STAT 870 Lecture 11

```
> 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

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

So: if $P(X_0 = D) = 1/3$ 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 the 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

Finding stationary initial distributions. Consider $\mathbf{P}$ above. The equation

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

is really

$$\begin{aligned} \alpha_D &= 3\alpha_D/5 + \alpha_W/5 \\ \alpha_W &= 2\alpha_D/5 + 4\alpha_W/5 \end{aligned}$$

The first can be rearranged to

$$\alpha_W = 2\alpha_D$$

and so can the second. If α is to be a probability vector then

$$\alpha_W + \alpha_D = 1$$

so we get

$$1 - \alpha_D = 2\alpha_D$$

leading to

$$\alpha_D = 1/3$$

Some more examples:

$$\mathbf{P} = \begin{bmatrix} 0 & 1/3 & 0 & 2/3 \\ 1/3 & 0 & 2/3 & 0 \\ 0 & 2/3 & 0 & 1/3 \\ 2/3 & 0 & 1/3 & 0 \end{bmatrix}$$

Set $\alpha\mathbf{P} = \alpha$ and get

$$\begin{aligned} \alpha_1 &= \alpha_2/3 + 2\alpha_4/3 \\ \alpha_2 &= \alpha_1/3 + 2\alpha_3/3 \\ \alpha_3 &= 2\alpha_2/3 + \alpha_4/3 \\ \alpha_4 &= 2\alpha_1/3 + \alpha_3/3 \\ 1 &= \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 \end{aligned}$$

First plus third gives

$$\alpha_1 + \alpha_3 = \alpha_2 + \alpha_4$$

so both sums 1/2. Continue algebra to get

$$(1/4, 1/4, 1/4, 1/4).$$

```
p:=matrix([[0,1/3,0,2/3],[1/3,0,2/3,0],
           [0,2/3,0,1/3],[2/3,0,1/3,0]]);

                [ 0      1/3      0      2/3]
                [                           ]
                [1/3      0      2/3      0 ]
           p := [                           ]
                [ 0      2/3      0      1/3]
                [                           ]
```

```
               [2/3     0     1/3     0 ]
> p2:=evalm(p*p);
              [5/9     0     4/9     0 ]
              [                        ]
              [ 0     5/9     0     4/9]
        p2:=  [                        ]
              [4/9     0     5/9     0 ]
              [                        ]
              [ 0     4/9     0     5/9]
> p4:=evalm(p2*p2):
> p8:=evalm(p4*p4):
> p16:=evalm(p8*p8):
> p17:=evalm(p8*p8*p):
> evalf(evalm(p16));
    [.5000000116 , 0 , .4999999884 , 0]
    [                                 ]
    [0 , .5000000116 , 0 , .4999999884]
    [                                 ]
    [.4999999884 , 0 , .5000000116 , 0]
    [                                 ]
    [0 , .4999999884 , 0 , .5000000116]
> evalf(evalm(p17));
    [0 , .4999999961 , 0 , .5000000039]
    [                                 ]
    [.4999999961 , 0 , .5000000039 , 0]
    [                                 ]
    [0 , .5000000039 , 0 , .4999999961]
    [                                 ]
    [.5000000039 , 0 , .4999999961 , 0]
> evalf(evalm((p16+p17)/2));
  [.2500, .2500, .2500, .2500]
  [                          ]
  [.2500, .2500, .2500, .2500]
  [                          ]
  [.2500, .2500, .2500, .2500]
  [                          ]
  [.2500, .2500, .2500, .2500]
```

$\mathbf{P}^n$ doesn't converges but$(\mathbf{P}^n + \mathbf{P}^{n+1})/2$ does. Next example:

$$p = \begin{bmatrix} \frac{2}{5} & \frac{3}{5} & 0 & 0 \\ \frac{1}{5} & \frac{4}{5} & 0 & 0 \\ 0 & 0 & \frac{2}{5} & \frac{3}{5} \\ 0 & 0 & \frac{1}{5} & \frac{4}{5} \end{bmatrix}$$

Solve $\alpha\mathbf{P} = \alpha$:

$$\begin{aligned} \alpha_1 &= \frac{2}{5}\alpha_1 + \frac{1}{5}\alpha_2 \\ \alpha_2 &= \frac{3}{5}\alpha_1 + \frac{4}{5}\alpha_2 \\ \alpha_3 &= \frac{2}{5}\alpha_3 + \frac{1}{5}\alpha_4 \\ \alpha_4 &= \frac{3}{5}\alpha_3 + \frac{4}{5}\alpha_4 \\ 1 &= \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 \end{aligned}$$

Second and fourth equations redundant. Get

$$\begin{aligned} \alpha_2 &= 3\alpha_1 \\ 3\alpha_3 &= \alpha_4 \\ 1 &= 4\alpha_1 + 4\alpha_3 \end{aligned}$$

Pick α_1 in $[0, 1/4]$; put $\alpha_3 = 1/4 - \alpha_1$.

$$\alpha = (\alpha_1, 3\alpha_1, 1/4 - \alpha_1, 3(1/4 - \alpha_1))$$

solves $\alpha\mathbf{P} = \alpha$. So solution is not unique.

