Stats 150, Fall 2019
Lecture 8, Tuesday, 9/24/2019

1 §1.11 Infinite State Space

This is starred in the text but is not optional for our course. We will discuss techniques for both finite and infinite state spaces, especially

- probability generating functions
- potential kernel (AKA) Green matrix

2 Review of Math Background

Know the following by heart (we'll need to use them on the midterm).

2.1 Binomial Theorem

This is to write:

$$(1+x)^n = \sum_{k=0}^{n} \binom{n}{k} x^k.$$

We should observe that $\binom{n}{k} = \frac{n!}{(n-k)!k!} = \boxed{\dfrac{n(n-1)\cdots(n-k+1)}{k(k-1)\cdots 1}}$, and it is important that the numerator is a polynomial in n. Pitman comments that no one realized why this is important until about 1670. The reason is that this form can be extended to other powers, namely $n := -1, \frac{1}{2}, \frac{-1}{2}$, or any real number $n \to r \in \mathbb{R}$.
Take r real and look at

$$(1+x)^r = \sum_{k=0}^{\infty} \binom{r}{k} x^k \qquad (|x| < 1)$$

which is valied for all real r and all real or complex x with $|x| < 1$. Notice that the combinatorial meaning of r 'choose' k makes sense only for $r = n$ a positive integer and k a non-negative integer. But the meaning of $\binom{r}{k}$ is extended to all real numbers r and all non-negative integers k by treating $\binom{n}{k}$ as a polynomial of degree k in n, then substituting r in place of n in this polynomial.
This is the instance with $f(x) = x^r$ of the Taylor expansion of a function f about the point 1:

$$f(1+x) = f(1) + f'(1)x + \frac{f''(1)}{2!}x^2 + \cdots = \sum_{k=0} f^{(k)}(1)k!x^k$$

which for suitable f is valid for $|x| < R$, where R is the radius of convergence. Usually for our purposes, $R \geq 1$. As another Taylor expansion (around 0 instead of 1)

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}.$$

We get exponentials arising as limit of binomial probabilities (e.g. the Poisson distribution). Also, recall that the geometric distribution converges to the exponential distribution with suitable scaling.

3 Probability Generating Functions

Suppose we have a non-negative integer-valued random variable $X \in \{0, 1, 2, \dots\}$.
We define the PGF (probability generating function) of X to be the function

$$\boxed{\phi_X(z) := \mathbb{E}z^X}$$

We usually take $0 \le z \le 1$. When discussing PGFs, we may push z to $|z| \le 1$, but in this course we will work entirely with PGFs defined as a function of an argument $z \in [0, 1]$. Then $\phi_X(z) \in [0, 1]$ too, and there are many contexts in which $\phi_X(z)$ acquires meaning as the probability of something. Now we can write the above as a power series. Recall that

$$\mathbb{E}g(X) = \sum_{n=0}^{\infty} \mathbb{P}(X = n)g(n),$$

so

$$\phi_X(z) := \mathbb{E}Z^X = \sum_{n=0}^{\infty} \mathbb{P}(X = n)z^n = \sum_{n=0}^{\infty} p_n z^n.$$

where $p_n := \mathbb{P}(X = n)$. We worked with PGFs very briefly in a previous lecture, for dice probabilities, namely taking X uniform on $\{1, 2, 3, 4, 5, 6\}$, and we looked at:

$$\phi_X(z) = \frac{1}{6}\left(z + \cdots + z^6\right).$$

Recall this is where Pitman asked us to look at powers of this expansion in Wolfram Alpha.
Notice that by convention, $0^0 = 1$, so $\phi_X(0) = \mathbb{P}(X = 0)$.
Now for any PGF, we have:

$$\begin{aligned}
\frac{d}{dx}\phi_X(z) &= \frac{d}{dz}\sum_n \mathbb{P}(X = n)z^n \\
&= \sum_n \mathbb{P}(X = n)\frac{d}{dz}z^n \\
&= \sum_n \mathbb{P}(X = n)nz^{n-1},
\end{aligned}$$

and so we see that

$$\mathbb{E}X = \frac{d}{dz}\phi_X(z)|_{z=1^-},$$

where we must approach $z = 1$ from the left if the radius of convergence R is exactly $R = 1$, but typically $R > 1$ and you can just evaluate the derivative at $z = 1$.
Perhaps we'd like to compute the variance. We ask, what happens if we differentiate twice?

$$\left(\frac{d}{dz}\right)^2 \phi_X(z) = \sum_{n=0}^{\infty} \mathbb{P}(X = n)n(n-1)z^{n-2}.$$

Again we'd like the z factor to go away, so we set $z := 1$ and we have:

$$\begin{aligned}
\mathbb{E}\left[X(X-1)\right] &= \sum_{n=0}^{\infty} \mathbb{P}(X = n)n(n-1) \\
&= \left(\frac{d}{dz}\right)^2 \phi_X(z)|_{z=1^-}
\end{aligned}$$

