Appendix B

Discrete-Space Markov Processes

by Robert J. Vanderbei

In this section we introduce the basic concepts of continuous-time Markov processes with discrete state space E. We assume that the reader has some familiarity with Markov processes and that this section serves mostly as a review. A good reference for this material is Dynkin [Dy].

B.1. Generators and Transition Semigroups

In the discrete-space setting, Markov processes are especially easy to grasp. Basically, if the process is at some state $x \in E$, then it will sit there until an exponential clock ticks, at which time it will jump to a new state y. The exponential clock that controls jumps from x to y ticks at rate $a(x, y) \geq 0, x \neq y$. To make an entire array out of $a(\cdot, \cdot)$, we define $a(x, x)$ so that

$$\sum_y a(x, y) = 0. \tag{B.1}$$

That is, $a(x, x) = -\sum_{y \neq x} a(x, y)$. We let L denote the array that these rates make: $L = [a(x, y)]$. To reduce technicalities, we assume throughout that each row of L has only a finite number of non-zero entries. This array is called the *generator* of the process. We will deal with the generator as a linear operator acting on functions defined on the state space. So, if $f(x)$ is a function on E, then $Lf(x)$ is the function given by

$$Lf(x) = \sum_y a(x, y) f(y) = \sum_y a(x, y)(f(y) - f(x)), \tag{B.2}$$

where the second formula follows from (B.1).

The transition semigroup.

For discrete-space continuous-time Markov processes, the generator gives the most natural and simplest description of the process. However, there is a related construct that is also important. It is called the *transition semigroup*. It consists of a family of linear operators $T_t = [p_t(x, y)]$, $t \geq 0$, that satisfy the following properties:

$$T_t T_s = T_{t+s} \tag{B.3}$$

$$T_t \geq 0 \tag{B.4}$$

$$T_t 1 = 1 \tag{B.5}$$

$$T_0 = I \tag{B.6}$$

$$\lim_{t \searrow 0} T_t = I \tag{B.7}$$

[where the first condition means that $\sum_y p_t(x, y) p_s(y, z) = p_{t+s}(x, z)$]. Equation (B.4) means the operator T_t maps positive functions to positive functions. Intuitively, the quantity $p_t(x, y)$ represents the probability that at time t the process will be at y given that at time zero it was at x.

There exists a one-to-one correspondence between generators and transition semigroups. Namely, given a generator L, the corresponding semigroup is given by

$$T_t = e^{tL} \triangleq I + tL + \ldots + \frac{(tL)^n}{n!} + \ldots$$

Conversely, given a transition semigroup T_t, the generator is

$$L = \lim_{t \searrow 0} \frac{T_t - I}{t} \tag{B.8}$$

Uniformization.

There is one more representation for discrete-space continuous-time Markov processes that works when the jump rates are uniformly bounded: $\max_x |a(x, x)| < \infty$. Let λ be a positive real number that dominates these rates and put

$$Q = \lambda^{-1} L + I.$$

It is easy to see that Q is nonnegative and has row sums equal to one (i.e., it is a stochastic matrix). A simple calculation shows that

$$T_t = e^{tL} = e^{-\lambda t} e^{\lambda t Q} = \sum_{k=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^k}{k!} Q^k. \tag{B.9}$$

This is called the uniformization theorem. The interpretation of the right-hand side is that there is a single exponential clock that ticks at rate λ and every time it ticks, a jump occurs according to the stochastic matrix Q. Hence, at time t, the number of times this clock has ticked is a Poisson random variable with parameter λt and so the chance that there have been k jumps is the Poisson probability that appears in front of Q^k in (B.9). If this result seems surprising (in particular the fact that λ can be chosen arbitrarily large), it might help to bear in mind that the stochastic matrix Q allows jumps from a state x to itself.

