STAT 380
Markov Chains

Richard Lockhart

Simon Fraser University

Spring 2021

Markov Chains

- Last names example has following structure:
- Suppose, at generation n there are m individuals.
- Number of sons in next generation has distribution of sum of m independent copies of rv X
- Recall X is number of sons in first generation.
- Distribution does not depend on n,
- Depends only on the value m of Z_n.
- We call Z_n a **Markov Chain**.

Ingredients of a Markov Chain

- A **state space** S.
- S will be finite or countable in this course.
- A sequence $X_0, X_1, \ldots$ of random variables whose values are all in S.
- Matrix **P** with entries $P_{i,j}$ for $i, j \in S$.
- **P** is required to be **stochastic**:

$$\sum_k P_{ik} = 1 \qquad \text{and} \qquad 0 \le P_{ij}$$

for all i, j.

Stochastic Process

- The **stochastic process** $X_0, X_1, \ldots$ is called a Markov chain if

$$P\left(X_{k+1} = j | X_k = i, A\right) = P_{i,j}$$

- Here A is *any* event defined in terms of $X_0, \ldots, X_{k-1}$.
- Formula must hold for all i, j, k.
- Usually used with

$$A = \{X_{k-1} = i_{k-1}, \ldots, X_0 = i_0\}$$

  for some $i_0, \ldots, i_{k-1}$.
- Matrix $\mathbf{P}$ is called *transition* matrix.

First Markov Chain Example

- Suppose X in the last names example has a Poisson(λ) distribution
- Given $Z_n = k$, Z_{n+1} is like sum of k independent Poisson(λ) rvs
- This has a Poisson($k\lambda$) distribution.
- So

$$\mathbf{P} = \begin{bmatrix} 1 & 0 & 0 & \cdots \\ e^{-\lambda} & \lambda e^{-\lambda} & \lambda^2 e^{-\lambda}/2 & \cdots \\ e^{-2\lambda} & (2\lambda)e^{-2\lambda} & \frac{1}{2}(2\lambda)^2 e^{-2\lambda} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

Second Markov Chain Example

- Weather: each day is dry (D) or wet (W).
- X_n is weather on day n.
- Suppose dry days tend to be followed by dry days, say 3 times in 5 and wet days by wet 4 times in 5.
- Markov assumption: yesterday's weather irrelevant to prediction of tomorrow's given today's.
- Transition Matrix:

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

Transition probabilities

- Suppose it's wet today.
- Compute P(wet in 2 days)?

$$
\begin{aligned}
P(X_2 =& W|X_0 = W) \\
=& P(X_2 = W, X_1 = D|X_0 = W) + P(X_2 = W, X_1 = W|X_0 = W) \\
=& P(X_2 = W|X_1 = D, X_0 = W)P(X_1 = D|X_0 = W) \\
& \quad + P(X_2 = W|X_1 = W, X_0 = W)P(X_1 = W|X_0 = W) \\
=& P(X_2 = W|X_1 = D)P(X_1 = D|X_0 = W) \\
& \quad + P(X_2 = W|X_1 = W)P(X_1 = W|X_0 = W) \\
=& P_{W,D}P_{D,W} + P_{W,W}P_{W,W} \\
=& \left(\frac{1}{5}\right)\left(\frac{2}{5}\right) + \left(\frac{4}{5}\right)\left(\frac{4}{5}\right)
\end{aligned}
$$

- Note all entries in last line are items in **P**.

Chapman Kolmogorov equations; matrix multiplication

- Look at the matrix product **PP**:

$$\begin{bmatrix} \frac{3}{5} & \frac{2}{5} \\ \\ \frac{1}{5} & \frac{4}{5} \end{bmatrix} \begin{bmatrix} \frac{3}{5} & \frac{2}{5} \\ \\ \frac{1}{5} & \frac{4}{5} \end{bmatrix} = \begin{bmatrix} \frac{11}{25} & \frac{14}{25} \\ \\ \frac{7}{25} & \frac{18}{25} \end{bmatrix}$$

- Notice that $P(X_2 = W | X_0 = W)$ is exactly the W, W entry in **PP**.
- General case. Define

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

- Then

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

Proof

- 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: for any x, y, u, v

$$P(Y = y|X = x, U = u, V = v) = P(Y = y|X = x)$$

- Then I claim

$$P(Y = y|X = x, U = u) = P(Y = y|X = x)$$

- 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

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

- This shows that

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

  where $\mathbf{P}^2$ means the matrix product $\mathbf{PP}$.
- Note: quantity does not depend on n
- Note: can compute by taking a power of $\mathbf{P}$.

