to prove the statement for the

case that Λ consists of a single element and hence $\mathcal{P}(\Lambda) \cong \{0, 1\}^\Lambda = \{0, 1\}$. We recall from (2.40) that

$$\begin{pmatrix} \psi_q(0, 0) & \psi_q(0, 1) \\ \psi_q(1, 0) & \psi_q(1, 1) \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & q \end{pmatrix}.$$

It is straightforward to check that

$$\begin{pmatrix} 1 & 1 \\ 1 & q \end{pmatrix}^{-1} = (1 - q)^{-1} \begin{pmatrix} -q & 1 \\ 1 & -1 \end{pmatrix}.$$

For a single site, the matrix of T_p is given by

$$\begin{pmatrix} T_p(0, 0) & T_p(0, 1) \\ T_p(1, 0) & T_p(1, 1) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 - p & p \end{pmatrix}.$$

Now

$$\psi_q \psi_{q'}^{-1} = (1 - q')^{-1} \begin{pmatrix} 1 & 1 \\ 1 & q \end{pmatrix} \begin{pmatrix} -q' & 1 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \frac{q - q'}{1 - q'} & \frac{1 - q}{1 - q'} \end{pmatrix} = T_p,$$

provided that $(1 - q)/(1 - q') = p \in [0, 1]$.

By (2.7), the fact that Y is a q -dual of X and a q' -dual of X' may algebraically be expressed as

$$\psi_q^{-1} G \psi_q = H^\dagger \quad \text{and} \quad G' = \psi_{q'} H^\dagger \psi_{q'}^{-1}.$$

Using (3.14), it follows that

$$G T_p = G \psi_q \psi_{q'}^{-1} = \psi_q H^\dagger \psi_{q'}^{-1} = \psi_q \psi_{q'}^{-1} G' = T_p G'.$$

The rest of the statements are now immediate from Proposition 3.4. ■

Remark In our proof, we have actually never used that H is a Markov generator. In view of this, all conclusions of Proposition 3.5 remain true if Y is only a *formal* dual. In particular, when we apply Theorem 2.8 to find three processes X, X' and Y such that Y is a q -dual of X and a q' -dual of X' , then it is not necessary that the rates a, b, c, d, e corresponding to Y are all nonnegative, even though this means, of course, that H is not a Markov generator.

Annihilating and coalescing random walks

For $0 \leq \alpha \leq 1$, let X^α be the process with generator

$$G_\alpha f(x) = \sum_{i \neq j} p(i, j) \{ \alpha (f(\text{ann}_{i,j}(x)) - f(x)) + (1 - \alpha) (f(\text{rw}_{i,j}(x)) - f(x)) \}.$$

as in (2.45), i.e., X^α is a collection of random walks with the property that if a particles lands on an occupied site, the two particles annihilates with probability α and coalesce with probability $1 - \alpha$. As we have seen near the end of Section 2.7, such a system is q -dual to the voter model, with $q = -\alpha$. Applying Proposition 3.5 we see that for any $0 \leq \alpha \leq \alpha' \leq 1$, the process X^α (started in an arbitrary initial law) can be coupled to a process $X^{\alpha'}$ in such a way that

$$\mathbb{P}[X_t^{\alpha'} \in \cdot \mid (X_s^\alpha)_{0 \leq s \leq t}] = T_{(1+\alpha)/(1+\alpha')}(X_t^\alpha, \cdot) \quad \text{a.s.} \quad (t \geq 0).$$

In particular, setting $\alpha = 0$ and $\alpha' = 1$, we see that we can couple a system of coalescing random walks to a system of annihilating random walks in such a way that the latter are a $1/2$ -thinning of the former.

3.4 Concluding remarks

Our discussion of duality has focussed on Markov processes with state space of the form $\{0, 1\}^\Lambda$. For these processes, we have seen that a large class of duals, that contains most of the known examples, consists of Lloyd and Sudbury's q -duals. In view of the close connection between q -duality and thinning, one might have the impression that most of the intertwining occurring in practice are thinning relations. Little could be further from the truth.

