THE DISCRETE FEYNMAN-KAC FORMULA

ETHAN NAEGELE

1. Introduction

We would like to understand differential equations of the following form:

$$\frac{\partial}{\partial t}u(t,i) = (R^E u(t,\cdot))_i = \sum_{k\in S} R^E_{ik} u(t,k), \tag{1.1}$$

where R^E is a matrix with entries R^E_{ik}, S is a finite or countably infinite set of positive integers, and u is a function from $[0,\infty)\times S$ to $\mathbb{R}$. For the cases that we will consider, R^E will be symmetric (not necessarily finite) with diagonal entries of the form $R^E_{ii} = -v_i + E(i)$ where $v_i > 0$ for each i, and E will be a function from S to $\mathbb{R}$, therefore giving rise to our choice of notation R^E. That is, the superscript in R^E is used to denote the choice of function E. We will write off-diagonal entries of R^E as $v_i P_{ik}$ where $P_{ik} \geq 0$. Without loss of generality, we may simply consider S to be $\{1,2,...,n\}$ for some n or simply $\mathbb{N}\setminus\{0\}$. Here we will treat the infinite case and finite case simultaneuously.

Now, although the equation presented in (1.1) has a familiar form that suggests an exponential solution, it may also seem somewhat opaque; it is initially unclear, for example, why it would be useful or interesting to have the diagonal entries of R^E written in such a way, or why one may want to impose symmetry of R^E. However, one can better understand such differential equations by using the theory of continuous time Markov chains. First, let us expand (1.1); we then we see that

$$\frac{\partial}{\partial t}u(t,i) = v_i\left(\sum_{k\neq i} P_{ik}u(t,k) - u(t,i)\right) + E(i)u(t,i). \tag{1.2}$$

Let us first assume for simplicity that $E = 0$ everywhere. That leaves us with

$$\frac{\partial}{\partial t}u(t,i) = v_i\left(\sum_{k\neq i} P_{ik}u(t,k) - u(t,i)\right). \tag{1.3}$$

We can examine the difference in the parentheses (assume for now that the possibly infinite sum makes sense). We can think of the summation as being some sort of weighted sum of the values that u takes on at different points k in space while at time t. Then the difference is this sort of weighted sum of values not at i, and the value at i. What the full expression tells us, therefore, is that the time derivative is proportional to such a difference.

Date: April 20, 2023.

Now, let us improve our heuristic; in order to make sense out of these weights, they should add to 1. Hence, this suggests we can treat the P_{ik} terms as probabilities. This would suggest that $\sum_{k \in S} P_{ik} u(t,k)$ is a sort of average of the values that $u(t,\cdot)$ takes on away from i.

When one views such an equation in the context of Markov chains, our interpretation of (1.1) begins to crystallize. We see that the R matrix we have described is a generator matrix for a continuous time Markov chain which we call X_t. The P_{ik}, therefore, are jump chain probabilities, which are the transition probabilities of X_t "when it moves". In Section 2.2, we will show that in the simplest case where we fix a state j and have the initial condition $\mathbf{1}_j$, $u(t,i) = P_{ij}(t)$ solves (1.3), where $P(t)$ is the transition matrix at time t of the chain. One can then interpret (1.3) as a diffusion equation; if $u(t,\cdot)$ has a larger value at i than at the neighbors of i, then one can imagine heat, for example, being distributed from i to its neighbors.

Now, just as we have examined the first term on the right hand side of (1.2) in isolation, let us do the same for the second term. If the first term were 0, but $E \neq 0$, we would have that

$$\frac{\partial}{\partial t} u(t,i) = E(i) u(t,i).$$

We then see that this is really just an equation involving a single variable, as the behavior of u in space is only considered at the point i. The function $u(t,i)$ will either increase or decrease at a rate proportional to $u(t,i)$, depending on if $E(i)$ is positive or negative. This equation is easily solved by the exponential; the solution is

$$u(t,i) = e^{tE(i)} u(0,i). \tag{1.4}$$

Thus, based on the results of (1.4) and the transition matrix solution to (1.3) discussed above, we would expect that the solution to (1.2) is some combination of an exponential in terms of $E(i)$ and transition matrix probabilities $P_{ij}(t)$. With this intuition in mind, we present the following statement, whose proof and intuitive picture will be the goal of our discussion.

Theorem 1.5 (Feynman-Kac Formula)**.** *A solution to (1.1) is*

$$u(t,i) = \mathbf{E}\left[\exp\left(\int_0^t E(X_s)ds\right) u_0(X_t) \mid X_0 = i\right]. \tag{1.6}$$

Note that we do not discuss questions of uniqueness here.

The formula in (1.6) tells us that the solution is of the following nature: look at the states that the chain could have ended up at at time t. Consider the initial condition function applied to these values, and consider the path that the chain took up to X_t; the function E will have some cumulative effect over that path, which will amount to either exponential growth or decay depending on the sign of E. Now, average over all such quantities–this is the value of $u(t,i)$.

We begin to see the role that E plays in this scenario; it determines whether certain paths within the state space will produce exponential growth or decay, and the magnitude of such growth or decay. For example, if one were to think of, say, bacteria growth, one can interpret E as signifying that there are regions of space where growth is favorable or unfavorable for such bacteria to grow. For this reason, we think of E as being the *environment function*. We will refer to it as such throughout our discussion.

FIGURE 1. Bacteria crawls through a tube with harsh (negative) environment in the red regions and favorable (positive) environment in the green regions, with intensity of red color describing the harshness.

2. The Simple Case

We would like to start by understanding the simplest case, before moving onto the general case in Section 3. Here, we will further develop our intuition for how continuous time Markov chains can relate to the differential equations of our study. To do so, we will first lay the groundwork by introducing the key results from Markov chain theory.

From this point forward, for any chain X_t that we discuss, we will assume that there exists $M < \infty$ such that $v_i \leq M$ for all i, and we will assume the discrete-time jump chain is irreducible and recurrent. This is for the purposes of ensuring that the Markov chain almost surely makes a finite number of jumps in any finite amount of time (see [1, pp. 418]).

2.1. Initial Results.

We will start this section with the results from [1].