Chapman-Kolmogorov

- 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.

Maple Output: rainfall example

1,2,4,8 and 16 step transition matrices:

```
> 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):
```

- Computes powers (`evalm` understands matrix algebra).
- Fact:

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

Maple powers

```
> evalf(evalm(p));
             [.6000000000    .4000000000]
             [                          ]
             [.2000000000    .8000000000]
> evalf(evalm(p2));
             [.4400000000    .5600000000]
             [                          ]
             [.2800000000    .7200000000]
> evalf(evalm(p4));
             [.3504000000    .6496000000]
             [                          ]
             [.3248000000    .6752000000]
```

Maple powers

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

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

Initial distributions

- Suppose: toss a coin $P(H) = \alpha_D$
- Start 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

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

- Last line is matrix multiplication of row vector α by matrix $\mathbf{P}$.

Stationary initial distribution

- 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.
- So X_0 and X_1 have the same distribution.

Formal definitions

- 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

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

Characterizing the stationary distribution

- 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

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

Eigenvectors

- **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.

Finding Stationary Initial Distributions

- Consider **P** for the weather example.
- The equation

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

is really

$$\alpha_D = 3\alpha_D/5 + \alpha_W/5$$
$$\alpha_W = 2\alpha_D/5 + 4\alpha_W/5$$

- The first can be rearranged to

$$\alpha_W = 2\alpha_D;$$

so can the second.

Finding Stationary Initial Distributions again

- 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

$$\alpha_D = 1/3$$

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

Maple examples

```
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 ]
```

Maple examples

```
> 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):
```

Maple examples

```
> 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]
```

Maple examples

```
> 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 converge but$(\mathbf{P}^n + \mathbf{P}^{n+1})/2$ does.

Maple examples

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

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

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

More

- 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.

Another Example