B.2. The Markov Process

So far we have talked about analytical objects that characterize a Markov process but we have yet to introduce the process itself. An actual Markov process consists of several objects. First we need an event space Ω. For discrete-space continuous-time Markov processes, the construction of Ω is quite simple. All we need is a space on which we have defined all the exponential clocks that were mentioned above. Next, we need to construct the trajectories $X_t(\omega)$ of the process. X_t represents the position in the state space at time t. Its construction simply follows the description that we gave above in terms of exponential clocks and jumps between states. Corresponding to each possible starting point $x \in E$, there is a probability measure P_x on Ω such that the probability of being at states $x_1, \ldots, x_k$ at the times $t_1 < t_2 < \ldots < t_k$ given that the process starts in state x is given by

$$P_x(X_{t_1} = x_1, \ldots, X_{t_k} = x_k) = p_{t_1}(x, x_1)p_{t_2-t_1}(x_1, x_2)\ldots p_{t_k-t_{k-1}}(x_{k-1}, x_k).$$

Finally, we need to introduce a family of σ-algebras $\mathcal{F}_t$, $t \geq 0$, representing the events that are observable up to time t. One should think of $\mathcal{F}_t$ as the σ-algebra generated by the random variables X_s, $s \leq t$ (however, to be precise, we want the completed right continuous modification of this family, whatever that means). The process X_t is quite simple in that it remains fixed at a point in E for an interval of time and then it jumps to another point in E. It will turn out to be important that we make the convention that at the instant when a jump occurs, the process is actually at the new point, not the old one. That is, we are assuming that the process is right continuous. It turns out that in the general theory of stochastic process, functions of time should be right continuous with left limits. We call such functions RCLL. All the functions of time that we will consider will be RCLL and in fact will generally be piecewise constant or an integral of a piecewise constant function. An important property of the family of σ-algebras $\mathcal{F}_t$ is that it is increasing: $\mathcal{F}_s \subset \mathcal{F}_t$, $s \leq t$. This captures the intuitive notion that as time progresses, the history of the process increases. Such a family of σ-algebras is called a *filtration*.

The Markov property.

Now that we have a Markov process to work with, we can begin writing probabilistic formulae. For example, the transition semigroup has the following interpretation:

$$T_t f(x) = \mathbb{E}_x f(X_t), \tag{B.10}$$

where $\mathbb{E}_x$ represents expectations calculated using the measure $\mathbb{P}_x$. As another example, the Markov property which was already visible in (B.3) has the following more sophisticated form:

$$\mathbb{E}_x[f(X_{t+u})|\mathcal{F}_t] = T_u f(X_t). \tag{B.11}$$

Although this form of the Markov property is sufficient for many purposes, it is not the most general. Indeed, even though we have conditioned on the entire past $\mathcal{F}_t$ in (B.11) we have considered only a very specific type of future event, namely the types that involve only a fixed time in the future. In order to formulate a more all-encompassing version of the Markov property, we need a convenient way to represent future events. To this end, it is convenient to assume that the probability space Ω is actually the space of RCLL paths. That is, a point $\omega \in \Omega$ is a function from $[0, \infty)$ into E that is right continuous and has left limits. Hence, the random variable X_t is simply the evaluation map at time t: $X_t(\omega) = \omega_t$. On this space Ω, we introduce a one-parameter family of shift operators:

$$X_t(\theta_s \omega) = X_{t+s}(\omega), \ s \in [0, \infty). \tag{B.12}$$

This transformation on Ω induces a transformation on random variables according to the following formula:

$$\theta_s Z(\omega) = Z(\theta_s \omega). \tag{B.13}$$

Using (B.10), (B.12), and (B.13), we see that (B.11) can be written as

$$\mathbb{E}_x[\theta_t f(X_u) | \mathcal{F}_t] = \mathbb{E}_{X_t} f(X_u).$$

A more general form of the Markov property says that we can replace $f(X_u)$ in the above formula by a general random variable:

$$\mathbb{E}_x[\theta_t Z | \mathcal{F}_t] = \mathbb{E}_{X_t} Z, \ Z \in \mathcal{F}. \tag{B.14}$$

By the definition of conditional expectation, (B.14) is equivalent to

$$\mathbb{E}_x Y \theta_t Z = \mathbb{E}_x Y \mathbb{E}_{X_t} Z, \quad Y \in \mathcal{F}_t, \ Z \in \mathcal{F}. \tag{B.15}$$

Stopping times and the strong Markov property.

We will at times need to apply the Markov Property not at fixed times t, but at random times τ. The property is not true at all random times but it is true if the random time does not look into the future. Such random times are called *stopping times* (Definition A.124). All of the obviously non-clairvoyant random times are actually stopping times.