Lemma 2.1 (Limiting Probabilities)**.** *For a continuous time Markov chain, the following limits hold:*

$$\lim_{t\to 0} \frac{1 - P_{ii}(t)}{t} = v_i$$

$$\lim_{t\to 0} \frac{P_{ij}(t)}{t} = q_{ij} := v_i P_{ij}.$$

This lemma tells us that, in a small window of time, the rate at which one leaves state i is v_i, and the rate at which one leaves from i to j is the rate at which one leaves i, multiplied by the probability of jumping from i to j at a given jump. Though we will not prove Lemma 2.1 here, the proof is in [1], and a generalized version is proven in Lemma 3.3. The driving idea behind the proof, which we will use extensively, is that there are two ways one can, say, start at i and be at i after t units of time: one can either not move for the whole time interval, or one can leave i, visiting at least one other state, then return to i and stay there up to time t. The latter case would involve jumping at least twice, and in a small time interval, one can imagine that such an event is highly unlikely, and turns out not to contribute to the limit.

Theorem 2.2 (Kolmogorov Backward Equations)**.** *For all $i, j \in S$ and $t \geq 0$,*

$$\frac{dP_{ij}}{dt}(t) = \sum_{k\neq i} q_{ik} P_{kj}(t) - v_i P_{ij}(t) \tag{2.3}$$

where q_{ik} is the transition rate from i to k and $P_{kj}(t)$ is the probability that the chain moves from k to j in t units of time.

If we denote R as the generator matrix, then we can write the backward equation in matrix form as

$$P'(t) = RP(t).$$

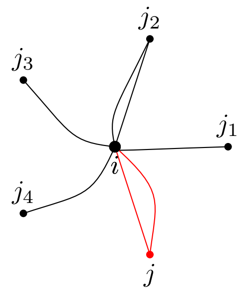

FIGURE 2. $u(t,i)$ looks at proportion of paths coming from i that end at j in the simplest case.

This exact statement is proven in [1]. We will not prove the statement here; we will instead prove a more general version of the backward equation in Theorem 3.8 from which this statement will follow.

Proposition 2.4. *The solution to the Kolmogorov backward equation is* $P(t) = e^{tR}$.

We omit the details here. However, Proposition 2.4 has a practically useful consequence for transition matrices, which we can prove now.

Proposition 2.5. *If R is a (real) symmetric matrix, possibly infinite dimensional, and $t \geq 0$, then e^{tR} is also symmetric. In particular, if R is the generator for a continuous time Markov chain on a finite or countable state space, and $P(t)$ is the transition matrix at an arbitrary time t, then $P(t)$ is symmetric.*

Proof. We prove symmetry entrywise. If R is symmetric, then for any pair of states i, j and for any $k \in \mathbb{N}$, we have that $(R^k)_{ij} = (R^k)_{ji}$. Thus, by Proposition 2.4, we have

$$\begin{aligned}(e^{tR})_{ij} &= \lim_{N\to\infty} \sum_{k=0}^{N} \frac{(tR)_{ij}^k}{k!} \\ &= \lim_{N\to\infty} \sum_{k=0}^{N} \frac{(tR)_{ji}^k}{k!} \\ &= (e^{tR})_{ji}.\end{aligned}$$

□

2.2. **Backward Equation as a Diffusion Equation.** Let $P(t)$ be a transition matrix which is symmetric for every time t. Fix a state j in an at most countable state space S. We can view $P_{ij}(t)$ as a function of time and space by defining $u(t,i) := P_{ij}(t)$. This is visualized in Figure 2. First, observe that $u(0,i) = \mathbf{1}_j$, that is, equal to 1 at $i = j$ and 0 otherwise. Next, observe that since $P(t)$ is a symmetric stochastic matrix, it is in fact doubly stochastic, or in other words all rows and columns of $P(t)$ sum to 1, for every time t.

At time 0, the function $u(t,i)$ simply denotes a point mass of 1 at $i = j$, but as time evolves, this mass will diffuse throughout the state space; as t increases, there will be more and more states i such that the chain could have started at i and reached j at time t. Since we assume the transition matrix is doubly stochastic for all t, we have $\sum_i P_{ij}(t) = 1$ for all $t \geq 0$, which mathematically describes a "conservation of mass" phenomenon occurring here. Time evolves, and the mass 1 gets distributed to more and more states i. To more clearly see this, we note that by the symmetry assumption, we have $P_{ij}(t) = P_{ji}(t)$ for any t, so that one can instead view $u(t, i)$ as the probability of a chain starting at j and being at i at time t. To further analyze how u changes with time, we can write, by using the backward equation,

$$\begin{aligned}
\frac{\partial}{\partial t} u(t,i) &= \frac{d}{dt} P_{ij}(t) \\
&= (RP(t))_{ij} \\
&= \sum_k R_{ik} P_{kj}(t) \\
&= \sum_k R_{ik} u(t,k) \\
&= \sum_{k \neq i} q_{ik} u(t,k) - v_i u(t,i) \\
&= v_i \left(\sum_{k \neq i} P_{ik} u(t,k) - u(t,i) \right).
\end{aligned}$$

First, this shows that $u(t, i) = P_{ij}(t)$ solves (1.1) with initial condition $\mathbf{1}_j$. This also shows that the time derivative of $u(t, i)$ is therefore proportional to the difference between the average value of u on the nearest neighbors of i, and the value of $u(t, i)$ itself. Here, the average is weighted by the jump chain probabilities.

Therefore, we would expect that the mass would be collecting at i at time t (in other words, $\frac{\partial}{\partial t} u(t, i) > 0$) if and only if the average mass at nearby points is larger than the mass at i. Conversely, if $u(t, i)$ is larger than the average value of u at the neighbors of i at time t, then the mass at i will be redistributing itself among the neighbors of i. This is, for example, akin to particles moving in a concentration gradient, as particles move from higher to lower concentration. The process is visualized in Figure 3.

Remark 2.6. We have started by assuming that the transition matrix is symmetric for every time t, which allowed us to conclude that $P(t)$ is doubly stochastic at every time t. Proposition 2.5 gives a sufficient condition for when $P(t)$ can be symmetric for all t, and this condition is not too restrictive.