```
> 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):
```

More

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

Limit Distributions

- 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 $\mathbf{P}$.

Classification of States

- State i **leads to** state j if $\mathbf{P}^n_{ij} > 0$ for some n.
- Convenient to agree say $\mathbf{P}^0 = \mathbf{I}$, the identity matrix.
- So i leads to i.
- Note i leads to j and j leads to k implies i leads to k (Chapman-Kolmogorov).
- States i and j **communicate** if i leads to j and j leads to i.
- The relation of communication is an **equivalence** relation.
- it is reflexive, symmetric and transitive: if i and j communicate and j and k communicate then i and k communicate.
- Group of communicating states called *Communicating Class*.

Example of communicating classes

- Example ($+$ signs indicate non-zero entries):

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

- For this example: $1 \rightsquigarrow 2$, $2 \rightsquigarrow 3$, $3 \rightsquigarrow 1$
- So $1, 2, 3$ are all in the same communicating class.
- $4 \rightsquigarrow 1, 2, 3$ but not vice versa.
- $5 \rightsquigarrow 1, 2, 3, 4$ but not vice versa.
- So the communicating classes are

$$\{1, 2, 3\} \quad \{4\} \quad \{5\}$$

Irreducible Chains, transience, recurrence

- A Markov Chain is **irreducible** if there is only one communicating class.
- Notation:

$$f_i = P(\exists n > 0 : X_n = i | X_0 = i)$$

- State i is **recurrent** if $f_i = 1$, otherwise **transient**.
- If $f_i = 1$ then Markov property (chain starts over when it gets back to i) means prob return infinitely many times (given started in i or given ever get to i) is 1.

Number of returns to transient state

- Consider chain started from transient i.
- Let N be number of visits to state i (including visit at time 0).
- To return m times must return once then starting over return $m-1$ times, then never return.
- So:

$$P(N = m|X_0 = i) = f_i^{m-1}(1 - f_i)$$

  for $m = 1, 2, \ldots$.
- N has a Geometric distribution and $\mathrm{E}(N|X_0 = i) = 1/(1 - f_i)$.

Condition for transience

- Another calculation:

$$N = \sum_{k=0}^{\infty} 1(X_k = i)$$

so

$$\mathrm{E}(N|X_0 = i) = \sum_{k=0}^{\infty} P(X_k = i|X_0 = i)$$

- If we start the chain in state i then this is

$$\mathrm{E}(N|X_0 = i) = \sum_{k=0}^{\infty} \mathbf{P}_{ii}^k$$

and i is transient if and only if

$$\sum_{k=0}^{\infty} \mathbf{P}_{ii}^k < \infty .$$

Continued

- For last example: 4 and 5 are transient.
- Claim: states 1, 2 and 3 are recurrent.
- Proof: argument above shows each transient state is visited only finitely many times.
- So: there is a recurrent state.
- (Note use of finite number of states.)
- It must be one of 1, 2 and 3.

Class properties

- **Proposition**: If one state in a communicating class is recurrent then all states in the communicating class are recurrent.
- Proof: Let i be the known recurrent state so

$$\sum_n \mathbf{P}_{ii}^n = \infty$$

- Assume i and j communicate. Find integers m and k such that

$$\mathbf{P}_{ij}^m > 0 \text{ and } \mathbf{P}_{ji}^k > 0$$

- Then

$$\mathbf{P}_{jj}^{m+n+k} \ge \mathbf{P}_{ji}^k \mathbf{P}_{ii}^n \mathbf{P}_{ij}^m$$

- Sum RHS over n get ∞ so

$$\sum_n \mathbf{P}_{jj}^n = \infty$$

- Proposition also means that if 1 state in a class is transient so are all.

Periodic chains

- State i has period d if d is greatest common divisor of

$$\{n : \mathbf{P}^n_{ii} > 0\}$$

- If i and j are in the same class then i and j have same period.
- If $d = 1$ then state i is called **aperiodic**.
- If $d > 1$ then i is periodic.

Periodic example

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

- 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.

Infinite State Spaces

- Example: sequence of independent coin tosses; probability p of Heads on a single toss.
- X_n is number of heads minus number of tails after n tosses.
- Put $X_0 = 0$.
- X_n is a Markov Chain.
