Stats 150, Fall 2019
Lecture 10, Tuesday, 10/1/2019

1 Comments on Homework 5

This week's homework is posted as a worksheet on bCourses, and there is one correction as posted on Piazza. The first two problems are quite easy. The first is on branching processes.

For the second problem, we'll be frustrated if we don't know about Poisson thinning. This is Poisson thinning in disguise within a Markov chain.

In words, this means that if we have a poisson number of independent binomial trials, the count of heads is poisson. We should use the 'obvious' generating function for this problem.

Now, a hint for the Kac identities, which are about a stationary process of 0s and 1s.
We claim that $\mathbb{P}(1\underbrace{000}_{n}) = \mathbb{P}(\underbrace{000}_{n}1)$. This is much weaker than assuming reversibility (this is only for one such pattern). Pitman gives the hint to check:

$$\mathbb{P}(1000) = \mathbb{P}(*000) - \mathbb{P}(0000),$$

where $*$ acts as a wildcard and can be a 0 or 1. Pitman gives that once we've worked with this idea, it takes about 4 lines to solve the problem.

The exercise on tail generating functions is simply routine for us to get to get practice with generating functions.

The renewal generating function problem is relatively easy from the result of #4. We'll discuss a bit of renewal theory today in-class.

2 Renewal Generating Functions

For our discussions today, we assume P is irreducible.

We look at the summation

$$\sum_{n=0}^{\infty}\left(u_n - \frac{1}{\mu}\right) = \sum_{n=0}^{\infty}\left[P^n(0,0) - \pi(0)\right],$$

where (S_n) is a Markov chain, P is our transition matrix with stationary probability π, and we assume irreducible and aperiodic.
We take the state 0 to be a special state. Then in the language of Renewal theory (by definition, returning to our initial state),

$$P^n(0,0) = \mathbb{P}_0(\text{return to 0 at time } n) = \mathbb{P}(\text{renewal at } n).$$

This formula comes naturally from looking at a generalized Potential kernel (Green matrix):

$$G := \sum_{n=0}^{\infty} P^n = (I-P)^{-1},$$

for a Markov matrix P. This is very useful for transient chains.
We may ask, what is the Green matrix if P is recurrent?

$$G(x,y) = \mathbb{E}_x \sum_{n=0}^{\infty} \mathbb{X} = \; = \sum_{n=0}^{\infty} P^n(x,y).$$

In other words, this is $\mathbb{E}_x(\#$ of hits on y with ∞ time horizon). Then we know

$$G(x,y) = \infty, \forall_{x,y \in S} \iff P \text{ is recurrent}$$

Hence finite state irreducible P are very interesting. They have stationary distribution π, where $\pi P = \pi$.
However, $(I - P)^{-1} = \infty$, which is of no interest. Thankfully, there is something nearly as informative as this ($G = (1 - P)^{-1}$) for a recurrent chain.
Assume for simplicity that S is finite, and that P is aperiodic. Then we know a lot about P^n. We know

$$P^n(x,y) \to \pi(y)$$

Informally, that is, we lose track of where we started and just see the limit distribution π.
As an aside, we invent notation $\mathbb{1}(x) = 1$, for all $x \in S$. Just as $\pi P = \pi$, we want $\pi \mathbb{1} = 1$.
A bit 'cuter': we write simply:

$$P^n \to \Pi,$$

where the limit matrix Π is the matrix with all rows equal to π. Pitman gives an 'awefully cute notation':

$$\Pi = \mathbb{1}\pi.$$

Notice that $P^n \to \Pi$ rapidly as $n \to \infty$. If we're careful about this,

$$\left|P^n(x,y) - \pi(y)\right| \le c\rho^n,$$

for one $\rho < 1$. This implies that

$$\sum_{n=1}^{\infty} \left|P^n(x,y) - \pi(y)\right| < \infty,$$

which then implies that

$$\sum_{n=0}^{\infty} (P^n - \Pi)$$

exists, entrywise as a limit matrix.
Pitman calls attention to when we square $(P - \Pi)$. Recall that matrix multiplication is not commutative; however, we can still perform the expansion:

$$\begin{aligned}(P - \Pi)^2 &= (P - \pi)(P - \pi) \\ &= P^2 - P\Pi - \Pi P + \Pi^2 \\ &= P^2 - \Pi - \Pi + \Pi \\ &= P^2 - \Pi\end{aligned}$$

