Stats 150, Fall 2019
Lecture 20, Tuesday, 11/12/2019

1 Remarks

We have a worksheet this week with a lot of problems, where some of these are due for homework. Pitman emphasizes that our time should be applied to problems and exercises.
Today we'll start with some remarks regarding the last problem of last week's homework. This is a very conceptual problem that gets us to think about Markov chains in the manner that Pitman would like us to view them.

Homework Problem. Suppose you have a recurrent chain in continuous time with transition rate $Q(0, j)$ from state 0 to state $j > 0$. Starting in state 0, let $H(0, j)$ be the total time spent in state 0 before the first jump from 0 to j.
a) What is the distribution of $H(0, j)$?
b) What is the joint distribution of $H(0, j)$ and $H(0, k)$ for different j and k?

Recall that the problem looks at a fixed state 0 of a chain in continuous time, and we look for each $y \neq 0$ at transitions from 0 to state y. Assume (otherwise this does not work) that the chain is recurrent.
To remind us, we have $P_t = e^{Qt} = \cdots$, and $Q(0, j)$ is the rate of transition from 0 to j. We have two meanings:
1) The small time meaning of transition rate, in words, is that if we start in state 0 and we run for a **small** time t,

$$P_t(0, j) = \mathbb{P}_0(X_t = j) = Q(0, j)t + o(t)$$

as our first-order approximation for the very small chance that we change states. This is,

$$\frac{P_t(0, j)}{t} = Q(0, j) + \underbrace{\frac{o(t)}{t}}_{\to 0}$$

We go to 1, float around in our state space, then come back to 0, and so on. The entire discussion is about these underbraced "successive" holds in state 0. Let $H_1, H_2, H_3, \ldots$ be the successive holds in state 0. We have a few observations:
1) Note that there are infinitely many of these because the problem gives that the chain is recurrent (we assume 0 was recurrent). That is, we keep coming back to 0 and have more opportunities to make transitions out of state 0.

2) These H_i are iid Exponential(λ_0) where $\lambda_0 = \sum_{j \neq 0} Q(0, j)$, the off-diagonal row sum.

3) Surely, $H_{0j} \geq H_1$, so we can write:

$$H_{0j} = H_1 + H_2 + H_3 + \cdots + H_{N(j)},$$

where $N(j)$ is the number of holds in state 0 before the first (direct) $0 \to j$ transition. In our diagram, we have $N(+1) = 1$ and $N(-1) = 3$. Then

$N(j) \sim$ Geometric($p = \frac{Q(0,j)}{\lambda_0}$) because of the hold-jump description of a Markov chain.

Let $J_1, J_2, J_3, \dots$ be the locations of jumps (that is, J_k is the state jumped to after the hold H_k).

4) Then $J_1, J_2, J_3, \dots$ are iid with $\mathbb{P}(J_i = j) = \frac{Q(0,j)}{\lambda_0}$.

5) By the Markov property, it is intuitive that $H_1, H_2, \dots$ and $J_1, J_2, \dots$ are independent (this was a part of the hold-jump description).

6) In terms of H_i and J_i, we have that:

$$N(j) := \min\{k \geq 1, J_k = j\},$$

where we start our indexing at 1.

7) We have the familiar fact that

$$\mathbb{P}(N(j) = k) = q^{k-1}p,$$

where $p = \frac{Q(0,j)}{\lambda_0}$ and $q := 1 - p$.
In summary, we have

$$H_{0j} = H_1 + \cdots + H_{N(j)}$$

where $N(j)$ is Geometric($Q(0,j)/\lambda_0$) independent of $H_1, H_2, \dots$. This implies that

$$H_{0j} \sim \text{Exponential}(Q(0,j)),$$

with

$$Q(0,j) = \lambda_0 \cdot \frac{Q(0,j)}{\lambda_0}.$$

Either by computation (e.g. of densities) or by analogy or instance of previous setup, we have a picture of the thinning of a Poisson Point Process.
In this picture, we see that the time until the first $\oplus$ is exponential. Now from the same picture, we can do part (b) of this problem.
(b) Because we have two independent (disjoint) regions in our PPP, H_{0+} and H_{0-} are independent by Poisson marking or thinning facts. This is also true for $H_{0,j}$ as j varies. To generalize this from 2 outcomes to 3, we simply another strip in our picture. Continue this as necessary.

2 The Laplace Transform

There was a question in the audience regarding how to show that the sum of a Geometric(p) number of Exponential(λ) is Exponential(λp) without the PPP picture we have drawn before.
Pitman gives two solutions. One is by computing the density by conditioning on some random variable N, but we will not pursue this because it is slightly boring.
The second method is to compute a suitable transform. Rrecall that if we are adding a random number N of iid positive-integer valued random numbers, we could use PGFs (e.g. as we have for branching processes).
We need a variant of PGFs for continuous variables, especially for nonnegative random variables $X \geq 0$. We have two related concepts:

1) MGF(θ) (Moment Generating Function)
2) Laplace transform, which is nothing more than MGF($-\theta$). The Laplace transform is nice for $X \geq 0$ for two reasons.
Let us define, for a random variable $X \geq 0$:

$$\boxed{\phi_X(\theta) := \mathbb{E}e^{-\theta X}}.$$

A leading example for which we can easily find this function is if $X \sim$ Exponential(λ) and $X \stackrel{d}{=} \mathcal{E}/\lambda$. In this case, then

$$\begin{aligned}
\phi_X(\theta) &= \mathbb{E}e^{-\theta\mathcal{E}/\lambda} \\
&= \mathbb{E}e^{-\frac{\theta}{\lambda}\mathcal{E}} \\
&= \int_0^\infty e^{-\frac{\theta}{\lambda}t}e^{-t}\ t \\
&= \left(\frac{\theta}{\lambda}+1\right)^{-1} = \boxed{\frac{\lambda}{\lambda+\theta}}.
\end{aligned}$$