Recall that $X_\lambda \sim \text{Poisson}(\lambda)$ if and only if:

$$\mathbb{P}(X_\lambda = n) = \frac{e^{-\lambda}\lambda^n}{n!},$$

which via the generating function implies:

$$\phi_{X_\lambda}(z) = \sum_{n=0}^{\infty} \frac{e^{-\lambda}\lambda^n z^n}{n!} = e^{-\lambda}e^{\lambda z} = e^{\lambda(z-1)}.$$

Easily from the above analysis by d/dz, or otherwise,

$$\begin{aligned} \mathbb{E}X_\lambda &= \lambda \\ \text{Var}(X_\lambda) &= \lambda \end{aligned}$$

A (good) question arises whether $\phi_X(z)$ is a probability. The answer is yes, because after all the range of values is between 0 and 1, and any such function can be interpreted as a probability. Notably, we have:

$$\phi_X(z) = \mathbb{P}(X \leq G_{1-z}),$$

where G_p for $0 \leq p \leq 1$ denotes a random variable independent of X with the geometric (p) distribution on $\{0, 1, \dots\}$: for $n \geq 0$
Then

$$\mathbb{P}(G_p = n) = (1-p)^n p, \text{ and } \mathbb{P}(G_p \geq n) = (1-p)^n.$$

In summary, we can think of a probability generating function as a probability, and we only need that G_{1-z} is independent of X.
Now if X, Y are independent, then

$$\begin{aligned} \mathbb{E}z^{X+Y} &= \mathbb{E}\left[z^X z^Y\right] \\ &= \left[\mathbb{E}z^X\right]\left[\mathbb{E}z^Y\right] \\ &= \phi_X(z)\,\phi_Y(z). \end{aligned}$$

Hence the PGF of a sum of independent variables is the product of their PGFs.

Example: Let $G_p \sim \text{Geometric(p)}$ on $\{0, 1, 2, \dots\}$. Then

$$\mathbb{P}(G_p = n) = (1-p)^n p, \text{ for } n = 0, 1, 2, \dots$$

Now if we want to look at the probability generating function, we have:

$$\mathbb{E}(z^{G_p}) = \sum_{n=0}^{\infty} q^n p z^n = \frac{p}{1-qz},$$

for $p + q = 1$ and $|z| < 1$. Now we look at $T_r := G_1 + G_2 + \cdots + G_r$, where $r = 1, 2, 3, \dots$, and G_i are all independent geometrically distributed with the same parameter p.
The interpretation is to see G_p as the number of failures before the first success. That is, the number of 0s before the first 1 in independent Bernoulli(p) 0/1 trials. Then similarly,

$$T_r = T_{r,p} = \text{ number of 0s before } r\text{th 1 in indep. Bernoulli(p) 0/1 trials.}$$

Looking at iid copies of G_p we use generating functions:

$$\begin{aligned}
\mathbb{E}z^{T_r} &= \left(\frac{p}{1-qz}\right)^r = p^r(1-qz)^{-r} \\
&= p^r\left(1+(-qz)\right)^{-r} \\
&= \sum_{n=0}^{\infty}\binom{-r}{n}(-qz)^n, \\
&= p^r\sum_{n=0}^{\infty}\frac{(r)_{n\uparrow}}{n!}q^n z^n,
\end{aligned}$$

where we simply plug into Newton's binomial formula and we define for convenience

$$(r)_{n\uparrow} := r(r+1)\cdots(r+n-1)$$

which is equivalent to

$$\frac{(r)_{n!}}{n!} = \binom{r+n-1}{n}.$$

From 134, we know this to be the negative binomial distribution. The above formula can be derived directly by counting: $\binom{r+n-1}{n}$ is the number of ways to place the n failures in the first $r+n-1$ trials, and the last $n+r$th trial must be a 1. But the generating function technique used above is instructive, and can be applied to more difficult problems.

4 Probability Generating Functions and Random Sums

Suppose we have $Y_1, Y_2, \ldots$ iid nonnegative integer random variables, with probability generating function $\phi_Y(z) = \mathbb{E}z^{Y_k} = \sum_{n=0}^{\infty}\mathbb{P}(Y_k = n)z^n$ (the same generating function for all Y_i). Now consider another random variable, $X \geq 0$, integer valued, and look at: $Y_1 + Y_2 + \cdots + Y_X =: S_X$, the sum of X independent copies of Y. Then

$$\begin{aligned}
S_n &= Y_1 + \cdots + Y_n \\
S_X &= Y_1 + \cdots + Y_X.
\end{aligned}$$

Now if $X = 0$ with 0 copies of Y, then our convention is to set the empty sum to give 0.
We wish to find the PGF of S_X. The random index X is annoying, so try conditioning on it: Ooops. I had no idea what aggrgate meant. I have edited to here. Can you splice the above into the aggregate please? And when you want me to edit in future point to the file in email? I have to quit now for a few hours, so OK for you to edit. thanks! – JP
Let me know if any of the diagrams I made need adjusting. Sounds good! I'll get to splicing..

$$\begin{aligned}
\mathbb{E}z^{S_X} &= \sum_{n=0}^{\infty}\mathbb{P}(X=n)\mathbb{E}\left(z^{S_n}\right) \\
&= \sum_{n=0}^{\infty}\mathbb{P}(X=n)\left[\phi_Y(z)\right]^n \\
&= \phi_X\left[\phi_Y(z)\right],
\end{aligned}$$

which is a composition of generating functions. In the middle line, Pitman notices this is a generating function, just evaluated at a different location. Notice that for this to hold, we needed to assume that X is independent of $Y_1, Y_2, \dots$.