3. The General Case: Environment with Arbitrary Initial Conditions

3.1. The Backward Equations Revisited.

Let S be a finite or countably infinite state space, and let R be the generator matrix of some continous time Markov chain X_t. Let $E : S \to \mathbb{R}$ be a real-valued function on the state space. E will essentially act as an environment effect which acts upon the diffusion process differently depending on the state. Define the modified generator matrix

$$R^E_{ij}(t) = \begin{cases} -v_i + E(i) & \text{if } i = j \\ v_i P_{ij} & \text{if } i \neq j. \end{cases}$$

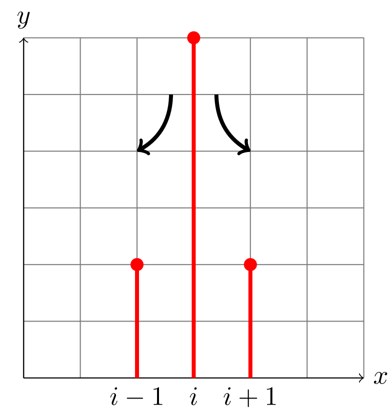

FIGURE 3. Graph of u for fixed t. Since there is more heat at i than the average amount of heat at its neighbors, heat will leave i.

Next, for every $t \geq 0$, we define the modified transition matrix

$$P_{ij}^E(t) := \mathbf{E}\left[\exp\left(\int_0^t E(X_s)ds\right)\mathbf{1}_{\{X_t=j\}} \mid X_0 = i\right].$$

We observe that the special case $E = 0$ gives us the usual transition matrix. Hence, when we prove the backward equations in Theorem 3.8 in a more general setting, it will prove Theorem 2.2 as well.

We will use the notation $\mathbf{E}_i[Y]$ to denote the expectation of a random variable Y conditioned on $X_0 = i$. It turns out that the Kolmogorov backward equation still holds when considering the modified generator and transition matrix, so long as we assume that the environment function E is bounded above. Luckily, the details of the proof will require only small modifications compared to the proof of Theorem 2.2 in [1]. First, however, we prove that the Chapman-Kolmogorov equations still hold for the modified transition matrix. This is a less obvious fact which takes a bit more work in this case, so we prove it in the Lemma below.

Lemma 3.1. *The Chapman-Kolmogorov equations hold for the modified transition matrix:*

$$P_{ij}^E(t+h) = \sum_{k\in S} P_{ik}^E(h)P_{kj}^E(t).$$

Proof. We have

$$\begin{aligned}
&\mathbf{E}_i\left[\exp\left(\int_0^{t+h} E(X_s)ds\right)\mathbf{1}_{\{X_{t+h}=j\}}\right]\\
&= \mathbf{E}_i\left[\exp\left(\int_0^{h} E(X_s)ds\right)\exp\left(\int_h^{t+h} E(X_s)ds\right)\mathbf{1}_{\{X_{t+h}=j\}}\right]\\
&= \sum_{k\in S}\mathbf{E}_i\left[\exp\left(\int_0^{h} E(X_s)ds\right)\exp\left(\int_h^{t+h} E(X_s)ds\right)\mathbf{1}_{\{X_{t+h}=j\}} \mid X_h = k\right]\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \cdot \mathbf{P}_i(X_h = k)
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{k\in S} \mathbf{E}_i \left[\exp\left(\int_0^h E(X_s)ds \right) \mid X_h = k \right] \\
&\qquad\qquad \cdot \mathbf{E}\left[\exp\left(\int_h^{t+h} E(X_s)ds \right) \mathbf{1}_{\{X_{t+h}=j\}} \mid X_h = k \right] \mathbf{P}_i(X_h = k) \\
&= \sum_{k\in S} \mathbf{E}_i \left[\exp\left(\int_0^h E(X_s)ds \right) \mathbf{1}_{\{X_h=k\}} \right] \\
&\qquad\qquad \cdot \mathbf{E}\left[\exp\left(\int_h^{t+h} E(X_s)ds \right) \mathbf{1}_{\{X_{t+h}=j\}} \mid X_h = k \right] \\
&= \sum_{k\in S} \mathbf{E}_i \left[\exp\left(\int_0^h E(X_s)ds \right) \mathbf{1}_{\{X_h=k\}} \right] \mathbf{E}\left[\exp\left(\int_0^t E(X_s)ds \right) \mathbf{1}_{\{X_t=j\}} \mid X_0 = k \right] \\
&= \sum_{k\in S} P^E_{ik}(h) P^E_{kj}(t),
\end{aligned}
$$

where the second to last equality holds from the time homogeneity of the chain. □

We can also prove a statement analagous to Lemma 2.1. The key assumption made in the statements that follow is that the function E is bounded from above. If this condition holds, then we may say that

$$
\begin{aligned}
\mathbf{E}_i \left[\exp\left(\int_0^t E(X_s)ds \right) \mathbf{1}_{\{X_t=j\}} \right] &\le \mathbf{E}_i[e^{Ct} \mathbf{1}_{\{X_t=j\}}] \\
&= e^{Ct} P_{ij}(t),
\end{aligned} \tag{3.2}
$$

which gives a bound in terms of the original transition matrix. We will use this bound repeatedly to adapt the results from the previous sections to the current situation.

Lemma 3.3. *Suppose there exists a constant $C > 0$ such that*

$$\sup_{i\in S} E(i) \le C.$$

Then we have the following limits:

$$\lim_{t\to 0} \frac{1 - P^E_{ii}(t)}{t} = v_i - E(i), \tag{3.4}$$

and for $i \neq j$,

$$\lim_{t\to 0} \frac{P^E_{ii}(t)}{t} = v_i P_{ij}. \tag{3.5}$$

Proof. Let J_t be the event that the chain has not jumped by time t. We have

$$
\begin{aligned}
&\mathbf{E}_i \left[\exp\left(\int_0^t E(X_s)ds \right) \mathbf{1}_{\{X_t=i\}} \right] \\
&\quad = \mathbf{E}_i \left[\exp\left(\int_0^t E(X_s)ds \right) \mathbf{1}_{\{X_t=i\}\cap J_t} \right] + \mathbf{E}_i \left[\exp\left(\int_0^t E(X_s)ds \right) \mathbf{1}_{\{X_t=i\}\cap J_t^c} \right].
\end{aligned}
$$