Notice that this is the (same answer as and hence exactly is) the race between two exponentials. In general, we have:

$$\phi_X(\lambda) := \mathbb{E}\left(e^{-\lambda X}\right) = \mathbb{P}\left(X < \frac{\mathcal{E}}{\lambda}\right) = \mathbb{P}(\lambda X < \mathcal{E}).$$

We arrive at this by first considering if X is constant. Now if we assume that (1) $\mathcal{E} \sim$ Exponential(1) where $\mathbb{P}(\mathcal{E} > t) = e^{-t}$ and (2) $\mathcal{E}$ and X are independent.
In summary, we can always interpret the Laplace transform $\phi_X(\theta)$ of a random variable as $\mathbb{P}(X < \frac{\mathcal{E}}{\lambda})$ for $\mathcal{E}$ independent of X.
Now for $X \sim$ Exponential(λ), we have

$$\begin{aligned}
\phi_X(\theta) &= \mathbb{E}e^{-\theta X} \\
&= \mathbb{P}(X < \frac{\mathcal{E}}{\theta}) \\
&= \mathbb{P}\left(\frac{\mathcal{E}'}{\lambda} < \frac{\mathcal{E}}{\theta}\right), \quad \mathcal{E}, \mathcal{E}' \text{ iid} \\
&= \frac{\lambda}{\lambda+\theta}.
\end{aligned}$$

3 Key Facts about Laplace Transforms

Recognizing that this is a way to morph PMFs to a way of looking at continuous variables, once we have found the Laplace transform, we should have some uniqueness to the probability distribution.
1) Uniqueness Theorem. For $X, Y \geq 0$, if $\phi_X(\theta) = \phi_Y(\theta)$ for all $\theta \geq 0$, then $X \stackrel{d}{=} Y$.
That is, $\mathbb{E}g(X) = \mathbb{E}g(Y)$ for any g such that either side is defined. Pitman defers this proof for later in the course.
2) Provided X, Y independent,

$$\phi_{X+Y}(\theta) = \phi_X(\theta)\phi_Y(\theta).$$

To see this, consider:

$$\begin{aligned}\phi_{X+Y}(\theta) &= \mathbb{E}e^{-\theta(X+Y)} \\ &= \mathbb{E}\quad \underbrace{e^{-\theta X}}_{\text{independent}} \quad \underbrace{e^{-\theta Y}}_{\text{independent}} \\ &= \left(\mathbb{E}e^{-\theta x}\right)\left(\mathbb{E}e^{-\theta Y}\right) \\ &= \phi_X(\theta)\phi_Y(\theta).\end{aligned}$$

3) Another thing that was quite helpful (as we have seen with Branching processes) is that this helps us with random sums. Suppose that $X_1, X_2, \dots$ are iid and that they all have the same $\phi_{X_i}(\lambda) = \lambda_X(\lambda)$ as a consequence of iid. Let N be a random index in $\{0, 1, 2, \dots\}$ independent of $X_1, X_2, \dots$. Let's try to calculate the Laplace transform. Condition to get:

$$\begin{aligned}\mathbb{E}e^{-\theta(X_1+X_2+\cdots+X_N)} &= \sum_{n=0}^{\infty} \mathbb{E}\left[e^{-\theta(X_1+\cdots+X_n)}\mathbb{1}(N=n)\right] \quad \text{(true with no assumptions)} \\ &= \sum_{n=0}^{\infty} \mathbb{P}(N=n)[\underbrace{\overset{z}{\phi_X(\theta)}}]^n,\end{aligned}$$

which is the same as $G_N(\phi_X(\theta))$ where $G_N(z) := \sum_n \mathbb{P}(N=n)z^n$.

4 Conclusion

The Laplace transform of $X_1 + \cdots + X_N$ of θ is $G_N(\phi_X(\theta))$. To apply this, let's go back to the geometric sum of exponentials with parameter λ.
From before, we have

$$\phi_X(\lambda) = \frac{\lambda}{\lambda+\theta} = \frac{1}{1+\theta/\lambda}$$

as the Laplace transform of an exponential. Then the generating function is:

$$G_N(z) = \frac{pz}{1-qz} = \sum_{n=1}^{\infty} pq^{n-1}z^n = pz\sum_{n=1}^{\infty}(qz)^{n-1}.$$

Of course, this does not work very nicely if we have a possibility of 0. Finishing the calculation, we see that

$$G_N(\phi_X(\theta)) = \frac{p\frac{\lambda}{\lambda+\theta}}{1-q\frac{\lambda}{\lambda+\theta}},$$

simply substituting in the Laplace transform into the argument of the probability generating function. Then simplifying the above gives:

$$\begin{aligned}G_N(\phi_X(\theta)) &= \frac{p\theta}{\lambda+\theta-q\lambda} \\ &= \frac{p\lambda}{p\lambda+\theta},\end{aligned}$$

which is precisely the Laplace transform at θ of Exponential($p\lambda$). This calculation is much easier and sweeter than having to work with the densities. Although this is a bit indirect, this is very helpful.
There are concepts like the Stone-Weierstrass theorem that justify these concepts.

Lecture ends here.