Pitman gives the exercise to check:

$$(P - \Pi)^n = P^n - \Pi,$$

where we need only use the binomial theorem to evaluate the coefficients. Then,

$$\begin{aligned}\sum_{n=0}^{\infty} (P^n - \Pi) &= I - \Pi + \sum_{n=1}^{\infty} (P^n - \Pi) \\ &= I - \Pi + \sum_{i=1}^{\infty} (P - \Pi)^n \\ &= \sum_{n=0}^{\infty} (P - \Pi)^n - \Pi.\end{aligned}$$

Recall that

$$\sum_{n=0}^{\infty} K^n = (I - K)^{-1},$$

for suitable K (like a sub-stochastic matrix). Now for our present expression, taking $K := \mathbb{P} - \Pi$, we have:

$$\boxed{\sum_{n=0}^{\infty}(P - \Pi)^n = (I - P + \Pi)^{-1}}$$

Then to summarize, our homework is to look at

$$\sum_{n=0}^{\infty} \left(u_n - \frac{1}{\mu} \right),$$

for a renewal sequence u_n. We can always write this, for a suitable Markov matrix P as

$$\begin{aligned}\sum_{n=0}^{\infty} \left(u_n - \frac{1}{\mu} \right) &= \sum_{n=0}^{\infty} (P^n(0,0) - \pi(0)) \\ &= \left[(I - P + \Pi)^{-1} - \Pi \right],\end{aligned}$$

for P, the transition matrix $\Pi = \mathbb{1}\pi$. (see “Renewal chain” early in the text).
What we learn is
For a recurrent P with finite state space S and $\pi P = \pi$, let $\Pi := \mathbb{1}\pi$. Then the matrix $I - P + \Pi$ is invertible.
Define

$$Z := (I - P + \Pi)^{-1} = \sum_{n=0}^{\infty}(P - \Pi)^n,$$

as before. Then there are lots of formulas for features of the Markov chain that pull / interpret entries of Z, which we call the **fundamental matrix**. For transient matrices, we pull entries out of the Green matrix. For recurrent chains, we pull entries out of Z.
We sketched a proof that Z exists for an aperiod chain, but this proof fails for a period chain where the series fails to converge. However, we can use the Abel sum and set:

$$Z := \lim_{S \uparrow 1} \sum_{n=0}^{\infty} (1 - P)^n S^n,$$

and everything works as before.

Example: An example would be to consider

$$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$

where $P^n(0,0)$ for $n = 0, 1, 2, \dots$ gives the sequence $\{1, 0, 1, 0, 1, 0, 1, 0, \dots)$. Then $\pi(0) = \frac{1}{2}$. Notice that

$$\left(P^n(0,0) - \frac{1}{2}\right) = \left(\frac{1}{2}, -\frac{1}{2}, \frac{1}{2}, -\frac{1}{2}, \cdots\right),$$

which is an oscillating sum if left untreated. Using an Abel sum, we should check that we get zero.

Break time.

Now we may ask, where else does Z come from? We consider the computation of variances in the Central Limit Theorem (CLT) for Markov chains. One such trivial special case is when $P = \Pi$. This means that under $\mathbb{P}_\lambda$ with $X_0 \sim \lambda, X_1, X_1, \dots,$ P times. Then all the variables are iid with $X_1 \sim \pi$, $X_2 \sim \pi$, and so on. That is, X_n for $n \geq 1$ are iid.
In this (iid) case with $P = \pi$ (or more generally), we look at:

$$S_n(f) := \sum_{k=1}^{n} f(X_k),$$

which we define as the reward from n steps of the chain if we are payed $\$f(x)$ for each value x. Then

$$\begin{aligned} \mathbb{E}S_n(f) &= n\mathbb{E}f(X_1) \\ &= n\pi f, \end{aligned}$$

where $\pi f = \sum_x \pi(x) f(x)$ is a number. Notice that X_i can itself be abstract, but we assume f to take on numerical values, so that we may discuss the expectation and variance. Then

$$\operatorname{Var}(S_n(f)) = n\operatorname{Var}(f(X_1)),$$

and $\operatorname{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$ equivalently gives, in matrix notation:

$$\operatorname{Var}(S_n(f)) = \pi f^2 - (\pi f)^2 = \sigma^2(f)$$

Then by the Central Limit Theorem,

$$\mathbb{P}\left(\frac{S_n(f) - n\pi f}{\sigma(f)\sqrt{n}} \leq z\right) \to \Phi(z),$$

the normal CDF. We ask, what about for a Markov chain?