In the event J_t where the chain doesn't jump, the integral just evaluates to $E(i)t$. Thus, the first term simplifies to $e^{E(i)t}\mathbf{P}_i(J_t)$. Therefore, the above is equal to

$$\begin{aligned} &e^{E(i)t}\mathbf{P}_i(J_t) + \mathbf{E}_i\left[\exp\left(\int_0^t E(X_s)ds\right)\mathbf{1}_{\{X_t=i\}\cap J_t^c}\right] \\ &\qquad = e^{t(E(i)-v_i)} + \mathbf{E}_i\left[\exp\left(\int_0^t E(X_s)ds\right)\mathbf{1}_{\{X_t=i\}\cap J_t^c}\right]. \end{aligned}$$

A Taylor expansion implies that

$$\lim_{t\to 0}\frac{1-e^{t(E(i)-v_i)}}{t} = v_i - E(i).$$

Next, (3.2) implies

$$\begin{aligned} \mathbf{E}_i\left[\exp\left(\int_0^t E(X_s)ds\right)\mathbf{1}_{\{X_t=i\}\cap J_t^c}\right] &\le e^{Ct}\mathbf{P}_i(X_t = i \cap J_t^{\mathsf{c}}) \\ &\le e^{Ct}\mathbf{P}_i(J_t^{\mathsf{c}}) \\ &\le e^{Ct}\mathbf{P}_i(T_2 < t), \end{aligned}$$

where T_2 is the number of time until the chain jumps twice. It was already shown in the proof of Lemma 2.1 in [1] that this probability is $o(t)$. (3.4) now follows.

We use a similar idea to prove (3.5). Let K_t be the event that there is only one jump in time t, and write

$$\begin{aligned} &\mathbf{E}_i\left[\exp\left(\int_0^t E(X_s)ds\right)\mathbf{1}_{\{X_t=j\}}\right] \\ &= \mathbf{E}_i\left[\exp\left(\int_0^t E(X_s)ds\right)\mathbf{1}_{\{X_t=j\}\cap K_t}\right] + \mathbf{E}_i\left[\exp\left(\int_0^t E(X_s)ds\right)\mathbf{1}_{\{X_t=j\}\cap K_t^c}\right]. \end{aligned}$$

In the event K_t, the chain only ever visits i and j, therefore

$$e^{\min\{E(i),E(j)\}} \le \exp\left(\int_0^t E(X_s)ds\right) \le \mathbf{E}_i[\mathbf{1}_{\{X_t=j\}\cap K_t}],$$

hence

$$e^{\min\{E(i),E(j)\}t}\mathbf{P}_i(\{X_t = j\}\cap K_t) \le \mathbf{E}_i\left[\exp\left(\int_0^t E(X_s)ds\right)\mathbf{1}_{\{X_t=j\}\cap K_t}\right] \tag{3.6}$$

$$\le e^{Ct}\mathbf{P}_i(\{X_t = j\}\cap K_t). \tag{3.7}$$

From the proof of Lemma 2.1 in [1], it follows that for any $r > 0$, we have

$$\lim_{t\to 0}\frac{e^{rt}\mathbf{P}_i(\{X_t = j\}\cap K_t)}{t} = v_i P_{ij}.$$

It now follows from the bounds in (3.6) and (3.7) that when we divide by t and take the limit $t \to 0$, we obtain

$$\lim_{t\to 0}\frac{\mathbf{E}_i\left[\exp\left(\int_0^t E(X_s)ds\right)\mathbf{1}_{\{X_t=j\}\cap K_t}\right]}{t} = v_i P_{ij}.$$

As before, we have by (3.2) that

$$\begin{aligned} \mathbf{E}_i\left[\exp\left(\int_0^t E(X_s)ds\right)\mathbf{1}_{\{X_t=i\}\cap K_t^c}\right] &\le e^{Ct}\mathbf{P}_i(X_t = i \cap K_t^{\mathsf{c}}) \\ &\le e^{Ct}\mathbf{P}_i(K_t^{\mathsf{c}}) \end{aligned}$$

$$\leq e^{Ct}\mathbf{P}_i(T_2 < t),$$

which is $o(t)$. (3.5) now follows, proving the statement. $\square$

We can now prove the backward equation in its more general form.

Theorem 3.8 (General Backward Equation). *Suppose there exists a constant $C < \infty$ such that*

$$\sup_{i \in S} E(i) \leq C.$$

Then we have the Kolmogorov backward equation

$$\frac{d}{dt}P^E(t) = R^E P^E(t).$$

Proof. By Lemma 3.1, we have

$$P^E_{ij}(t+h) - P^E_{ij}(t) = \sum_{k \neq i} P^E_{ik}(h)P^E_{kj}(t) - (1 - P^E_{ii}(h))P^E_{ij}(t).$$

Dividing by h and taking the limit as $h \to 0$, we have

$$P'^E_{ij}(t) = \lim_{h \to 0} \sum_{k \neq i} \frac{P^E_{ik}(h)}{h} P^E_{kj}(t) - \lim_{h \to 0} \frac{1 - P^E ii(h)}{h} P_{ij}(t).$$

From Lemma 3.3, we know that the second term is

$$(v_i - E(i))P^E_{ij}(t) = -R^E_{ii}P_{ij}(t).$$

We focus on the first term, which is the sum of two parts:

$$\lim_{h \to 0} \sum_{k > N, k \neq i} \frac{P^E_{ik}(h)}{h} P^E_{kj}(t) \tag{3.9}$$

$$\lim_{h \to 0} \sum_{k \leq N, k \neq i} \frac{P^E_{ik}(h)}{h} P^E_{kj}(t). \tag{3.10}$$

Since the first expression is a finite sum, we have no problem switching the limit and the sum:

$$\lim_{h \to 0} \sum_{k \leq N, i \neq k} \frac{P^E_{ik}(h)}{h} P^E_{kj}(t) = \sum_{k \leq N, i \neq k} \lim_{h \to 0} \frac{P^E_{ik}(h)}{h} P^E_{kj}(t) = \sum_{k \leq N, i \neq k} q_{ik}P^E_{kj}(t) \tag{3.11}$$

by applying Lemma 3.3.