The strong Markov property says that (B.14) and (B.15) are true even when t is replaced by a stopping time:

$$\mathbb{E}_x[\theta_\tau Z | \mathcal{F}_\tau] = \mathbb{E}_{X_\tau} Z, \ Z \in \mathcal{F}. \tag{B.16}$$

There is a similar form analogous to (B.15).

Semigroup calculus.

Formula (B.8) says that the generator is the derivative at time zero of the semigroup. Derivatives at later times are almost as simple:

$$\frac{d}{ds} T_s f(x) = \lim_{u \searrow 0} \frac{T_{s+u} f(x) - T_s f(x)}{u} = \lim_{u \searrow 0} T_s \frac{T_u - I}{u} f(x) = T_s L f(x).$$

Hence, the fundamental theorem of calculus gives us the following important semigroup identity:

$$T_t f(x) - f(x) = \int_0^t T_s L f(x) \, ds. \tag{B.17}$$

Written probabilistically, (B.17) becomes

$$\mathbb{E}_x f(X_t) - f(x) = \mathbb{E}_x \int_0^t Lf(X_s)\,ds. \qquad (B.18)$$

Fix a function $q(x)$ defined on the state space and define a family of linear operators $\tilde{T}_t$ according to the following formula:

$$\tilde{T}_t f(x) = \mathbb{E}_x f(X_t) \exp\left(-\int_0^t q(X_u)\,du\right).$$

Using the Markov property (B.15), it is easy to see that $\tilde{T}_t$ satisfies (B.3). In fact, it also satisfies (B.4), (B.6), and (B.7). Hence, it makes sense to define a generator $\tilde{L}$ for $\tilde{T}_t$ just like we would for a transition semigroup. A simple calculation using the fact that

$$\exp\left(-\int_0^t q(X_u)du\right) = 1 - \int_0^t q(X_u)\,du + o(t)$$

shows that

$$\tilde{L}f(x) = Lf(x) - q(x)f(x).$$

Now, given any function $g(x)$, if we pick $q(x) = Lg(x)/g(x)$ and $f(x) = g(x)$, then $\tilde{L}f(x) = 0$ and so (B.17) implies that $\tilde{T}_t g(x) = g(x)$. Rewriting this probabilistically, we get

$$g(x) = \mathbb{E}_x g(X_t) \exp\left(-\int_0^t \frac{Lg(X_u)}{g(X_u)}du\right). \qquad (B.19)$$

This formula can be thought of as the multiplicative counterpart to (B.18).

B.3. Birth-Death Processes

Suppose that $E = \mathbb{Z}^d$, the d-dimensional integer lattice, and that there is a finite collection of step directions e_j, $j = 1, \ldots, m$, and that the transition rates are

$$a(x,y) = \begin{cases} \lambda_j(x) & \text{if } y = x + e_j \text{ for some } j; \\ 0 & \text{otherwise.} \end{cases}$$

In this case the generator takes the following simple form:

$$Lf(x) = \sum_j \lambda_j(x)\left(f(x+e_j) - f(x)\right). \qquad (B.20)$$

Eventually, we will only consider Markov processes that are multidimensional birth-death processes.

For these processes, we introduce a little more notation. Let Y_t^j be the number of times up to and including time t that the process X_t steps in direction e_j. The process Y_t^j is called the *counting process* for the steps in the e_j direction. It is easy to recover X_t from knowledge of all the counting processes:

$$X_t = X_0 + \sum_j Y_t^j e_j.$$

Most Markov processes that arise in queueing models are of this type. For example, the queue length process in an $M/M/1$ queue is the process on the non-negative integers that takes steps to the right at rate λ and steps to the left at rate μ as long as $x > 0$. Hence, for this case, we may take $e_1 = 1, e_2 = -1, \lambda_1 = \lambda$, and $\lambda_2 = \mu 1_{x>0}$. Another example from queueing theory is two queues in tandem. In this case the state space is the non-negative quadrant of $\mathcal{Z}^2$, $e_1 = (1, 0)$ represents arrivals to the first queue which occur at rate $\lambda_1(x) = \lambda_1$, $e_2 = (-1, 1)$ represents transfers from the first queue to the second which occur at rate $\lambda_2(x) = \lambda_2 1_{x_1>0}$, and $e_3 = (0, -1)$ represents service completions at the second queue which occur at rate $\lambda_3(x) = \lambda_3 1_{x_2>0}$.