2.1 Mean

$$\begin{aligned} \mathbb{E}_\lambda \sum_{k=1}^{n} f(X_k) &= \lambda \left(\sum_{k=1}^{n} P^k\right) f \\ &= n\lambda \left(\frac{1}{n}\sum_{k=1}^{n} P^k\right) f \\ &\sim n \underbrace{\lambda\Pi}_{=\pi} f = \boxed{n\pi f}, \quad \text{(asymptotically)} \end{aligned}$$

Notice that this is the same as the iid case, except this holds asympotitically at the limit instead of exactly.

2.2 Variance

A quick fact: asympotitically, it doesn't matter what λ is. We certainly see this for the mean, and it holds here for the variance as well.
So we try $\lambda := \pi$ and compute the variance for starting distribution:

$$\begin{aligned}\mathrm{Var}_\pi\left(S_n(f)\right) &= \mathrm{Var}_\pi\left(f(X_1)+f(X_2)+\cdots+f(X_n)\right)\\ &= \sum_{k=1}^{n}\mathrm{Var}_\pi f(X_k) + 2\cdot\sum_{1\le j<k\le n}\mathrm{Cov}\left[f(X_j),f(X_k)\right]\\ &= n\cdot\mathrm{Var}_\pi f(X_1) + 2\sum_{l=1}^{n-1}(n-l)\mathrm{Cov}\left[f(X_0),f(X_l)\right]\end{aligned}$$

Now we are interested in what happens for large n. So we take $\frac{\mathrm{Var}_p i(S_n(f))}{n}$ and get:

$$\frac{\mathrm{Var}_p i(S_n(f))}{n} = \mathrm{Var}_\pi f(X_1) + 2\sum_{l=1}^{n-1}\left(1-\frac{l}{n}\right)\mathrm{Cov}\left[f(X_0),f(X_l)\right]$$

Simplify any assume that $\pi f = 0 \iff \mathbb{E}_\pi f(X_k) = 0$ because $X_k \sim \pi$ under $\mathbb{P}_\pi$, the probability with the stationary measure. Then

$$\frac{\mathrm{Var}_p i(S_n(f))}{n} = \pi f^2 + 2\sum_{l=1}^{n-1}\left(1-\frac{l}{n}\right)\mathbb{E}_\pi f(X_0)f(X_l),$$

where we get rid of the subtraction terms by our assumption. Now we ask how to compute $\mathbb{E}_\pi\left[f(X_0)f(X_l)\right]$, where we must respect the joint distribution between X_0 and X_l, which involves P^l. Pitman notes that we need to use Markov chain properties and condition:

$$\begin{aligned}\mathbb{E}_\pi\left[f(X_0)f(X_l)\right] &= \mathbb{E}_\pi\left[f(X_0)\overbrace{\mathbb{E}_\pi f(X_l)|X_0)}^{=(P^l f)(x)}\right]\\ &= \mathbb{E}_\pi f(X_0)(P^l f)(X_0)\\ &= \mathbb{E}_\pi\left(f\cdot P^l f\right)(X_0)\\ &= \pi(f\cdot P^l f),\end{aligned}$$

where $(f\cdot g)(x) = f(x)g(x)$; in other words, don't perform the matrix multiplication (column vector times column vector).

3 Conclusion

We have derived the following exact formula

$$\frac{\mathrm{Var}_\pi\left[S_n(f)\right]}{n} = \pi f^2 + 2\sum_{l=1}^{n-1}\left(1-\frac{l}{n}\right)\pi\left(f\cdot P^l f\right).$$

Now we would like to see what happens when n is large (a la Central Limit Theorem). Let $n\to\infty$. Assume row P is aperiodic, so that $P^l\to\pi$, and hence $1-\frac{l}{n}\to 1$.

Then

$$
\begin{aligned}
\frac{\mathrm{Var}_\pi\left[S_n(f)\right]}{n} &\xrightarrow{n\to\infty} \pi f^2 + 2\sum_{l=1}^{\infty} \pi\left(f \cdot P^l f\right) \\
&= \pi f^2 + 2\pi f\left(\sum_{l=1}^{\infty} P^l f\right) \\
&= \pi f^2 + 2\pi f \sum_{l=1}^{\infty}\left(P^l - \pi\right) f \\
&= \pi f^2 + 2\pi f\,(Z - I)\,f \\
&= \boxed{2\pi f \cdot Z f - \pi f^2}.
\end{aligned}
$$

Recall that $Z := \sum_{l=0}^{\infty}(P - \Pi)^l$.

Theorem 3.1. (Central Limit Theorem for Markov Chains)
The CLT works with this evaluation of mean and variance.

Lecture ends here.