Stats 150, Fall 2019

Lecture 2, Tuesday, 9/3/2019

Pitman reminds us that Wikipedia serves as a valuable resource for clarifying definitions.
Recall from Lecture 1 we worked with a transition matrix P with columns y and row x. The xth row and yth column entry is $P(x, y)$. All entries are non-negative and row sums are 1.
For the first step in the Markov chain, we have:

$$P(x, y) = \mathbb{P}(X_1 = y \mid X_0 = x).$$

With many steps and homogeneous transition probabilities, we have:

$$P(x, y) = \mathbb{P}(X_{n+1} = y \mid X_n = x).$$

Pitman notes that the first problem on homework 1 is very instructional, which gets us to think about what exactly is the Markov property.

0.1 Action of a transition matrix on a row vector

Take an initial distribution $\lambda(x) = \mathbb{P}(X_0 = x)$. If we write $P(x, \cdot)$, we're taking the row of numbers in the matrix. With N states we can simply consider sequences of length N rather than N-dimensional space.
To ensure we really know what's going on here, consider 2 steps (indexed 0 and 1). What is the distribution of X_0? Trivially, it's λ. Now what is the distribution of X_1? We need to do a little more. We don't know how we started, and we want to think of all the ways we could have ended up at our final state X_1.
To do this, we use the **law of total probability**, which gives:

$$\boxed{\mathbb{P}(X_1 = y) = \sum_{x \in S} \mathbb{P}(X_0 = x, X_1 = y).}$$

Now it just takes a little bit of calculation to go forward. Conditioning on X_0 (turning a joint probablity into a marginal for the first and a conditional given the first) gives:

$$\begin{aligned}
\mathbb{P}(X_1 = y) &= \sum_{x \in S} \mathbb{P}(X_0 = x, X_1 = y) \\
&= \sum_{x \in S} \mathbb{P}(X_0 = x) \cdot \mathbb{P}(X_1 = y \mid X_0 = x) \\
\mathbb{P}(X_1 = y) &= \sum_{x \in S} \lambda(x) \underbrace{P(x, y)}_{\text{matrix}} \\
&= (\lambda P)(y) \text{ or equivalently, } \quad = (\lambda P)_y
\end{aligned}$$

To have this fit with our convention of matrix multiplication, we take $\lambda(x)$ to be a ROW VECTOR. Back in our picture going from one step to the next of a Markov chain, we use x (the nth state) to index the row of the matrix and y (the $n + 1$th state) to index the column of the matrix $P(x, y)$.

0.2 Conclusion

There is a happy coincidence between the rules of probability and the rules of matrices, which implies that if a Markov Chain has $X_0 \sim \lambda$ (meaning random variable X_0 has distribution λ), then at the next step we have the following distribution:

$$\boxed{X_1 \sim \lambda P},$$

where argument y is hidden. If we evaluate the row vector λP at entry y, we get:

$$(\lambda P)_y = \mathbb{P}(X_1 = y).$$

Although this may not be terribly exciting, Pitman notes this is fundamental and important to understand the connection between linear algebra and rules of matrices with probability. We will maintain and strengthen this connection throughout the course.

1 Action of a transition matrix on a column vector:

Suppose f is a function on S. Think of it as a **reward** in that if $X_1 = x$, then you get $\$f(x)$ (random monetary reward $f(X_1)$ where $X_1 \in S$ as an abstract object; these can be partitions or something very abstract). Pitman notes some applications to Google's PageRank with a Markov property and others. Without being scared about the potential size of the **state space**, we open to some abstraction in our immediate example.
Consider the Markov Chain from X_0 to X_1 and the conditional expectation:

$$\mathbb{E}\left(f(X_1) \mid X_0 = x\right) = \sum_y \underbrace{P(x,y)}_{\text{matrix}} \underbrace{f(y)}_{\text{col.vec.}}$$

where we could make some concrete financial definitions to apply our abstract problem if we wish.
Starting at state x, we move to the next state according to the row $P(x, \cdot)$. Recognize this as a matrix operation and we have, for the above:

$$\mathbb{E}\left(f(X_1) \mid X_0 = x\right) = (Pf)(x)$$

Remark: f can be signed (there is no difficulty if we are losing money as opposed to gaining); it is only difficult to interpret if λ is signed.

2 Two Steps

Now consider two steps:

$$X_0 \xrightarrow{P} X_1 \xrightarrow{Q} X_2.$$

Assume the Markov property. Now let's discuss the probability of X_2, knowing $X_0 = x$. That is,

$$\mathbb{P}(X_2 = z \mid X_0 = x),$$

where we have some mystery intermediate X_1. The row out of the matrix which we use for the intermediate is random.