Random walk.

Now suppose that $E = \mathcal{Z}^d$ and that $a(x, y) = \lambda(y-x)$, $y \neq x$, for some function λ. In this case, $Lf(x)$ has the following form:

$$Lf(x) = \sum_z \lambda(z)(f(x+z) - f(x)).$$

This process has a particularly simple description. Indeed, from the uniformization theorem, we see that $X_t - X_0$ is a Poisson sum of independent random increments:

$$X_t = X_0 + \sum_{k=1}^{N_t} \xi_k.$$

Here N_t is a Poisson random variable with parameter $\tilde{\lambda} = \sum_{z\neq 0} \lambda(z)$ and ξ_k is a random increment that takes value z with probability $\lambda(z)/\tilde{\lambda}$.

Specializing even further, suppose that the state space is the integers and that $\lambda(z) = 1/2$ for $z = \pm 1$ and zero for other values of z. Then,

$$Lf(x) = \frac{f(x+1) - 2f(x) + f(x-1)}{2}.$$

This process is called a *simple random walk*. For a simple random walk, N_t is Poisson with parameter one and the ξ_k take values ± 1 with probability $1/2$.

B.4. Martingales

Recall that, for a process $\{M_t, t \geq 0\}$ to be a martingale with respect to a filtration $\{\mathcal{F}_t, t \geq 0\}$ and a measure P, each M_t must be integrable against the measure P and measurable with respect to $\mathcal{F}_t$ and the process $\{M_t, t \geq 0\}$ must satisfy the martingale property:

$$\mathbb{E}[M_{t+u}|\mathcal{F}_t] = M_t, \quad t, u \geq 0.$$

Generally when we claim something is a martingale, the integrability property will be left to the reader to check, the measurability property will be completely obvious, and the martingale property will be shown.

In this section, we will introduce several martingales associated with a Markov process. In later sections, we will then apply certain martingale theorems to these martingales to help us prove large deviations results. The results that we will need about martingales are collected in §A.5.

Additive functionals.

An *additive functional* is a real-valued process M_u^t, $0 \leq u \leq t$, that satisfies the following properties:

$$M_u^t + M_t^v = M_u^v \tag{B.21}$$
$$\theta_s M_u^t = M_{u+s}^{t+s} \tag{B.22}$$
$$\mathbb{E}_x M_0^t = 0 \qquad \text{for all } x \tag{B.23}$$
$$M_0^t \in \mathcal{F}_t. \tag{B.24}$$

The simplest example of an additive functional is

$$M_u^t = f(X_t) - f(X_u) - \int_u^t Lf(X_s)\,ds.$$

In this case, properties (B.21)–(B.24) are all perfectly straightforward except perhaps (B.23) which follows from (B.17).

Theorem B.1. *If M_u^t is an additive functional, then M_0^t is a martingale with respect to every measure P_x.*

Proof. The fact that M_0^t is a martingale follows from the defining properties of an additive functional and the Markov property of the filtration $\mathcal{F}_t$:

$$\begin{aligned}
\mathbb{E}_x[M_0^{t+u}|\mathcal{F}_t] &= \mathbb{E}_x[M_0^t + M_t^{t+u}|\mathcal{F}_t] \\
&= M_0^t + \mathbb{E}_x[M_t^{t+u}|\mathcal{F}_t] \\
&= M_0^t + \mathbb{E}_x[\theta_t M_0^u|\mathcal{F}_t] \\
&= M_0^t + \mathbb{E}_{X_t} M_0^u \\
&= M_0^t.
\end{aligned}$$

Multiplicative functionals.

A *multiplicative functional* is a real-valued process M_u^t, $0 \le u \le t$, that satisfies the following properties:

$$M_u^t M_t^v = M_u^v \tag{B.26}$$
$$\theta_s M_u^t = M_{u+s}^{t+s} \tag{B.27}$$
$$\mathbb{E}_x M_0^t = 1 \tag{B.28}$$
$$M_u^t > 0 \tag{B.29}$$
$$M_0^t \in \mathcal{F}_t. \tag{B.30}$$

The simplest example of a multiplicative functional is

$$M_u^t = \frac{g(X_t)}{g(X_u)} \exp\left(-\int_u^t \frac{Lg(X_s)}{g(X_s)} ds\right). \tag{B.31}$$

In this case, properties (B.26)–(B.30) are all trivial except (B.28) which follows from (B.19).

