Stats 150, Fall 2019
Lecture 19, Thursday, 11/7/2019

1 A Remark on Homework

Pitman mentions that we may find the last problem to be challenging. If we find ourselves stuck in a fairly general scenario, we can give up, or we can break this down into a specific example. Pitman recommends "How to Solve It" by Polya, which is to take a special case of some interest. Then if we understand the structure of the special case, we may get an idea for the general case. In this problem, Pitman suggests to try the case if there are only two ways out of the initial state x. That is, from $x = 0$, we can only move to ± 1 (this is the typical setup for a B/D process). Even in this simple scenario, the problem is highly nontrivial.

2 Review of Hold-Jump Description

As this is tremendously important, we'd like to look deeper into the Hold-Jump description of a Continuous-time Markov chain. Take a finite state space S and a Markov chain $X_t, t \geq 0)$ and $X_t \in S$. We typically make the implicit assumption (and we make this explicit now) that the paths of X are right-continuous step functions. That is, if $X_0 = x$ fo
That is (as we have covered in §16.1),

$$P_t(x, y) := \mathbb{P}_x(X_t = y)$$

gives a semigroup of transition matrices with $P_{s+t} = P_s P_t$. In matrix form, we have

$$P_{s+t} = \sum_{y \in S} P_s(x, y) P_t(y, z)$$

Then in probability form, by conditioning on X_s and the time-homogeneous Markov property, the above is equivalent to

$$\mathbb{P}_x(X_{s+t} = z) = \sum_{y \in S} \mathbb{P}_x(X_s = y, X_{s+t} = z). \tag{1}$$

We know by analysis that every such semigroup $(P_t, t \geq 0)$ is of the form

$$P_t = \exp(Qt), \text{ where } Q = \frac{d}{dt} P_t \Big|_{t=0+} = \lim_{t \to 0} \frac{P_t - I}{t}.$$

Then to get the backwards equation (essentially that the Q part comes first), we differentiate to get

$$\frac{d}{dt} P_t = Q P_t.$$

Alternatively, we can get the forward equations

$$\frac{d}{dt} P_t = P_t Q.$$

Pitman notes that this is the limit of equation (1) for $t := \delta + (t - \delta)$, and this result is truly driven by (1).

2.1 Probability Interpretation

Now the probability interpretation from this is how we may make such a Markov chain. We have constructed this before by using a PPP, which implies a "hold-jump" description. Suppose that we start in state x, so that we must hold there for H_x amount of time. Then $H_x \sim \text{Exponential}(\lambda_x)$ for $\lambda_x - Q(x,x)$.
Let's take a state-space $x \in \{1,2,3\}$ and consider $\pi = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. Then we have $P_t(1,1) \geq \mathbb{P}_1(H_1 > t)$. Essentially, in a short amount of time, in terms of the hold-jump process, there is a small chance that we've left a state, and given if we have left a state, there is an extremely small likelihood that we come back to that same state.
If we want to think of all elements of the transition matrix, it may be helpful to graph the following picture of all three curves in terms of the differential equations.
Pitman would like that we see how the curves of the probabilities reflect the dynamics of the chain. Bringing this back to our present discussion, at time H_x, the chain jumps to some other state that is not x. That is, we go to state J_x, which is:

$$\mathbb{P}(J_x = z) = \frac{Q(x,z)}{\lambda_x},$$

where $z \neq x$ and $\lambda_x = -Q(x,x) = \sum_{z \neq x} Q(x,z)$.
This uses the fact that $\sum_z P_t(x,z) = 1$, and therefore differentiating gives

$$\sum_z \frac{d}{dt} P_t(x,z) = 0 \implies \sum_z Q(x,z) = 0, \forall_x.$$

Essentially, if one curve 'goes down', another must 'go up' to compensate. That is, $Q\mathbb{1} = \vec{1}$.
Next, we condition on the case $H_x = t$ and $J_x = z$. We've held and jumped, so the next thing is to hold: we hold in state z for an Exponential(λ_z) time, where $\lambda_z = -Q(z,z)$. Then at this time, because we conditioned on arriving at z, we jump according to the jump distribution $\frac{Q(z,\cdot)}{\lambda_z}$, where $\cdot \neq z$.
We introduce the following notation.

Definition: Embedded Jumping Chain -

Let us say

$$\begin{aligned} Z_0 &= x \\ Z_1 &= J_x \\ Z_2 &= \text{next state after } J_x, \end{aligned}$$

and so on, where $Z_0 \neq Z_1 \neq Z_2 \neq Z_3 \neq \cdots$.
We call this the **embedded jumping chain**. This has the transition matrix

$$(x,y) \to \frac{Q(x,y)}{\sum_{z \neq x} Q(x,z)},$$

where we use the off-diagonal part of Q.