Next, let $\varepsilon > 0$. Since $\sum_k P_{ik}$ is a convergent sum, we may choose N so large that $N > i$ and $\sum_{k>N} P_{ik} \leq \frac{\varepsilon}{v_i e^{2Ct}}$. Then, as in the proof of Lemma 2.1 in the text, we know that

$$\begin{aligned}\sum_{k > N} \frac{P^E_{ik}(h)}{h} &\leq e^{Ct} \sum_{k > N} \frac{P_{ik}(h)}{h} \\ &\leq e^{Ct}\left(\frac{\mathbf{P}_i(T_1 < h \cap X_{T_1} > N)}{h} + \frac{\mathbf{P}_i(T_2 < h)}{h}\right),\end{aligned}$$

where T_1 is the time of the first jump of the chain. Since we know from the proof of Lemma 2.1 (in [1]) that $\mathbf{P}_i(T_2 < h) = o(h)$, we may focus on the first term on the right hand side of the inequality. We have:

$$\begin{aligned}\lim_{h\to 0}\frac{\mathbf{P}_i(T_1 < h \cap X_{T_1} > N)}{h} &\le e^{Ct}\lim_{h\to 0}\frac{1-e^{-v_i h}}{h}\sum_{k>N}P_{ik}\\ &= e^{Ct}v_i\sum_{k>N}P_{ik}\\ &\le \frac{\varepsilon}{e^{Ct}},\end{aligned}$$

where the equality follows from the Taylor expansion of $1-e^{-v_i h}$. This shows that for all $\varepsilon > 0$, there exists N such that

$$\lim_{h\to 0}\sum_{k>N}\frac{P^E_{ik}(h)}{h} \le \frac{\varepsilon}{e^{Ct}}$$

hence

$$\lim_{h\to 0}\sum_{k>N}\frac{P^E_{ik}(h)}{h}P^E_{ik}(t) \le \varepsilon.$$

Furthermore, for any such N, we have that

$$\sum_{k\le N, i\ne k} q_{ik}P^E_{kj}(t) \le \lim_{h\to 0}\sum_{k\ne i}\frac{P^E_{ik}(h)}{h}P^E_{kj}(t).$$

Thus in summary, for every $\varepsilon > 0$, there exists an N such that

$$\sum_{k\le N, i\ne k} q_{ik}P^E_{kj}(t) \le \lim_{h\to 0}\sum_{k\ne i}\frac{P^E_{ik}(h)}{h}P^E_{kj}(t) \le \sum_{k\le N, i\ne k} q_{ik}P^E_{kj}(t) + \varepsilon.$$

Taking $N \to \infty$ gives the desired result that

$$\lim_{h\to 0}\sum_{k\ne i}\frac{P^E_{ik}(h)}{h}P^E_{kj}(t) = \sum_{k\ne i} q_{ik}P^E_{kj}(t),$$

and now the theorem follows. □

Just as we have seen that the exponential of the generator solves the backward equations in Proposition 2.4, so too does the modified generator R^E solve the generalized backward equations. We will use this fact in Section 4.

3.2. **Arriving at the Feynman-Kac Formula.** Let $u(t, i) : [0, \infty) \times S \to \mathbb{R}$ be a function and let $u_0(i) : S \to \mathbb{R}$ denote $u(0, t)$. Previously, we only considered a specific case for the function u_0, where we had $u_0(i) = \mathbf{1}_j$. We now consider a general case. Our goal in this section is to find a solution to the differential equation

$$\frac{\partial}{\partial t}u(t,i) = (R^E u(t,i))_i = \sum_{j\in S} R^E_{ij}u(t,j) \tag{3.12}$$

with initial condition u_0. In particular, we can show that the solution is

$$u(t,i) = \sum_{j\in S} P^E_{ij}(t)u_0(j).$$

As before, we assume that E is bounded from above. We go about solving (3.12) using several lemmas. We wish to use our results for the backward equations discussed in the last section, but to ensure our arguments hold for infinite state spaces, some care is needed, so we first need to establish some preliminary results on when the sums we are dealing with converge, and when certain operations can be exchanged with infinite sums in our context.

First, we see that the boundedness of the function E allows us to have well-defined entries of certain products. We will be particularly interested in applying the following lemma when $p = 1$, but we give a general statement for completeness.

Lemma 3.13. *Suppose that the function u_0 is such that $\|u_0\|_{\ell^p} < \infty$ for some $1 \le p \le \infty$. Then if $P(t)$ is a doubly stochastic matrix for all time t, we have $\|P^E(t)u_0\|_{\ell^p} < \infty$.*

Proof. First assume that $p < \infty$. Then

$$\begin{aligned}
\|P^E(t)u_0\|_{\ell^p} &= \left(\sum_{i\in S}|(P^E(t)u_0)_i|^p\right)^{1/p} \\
&= \left(\sum_{i\in S}\left|\sum_{j\in S}P^E_{ij}(t)u_0(j)\right|^p\right)^{1/p} \\
&\le e^{Ct}\left(\sum_{i\in S}\left|\sum_{j\in S}P_{ij}(t)u_0(j)\right|^p\right)^{1/p} \\
&= e^{Ct}\left(\sum_{i\in S}|\mathbf{E}(u_0(X_t)\mid X_0=i)|^p\right)^{1/p} \\
&\le e^{Ct}\left(\sum_{i\in S}\mathbf{E}(|u_0(X_t)|^p\mid X_0=i)\right)^{1/p} \\
&= e^{Ct}\left(\sum_{i\in S}\sum_{j\in S}P_{ij}(t)|u_0(j)|^p\right)^{1/p} \\
&= e^{Ct}\left(\sum_{j\in S}\sum_{i\in S}P_{ij}(t)|u_0(j)|^p\right)^{1/p} \\
&= e^{Ct}\left(\sum_{j\in S}|u_0(j)|^p\sum_{i\in S}P_{ij}(t)\right)^{1/p} \\
&= e^{Ct}\left(\sum_{j\in S}|u_0(j)|^p\right)^{1/p} \\
&< \infty.
\end{aligned}$$

The inequality is from Jensen's inequality. Tonelli's theorem allows us to switch the sums since the terms are nonnegative. The last equality uses that the matrix is doubly stochastic.