Theorem B.3. *If M_u^t is a multiplicative functional, then M_0^t is a martingale with respect to every measure P_x.*

Proof. The result follows easily from the defining properties of a multiplicative functional and the Markov property:

$$\begin{aligned}
\mathbb{E}_x[M_0^{t+u}|\mathcal{F}_t] &= \mathbb{E}_x[M_0^t M_t^{t+u}|\mathcal{F}_t] \\
&= M_0^t \mathbb{E}_x[M_t^{t+u}|\mathcal{F}_t] \\
&= M_0^t \mathbb{E}_x[\theta_t M_0^u|\mathcal{F}_t] \\
&= M_0^t \mathbb{E}_{X_t} M_0^u \\
&= M_0^t.
\end{aligned}$$

This completes the proof. ∎

From this result, we see that

$$M_0^t = \frac{g(X_t)}{g(X_0)} \exp\left(-\int_0^t \frac{Lg(X_s)}{g(X_s)} ds\right) \tag{B.32}$$

is a martingale. We will call such martingales the *exponential martingales* associated with the Markov process X_t.

The Dirichlet problem.

As an example of the utility of the linear martingale, we briefly describe its connection with the Dirichlet problem. Let D be a subset of E and consider the following boundary value problem:

$$\begin{aligned} Lf(x) &= 0, \qquad && x \in D, \\ f(x) &= \phi(x), && x \in D^c. \end{aligned}$$

We call this the Dirichlet problem. When the space is Euclidean and the operator L is the Laplacian, this boundary value problem is the classical Dirichlet problem (that's what we'd get if the Markov process were Brownian motion — it's also why the generator is denoted by L). Suppose we have a function f that solves this problem. Let's see what the optional sampling theorem A.129 says when applied to the linear martingale, and τ is the first time X_t exits the domain D:

$$\begin{aligned} f(x) &= \mathbb{E}_x f(X_\tau) - \mathbb{E}_x \int_0^\tau Lf(X_s)\,ds \\ &= \mathbb{E}_x \phi(X_\tau). \end{aligned} \tag{B.33}$$

The integral vanishes since, before time τ, $Lf(X_s) = 0$. Also, at time τ, the process is in the complement of D and so f can be replaced by ϕ in the first term (we have used the right continuity of the process here). The right-hand side in (B.33) does not involve the function f. This proves that the solution to the Dirichlet problem is unique. To prove existence, it suffices to check that the function defined by the right-hand side of (B.33) actually is a solution of the problem. This is an easy exercise using the Markov property that we leave to the reader.

If we choose $\phi(x) = 1_y(x)$, then (B.33) reduces to

$$f(x) = \mathbb{P}_x(X_\tau = y).$$

Hence we see that the Dirichlet problem is closely related to finding exit distributions from domains.

A related problem is to find the expected exit time $g(x) = \mathbb{E}_x \tau$. An analysis similar to the one above shows that $g(x)$ is the unique solution of the following inhomogeneous Dirichlet problem:

$$\begin{aligned} Lg(x) &= -1, \qquad && x \in D, \\ g(x) &= 0, && x \in D^c. \end{aligned} \tag{B.34}$$

Sometimes these Dirichlet problems can be solved explicitly, but even when they can't, they yield efficient numerical methods.

This kind of probabilistic analysis of boundary value problems is the basis of a branch of mathematics called probabilistic potential theory. We have only given the briefest exposure just to give a quick appreciation for the value of the linear martingale.

The Feynman-Kac formula.

For a moment let's consider the Shroedinger equation:

$$Lf(x) = \psi(x)f(x), \quad x \in D, \tag{B.35}$$

$$f(x) = \phi(x), \quad x \in D^c. \tag{B.36}$$

Again, let τ be the first exit time from D and apply the optional sampling theorem to the exponential martingale (B.32) to get that any solution to (B.35), (B.36) must be given by

$$f(x) = \mathbb{E}_x\phi(X_\tau)\exp\left(-\int_0^\tau \psi(X_s)\,ds\right). \tag{B.37}$$

This proves uniqueness and as before existence is proved by verifying that the formula given by (B.37) is indeed a solution.