Intertwining relations between Markov processes are very common. Indeed, this is such a diverse subject, and at the same time such a young subject, that it is hard to give anything like a comprehensive overview. A thorough discussion of the matter should at least discuss the interesting intertwining for birth-and-death processes discovered by Diaconis and Miclo [DM09] and elaborated on in [Swa11], as well as some examples of how intertwining can be used to construct strong stationary times and give bounds on mixing times. For this subject, see [DF90, Fil92, DM09] and also the evolving-set process discussed in [LPW09, Chap. 17] (even though the latter do not formalize the concept of intertwining). Intertwining can also nicely be applied to give lower bounds on the time till absorption, as is done in [AS10]. To give the reader at least some idea of the various sorts of intertwining that can be constructed and made use of, we have tried to include a wide range of examples in the exercises.

3.5 Exercises

Exercise 3.1 (Autonomous Markov chain) Let $X = (X_k)_{k \geq 0}$ be a Markov chain with finite state space S and transition kernel P , let $f : S \rightarrow R$ be a surjective function from S onto some other space R , and let $Y_k := f(X_k)$ ($k \geq 0$). Recall that a random mapping representation for P is a probability space (E, \mathcal{E}, μ) together with a measurable map $\phi : S \times E \rightarrow S$ such that (compare (1.10))

$$P(x, x') = \mu(\{z \in E : \phi(x, z) = x'\}) \quad (x, x' \in S).$$

Show that $Y = (Y_k)_{k \geq 0}$ is an autonomous Markov chain if and only if there exists such a random mapping representation for P with the additional property that

$$f(x) = f(x') \quad \text{implies} \quad f(\phi(x, z)) = f(\phi(x', z)) \quad (x, x' \in S, z \in E).$$

Note that this says that when we construct X inductively as $X_k = \phi(X_{k-1}, Z_k)$, where the $(Z_k)_{k \geq 1}$ are i.i.d. with common law μ , then $Y_k = f(\phi(X_{k-1}, Z_k))$ is a function of Y_{k-1} and Z_k only.

Exercise 3.2 (Thinning semigroup) Let Λ be a finite set and let X be the Markov process with state space $\mathcal{P}(\Lambda)$ and generator

$$Gf(x) := \sum_{i \in \Lambda} (f(x \setminus \{i\}) - f(x)).$$

Let $(P_t)_{t \geq 0}$ be the transition kernels of X . Show that

$$P_t = T_{e^{-t}} \quad (t \geq 0),$$

where T_p is the thinning kernel defined in (3.12).

Exercise 3.3 (Conditioning on the future) Fix $N \geq 2$ and let $X = (X_t)_{t \geq 0}$ be the Markov process with state space $\{0, \dots, N\}$ and generator

$$Gf(x) := \begin{cases} \frac{1}{2}(f(x+1) + f(x-1) - 2f(x)) & \text{if } 0 < x < N, \\ 0 & \text{if } x = 0, N. \end{cases}$$

Note that X is a nearest-neighbor random walk on $\{0, \dots, N\}$ that gets trapped in $0, N$. Let p be the function

$$p(x) := \frac{x}{N}.$$

Define a probability kernel K from $\{0, \dots, N\}$ to the space

$$R := \{(x, 1) : 1 \leq x \leq N\} \cup \{(x, 2) : 0 \leq x \leq N-1\}$$

by

$$K(x; x', a) := \begin{cases} 1_{\{x=x'\}}p(x) & \text{if } a = 1, \\ 1_{\{x=x'\}}(1 - p(x)) & \text{if } a = 2. \end{cases}$$

Show that there exists a Markov process Y with state space R and generator H satisfying the intertwining relation

$$GK = KH.$$

The process Y corresponds to a ‘richer’ version of X where the process knows in advance in which trap it will get absorbed.

