Stats 150, Fall 2019
Lecture 4, Tuesday, 10/9/2019

CLASS ANNOUNCEMENTS: Pitman announces he added a Mathematica notebook on bCourses to illustrate some examples.

Topics Today:

- Symmetries in Distributions
  - Exchangeability
  - Reversibility
  - Stationarity

1 Sampling Without Replacement (SWOR)

To do this, suppose we have a population of tickets in a box, say 3 tickets labeled 1, and 7 tickets labeled 0. These tickets are all shuffled within the box, and we pull them out via 'random draws' in a sequence:

$$(X_1, X_2, \ldots, X_{10}) = (100|0110000).$$

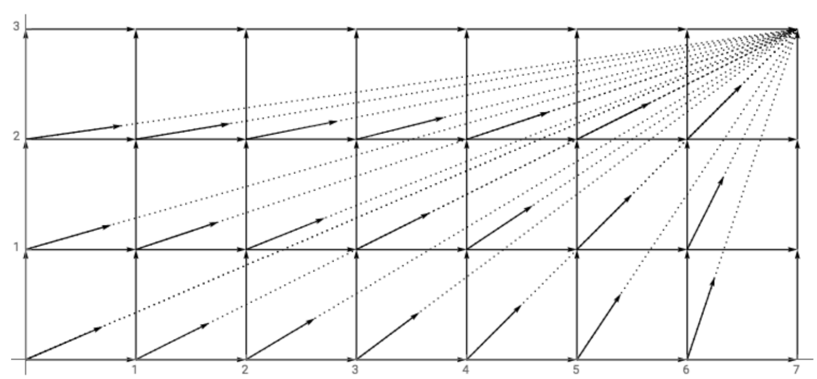
Now, writing the sequence like this helps us think about conditioning, as per usual combinatorial counting. We can make a Markov chain out of this. Let S_n be the number of successes (1) in the first n draws. Let F_n be the number of fails (0) in the first n draws. By convention, we write the fails first:

$$W_n = (F_n, S_n),$$

and note obviously $F_n + S_n = n$. Now if we say this is a chain, then very easily,

$$(W_n, 0 \le n \le N)$$

is a Markov chain. For example, we could have a pair in the state space like $(B, A) = (7, 3)$.



We have a nice grid diagram starting at (0,0) and ending at (B, A). Now, this gives $\binom{10}{3} = \binom{10}{7}$ such paths via basic counting.

1.1 Transition Probabilities

Pitman wants us to think about vectors here. Consider the 2-state probability vector, denoted by (q, p) (or similarly denoted as (λ_0, λ_1) or (π_0, π_1)), where q, p are probabilities of the assignment (of 0 and 1, respectively). Consider the probability 'vectors' from the origin (0,0) to some point on the graph as a way to think about the transitional probability.
From each state, we draw the line connecting the current state with our final target, as can be seen in the earlier diagram.
Now, we look at (X_1, X_2) for $X_1, X_2, \ldots, X_N$ on sampling without replacement (SWOR). Consider, in terms of the probability of getting 1 in the first draw is:

$$\begin{aligned}\mathbb{P}(X_1 = 1, X_2 = 1) &= \frac{A}{A+B} \cdot \frac{A-1}{A+B-1} \\ \mathbb{P}(X_1 = 0, X_2 = 0) &= \frac{B}{A+B} \cdot \frac{B-1}{A+B-1} \\ \mathbb{P}(X_1 = 1, X_2 = 0) = \mathbb{P}(X_1 = 0, X_2 = 1) &= \frac{BA}{(A+B)(A+B-1)} \\ &= \frac{AB}{(A+B)(A+B-1)},\end{aligned}$$

where the left factor is the π part, and the right factor is the P part.

Definition: Reversible -

We say that the pair (X_1, X_2) 'reversible' if (and only if)

$$(X_1, X_2) \stackrel{d}{=} (X_2, X_1)$$

; that is, they are equal in distribution.

Remark: If we have X has the same distribution as Y, then $\psi(X) \stackrel{d}{=} \psi(Y)$. To see this, consider:

$$\begin{aligned}\mathbb{P}(\psi(X) \leq 3) &= \mathbb{P}(\{x : \psi(x) \leq 3\}) \\ &= \mathbb{P}(X \text{ has property } \psi(X) \leq 3) \\ &= \mathbb{P}(Y \text{ has property } \psi(Y) \leq 3).\end{aligned}$$

That is, equality in distribution pushes forward through functions.