- State space is $\mathbb{Z}$, the integers and

$$\mathbf{P}_{ij} = \begin{cases} p & j = i+1 \\ 1-p & j = i-1 \\ 0 & \text{otherwise} \end{cases}$$

Analysis of heads minus tails

- Chain has one communicating class (for $p \neq 0, 1$).
- All states have period 2.
- According to the strong law of large numbers X_n/n converges to $2p - 1$.
- If $p \neq 1/2$ this guarantees that for all large enough n $X_n \neq 0$, that is, the number of returns to 0 is not infinite.
- So state 0 is transient
- So **all** states must be transient.

Fair coin case

- For $p = 1/2$ the situation is different.
- It is a fact that

$$\mathbf{P}_{00}^{n} = P(\#\text{ H} = \#\text{ T at time } n)$$

- For n even this is the probability of exactly $n/2$ heads in n tosses.
- Local Central Limit Theorem (normal approximation to $P(-1/2 < X_n < 1/2)$) (or Stirling's approximation) shows

$$\sqrt{2m}P(\text{Binomial}(2m, 1/2) = m) \to (2/\pi)^{1/2}$$

so:

$$\sum_n \mathbf{P}_{00}^{n} = \infty$$

- That is: 0 is a recurrent state.

Hitting Times

- Start irreducible recurrent chain X_n in state i.
- Let T_j be first $n > 0$ such that $X_n = j$.
- Define
$$m_{ij} = \mathrm{E}(T_j | X_0 = i)$$
- First step analysis:
$$\begin{aligned} m_{ij} &= 1 \cdot P(X_1 = j | X_0 = i) \\ &\qquad + \sum_{k \neq j} (1 + \mathrm{E}(T_j | X_0 = k)) P_{ik} \\ &= \sum_k P_{ik} + \sum_{k \neq j} P_{ik} m_{kj} \\ &= 1 + \sum_{k \neq j} P_{ik} m_{kj} \end{aligned}$$

Example

- Example

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

- The equations are

$$\begin{aligned} m_{11} &= 1 + \frac{2}{5} m_{21} \\ m_{12} &= 1 + \frac{3}{5} m_{12} \\ m_{21} &= 1 + \frac{4}{5} m_{21} \\ m_{22} &= 1 + \frac{1}{5} m_{12} \end{aligned}$$

Example Continued

- The second and third equations give immediately

$$m_{12} = \frac{5}{2}$$
$$m_{21} = 5$$

- Then plug in to the others to get

$$m_{11} = 3$$
$$m_{22} = \frac{3}{2}$$

Relation to Stationary Initial Distribution

- Notice stationary initial distribution is

$$\left(\frac{1}{m_{11}}, \frac{1}{m_{22}}\right)$$

- Consider fraction of time spent in state j:

$$\frac{1(X_0 = j) + \cdots + 1(X_n = j)}{n+1}$$

- Imagine chain starts in state i; take expected value.

$$\frac{\sum_{r=1}^{n} \mathbf{P}_{ij}^{r} + 1(i=j)}{n+1} = \frac{\sum_{r=0}^{n} \mathbf{P}_{ij}^{r}}{n+1}$$

- If rows of $\mathbf{P}^n$ converge to stationary α then fraction converges to α_j; i.e. limiting fraction of time in state j is α_j.

Heuristics

- Heuristic: start chain in i.
- Expect to return to i every m_{ii} time units.
- So are in state i about once every m_{ii} time units; i.e. limiting fraction of time in state i is $1/m_{ii}$.
- Conclusion: for an irreducible recurrent finite state space Markov chain

$$\alpha_i = \frac{1}{m_{ii}}.$$

Infinite State Spaces

- Conclusion still right **if** $\exists$ a stationary initial distribution.
- Example: $X_n = \text{Heads} - \text{Tails}$ after n tosses of fair coin.
- Equations are

$$m_{0,0} = 1 + \frac{1}{2} m_{1,0} + \frac{1}{2} m_{-1,0}$$
$$m_{1,0} = 1 + \frac{1}{2} m_{2,0}$$

  and many more.
- You have to go through 1 to get to 0 from 2 so

$$m_{2,0} = m_{2,1} + m_{1,0}$$

- Symmetry (switching H and T):

$$m_{1,0} = m_{-1,0}$$

- The transition probabilities are **homogeneous**:

$$m_{2,1} = m_{1,0}$$

Example continued

- Conclusion:

$$
\begin{aligned}
m_{0,0} &= 1 + m_{1,0} \\
&= 1 + 1 + \frac{1}{2} m_{2,0} \\
&= 2 + m_{1,0}
\end{aligned}
$$

- Notice that there are **no** finite solutions!
- Summary of the situation:
- Every state is recurrent.
- All the expected hitting times m_{ij} are infinite.
- All entries $\mathbf{P}^n_{ij}$ converge to 0.
- Jargon: The states in this chain are null recurrent.

One Example

- Page 229, question 21 in old edition.
- Runner goes from front or back door, prob $1/2$ each.
- Returns front or back, prob $1/2$ each.
- Has k pairs of shoes, wears pair if any at departure door, leaves at return door.
- No shoes? Barefoot.
- Long run fraction of time barefoot?

Solution