Exercise 3.4 (Leading eigenvector) Let X be the Markov process from the previous exercise, assume that N is even and let f be the function

$$f(x) := \sin\left(\frac{\pi x}{N}\right) \quad (0 \leq x \leq N).$$

Define a probability kernel K from $\{0, \dots, N\}$ to the space $\{1, 2\}$ by

$$K(x, a) := \begin{cases} f(x) & \text{if } a = 1, \\ 1 - f(x) & \text{if } a = 2. \end{cases}$$

Show that there exists a Markov process Y with state space $\{1, 2\}$ and generator H satisfying the intertwining relation

$$GK = KH.$$

Use this to give a lower bound on the probability

$$\mathbb{P}^x[X_t \notin \{0, N\}] \quad (0 < x < N, t \geq 0).$$

Exercise 3.5 (Look-down construction) Let M be a Markov process with state space $\{0, \dots, N\}$ and generator

$$Gf(m) := \frac{1}{2}m(N - m)(f(m + 1) + f(m - 1) - 2f(m)).$$

(Compare Exercise 2.11.) Let $\Lambda = \{1, \dots, N\}$ and let Y be the Markov process with state space $\mathcal{P}(\Lambda)$ and generator

$$Hf(y) := \sum_{i < j} (f(\text{vot}_{i,j}(y)) - f(y)),$$

which is a voter model in which a site j can only adopt the type of a site i if i lies below j . For $m \in \{0, \dots, N\}$, write $[1 : m] := \{0, \dots, m\}$ (with $[1 : 0] := \emptyset$) and define a probability kernel K from $\{0, \dots, N\}$ to $\mathcal{P}(\Lambda)$ by

$$K(m, y) := \frac{1}{N!} \sum_{\pi} 1_{\{\pi([0:m])=y\}} \quad (m \in [0 : N], y \in \mathcal{P}(\Lambda)),$$

where the sum runs over all permutations π of $\Lambda = \{1, \dots, N\}$ and $\pi([0 : m]) := \{\pi(k) : k \leq m\}$ denotes the image of the discrete interval $[1 : m]$ under π . Show that G, H and K satisfy the intertwining relation

$$GK = KH.$$

The process Y is a ‘richer’ version of the process X in which a lot of information is already known from the start (such as in which trap X will end up!). Note that it is easy to define the limit as $N \rightarrow \infty$ of the process Y . This may be used to study the scaling limit as $N \rightarrow \infty$ of the process X , which is the Wright-Fisher diffusion. This ‘look-down’ construction was invented by Donnelly and Kurtz [DK99].

Exercise 3.6 (Thinnings of the biased voter model) Fix $\beta > 0$ and in the set-up of Theorem 2.8, let X be the biased voter model defined by the rates

$$a = 0, \quad b = 1 + \beta, \quad c = 0, \quad d = 1, \quad e = 0.$$

Let Y be its additive dual ($q = 0$), defined by the rates

$$a' = 0, \quad b' = \beta, \quad c = 1, \quad d = 0, \quad e = 1,$$

which corresponds to a system of branching and coalescing random walks with branching rate β . Show that X and Y can be coupled such that

$$\mathbb{P}[Y_t \in \cdot \mid (X_s)_{0 \leq s \leq t}] = \mathbb{P}[\text{Thin}_{\beta/(1+\beta)}(X_t) \in \cdot] \quad \text{a.s.} \quad (t \geq 0).$$

Use your result to show that product measure with intensity $\beta/(1 + \beta)$ is an invariant law for the process Y . *Hint* Use Theorem 3.5 and the self-duality of the biased voter model (see Exercise 2.12). *Note* This intertwining can be used to give bounds on the time needed by the process Y started with a single particle to converge to its invariant law.