Now, consider the projection function $(X_1, X_2) \stackrel{p}{\mapsto} X_1$ and $(X_2, X_1) \stackrel{p}{\mapsto} X_2$ (that is, take the first element). Now because $(X_1, X_2) \stackrel{d}{=} (X_2, X_1)$, via this projection mapping, we conclude $X_1 \stackrel{d}{=} X_2$.
Hence we proved that as we are along the 'line of symmetry', the two marginal probabilities are the same.

2 Reversibility

Take a distribution $X_1 \sim \pi$, and take a conditional distribution $X_2 | X_1 \sim P(X_1, \cdot)$, which is the blackboard shorthand for:

$$\begin{aligned}\mathbb{P}(X_1 = x) &= \pi(x) \\ \mathbb{P}(X_2 = y \mid X_1 = x) &= P(x, y),\end{aligned}$$

where we take the element of row x and column y of transition matrix P. Now this shows that the joint distribution of X_1, X_2 is

$$\mathbb{P}(X_1 = x, X_2 = y) = \pi(x)P(x, y)$$
$$\mathbb{P}(X_2 = x, X_1 = y) = \pi(y)P(y, x),$$

because intersection is a commutative operation. Notice:

$$(X_1, X_2) \stackrel{d}{=} (X_2, X_1) \iff \pi(x)P(x, y) = \pi(y)P(y, x), \forall_{x,y\in S},$$

where obviously these must have the same space of values. We say that X_1, X_2 are **reversible** (or exchangeable).

Remark: Consider the space of 0s and 1s. From the beginning of class, take (X_1, X_2) to be a sample of size 2 **without replacement** (the case with replacement is trivially true) is in fact, reversible.
To show that the pair of indicators is exchangeable, we need only show that the off-diagonals are equal.

Now, for a typical example of a two-step transition, first look at the matrix P for sampling without replacement. It has some form like:

$$\frac{1}{A+B-1}\begin{bmatrix} * & 0 & 1 \\ 0 & B-1 & A \\ 1 & B & A-1 \end{bmatrix}.$$

Now to claim this is a transition matrix, we need to check that all entries are nonnegative, so take integers $A, B \geq 1$. Then we check that the row sums are equal to 1. These can be seen easily via inspection. Now for sampling without replacement for (B, A) for 0 and 1.

2.1 Generalizing

Pitman notes that a mathematician may wish to generalize this past the positive integers. That is, what pairs of reals (B, A) is this P a (valid) transition matrix? Well, of course we want $A + B - 1 \neq 0$ to avoid division by zero. Then, notice that the row sum requirement (identically equal to 1) is completely independent of our choices for A, B. Finally, we require that all entries are nonnegative. Considering a (B, A)-space, take $B \geq 1$ and $A \geq 1$, and notice that this will satisfy this nonnegativity requirement. Moreover, notice that if we take $A \leq 0$ and $B \leq 0$, then we have division of a nonpositive number by a negative denominator.
We claim that every 2×2 transition matrix $P = \begin{bmatrix} P_{00} & P_{01} \\ P_{10} & P_{11} \end{bmatrix}$ is of the form:

$$P = \frac{1}{A+B-1}\begin{bmatrix} B-1 & A \\ B & A-1 \end{bmatrix}$$

for a unique pair of (B, A). Of course, to prove this, we need to play with a system of equations and unknowns. Pitman recalls the result is essentially:

$$\begin{bmatrix} B-1 & A \\ B & A-1 \end{bmatrix} = \frac{1}{P_{01} - P_{10}}\begin{bmatrix} P_{00} & P_{01} \\ P_{10} & P_{11} \end{bmatrix},$$

so let $B := \frac{P_{10}}{P_{01}-P_{11}}$, so that

$$B - 1 = \frac{P_{10} - P_{01} + P_{11}}{P_{01} - P_{11}} = \frac{1 - P_{01}}{P_{01} - P_{11}} = \frac{P_{00}}{P_{01} - P_{11}}.$$

Now it only remains to check our result (more or less following from properties of a transition matrix). Generally, what we showed is that a two-state Markov chain is just like a pair of indicators and sampling without replacement, where we changed only a few things.
Notice that the stationary distribution π is

$$(\pi_0, \pi_1) = \left(\frac{B}{A+B}, \frac{A}{A+B} \right),$$

which in general gives that the stationary π for P: $\pi P = \pi$ is:

$$(\pi_0, \pi_1) = \frac{(P_{10}, P_{01})}{P_{10} + P_{01}},$$

which is a generalization of our A, B formula earlier for sampling without replacement.