We condition upon what we don't know in order to reach a solution. It should become instinctive to us soon to do such a thing: condition on X_1. This gives:

$$\begin{aligned}\mathbb{P}(X_2 = z \mid X_0 = x) &= \sum_y \mathbb{P}\left(X_1 = y, X_2 = z \mid X_0 = x\right) \\ &= \sum_y P(x,y)Q(y,z),\end{aligned}$$

where in the homogeneous case, $P = Q$; however, here we prefer the more clear notation as above. Pitman jokes that generations of mathematicians developed a surprisingly compact form for this, namely matrix multiplication. If P, Q are matrices, this is simply:

$$\sum_y P(x,y)Q(y,z) = PQ(x,z),$$

where we take the x, zth element of the resulting matrix PQ.

2.1 Review: Matrix Multiplication

Assuming P, Q, R are $S \times S$ matrices, where S is the label set of indices, then Pitman notes that indeed,

$$PQR := (PQ)R = P(QR),$$

via the associativity of matrix multiplication. This is true for all finite matrices. As a side comment, this is also true for infinite matrices, provided they are nonnegative ≥ 0 (of course, if we have signed things, then summing infinite arrays in different orders may cause issues). For our purposes, all our entries are nonnegative, so we have no issues.
Now, recall that typically, matrix multiplication is not commutative; that is,

$$PQ \neq QP.$$

However, one easy (and highly relevant) case:
If our chain has homogeneous transition probabilities: P, P, P, P, then Pitman may ask us what is the probability that $X_n = z$ if we knew $X_0 = x$, then we iterate what we found for 2 steps:

$$\mathbb{P}(X_n = z \mid X_0 = x) = \underbrace{PPP\cdots P}_{n \text{ times}}(x,z) =: \boxed{P^n(x,z)}.$$

Again, Pitman notes we have a very happy 'coinkidink' (coincidence): If we take an n-step transition matrix (TM) of a Markov chain (MC) with homogeneous probabilities P, this is equivalent to simply P^n, the nth power of matrix P. We can bash this out with computers, but Pitman notes there are techniques of diagonalizing and spectral theory to perform high powers of matrices (minimizing numerical error). Realize that every technique here has an **immediate application** to Markov chains (with very many steps).
Note the Chapman-Kolmogorov equations:

$$P^{m+n} = P^m P^n = P^n P^m,$$

which shows that powers of a single matrix do in fact commute. These equations are easily justified either by algebra, or by probabilistic reasoning. See text Section 1.2 for details of the probabilistic reasoning.

Break time.

3 Techniques for finding P^n for some P

Pitman wants to warn us that these ideas will be coming and eventually will be useful for this course. Especially, we consider matrix P related to sums of independent random variables. The most basic example is a **Random Walk** on $\mathbb{N}_0 := \{0, 1, 2, \dots\}$.
In this problem, one usually writes S_n for the state instead of X_n. Our basic X has $X_0, X_1, X_2, \dots$ i.i.d. according to some transition matrix P. This is truly a trivial MC. All rows of P are equivalent to some $p = (p_0, p_1, \dots)$. We consider:

$$S_n = X_0 + X_1 + \cdots + X_n = \text{ cummulated winnings in a gambling game}$$

(Pitman adds that we ignore costs or losses for convenience, so that natural state space of S_n is $\mathbb{N}_0$).

4 First Example:

Let $p \sim$ Bernoulli(p) where values $0, 1$ have probabilities q, p, respectively. Then $S_n := X_0 + X_1 + \cdots + X_n$.
This admits the following (infinite) matrix:

$$\begin{bmatrix} * & 0 & 1 & 2 & 3 & 4 & 5 & \cdots \\ 0 & q & p & 0 & 0 & 0 & 0 & \cdots \\ 1 & 0 & q & p & 0 & 0 & 0 & \cdots \\ 2 & 0 & 0 & q & p & 0 & 0 & \cdots \\ 3 & 0 & 0 & 0 & q & p & 0 & \cdots \\ 4 & 0 & 0 & 0 & 0 & q & p & \cdots \\ 5 & 0 & 0 & 0 & 0 & 0 & q & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

Because we can only win \$1 at a time, we fill in the first row trivially.
Pitman asks us now to write down a formula for P^n. As a hint, he says to start with the top row.

$$P^n(0, k) = \mathbb{P}(\underbrace{X_1 + \cdots + X_n}_{n \text{ iid Bernoulli(p)}} = k)$$

If this doesn't come quickly to us (the answer is trivial according to Pitman), then we should re-visit our 134 probability text (which for me happens to be by Pitman).
To find P^n, we note $n = 1$ is known, so taking $n = 2$ for a state space of X_0, X_1, X_2 gives the probabilities:

$$\begin{aligned} P^2(0,0) &= q^2 \\ P^2(0,2) &= p^2 \\ P^2(0,1) &= 2pq, \end{aligned}$$

and this is the familiar **binomial distribution**. Our formula is:

$$\begin{aligned} P^n(0, k) &= \mathbb{P}(\underbrace{X_1 + \cdots + X_n}_{n \text{ iid Bernoulli(p)}} = k) \\ &= \boxed{\binom{n}{k} p^k q^{n-k}}. \end{aligned}$$