Bibliography

- [AS05] S.R. Athreya and J.M. Swart. Branching-coalescing particle systems. *Prob. Theory Relat. Fields*. 131(3), 376–414, 2005.
- [AS10] S.R. Athreya and J.M. Swart. Survival of contact processes on the hierarchical group. *Prob. Theory Relat. Fields* 147(3) (2010), 529–563.
- [CGGR13] G. Carinci, C. Giardinà, Claudio Giberti, and Frank Redig. Dualities in population genetics: a fresh look with new dualities. Preprint (2013), ArXiv:1302.3206v1.
- [CR83] J.T. Cox and U. Rösler. A duality relation for entrance and exit laws for Markov processes. *Stochastic Processes Appl.* 16 (1984), 141–156.
- [DK99] P. Donnelly and T.G. Kurtz. Particle representations for measure-valued population models. *Ann. Probab.* 27(1) (1999), 166–205.
- [DF90] P. Diaconis and J. Fill. Strong stationary times via a new form of duality. *Ann. Probab.* 18(4), 1483–1522, 1990.
- [DM09] P. Diaconis and L. Miclo. On times to quasi-stationarity for birth and death processes. *J. Theor. Probab.* 22, 558–586, 2009.
- [Fil92] J.A. Fill. Strong stationary duality for continuous-time Markov chains. I. Theory. *J. Theor. Probab.* 5(1), 45–70, 1992.
- [GKR07] C. Giardinà, J. Kurchan, and Frank Redig. Duality and exact correlations for a model of heat conduction. *J. Math. Phys.* 48 (2007), 033301 (15 pages).
- [Gri79] D. Griffeath. *Additive and Cancellative Interacting Particle Systems*. Lecture Notes in Math. 724, Springer, Berlin, 1979.
- [JK12] S. Jansen and N. Kurt. On the notion(s) of duality for Markov processes. Preprint, 50 pages. ArXiv:1210.7193v1.

- [KM57] S. Karlin and J. McGregor. The classification of birth and death processes. *Trans. Am. Math. Soc.* 86 (1957), 366–400.
- [Lig85] T.M. Liggett. *Interacting Particle Systems*. Springer-Verlag, New York, 1985.
- [LPW09] D.A. Levin, Y. Peres and E.L. Wilmer. *Markov Chains and Mixing Times*. AMS, Providence, 2009.
- [LS95] A. Sudbury and P. Lloyd. Quantum operators in classical probability theory. II: The concept of duality in interacting particle systems. *Ann. Probab.* 23(4) (1995), 1816–1830.
- [LS97] A. Sudbury and P. Lloyd. Quantum operators in classical probability theory. IV: Quasi-duality and thinnings of interacting particle systems. *Ann. Probab.* 25(1) (1997), 96–114.
- [Myt98] L. Mytnik. Uniqueness for a mutually catalytic branching model. *Prob. Theory Relat. Fields* 112(2) (1998), 245–253.
- [RP81] L.C.G. Rogers and J.W. Pitman. Markov functions. *Ann. Probab.* 9(4) (1981), 573–582.
- [Sie76] D. Siegmund. The equivalence of absorbing and reflecting barrier problems for stochastically monotone Markov processes. *Ann. Probab.* 4 (1976), 914–924.
- [SS13] A. Sturm and J.M. Swart. A particle system with cooperative branching and coalescence. *Preprint*, 2013.
- [Sud00] A. Sudbury. Dual families of interacting particle systems on graphs. *J. Theor. Probab.* 13(3) (2000), 695–716.
- [Swa06] J.M. Swart. Duals and thinnings of some relatives of the contact process. Pages 203–214 in: *Prague Stochastics 2006*, M. Hušková and M. Janžura (eds.), Matfyzpress, Prague, 2006. ArXiv:math.PR/0604335.
- [Swa11] J.M. Swart. Intertwining of birth-and-death processes. *Kybernetika* 47(1), 1–14, 2011.
- [Swa13] J.M. Swart. Noninvadability implies noncoexistence for a class of cancellative systems. *Electron. Commun. Probab.* 18 (2013), paper No. 38, 1–12.

- [Yor88] M. Yor. Intertwinings of Bessel processes. Technical report, University of California, Berkeley, 1988.