Now we consider $p = \infty$. We have

$$\|P^E(t)u_0\|_{\ell^\infty} = \sup_{i\in S}|(P^E(t)u_0)_i| = \sup_{i\in S}\left|\sum_{j\in S} P^E_{ij}(t)u_0(j)\right|.$$

But

$$\begin{aligned}\left|\sum_{j\in S} P^E_{ij}(t)u_0(j)\right| &\le \sum_{j\in S} P^E_{ij}(t)|u_0(j)| \\ &\le e^{Ct}\sum_{j\in S} P_{ij}(t)\|u_0\|_{\ell^\infty} \\ &= e^{Ct}\|u_0\|_{\ell^\infty},\end{aligned}$$

which proves the statement. □

Lemma 3.14. *Assume that $\|u_0\|_{\ell^1} < \infty$. Then*

$$\frac{d}{dt}\sum_{j\in S} P^E_{ij}(t)u_0(j) = \sum_{j\in S}\frac{d}{dt}P^E_{ij}(t)u_0(j).$$

Proof. Of course, if S is finite, the result is immediate. The question of when the interchange of derivatives and infinite sums is allowed, however, gives rise to a well-known result in real analysis. Our goal, therefore, is to be able to apply Theorem 7.17 of [2] to this particular situation, which allows us to interchange the sum and derivative when the sum is infinite.

Define

$$f_n(t) := \sum_{j=1}^{n} P^E_{ij}(t)u_0(j).$$

Observe that, by the general backward equations Theorem 3.8 and linearity of the derivative, f_n is differentiable and

$$f'_n(t) = \sum_{j=1}^{n} P'^E_{ij}(t)u_0(j).$$

Next, we would like to show that $\{f_n\}$ converges pointwise for at least one point $t \in [0,\infty)$. If $u_0 \in \ell^1$, then we have

$$\begin{aligned}\lim_{n\to\infty} f_n(t) &= \lim_{n\to\infty}\sum_{j=1}^{n} P^E_{ij}(t)u_0(j) \\ &\le e^{Ct}\lim_{n\to\infty}\sum_{j=1}^{n} u_0(j) \\ &\le e^{Ct}\lim_{n\to\infty}\sum_{j=1}^{n}|u_0(j)| \\ &= e^{Ct}\|u_0\|_{\ell^1} \\ &< \infty.\end{aligned}$$

Hence, $\{f_n\}$ converges pointwise. Next, to be able to use the theorem, we must show that $\{f'_n\}$ converges uniformly. First, we note that, by Theorem 3.8, we have

$$
\begin{aligned}
|P'^E_{ij}(t)u_0(j)| &= \left|\sum_{k=1}^{\infty} R^E_{ik}P^E_{kj}(t)\right||u_0(j)| \\
&\leq |u_0(j)|\sum_{k=1}^{\infty}|R^E_{ik}P^E_{kj}(t)| \\
&\leq |u_0(j)e^{Ct}|\sum_{k=1}^{\infty}|R^E_{ik}| \\
&\leq e^{Ct}|u_0(j)|\left(\sum_{k\neq i}(v_iP_{ik}) + v_i + |E(i)|\right) \\
&= e^{Ct}(2v_i + |E(i)|)|u_0(j)|.
\end{aligned}
$$

It then follows from the Weierstrass M-test that

$$\lim_{n\to\infty} f'_n(t) = \sum_{j=1}^{\infty} P'^E_{ij}(t)u_0(j)$$

indeed converges uniformly on $[0,\infty)$ hence on any closed interval in the positive reals. We then apply Theorem 7.17 from [2] to conclude that on any $[a,b]\subset[0,\infty)$, we have

$$\frac{d}{dt}\sum_{j=1}^{\infty} P^E_{ij}(t)u_0(j) = \sum_{j=1}^{\infty}\frac{d}{dt}P^E_{ij}(t)u_0(j).$$

The result now follows for all of $[0,\infty)$. □

Lemma 3.15. *We may switch the following order of summation:*

$$\sum_{j\in S}\sum_{k\in S} R^E_{ik}P^E_{kj}(t) = \sum_{k\in S}\sum_{j\in S} R^E_{ik}P^E_{kj}(t).$$

Proof. By Fubini's theorem (or equivalently Theorem 8.3 in [2]), we know that if the double sum converges absolutely, we can switch the order of summation. By Tonelli's Theorem, we have

$$
\begin{aligned}
\sum_{j=1}^{\infty}\sum_{k=1}^{\infty}|R^E_{ik}|P^E_{kj}(t) &= \sum_{k=1}^{\infty}\sum_{j=1}^{\infty}|R^E_{ik}|P^E_{kj}(t) \\
&= \sum_{k=1}^{\infty}|R^E_{ik}|\sum_{j=1}^{\infty}P^E_{kj}(t) \\
&\leq e^{Ct}\sum_{k=1}^{\infty}|R^E_{ik}| \\
&\leq e^{Ct}(2v_i + |E(i)|) \\
&< \infty,
\end{aligned}
$$

so that the double sum converges. The result then follows from Fubini's theorem. □

Theorem 3.16 (Feynman-Kac Formula)**.** *Assume that* $u_0 \in \ell^1$. *A solution to (3.12) is*

$$u(t,i) = \sum_{j \in S} P^E_{ij}(t) u_0(j). \tag{3.17}$$

That is,

$$u(t,i) = \mathbf{E}_i \left[\exp\left(\int_0^t E(X_s) ds \right) u_0(X_t) \right]. \tag{3.18}$$

Proof. We check that the expression indeed solves the equation. By Theorem 3.8 and the previous lemmas, we have

$$\begin{aligned}
\frac{\partial}{\partial t} u(t,i) &= \frac{d}{dt} \sum_{j \in S} P^E_{ij}(t) u_0(j) \\
&= \sum_{j \in S} \frac{d}{dt} P^E_{ij}(t) u_0(j) \\
&= \sum_{j \in S} \left(\sum_{k \in S} R^E_{ik} P^E_{kj}(t) \right) u_0(j) \\
&= \sum_{k \in S} R^E_{ik} \left(\sum_{j \in S} P^E_{kj}(t) u_0(j) \right) \\
&= \sum_{k \in S} R^E_{ik} u(t,k).
\end{aligned}$$

This shows that (3.17) solves (3.12). Then we have

$$\begin{aligned}
u(t,i) &= \sum_{j \in S} P^E_{ij}(t) u_0(j) \\
&= \sum_{j \in S} \mathbf{E}_i \left[\exp\left(\int_0^t E(X_s) ds \right) \mathbf{1}_{\{X_t = j\}} \right] u_0(j) \\
&= \sum_{j \in S} \mathbf{E}_i \left[\exp\left(\int_0^t E(X_s) ds \right) \mid X_t = j \right] P_{ij}(t) u_0(j) \\
&= \mathbf{E}_i \left[\mathbf{E}_i \left[\exp\left(\int_0^t E(X_s) ds \right) \mid X_t \right] u_0(X_t) \right] \\
&= \mathbf{E}_i \left[\exp\left(\int_0^t E(X_s) ds \right) u_0(X_t) \right],
\end{aligned}$$

where the last equality holds by the law of total expectation. Thus, we see that the forms (3.17) and (3.18) are equivalent. □

The form presented in (3.18) is a particularly helpful expression for understanding $u(t,i)$, as it allows us to properly think of, say, heat diffusion, as an average based on heat at nearby states and paths leading to them.