There is one special case of (B.37) that deserves mentioning. When the functions ϕ and ψ are actually the constants one and λ, respectively, then (B.37) becomes

$$f(x) = \mathbb{E}_x e^{-\lambda\tau}.$$

Therefore, solving (B.35), (B.36) gives us the Laplace transform of the first exit time from a domain. This result is often useful.

Change of measure.

Multiplicative functionals are useful for changing measure in such a way that the Markov property is preserved under the new measure. Let M_u^t be a multiplicative functional. We fix a finite time horizon T and consider the process X_t only on the interval $0 \le t \le T$. For each $x \in E$, let $\tilde{P}_x$ be a new measure on $(\Omega, \mathcal{F}_T)$ defined by

$$\tilde{\mathbb{E}}_x Z = \mathbb{E}_x M_0^T Z, \quad Z \in \mathcal{F}_T, \tag{B.38}$$

where $\tilde{\mathbb{E}}_x$ denotes expectation calculated using the measure $\tilde{P}_x$. From (B.38), we see that $\tilde{P}_x$ is absolutely continuous with respect to P_x, and the Radon-Nikodym derivative is simply M_0^T (§A.4). Now for $0 \le t \le T$, define a linear operator $\tilde{T}_t$ as follows:

$$\tilde{T}_t f(x) = \tilde{\mathbb{E}}_x f(X_t). \tag{B.39}$$

Proposition B.4. *If M_u^t is a multiplicative functional, then the operators $\tilde{T}_t$, $t \ge 0$, defined by (B.39) form a transition semigroup and*

$$\tilde{T}_t f(x) = \mathbb{E}_x M_0^t f(X_t). \tag{B.40}$$

Proof. We start by proving (B.40). Using the fact that M_0^t is a martingale, we see that

$$\begin{aligned}
\tilde{T}_t f(x) &= \mathbb{E}_x M_0^T f(X_t) \\
&= \mathbb{E}_x \mathbb{E}_x [M_0^T f(X_t) | \mathcal{F}_t] \\
&= \mathbb{E}_x f(X_t) \mathbb{E}_x [M_0^T | \mathcal{F}_t] \\
&= \mathbb{E}_x f(X_t) M_0^t.
\end{aligned}$$

All the properties that have to be satisfied to be a transition semigroup follow trivially from the fact that M_0^t is positive and has mean one except the most important property, $\tilde{T}_t\tilde{T}_s = \tilde{T}_{t+s}$, which we now verify:

$$\begin{aligned}
\tilde{T}_t\tilde{T}_s f(x) &= \mathbb{E}_x M_0^t \mathbb{E}_{X_t} M_0^s f(X_s) \\
&= \mathbb{E}_x M_0^t \theta_t [M_0^s f(X_s)] \\
&= \mathbb{E}_x M_0^t M_t^{s+t} f(X_{s+t}) \\
&= \mathbb{E}_x M_0^{s+t} f(X_{s+t}) \\
&= \tilde{T}_{s+t} f(x).
\end{aligned}$$

Here, the first equality follows from (B.40), the second from the Markov property (B.15), the third from the definitions (B.12) and (B.13) of the shift operator, and the fourth from the multiplicative property (B.26) of M_u^t. This completes the proof. ∎

Since starting with a semigroup, say $\tilde{T}_t$, there is, for each $x \in E$, a unique measure on the space of right-continuous paths such that (B.39) holds, it follows from Proposition B.4 that $\tilde{P}_x$ is this unique measure. The next theorem shows that multidimensional birth-death processes are invariant under this change of measure.

Theorem B.5. *If, under the measure P_x, X_t is a multidimensional birth-death process with step directions e_j and step rates $\lambda_j(x)$, then under the measure $\tilde{P}_x$ the process X_t is again a multidimensional birth-death process with the same step directions and with the rates changed to*

$$\mu_j(x) = \lambda_j(x)\frac{g(x+e_j)}{g(x)}. \tag{B.41}$$

Proof. To prove this, we calculate the generator of $\tilde{T}_t$. With overwhelming probability, the process X_t (relative to the original measure P_x) will have taken at most one jump in a small amount of time t. Hence, using (B.40) and (B.31), we get