Now being at an initial fortune i, we have:

$$P^n(i, k) = \binom{n}{k-i} p^{k-i} q^{n-(k-i)}.$$

4.1 More Challenging:

Now consider the same problem, same setup, but now with $X_1, X_2, \dots$ are i.i.d. with the distribution $(p_0, p_1, p_2, \dots)$ (perhaps all strictly positive) instead of $(q, p, 0, 0, 0, \dots)$. We are interested in the distribution of our Markov chain after n steps. Taking the same method, it's enough to discuss the distribution of $S_n = X_1 + \cdots + X_n$, because we just shift i to $S_0 = i$.
Our matrix is now:

$$\begin{bmatrix} * & 0 & 1 & 2 & 3 & 4 & 5 & \cdots \\ 0 & p_0 & p_1 & p_2 & \cdots & \cdots & \cdots & \cdots \\ 1 & 0 & p_0 & p_1 & p_2 & \cdots & \cdots & \cdots \\ 2 & 0 & 0 & p_0 & p_1 & p_2 & \cdots & \cdots \\ 3 & 0 & 0 & 0 & p_0 & p_1 & p_2 & \cdots \\ 4 & 0 & 0 & 0 & 0 & p_0 & p_1 & \cdots \\ 5 & 0 & 0 & 0 & 0 & 0 & p_0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

Again, to get closer to induction, we take $n = 1$ to $n = 2$ steps (with $S_0 = 0$). In matrix notation, we have:

$$\mathbb{P}_0(S_2 = k) = \sum_{j=0}^{k} P(0, j)P(j, k),$$

where we stop at k because we are only adding nonnegative variables. And in probability notation, where we start with j and need to get to k (so we move $k - j$) we have:

$$\mathbb{P}_0(S_2 = k) = \sum_{j=0}^{k} \mathbb{P}(X = j)\mathbb{P}(X = k - j),$$

and either way (of the above two), this ends up being equal to:

$$\mathbb{P}_0(S_2 = k) = \sum_{j=0}^{k} P_j P_{k-j}.$$

So in conclusion, we have found:

$$P^2(0, k) = \sum_{j=0}^{k} P_j P_{k-j}$$

We may want to know the name of this operation: **discrete convolution** (so that we know what to look up!). This gets us from a distribution of random variables to the distribution of their sum. There is a "brilliant idea" (as termed by Pitman):
Consider the power series (of the generating function) $G(z) := \sum_{n=0}^{\infty} p_n z^n$, where taking

$$\left(p_0 + p_1 z + p_2 z^2 + \cdots\right)\left(p_0 + p_1 z + p_2 z^2 + \cdots\right)$$

yields that $\sum_{j=0}^{k} P_j P_{k-j}$ is simply the coefficient of a particular term. Pitman gives us a slick notation:

$$\begin{aligned} P^2(0,k) &= \sum_{j=0}^{k} P_j P_{k-j} \\ &= \left[z^k\right] \underbrace{\left(\sum_{n=0}^{\infty} p_n z^n\right)^2}, \end{aligned}$$

which is just the coefficient of z^k in the underbraced expression.
Repeating this convolution, we move forward from $n = 2$:

$$P^n(0,k) = [z^k][G(z)]^n$$

Example: Pitman asks us to simulate via Wolfram Alpha dice rolls $(p_0, p_1, \dots) = \left(\underbrace{\frac{1}{6}, \frac{1}{6}, \dots, \frac{1}{6}}_{6}, 0, \dots\right)$ and want to find: $P^4(0,5)$ for dice rolls $= \mathbb{P}(S_4 = 5)$.
We have:

$$\begin{aligned} \left[\frac{1}{6}(z + z^2 + z^3 + z^4 + z^5 + z^6)\right]^4 = \frac{1}{6^4}(&z^{24} + 4z^{23} + 10z^{22} + 20z^{21} \\ &+ 35z^{20} + 56z^{19} + 80z^{18} \\ &+ 104z^{17} + 125z^{16} + 140z^{15} \\ &+ 146z^{14} + 140z^{13} + 125z^{12} \\ &+ 104z^{11} + 80z^{10} + 56z^9 + 35z^8 \\ &+ 20z^7 + 10z^6 + \underbrace{4z^5} + z^4) \end{aligned}$$

which implies

$$P^4(0,5) = \frac{4}{6^4},$$

where we took the coefficient of the underbraced term (power of 5). Pitman credits the inventor of this method, Laplace. This is unusually simple but demonstrates the general method.
Of course, $P^4(0,5) = \frac{4}{6^4}$ is rather trivial because we can count the number of dice patterns on one hand; however, the evaluations of $P^4(0,k)$ for $4 \le k \le 24$ above are not so trivial (outside of this method by Laplace). This method can be used to prove all the familiar properties of sums of independent discrete variables (e.g. sums of Poissons is Poisson). We should try it for this purpose.

Lecture ends here.