We have previously discussed how one can think of Theorem 3.16 in general terms. When the environment term comes into play, the diffusion analogy is less clear. However, both $E = 0$ and $E \neq 0$ are interesting cases to consider, so let us re-examine them more closely.

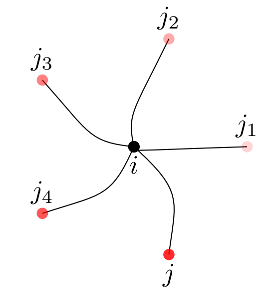

FIGURE 4. With general initial conditions, $u(t, i)$ averages over the initial heat at the different points, distinguished by their opacity here.

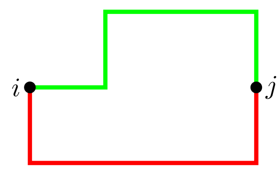

FIGURE 5. A path from i to j through a region where the environment is positive (green) will be weighted more in the average in (3.18) compared to a path through a region where the envirionment is negative (red).

If $E = 0$, Theorem 3.16 gives us

$$u(t, i) = \sum_{j \in S} P_{ij}(t) u_0(j) = \mathbf{E}_i[u_0(X_t)].$$

In terms of diffusion of, say, heat, this says the following: the amount of heat one has at i is an average calculated in the following way. A particle of heat that starts at i and travels for time t will end up at some other location j. The proportion of such paths that run for time t and end up at j is $P_{ij}(t)$. By symmetry, this is equal to $P_{ji}(t)$, which we interpret as the proportion of all paths that start at j and end up at i at time t. Next, we take the amount of heat that was at j to begin with into consideration; even if heat coming from j doesn't end up at i that often, it may still contribute a lot of heat to i if there is a lot of heat at j to begin with. Therefore, we are taking an average over all states of how much heat there is initially, something like an "intensity" of a certain state, weighted by how likely it is for heat to go from j to i by time t.

Now consider $E \neq 0$. A similar averaging takes place, except now one additionally considers the path that the chain X_s takes up to time t. One can imagine that there are, say, two paths that connect j and i, say p_1 and p_2. E creates a favorable environment on p_1 by having $E > 0$ on p_1, and a hostile environment on p_2 by having $E < 0$ on p_2. In terms of diffusion, the interpretation is less clear but can amount to heat being added into the system exponentially on p_1 and removed from the system exponentially at p_2. Thus, the path p_2 contributes less to the heat at i than p_1.

4. Examples

We now apply the results from previous sections to see the Feynman-Kac formula in action. Now, as one may suspect from the formulas given in Theorem 3.16, it is not feasible in general to compute the desired expectations directly. However, we call upon a fact alluded to in Proposition 2.4, where we saw that the solution to the backward equations is $P(t) = e^{tR}$. It follows (proof omitted) from the generalized backward equations in Theorem 3.8 that we have $P^E(t) = e^{tR^E}$. Therefore, given a generator R and an environment function E, one can find $P^E(t)$ with the assistance of a computer.

We consider the following example: we would like to see how heat diffuses on a cycle graph with seven states. Let us say that a particle at state i moves either to state $i+1$ or to $i-1$, with rate $1/2$ for each, and addition is done modulo 7. That is, we have the generator matrix

$$R = \begin{bmatrix} -1 & 1/2 & 0 & 0 & 0 & 0 & 1/2 \\ 1/2 & -1 & 1/2 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & -1 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & -1 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & -1 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & -1 & 1/2 \\ 1/2 & 0 & 0 & 0 & 0 & 1/2 & -1 \end{bmatrix}.$$

First, we assume in Figure 6 that the initial condition is $\mathbf{1}_{\{i=1\}}$. We see that this gives us a transition matrix $P_{ij}(t)$ as our solution for $u(t,i)$. Furthermore, applying the formula discussed in Section 2.2, we see that

$$\frac{\partial}{\partial t} u(t,i) = \frac{u(t,i-1) + u(t,i+1)}{2} - u(t,i).$$

In a short amount of time, not much heat diffuses to the middle states, because this would be the same as the probability that a continuous time random walk starting at 1 makes several jumps. In small time, this is unlikely, as seen by the small heights of states 3, 4, and 5 at time $t = 0.5$ and $t = 1$. This showcases the $o(t)$ probability of two or more jumps in a time interval of length t that we discussed in Lemma 2.1 and Lemma 3.3. As time increases, heat spreads throughout the seven states. We see that heat will distribute itself evenly in this case; in the language of Markov chains, its stationary distribution is uniform. Indeed, one can check that the vector given by $\pi_i = 1/7$ for $i = 1, ..., 7$ satisfies $\boldsymbol{\pi} R = \mathbf{0}$.

Figure 7 considers a nontrivial initial condition, where we have $u_0(j) = j$ for $j = 1, ..., 7$. Consider the state $i = 1$. There is not much heat at i compared to other states. However, to reason about what happens as time passes, we refer to our rule $u(t,i) = \mathbf{E}_i[u_0(X_t)]$. As time passes, it is more likely that a chain starting at state 1 can travel to states where there is more heat, so we expect that $u_0(X_t)$ will increase. Likewise, if the chain starts at a state with more initial heat, it is more likely for X_t to be at states like 1, 2, or 3 that had less heat initially, so we expect $u_0(X_t)$ will decrease. After a long time, we have $u(t,i) = 4$ for each i. Thus, the total initial heat of $1 + \cdots + 7 = 28$ was able to distribute itself evenly across the 7 states as before.