5 Application to Galton-Watson Branching Process

Assume that we're given some probability distribution (offspring distribution) $p_0, p_1, p_2, \dots$. Start with some fixed number k of individuals in generation 0, where each of these k individuals has offspring with distribution according to X. Our common notation is

$$Z_n := \# \text{ of individuals in generation } n,$$

and so we have the following equality in distribution:

$$(Z_1 \mid Z_0 = k) \stackrel{d}{=} X_1 + X_2 + \cdots + X_k,$$

where the X_i are iid $\sim p$.
Continuing the problem, given $Z_0, Z_1, \dots, Z_n$ with $Z_n = k$, and $Z_{n+1} \sim X_1 + \cdots + X_k$. It's intuitive to draw this as a tree, where individuals of generation 0 have some number of offspring and some have none. We create a branching tree from one stage to the next. Clearly, this is a Markov chain on $\{0, 1, 2, \dots\}$. Pitman notes that there is a conspicuous detail, in that $k = 0$ is absorbing, which happens to fit with the convention of empty sums (that summing 0 copies of nothing gives nothing).
Now, we should expect that generating functions should be helpful, as we are iterating random sums. We'll iterate the composition of generating functions. For simplicity, start with $z_0 = 1$. Let $\phi_n(s) = \mathbb{E}\left(s^{Z_n}\right)$ for $0 \le s \le 1$. We see that

$$Z_{n+1} = \text{ sum of } Z_n \text{ copies of } X.$$

Hence

$$\phi_1(s) = \sum_{n=0}^{\infty} p_n s^n = \mathbb{E} s^X,$$

which we define as the **offspring generating function**. To find ϕ_2, we look at $\phi_1(\phi_1(s))$. That is,

$$\begin{aligned} \phi_2(s) &= \text{ PGF of sum of } Z_1 \text{ copies of } X \\ &= \phi_1\left[\phi_1(s)\right]. \end{aligned}$$

Continuing, we similarly have:

$$\begin{aligned} \phi_3(s) &= \text{ PGF of sum of } Z_2 \text{ copies of } X \\ &= \phi_1(\phi_1(\phi_1(s))), \end{aligned}$$

and so on. Now Pitman presents the famous problem of finding the probability of extinction:

$$\begin{aligned} \mathbb{P}_1(\text{extinction}) &= \mathbb{P}_1(Z_n = 0 \text{ for large } n) \\ &= \lim_{n\to\infty} \mathbb{P}_1(Z_n = 0). \end{aligned}$$

Now we ask, how do we find $Z_n = 0$? We basically have a formula for this. What is the probability that $Z_1 = 0$? This is simply

$$\mathbb{P}_1(Z_1 = 0) = p_0.$$

Then

$$\mathbb{P}_1(Z_2 = 0) = \phi(\phi(0)) = \phi(p_0),$$

and similarly,

$$\mathbb{P}_1(Z_3 = 0) = \phi(\phi(\phi(0))) = \phi(\phi(p_0)),$$

and so on. Pitman gives that there is a very nice picture we can draw for intuition here. As an example, we sketch ($\phi(s)$ with respect to s) the generating function of Poisson(3/2). This gives a fixed point iteration returning the unique root s of $s = \phi(s)$ with $s < 1$. We draw the special case where the mean is large than 1 ($\phi'(1) = \mu > 1$).
We see that if $p_0 > 0$ and $\mu := \sum_n np_n > 1$, then the probability generating function is a convex curve with slope > 1 at 1 and value $p_0 > 0$ at 0. Then by elementary analysis, we have that there is a unique root $0 < s < 1$ of $\phi(s) = s$, and

$$\phi(\phi(\phi(\cdots(0)))) \xrightarrow{n\to\infty} \text{ the unique root }.$$

Now, even if we aren't a fan of generating functions, we should note that they are inescapable in the solution to the branching extinction problem.
(By the way, there is a very annoying case for branching processes that we should not forget, which is where $\mathbb{P}(X = 1) = 1$, just makes the population stay at 1, and extinction probability is 0. There is no random fluctuation in it.)
Pitman notes that there is another interesting case, where $\mu := 1$ and $p_0 > 0$ (to avoid the above boring function). Then by our generating function convex, the only root returned from fixed point iteration is precisely at 1. This implies that the probability of extinction $\mathbb{P}_1(\text{extinction}) = 1$. The nonzero random probability eventually leads to extinction in this case, which according to Pitman is nonobvious. The book presents this conclusion in different ways.

Lecture ends here.