- Let X_n be number of shoes at front door on day n.
- Then X_n is a Markov Chain.
- Transition probabilities?
- k pairs at front door on day n.
- X_{n+1} is k if goes out back door (prob is $1/2$) or out front door and back in front door (prob is $1/4$).
- Otherwise X_{n+1} is $k-1$.

Solution

- $0 < j < k$ pairs at front door on day n.
- X_{n+1} is $j+1$ if out back, in front (prob is $1/4$). X_{n+1} is $j-1$ if out front, in back.
- Otherwise X_{n+1} is j.
- 0 pairs at front door on day n.
- X_{n+1} is 0 if out front door (prob $1/2$) or out back door and in back door (prob $1/4$) otherwise X_{n+1} is 1.

Solution

- Transition matrix **P**:

$$\begin{bmatrix} \frac{3}{4} & \frac{1}{4} & 0 & 0 & \cdots & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 & \cdots & 0 \\ 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{1}{4} & \frac{3}{4} \end{bmatrix}$$

- Doubly stochastic: row sums and column sums are 1.
- So $\alpha_i = 1/(k+1)$ for all i is stationary initial distribution.

Solution

- Solution to problem: 1 day in $k+1$ no shoes at front door.
- Half of those go barefoot.
- Also 1 day in $k+1$ all shoes at front door; go barefoot half of these days.
- Overall go barefoot $1/(k+1)$ of the time.

Gambler's Ruin

- Insurance company's reserves fluctuate: sometimes up, sometimes down.
- Ruin is event they hit 0 (company goes bankrupt).
- General problem.
- For given model of fluctuation compute probability of ruin either eventually or in next k time units.

Gambler's Ruin, simple example

- Simplest model: gambling on Red at Casino.
- Bet \$1 at a time.
- Win \$1 with probability p, lose \$1 with probability $1 - p$.
- Start with k dollars.
- Quit playing when down to \$0 or up to N.
- Compute

$$P_k = P(\text{reach } N \text{ before } 0 | X_0 = k)$$

Gambler's Ruin, simple example

- X_n = fortune after n plays.
- $X_0 = k$.
- Transition matrix:

$$\mathbf{P} = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ 1-p & 0 & p & 0 & \cdots & 0 \\ 0 & 1-p & 0 & p & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix}$$

Gambler's Ruin, simple example

- First step analysis:

$$
\begin{aligned}
P_0 &= 0 \\
P_i &= (1-p)P_{i-1} + pP_{i+1} \\
P_N &= 1
\end{aligned}
$$

Gambler's Ruin, simple example

- Middle equation is

$$pP_i + (1-p)P_i = (1-p)P_{i-1} + pP_{i+1}$$

or

$$\begin{aligned} P_{i+1} - P_i &= \frac{1-p}{p}(P_i - P_{i-1}) \\ &= \left(\frac{1-p}{p}\right)^2 (P_{i-1} - P_{i-2}) \\ &\vdots \\ &= \left(\frac{1-p}{p}\right)^i (P_1 - P_0) \\ &= \left(\frac{1-p}{p}\right)^i P_1 \end{aligned}$$

Gambler's Ruin, simple example

- Sum from $i = 0$ to $i = k - 1$ to get

$$P_k = \sum_{i=0}^{k-1} \left(\frac{1-p}{p} \right)^i P_1$$

or

$$P_k = \frac{1 - \{(1-p)/p\}^k}{1 - \{(1-p)/p\}} P_1$$

Gambler's Ruin, simple example

- For $k = N$ we get

$$1 = \frac{1 - \{(1-p)/p\}^N}{1 - \{(1-p)/p\}} P_1$$

so that

$$P_k = \frac{1 - \{(1-p)/p\}^k}{1 - \{(1-p)/p\}^N}$$

- Notice that if $p = 1/2$ our formulas for the sum of the geometric series are wrong.
- But for $p = 1/2$ we get

$$P_k = kP_1$$

so

$$P_k = \frac{k}{N}.$$

Mean time in transient states

$$\mathbf{P} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{4} & \frac{3}{4} & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{3}{8} & \frac{1}{8} \end{bmatrix}$$

- States 3 and 4 are transient.
- Let $m_{i,j}$ be the expected total number of visits to state j for chain started in i.
- For $i = 1$ or $i = 2$ and $j = 3$ or 4:

$$m_{ij} = 0$$

- For $j = 1$ or $j = 2$

$$m_{ij} = \infty$$

Mean time in transient states

- For $i, j \in \{3, 4\}$ first step analysis:

$$m_{3,3} = 1 + \frac{1}{4} m_{3,3} + \frac{1}{4} m_{4,3}$$
$$m_{3,4} = 0 + \frac{1}{4} m_{3,4} + \frac{1}{4} m_{4,4}$$
$$m_{4,3} = 0 + \frac{3}{8} m_{3,3} + \frac{1}{8} m_{4,3}$$
$$m_{4,4} = 1 + \frac{3}{8} m_{3,4} + \frac{1}{8} m_{4,4}$$

- In matrix form