```
> p:=matrix([[2/5,3/5,0,0],[1/5,4/5,0,0],
            [0,0,2/5,3/5],[0,0,1/5,4/5]]);

                 [2/5    3/5     0       0 ]
                 [                         ]
                 [1/5    4/5     0       0 ]
            p := [                         ]
                 [ 0      0     2/5    3/5]
                 [                         ]
```

```
                [ 0        0       1/5     4/5]
> p2:=evalm(p*p):
> p4:=evalm(p2*p2):
> p8:=evalm(p4*p4):

> evalf(evalm(p8*p8));
        [.2500000000 , .7500000000 , 0 , 0]
        [                                 ]
        [.2500000000 , .7500000000 , 0 , 0]
        [                                 ]
        [0 , 0 , .2500000000 , .7500000000]
        [                                 ]
        [0 , 0 , .2500000000 , .7500000000]
```

Notice that rows converge but to two different vectors:

$$\alpha^{(1)} = (1/4, 3/4, 0, 0)$$

and

$$\alpha^{(2)} = (0, 0, 1/4, 3/4)$$

Solutions of $\alpha\mathbf{P} = \alpha$ revisited? Check that

$$\alpha^{(1)}\mathbf{P} = \alpha^{(1)}$$

and

$$\alpha^{(2)}\mathbf{P} = \alpha^{(2)}$$

If $\alpha = \lambda\alpha^{(1)} + (1-\lambda)\alpha^{(2)}$ $(0 \le \lambda \le 1)$ then

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

so again solution is not unique. Last example:

```
> p:=matrix([[2/5,3/5,0],[1/5,4/5,0],
             [1/2,0,1/2]]);

                 [2/5    3/5     0 ]
                 [                 ]
            p := [1/5    4/5     0 ]
                 [                 ]
```