$$\begin{aligned}
\tilde{T}_t f(x) &= \exp\left(-t\sum_j \lambda_j(x)\right)\exp\left(-t\frac{Lg(x)}{g(x)}\right) f(x) \\
&+ \int_0^t \sum_k \exp\left(-s\sum_j \lambda_j(x)\right)\lambda_k(x)\exp\left(-(t-s)\sum_j \lambda_j(x+e_k)\right) \\
&\times \frac{g(x+e_k)}{g(x)}\exp\left(-s\frac{Lg(x)}{g(x)} - (t-s)\frac{Lg(x+e_k)}{g(x+e_k)}\right) f(x+e_k)\,ds + o(t).
\end{aligned}$$

In the second summand we have a jump at s in direction k, and no further jumps until t. The first two terms give the probability of this event. From this simple calculation, using (B.20) we see that

$$\begin{aligned}
\tilde{L} f(x) &= \lim_{t \searrow 0} \frac{1}{t} (\tilde{T}_t f(x) - f(x)) \\
&= -\left(\sum_j \lambda_j(x) + \frac{Lg(x)}{g(x)} \right) f(x) + \sum_j \lambda_j(x) \frac{g(x+e_j)}{g(x)} f(x+e_j) \\
&= \sum_j \lambda_j(x) \frac{g(x+e_j)}{g(x)} (f(x+e_j) - f(x)).
\end{aligned}$$

Hence, the new process is a multidimensional birth-death process with the same jump directions and with jump rates given by (B.41). ■

It is interesting to note that there are generally not enough degrees of freedom in (B.41) to be able to get all sets of rates $\mu_j(x)$ by appropriate choices of $g(x)$. But it turns out that we can write the Radon-Nikodym derivative M_0^T purely in terms of $\mu_j(x)$ and that, once we do this, we get a derivative that is correct even if there is no function $g(x)$ connecting the old and the new rates. So, start by assuming that μ_j and λ_j are related through (B.41) for some g. Since $g(X_t)$ is piecewise constant, we can write the difference $\log g(X_T) - \log g(X_0)$ as the sum of the changes over the jumps:

$$\begin{aligned}
\log g(X_T) - \log g(X_0) &= \int_0^T \sum_j (\log g(X_s) - \log g(X_{s-}))\, dY_s^j \\
&= \int_0^T \sum_j \log \frac{g(X_{s-} + e_j)}{g(X_{s-})} dY_s^j \\
&= \int_0^T \sum_j \log \frac{\mu_j(X_{s-})}{\lambda_j(X_{s-})} dY_s^j, \qquad (B.42)
\end{aligned}$$

where Y_t^j is the counting process that is incremented every time a jump occurs in the e_j direction. Now, using (B.41), we see that

$$\begin{aligned}
\frac{Lg(x)}{g(x)} &= \sum_j \lambda_j(x) \left(\frac{g(x+e_j)}{g(x)} - 1 \right) \\
&= \sum_j (\mu_j(x) - \lambda_j(x)). \qquad (B.43)
\end{aligned}$$

From (B.43) and (B.42), we see that

$$
\begin{aligned}
M_0^T &= \exp\left(\log g(X_T) - \log g(X_0) - \int_0^T \frac{Lg(X_s)}{g(X_s)} ds\right) \\
&= \exp\left(\int_0^T \sum_j \log \frac{\mu_j(X_{s-})}{\lambda_j(X_{s-})} dY_s^j - \int_0^T \sum_j (\mu_j(X_s) - \lambda_j(X_s))\, ds\right).
\end{aligned}
\tag{B.44}
$$

Equation (B.44) defines a change of measure that maps the multidimensional birth-death process with rates $\lambda_j(x)$ into a process of the same type but having rates $\mu_j(x)$.