Then in terms of holds, we summarize:

$$\begin{aligned} H_1 &: \text{ in state } Z_1 \\ H_2 &: \text{ in state } Z_2 \\ &\vdots \end{aligned}$$

Now given $Z_0, Z_1, \dots, Z_n$ and $H_0, H_1, \dots, H_n$ with $Z_n = z$,

$$\begin{aligned} H_{n+1} &\sim \exp(\lambda_z) \\ Z_{n+1} &\sim Q(z, \cdot)/\lambda_3, \text{ where } \cdot \neq z. \end{aligned}$$

3 Exit Distributions/Times (Durett §4.4)

Pitman would like to first sketch some ideas. Suppose that we have an irreducible chain, and that we have some set of states C (the 'interior') and some target sets A, B, where A, B, C are disjoint. We are interested in two things:
(1) the probability of starting at x of hitting, for example, A before B ($\mathbb{P}_x$(hit B before A)), and
(2) the expectation $\mathbb{E}_x$(time to hit $A \cup B$).
We discussed and solved this problem before in the discrete case, and now we would like to generalize this to continuous-time. Notice that (1) is easy, because there is no new difficulty from working in continuous-time. Recall that our Q matrix tells us how long we need to hold and when we need to jump. Then (1) is simply a question about what the jumping chain (which we previously called $Z_0, Z_1, Z_2, \dots$) does. That is,

$$\mathbb{P}_x(x \text{ hits } B \text{ before } A) = \mathbb{P}_x(Z \text{ hits } B \text{ before } A).$$

We can solve this with the same old methods, and if we see in the text, we can wrap this up with a solution in terms of the Q matrix. Recall this for a discrete-time Markov chain was to take $h = Ph$ with some boundary conditions. Notice this is equivalent to $(I - P)h = 0$. Now in this case, we can neatly write

$$\frac{(I - P_t)}{t} h = 0 \implies \boxed{Qh = 0},$$

but this is not a new result.
Let's look at problem (2), which is a little more interesting. Recall from discrete time in that if we take a set C of interior states, we have a matrix P like:

$$P: \qquad \begin{bmatrix} * & C & A \cup B \\ C & P_C & \cdot \\ A \cup B & \cdot & \cdot \end{bmatrix}$$

where $P_c(x, y) := P(x, y)$ for $x, y \in C$. This is simply the $C \times C$ restriction of P. See the sub-stochastic discussion in our midterm exam.
The fact that the row sums are less than 1 tells us that

$$I + P_C + P_C^2 + P_C^3 + \cdots = (I - P_C)^{-1},$$

because we have (rapid) convergence. Recall this is the Green matrix:

$$\begin{aligned} (I - P_c)^{-1}(x, y) &= \mathbb{E}_x \sum_{n=0}^{\infty} \mathbb{1}(X_n = y, T > n) \\ &= \mathbb{E}_x \sum_{n=0}^{\infty} \underbrace{\mathbb{P}_x(X_n = y, T > n)}_{= P_C^n(x,y)} \end{aligned}$$

where $T = \min\{n \geq 0 : X_n \notin C\}$.
If $\mathbb{E}_x T$ is the expected hitting time, we can extract this from our Green matrix by 'covering up' $X_n = y$ by summing out the cases. That is,

$$\mathbb{E}_x T = \sum_y (I - P_C)^{-1}(x, y)$$

Let's do the same thing but in the continuous parameter case. Take

$$T := \min\{t \geq 0 : X_t \notin C\}.$$

Let C be fixed. We want something similar to our discrete case, where for $n = 1, 2, \dots,$ we have

$$P_C^n(x, y) = \mathbb{P}_x(X_n = y, T > n).$$

We should expect the continuous-time analog to be

$$P_{C,t}(x, y) = \mathbb{P}_x(X_t = y, T > t).$$

Then

$$\begin{aligned} \mathbb{E}_x \int_0^\infty \mathbb{1}(X_t = y, T > t) &= \mathbb{E}_x(\text{ time in } y \text{ before } T) \\ &= \int_0^\infty P_{C,t}(x, y)\ dt. \end{aligned}$$

We should guess, expect, or hope that

$$\boxed{P_{C,t} = \exp(Q_c t)},$$

where Q_C is simply Q restricted to C.
Now suppose we trust that $P_C = \exp(Qt)$, and that $\int_0^\infty \exp(Qt)\ dt < \infty$. In discrete time, we summed and we had the geometric series. We might guess that

$$\begin{aligned} \sum_{n=0}^\infty r^n = \frac{1}{1-r} = (1-r)^{-1} &\implies \int_0^\infty e^{-\lambda t}\ dt = \frac{1}{\lambda} \\ &\implies \boxed{\int_0^\infty e^{Qt}\ dt = \frac{1}{-Q} = (-Q)^{-1}}. \end{aligned}$$

Durett uses the term R for Q_C.

Lecture ends here.