Figure 8 now considers the case where we have a nonzero environment function. Here we have

$$R^E = R + diag(-3, -2, -1, 0, 1, 2, 3),$$

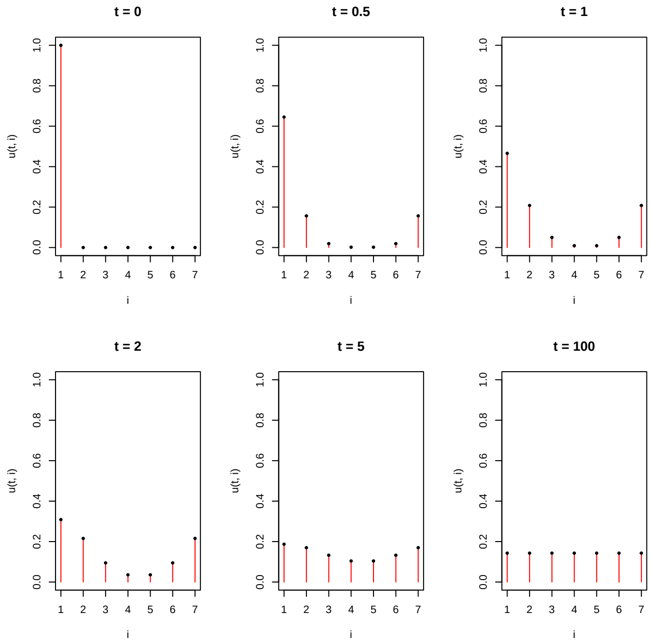

FIGURE 6. Seven-state example with trivial initial condition and no environment term.

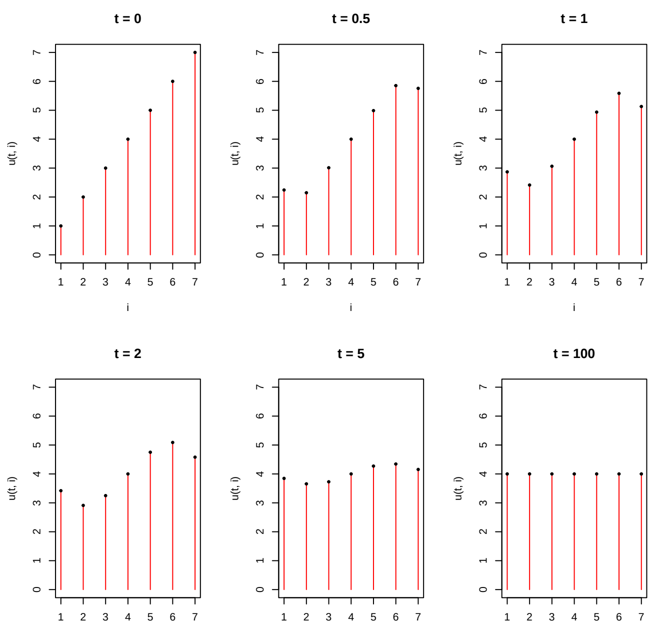

FIGURE 7. Same example with $u(0, j) = j$.

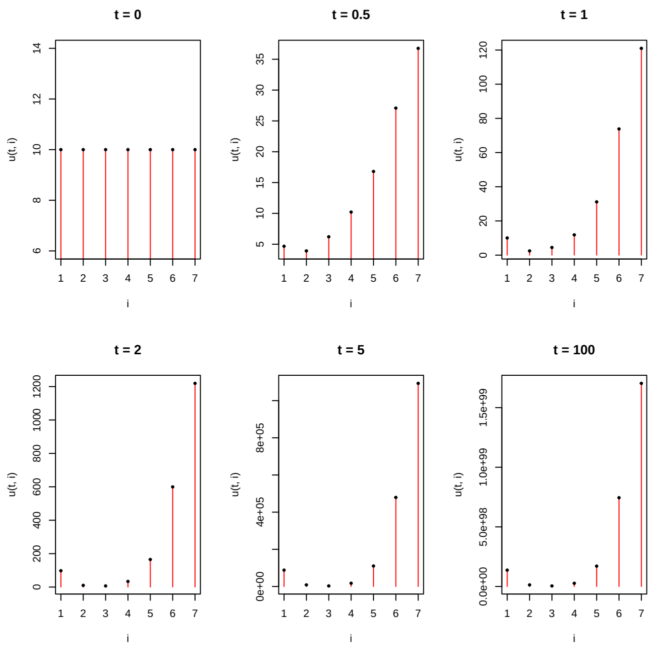

FIGURE 8. Example with Environment and $u(0, j) = 10$.

where R is as before. We start with equal heat at each state and we see the exponential growth occur. This growth is more pronounced at some states than others. Figure 9 shows how the environment affects the middle three states slightly differently, but explosion occurs in any case. In state 3, for instance, decay occurs initially, but exponential growth begins to occur after small time.

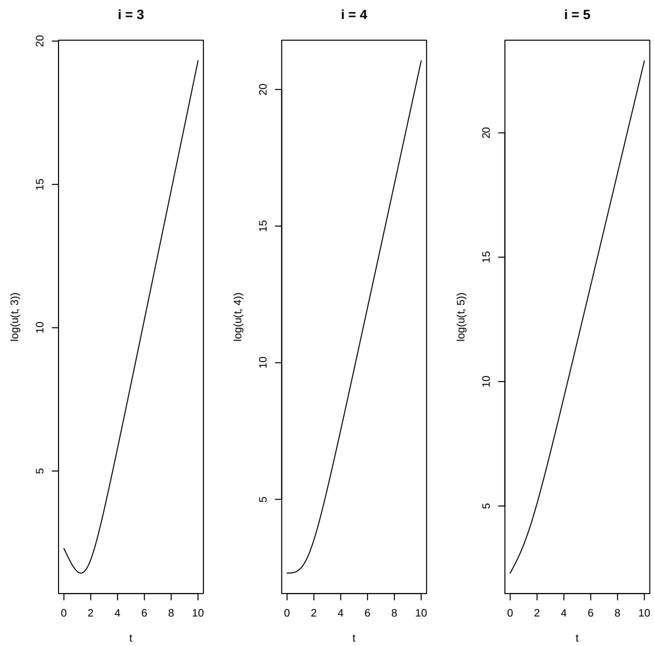

FIGURE 9. States 3, 4, and 5 over time with log scale.

REFERENCES

[1] Florescu, Ionuţ. Probability and Stochastic Processes. Hoboken, NJ: Wiley, 2015.
[2] Rudin, Walter. Principles of Mathematical Analysis. 3rd ed. New York: McGraw-Hill, 1976.