Definition: Stationary Distribution -

We say that π is a stationary distribution for P if and only if $\pi P = \pi$, which is precisely when we have the following 'balance equation':

$$X \sim \pi \text{ and } X_2 \mid X_1 \sim P(X, \cdot) \implies X_2 \sim \pi.$$

We call π REVERSIBLE if additionally $(X_1, X_2) \stackrel{d}{=} (X_2, X_1)$, which we've showed is equivalent to the property:

$$\pi(x)P(x,y) = \pi(y)P(y,x), \forall_{x \neq y}.$$

We call this the 'detailed balance' equations.

2.2 Does a solution π of the Detailed Balance Equation imply the solution π of the Balance Equation?

If we have found a solution π of the DBE, this means that

$$(X_1, X_2) \stackrel{d}{=} (X_2, X_1).$$

Now, having the solution π for BE is that:

$$X_1 \stackrel{d}{=} X_2.$$

It is completely trivial that if we can flip the pair across the diagonal, then the two entries must be equal.

3 Checking by Algebra

We know that $\pi(x)P(x,y) = \pi(y)\pi(y,x)$. We want to show

$$\sum_x \underbrace{\pi(x)P(x,y)}_{\pi(y)P(y,x)} = \pi(y).$$

Now within the summation, y is fixed, and so this equation is true for all states y.

4 Some Easy Facts:

(1) If π is stationary for P, then π is stationary for P^n. We can show this easily for $n = 2$:

$$\pi P^2 = \pi P P = \pi P = \pi.$$

Mathematical (strong) induction delivers the result similarly.
(2) If π is a reversible equilibrium, then for a Markov chain with initial distribution π and transition matrix p (π, P), then we have the same probability distribution. That is,

$$(X_N, X_{N-1}, X_{N-2}, \dots, X_0) \stackrel{d}{=} (X_0, X_1, \dots, X_N).$$

(3) Not all equilibriums are reversible. Consider states rotating around a circle, so that the transition matrix is

$$P = \begin{bmatrix} * & 0 & 1 & 2 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 2 & 1 & 0 & 0 \end{bmatrix}.$$

Notice that column sums of P also equal 1, so the transpose of P, P^{-1}, also has row sums 1. We say that P is a 'doubly-stochastic' matrix. This shows that if π is constant, then $\pi_i = \frac{1}{N}$, where N is the number of stationary vectors. We may ask: is this reversible?
We answer that clearly not, where the plot is not symmetric across the diagonal.
Pitman concludes the lecture with some interesting and much less trivial facts, consequences of the above definitions. These are accepted as known and quotable for homeworks.
In general, Pitman allows us to cite anything in the text.

5 Key Theorems

Recall that we've talked about irreducible states. If P has a finite number of states and P is irreducible (that is, $\forall_{i,j} \exists_n : P^n(i,j) > 0.$).
Then $\exists!$ stationary probability of $\pi : \pi P = \pi$. A formula for this is:

$$\pi_i = \frac{1}{\mathbb{E}_i T_i} > 0,$$

where $\mathbb{E}_i T_i$ is the mean return time. The same is true for any irreducible positive recurrent P, then $\mathbb{E}_i T_i < \infty$, for all i. Then we say that

$$\pi_i := \frac{1}{\mathbb{E}_i T_i}$$

is the unique stationary probability distribution.
Now define $d(x) :=$ gcd of $n : P^n(x,x) > 0$ (where intuitively we are moving around a circle). If P is irreducible, then $d(x) \equiv d$ for all x, and if it is **aperiodic** (namely $d = 1$), then $P^n(i,j) \to \pi$ as $n \to \infty$. That is, when we take higher powers, we approach equilibrium with:

$$P^n \to \begin{pmatrix} \pi \\ \pi \\ \pi \\ \vdots \end{pmatrix}$$

Pitman urges us to check the claims in the lecture notes on our own as an exercise.

Lecture ends here.