```
                  [1/2      0      1/2]
> p2:=evalm(p*p):
> p4:=evalm(p2*p2):
> p8:=evalm(p4*p4):
> evalf(evalm(p8*p8));
  [.2500000000 .7500000000        0       ]
  [                                       ]
  [.2500000000 .7500000000        0       ]
  [                                       ]
  [.2500152588 .7499694824 .00001525878906]
```

Interpretation of examples

- For some $\mathbf{P}$ all rows converge to some α. In this case this α is a stationary initial distribution.
- For some $\mathbf{P}$ the locations of zeros flip flop. $\mathbf{P}^n$ does not converge. Observation: average

$$\frac{\mathbf{P}+\mathbf{P}^2+\cdots+\mathbf{P}^n}{n}$$

  does converge.
- For some $\mathbf{P}$ some rows converge to one α and some to another. In this case the solution of $\alpha\mathbf{P}=\alpha$ is not unique.

Basic distinguishing features: pattern of 0s in matrix $\mathbf{P}$.

The ergodic theorem

Consider a finite state space chain. If x is a vector then the ith entry in $\mathbf{P}x$ is

$$\sum_j \mathbf{P}_{ij}x_j$$

Rows of $\mathbf{P}$ probability vectors, so a weighted average of the entries in x.

If the weights are strictly between 0 and 1 and the largest and smallest entries in x are not the same then $\sum_j \mathbf{P}_{ij}x_j$ is strictly between the largest and smallest entries in x. In fact

$$\sum_j \mathbf{P}_{ij}x_j - \min(x_k) \ge \min_j\{p_{ij}\}(\max\{x_k\}-\min\{x_k\})$$

and

$$\max\{x_j\} - \sum_j \mathbf{P}_{ij} x_j \geq \min_j\{p_{ij}\}(\max\{x_k\} - \min\{x_k\})$$

Now multiply $\mathbf{P}^r$ by $\mathbf{P}^m$.

ijth entry in $\mathbf{P}^{r+m}$ is a weighted average of the jth column of $\mathbf{P}^m$.

So, if all the entries in row i of $\mathbf{P}^r$ are positive and the jth column of $\mathbf{P}^m$ is not constant, the ith entry in the jth column of $\mathbf{P}^{r+m}$ must be strictly between the minimum and maximum entries of the jth column of $\mathbf{P}^m$.

In fact, fix a j.

$\overline{x}_m$ = maximum entry in column j of $\mathbf{P}^m$

$\underline{x}_m$ the minimum entry.

Suppose all entries of $\mathbf{P}^r$ are positive.

Let $\delta > 0$ be the smallest entry in $\mathbf{P}^r$. Our argument above shows that

$$\overline{x}_{m+r} \leq \overline{x}_m - \delta(\overline{x}_m - \underline{x}_m)$$

and

$$\underline{x}_{m+r} \geq \underline{x}_m + \delta(\overline{x}_m - \underline{x}_m)$$

Putting these together gives

$$(\overline{x}_{m+r} - \underline{x}_{m+r}) \leq (1 - 2\delta)(\overline{x}_m - \underline{x}_m)$$

In summary the column maximum decreases, the column minimum increases and the gap between the two decreases exponentially along the sequence $m, m+r, m+2r, \ldots$.

This idea can be used to prove

Proposition 1 *Suppose X_n finite state space Markov Chain with stationary transition matrix $\mathbf{P}$. Assume that there is a power r such that all entries in $\mathbf{P}^r$ are positive. Then for $\mathbf{P}^k$ has all entries positive for all $k \geq r$ and $\mathbf{P}^n$ converges, as $n \to \infty$ to a matrix $\mathbf{P}^\infty$. Moreover,*

$$(\mathbf{P}^\infty)_{ij} = \pi_j$$

where π is the unique row vector satisfying

$$\pi = \pi \mathbf{P}$$

whose entries sum to 1.

Proof: First for $k > r$

$$(\mathbf{P}^k)_{ij} = \sum_k (\mathbf{P}^{k-r})_{ik} (\mathbf{P}^r)_{kj}$$

For each i there is a k for which $(\mathbf{P}^{k-r})_{ik} > 0$ and since $(\mathbf{P}^r)_{kj} > 0$ we see $(\mathbf{P}^k)_{ij} > 0$. The argument before the proposition shows that

$$\lim_{j \to \infty} \mathbf{P}^{m+jk}$$

exists for each m and $k \ge r$. This proves $\mathbf{P}^n$ has a limit which we call $\mathbf{P}^\infty$. Since $\mathbf{P}^{n-1}$ also converges to $\mathbf{P}^\infty$ we find

$$\mathbf{P}^\infty = \mathbf{P}^\infty \mathbf{P}$$

Hence each row of $\mathbf{P}^\infty$ is a solution of $x\mathbf{P} = x$. The argument before the statement of the proposition shows all rows of $\mathbf{P}^\infty$ are equal. Let π be this common row.

Now if α is any vector whose entries sum to 1 then $\alpha \mathbf{P}^n$ converges to

$$\alpha \mathbf{P}^\infty = \pi$$

If α is any solution of $x = x\mathbf{P}$ we have by induction $\alpha \mathbf{P}^n = \alpha$ so $\alpha \mathbf{P}^\infty = \alpha$ so $\alpha = \pi$. That is exactly one vector whose entries sum to 1 satisfies $x = x\mathbf{P}$. •

Note conditions:

There is an r for which all entries in $\mathbf{P}^r$ are positive.

The chain has a finite state space.

Consider finite state space case: $\mathbf{P}^n$ need not have limit. Example:

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

Note $\mathbf{P}^{2n}$ is the identity while $\mathbf{P}^{2n+1} = \mathbf{P}$. Note, too, that

$$\frac{\mathbf{P}^0 + \cdots + \mathbf{P}^n}{n+1} \to \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

Consider the equations $\pi = \pi \mathbf{P}$ with $\pi_1 + \pi_2 = 1$. We get

$$\pi_1 = \frac{1}{2}\pi_1 + \frac{1}{2}(1 - \pi_1) = \frac{1}{2}$$

so that the solution to $\pi = \pi \mathbf{P}$ is again unique.

Definition: The period d of a state i is the greatest common divisor of

$$\{n : (\mathbf{P}^n)_{ii} > 0\}$$

Lemma 1 *If $i \leftrightarrow j$ then i and j have the same period.*

Definition: A state is **aperiodic** if its period is 1.
Proof: I do the case $d = 1$. Fix i. Let

$$G = \{k : (\mathbf{P}^k)_{ii} > 0\}$$

If $k_1, k_2 \in G$ then $k_1 + k_2 \in G$.

This (and aperiodic) implies (number theory argument) that there is an r such that $k \geq r$ implies $k \in G$.

Now find m and n so that

$$(\mathbf{P}^m)_{ij} > 0 \text{ and } (\mathbf{P}^n)_{ji} > 0$$

For $k > r + m + n$ we see $(\mathbf{P}^k)_{jj} > 0$ so the gcd of the set of k such that $(\mathbf{P}^k)_{jj} > 0$ is 1. •

The case of period $d > 1$ can be dealt with by considering $\mathbf{P}^d$.

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

For this example $\{1, 2, 3\}$ is a class of period 3 states and $\{4, 5\}$ a class of period 2 states.

$$\mathbf{P} = \begin{bmatrix} 0 & 1/2 & 1/2 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

has a single communicating class of period 2.

A chain is **aperiodic** if all its states are aperiodic.