Note that formula (B.44) for M_0^T involves the function g only indirectly through formula (B.41). This leads us to the question: Does (B.44) define a change of measure that maps the process with rates $\lambda_j(x)$ to the one with rates $\mu_j(x)$, even when (B.41) does not hold? The answer to this question is yes. Indeed, instead of specifying a function g, we start with desired rates $\mu_j(x)$. We formulate this result as a theorem:

Theorem B.6. *Let X_t be a multidimensional birth-death process with step directions e_j and step rates $\lambda_j(x)$ and let P_x denote the corresponding measure on the space of trajectories. Let $\mu_j(x)$ be an arbitrary set of transition rates and let $\tilde{P}_x$ be a new measure defined by (B.38) with M_0^T defined by (B.44). Then, under this new measure, the process X_t is again a multidimensional birth-death process with the same step directions and with the rates changed to $\mu_j(x)$.*

Proof. Define M_u^t by the formula:

$$
M_u^t = \exp\left(\int_u^t \sum_j \log \frac{\mu_j(X_{s-})}{\lambda_j(X_{s-})} dY_s^j - \int_u^t \sum_j (\mu_j(X_s) - \lambda_j(X_s))\, ds\right).
\tag{B.45}
$$

We need to show that M_u^t is a multiplicative functional. All of the defining properties are trivial except (B.28) which can be checked by direct calculation. Indeed, using the law of total probability, we get

$$
\begin{aligned}
\mathbb{E}_x M_0^t = \sum_n \int \cdots \int_{0<s_1<s_2<\cdots<s_n\le t} & \sum_{j_1=1}^{m} \cdots \sum_{j_n=1}^{m} \\
& \exp\left(-s_1 \sum_j \lambda_j(x_0)\right) \lambda_{j_1}(x_0)\, ds_1 \\
& \qquad \times \exp\left(-(s_2 - s_1) \sum_j \lambda_j(x_1)\right) \lambda_{j_2}(x_1)\, ds_2 \\
& \vdots \\
& \exp\left(-(s_n - s_{n-1}) \sum_j \lambda_j(x_{n-1})\right) \\
& \qquad \times \lambda_{j_n}(x_{n-1})\, ds_n \ \exp\left(-(t - s_n) \sum_j \lambda_j(x_n)\right) \\
& \exp\left(\sum_{k=1}^{n} \log \frac{\mu_{j_k}(x_{k-1})}{\lambda_{j_k}(x_{k-1})}\right. \\
& \qquad \left. - \sum_{k=1}^{n+1} \sum_j (\mu_j(x_{k-1}) - \lambda_j(x_{k-1}))(s_k - s_{k-1})\right),
\end{aligned}
$$

where, for notational efficiency, we have put $x_0 = x$, $x_i = x_{i-1} + e_{j_i}$, $s_0 = 0$, and $s_{n+1} = t$. This formula may look messy, but it is easy to explain. The first line corresponds to partitioning the sample space into small pieces. One piece is a trajectory that makes n steps up to time t, with the steps occurring at times $s_1, s_2, \ldots, s_n$, and the i^{th} step being in the direction e_{j_i}. The complicated product on the second through fourth lines is the likelihood of this trajectory. It too is easy to explain. The first factor, $\exp(-s_1 \sum_j \lambda_j(x_0))$, represents the probability that nothing happens in the interval $[0, s_1)$, the second factor, $\lambda_{j_1}(x_0)\, ds_1$, represents the probability that a step in direction e_{j_1} occurs exactly at time s_1, etc. Finally, the last line is the integrand evaluated on this specific trajectory. Now, note that the last line cancels with the previous lines in such a way as to change the previous lines into a formula of exactly the same type but with the λs replaced by μs:

$$
\begin{aligned}
\mathbb{E}_x M_0^t = \sum_n \int \cdots \int_{0<s_1<s_2<\cdots<s_n\le t} & \sum_{j_1=1}^{m} \cdots \sum_{j_n=1}^{m} \\
& \exp\left(-s_1 \sum_j \mu_j(x_0)\right) \mu_{j_1}(x_0)\, ds_1 \\
& \qquad \times \exp\left(-(s_2 - s_1) \sum_j \mu_j(x_1)\right) \mu_{j_2}(x_1)\, ds_2 \\
& \vdots \\
& \exp\left(-(s_n - s_{n-1}) \sum_j \mu_j(x_{n-1})\right) \\
& \qquad \times \mu_{j_n}(x_{n-1})\, ds_n \exp\left(-(t - s_n) \sum_j \mu_j(x_n)\right).
\end{aligned}
$$

But, again by the law of total probability (now using μs), we see that this is exactly equal to one. Even though this calculation is rather tedious, we have included it so that the reader will see that for multidimensional birth-death processes everything can in principle be calculated explicitly. ■