$$\begin{bmatrix} m_{3,3} & m_{3,4} \\ m_{4,3} & m_{4,4} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} \frac{1}{4} & \frac{1}{4} \\ \frac{3}{8} & \frac{1}{8} \end{bmatrix} \begin{bmatrix} m_{3,3} & m_{3,4} \\ m_{4,3} & m_{4,4} \end{bmatrix}$$

Mean time in transient states

- Translate to matrix notation:

$$\mathbf{M} = \mathbf{I} + \mathbf{P}_T\mathbf{M}$$

- where $\mathbf{I}$ is the identity,
- $\mathbf{M}$ is the matrix of means and
- $\mathbf{P}_T$ the part of the transition matrix corresponding to transient states.
- Solution is

$$\mathbf{M} = (\mathbf{I} - \mathbf{P}_T)^{-1}$$

In our case

$$\mathbf{I} - \mathbf{P}_T = \begin{bmatrix} \frac{3}{4} & -\frac{1}{4} \\ -\frac{3}{8} & \frac{7}{8} \end{bmatrix}$$

so that

$$\mathbf{M} = \begin{bmatrix} \frac{14}{9} & \frac{4}{9} \\ \frac{2}{3} & \frac{4}{3} \end{bmatrix}$$

Data Analysis

- Imagine we have data $X_0, \ldots, X_N$ which we model as coming from a Markov Chain.
- Simplest case: K states.
- If we observe $X_0 = x_0, \ldots, X_N = x_N$ how should we estimate the transition probabilities?
- Two kinds of models: parametric and 'empirical'?
- Second kind: P_{ij} can be *any* probabilities subject to $\sum_j P_{ij} = 1$.
- First kind: each P_{ij} is function of smaller number of parameters, θ.
- The θ or the P_{ij} are *parameters*.

Small likelihood example

- Two states $\{0, 1\}$.
- Suppose we observe the sequence $0, 0, 1, 1, 1, 0, 1, 0, 1$.
- So $N = 8$ and $x_0 = 0, x_1 = 0, \ldots, x_8 = 1$.
- The likelihood is

$$L = P(X_0 = 0, X_1 = 0, X_2 = 1, X_3 = 1, \\ X_4 = 1, X_5 = 0, X_6 = 1, X_7 = 0, X_8 = 1)$$

- Use Markov property to get

$$L = P(X_0 = 0) P_{00} P_{01} P_{11} P_{11} P_{10} P_{01} P_{10} P_{01}$$

Small likelihood example, II

- Collect terms:
$$L = P(X_0 = 0)P_{00}P_{01}^3 P_{10}^2 P_{11}^2$$
- Power on P_{ij} is number of times we observe a transition from i to j.
- Let N_{ij} be number of transitions from i to j.
- Likelihood is generally
$$L = P(X_0 = x_0)\prod_{ij} P_{ij}^{N_{ij}}.$$
- Usually: we condition on X_0 and use *conditional likelihood*
$$L_c = \prod_{ij} P_{ij}^{N_{ij}}.$$
- Conditional log-likelihood is
$$\ell_c = \sum_{ij} N_{ij} log(P_{ij})$$

Maximum likelihood

- Maximize ℓ_c over all P_{ij}
- Remember K constraints

$$\sum_j P_{ij} = 1$$

- Lagrange multipliers:

$$\ell_P - \sum_{i=1}^{k} \lambda_i \left[\sum_j P_{ij} - 1 \right]$$

- Take derivative wrt P_{kl} to get

$$\frac{N_{kl}}{P_{kl}} - \lambda_k$$

Maximum likelihood

- Set equal to 0 to find

$$P_{kl} = \frac{N_{kl}}{\lambda_k}$$

- Use constraint to get

$$1 = \sum_l P_{kl} = \frac{\sum_l N_{kl}}{\lambda_k}$$

- So

$$\hat{P}_{kl} = \frac{N_{kl}}{\sum_l N_{kl}}$$

Maximum likelihood, parametric models

- If each $P_{ij} = P_{ij}(\theta)$ then the likelihood equations become

$$\sum_{ij} \frac{N_{ij}}{P_{ij}} \frac{\partial P_{ij}(\theta)}{\partial \theta_r} = 0$$

  for $r = 1, \ldots, q$ if $\theta = (\theta_1, \ldots, \theta_q)$.
- In parametric models we usually parametrize the functions P_{ij} in such a way that the constraints are automatic.

Data example from Guttorp, 1995

- Rock layers (*strata*), $N = 606$.
- X_n is one of 6 *types* (numbered 0 to 5).
- Markov chain model.
- The N_{ij} are *sufficient* statistics.
- Simple parametric model: $\theta = (\theta_0, \ldots, \theta_5)$ and

$$P_{ij} = \begin{cases} 0 & i = j \\ \frac{\theta_j}{\sum_{k \neq i} \theta_k} & i \neq j \end{cases}$$

- Best approximation to *independent*.
- Independence of layers not possible because $P_{ii} = 0$ by definition.

Data example from Guttorp, 1995

- Needed derivatives

$$\frac{\partial P_{ij}(\theta)}{\partial \theta_r} = \begin{cases} 0 & i = j \\ 0 & r = i \\ \frac{1}{\sum_{k \neq i} \theta_k} - \frac{\theta_r}{\left(\sum_{k \neq i}\right)^2} & i \neq j, r = j \\ -\frac{\theta_j}{\left(\sum_{k \neq i}\right)^2} & i \neq j, r \neq j \end{cases}$$

- Equations tedious to write out.

The Data

Facies	0	1	2	3	4	5	
0	0	2	0	2	2	0	6
1	5	0	23	31	17	8	84
2	0	21	0	45	27	8	101
3	1	54	44	0	66	25	190
4	0	6	24	81	0	38	149
5	0	5	8	31	32	0	76
	6	88	99	190	144	79	606