STAT 380 Lecture 8

Reading for Today's Lecture: Chapter 4 sections 1-3.
Goals of Today's Lecture:

- Define Markov Chain.
- Introduce transition matrix
- Derive Chapman Kolmogorov equations.

Look at the matrix product $\mathbf{PP}$:

$$\begin{bmatrix} \frac{3}{5} & \frac{2}{5} \\ \frac{1}{5} & \frac{4}{5} \end{bmatrix} \begin{bmatrix} \frac{3}{5} & \frac{2}{5} \\ \frac{1}{5} & \frac{4}{5} \end{bmatrix}$$

Notice that the probability $P(X_2 = W|X_0 = W)$ is exactly the formula for the W, W entry in $\mathbf{PP}$.

General case. Define

$$P_{i,j}^{(n)} = P(X_n = j|X_0 = i)$$

Then

$$\begin{aligned} P(X_{m+n} = j|X_m = i, X_{m-1} = i_{m-1}, \ldots) \\ &= P(X_{m+n} = j|X_m = i) \\ &= P(X_n = j|X_0 = i) \\ &= P_{i,j}^{(n)} \end{aligned}$$

Proof of these assertions by induction on m, n.
Example for $n = 2$. Two bits to do:
First suppose U, V, X, Y are discrete variables. Assume

$$\begin{aligned} P(Y = y|X = x, &U = u, V = v) \\ &= P(Y = y|X = x) \end{aligned}$$

for any x, y, u, v. Then I claim

$$\begin{aligned} P(Y = y|X = x, U = u) \\ &= P(Y = y|X = x) \end{aligned}$$

In words, if knowing both U and V doesn't change the conditional probability then knowing U alone doesn't change the conditional probability.

Proof of claim:

$$A = \{X = x, U = u\}$$

Then

$$\begin{aligned}
P(Y = y|X = x, U = u) & \\
&= \frac{P(Y = y, A)}{P(A)} \\
&= \frac{\sum_v P(Y = y, A, V = v)}{P(A)} \\
&= \frac{\sum_v P(Y = y|A, V = v)P(A, V = v)}{P(A)} \\
&= \frac{\sum_v P(Y = y|X = x)P(A, V = v)}{P(A)} \\
&= \frac{P(Y = y|X = x)\sum_v P(A, V = v)}{P(A)} \\
&= \frac{P(Y = y|X = x)P(A)}{P(A)} \\
&= P(Y = y|X = x)
\end{aligned}$$

Second step: consider

$$\begin{aligned}
P(X_{n+2} = k|X_n = i) & \\
&= \sum_j P(X_{n+2} = k, X_{n+1} = j|X_n = i) \\
&= \sum_j P(X_{n+2} = k|X_{n+1} = j, X_n = i) \\
&\qquad \times P(X_{n+1} = j|X_n = i) \\
&= \sum_j P(X_{n+2} = k|X_{n+1} = j) \\
&\qquad \times P(X_{n+1} = j|X_n = i) \\
&= \sum_j \mathbf{P}_{i,j}\mathbf{P}_{j,k}
\end{aligned}$$

This shows that

$$P(X_{n+2} = k|X_n = i) = (\mathbf{P}^2)_{i,k}$$

where $\mathbf{P}^2$ means the matrix product $\mathbf{PP}$. Notice both that the quantity does not depend on n and that we can compute it by taking a power of $\mathbf{P}$.

More general version

$$P(X_{n+m} = k | X_n = j) = (\mathbf{P}^m)_{j,k}$$

Since $\mathbf{P}^n\mathbf{P}^m = \mathbf{P}^{n+m}$ we get the Chapman-Kolmogorov equations:

$$P(X_{n+m} = k | X_0 = i) = \sum_j P(X_{n+m} = k | X_n = j) P(X_n = j | X_0 = i)$$

Summary: A Markov Chain has *stationary* n step transition probabilities which are the nth power of the 1 step transition probabilities.

Here is Maple output for the 1,2,4,8 and 16 step transition matrices for our rainfall example:

```
> p:= matrix(2,2,[[3/5,2/5],[1/5,4/5]]);
                          [3/5    2/5]
                  p  := [           ]
                          [1/5    4/5]
> p2:=evalm(p*p):
> p4:=evalm(p2*p2):
> p8:=evalm(p4*p4):
> p16:=evalm(p8*p8):
```

This computes the powers (`evalm` understands matrix algebra).

Fact:

$$\lim_{n\to\infty} \mathbf{P}^n = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} \\ \\ \frac{1}{3} & \frac{2}{3} \end{bmatrix}.$$

```
> evalf(evalm(p));
            [.6000000000    .4000000000]
            [                          ]
            [.2000000000    .8000000000]
> evalf(evalm(p2));
            [.4400000000    .5600000000]
            [                          ]
            [.2800000000    .7200000000]
```

```
> evalf(evalm(p4));
            [.3504000000    .6496000000]
            [                          ]
            [.3248000000    .6752000000]
> evalf(evalm(p8));
            [.3337702400    .6662297600]
            [                          ]
            [.3331148800    .6668851200]
> evalf(evalm(p16));
            [.3333336197    .6666663803]
            [                          ]
            [.3333331902    .6666668098]
```

Where did 1/3 and 2/3 come from?

Suppose we toss a coin $P(H) = \alpha_D$ and start the chain with Dry if we get heads and Wet if we get tails.

Then

$$P(X_0 = x) = \begin{cases} \alpha_D & x = \text{Dry} \\ \alpha_W = 1 - \alpha_D & x = \text{Wet} \end{cases}$$

and

$$\begin{aligned} P(X_1 = x) &= \sum_y P(X_1 = x | X_0 = y) P(X_0 = y) \\ &= \sum_y \alpha_y P_{y,x} \, . \end{aligned}$$

Notice last line is a matrix multiplication of row vector α by matrix $\mathbf{P}$.

A special α: if we put $\alpha_D = 1/3$ and $\alpha_W = 2/3$ then

$$\begin{bmatrix} \frac{1}{3} & \frac{2}{3} \end{bmatrix} \begin{bmatrix} \frac{3}{5} & \frac{2}{5} \\ \\ \frac{1}{5} & \frac{4}{5} \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} \end{bmatrix} .$$

In other words if we start off $P(X_0 = D) = 1/3$ then $P(X_1 = D) = 1/3$ and analogously for W. This means that X_0 and X_1 have the same distribution.

A probability vector α is called an initial distribution for the chain if

$$P(X_0 = i) = \alpha_i \, .$$

A Markov Chain is **stationary** if

$$P(X_1 = i) = P(X_0 = i)$$

for all i.

An initial distribution is called **stationary** if the chain is stationary. We find that α is a stationary initial distribution if

$$\alpha\mathbf{P} = \alpha\,.$$

Suppose $\mathbf{P}^n$ converges to some matrix $\mathbf{P}^\infty$. Notice that

$$\lim_{n\to\infty} \mathbf{P}^{n-1} = \mathbf{P}^\infty$$

and

$$\begin{aligned}\mathbf{P}^\infty &= \lim \mathbf{P}^n \\ &= \left[\lim \mathbf{P}^{n-1}\right]\mathbf{P} \\ &= \mathbf{P}^\infty\mathbf{P}\,.\end{aligned}$$

This proves that each row α of $\mathbf{P}^\infty$ satisfies

$$\alpha = \alpha\mathbf{P}\,.$$

Def'n: A row vector x is a left eigenvector of A with eigenvalue λ if

$$xA = \lambda x\,.$$

So each row of $\mathbf{P}^\infty$ is a left eigenvector of $\mathbf{P}$ with eigenvalue 1.