Markov Chains Basics

Stochastic processes

- Examples:
  - Price of a stock or portfolio of stocks
  - Inventory level of a good in a warehouse
  - Career path (workforce planning)
  - Number of customers present in a store or bank
  - etc.

Stochastic processes

- We will start by looking at systems that we observe at *discrete* points in time
  - E.g., we evaluate the value of a stock portfolio at the end of each trading day
- Denote the time points by t=0,1,2,…
- Denote the value (at time t) of the characteristic of the system that we are interested in by X_t
  - X_t is a *random variable*
  - The sequence X_0, X_1, X_2,… is called a *stochastic process*

Example 1

- We are interested in tracking the stock market on a daily basis
- Let X_t denote the value of the Dow Jones index at the end of trading day t
  - We are currently at the end of trading day 0, and observe $X_0=x_0$
- Can we model/study the relationship between the random variables X_t (t=0,1,2,…)?

Example 2

- You are visiting Las Vegas, and have a gambling budget of \$ x_0
- You participate in a game in which you repeatedly
  - bet \$1
    - if you win (which happens with probability p) you receive \$2
    - if you loose (which happens with probability 1-p) you receive nothing
  - you stop playing as soon as you are broke or have doubled your initial budget
- Denote your total *wealth* at time t by X_t

States and state space

- The set of values that the random variable X_t can take is called the *state space* of the stochastic process
  - Example 1: $[0,\infty)$
  - Example 2: $\{0,1,\ldots,2x_0\}$
  - We will restrict ourselves to situations in which the state space consists of a finite number of elements only
    - often: $S = \{1,2,\ldots,s\}$
- If $X_t=i$ we say that the stochastic process is in state i at time t.

Dynamics of the stochastic process

- Our goal is to describe and study the behavior of the sequence of random variables $X_0, X_1, X_2, \ldots$
- More specifically, suppose we are currently at time t
  - The stochastic process is characterized by how the next observed value, X_{t+1}, depends on the past observations

Markov Chains

- An important type of stochastic processes is called a *Markov chain*
- In a Markov chain, the *future* value of the random variables only depend on the *current* value of the random variable, but not on past values
  - In other words:
    - all past observations can be *summarized* by the current value of the stochastic process
    - if we know the current state of the stochastic process, it is irrelevant to know how that state was reached

Markov Chains

- Example 1:
  - To predict future values of the DJIA [Dow Jones Industrial Average (Dow Jones Indices)], is it sufficient to know the current value of the DJIA or do we need to know some or all past observations?
  - Under the so-called *efficient market hypothesis* past observations are irrelevant if we know the current value of the DJIA, so the sequence of stock prices is a Markov chain

Markov Chains

- Example 2:
  - To predict your wealth after the next play, is it sufficient to know what your current wealth is or do we need to know what happened on all previous plays?
  - Given that the outcome of the next play is independent of the outcome of the previous plays it is sufficient to know your current wealth

Markov Chains

- Formally:
  - A stochastic process is a *Markov chain* if for t=0,1,2,… and all states

$$P(X_{t+1} = j \mid X_t = i, X_{t-1} = i_{t-1}, \ldots, X_0 = i_0) = P(X_{t+1} = j \mid X_t = i)$$

- Furthermore, we will assume that this probability is *independent of t:*

$$P(X_{t+1} = j \mid X_t = i) = p_{ij}$$

Transition probabilities

- We call the numbers p_{ij} the *transition probabilities*
- The assumption that these are independent of t is called the *stationarity assumption*
  - The Markov chain is called *stationary* or *homogeneous*

Transition probability matrix

- It is common and useful to arrange the transition probabilities in a *matrix*:

$$P = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1s} \\ p_{21} & p_{22} & \cdots & p_{2s} \\ \vdots & \vdots & & \vdots \\ p_{s1} & p_{s2} & \cdots & p_{ss} \end{pmatrix}$$

Transition probability matrix

- Each *row* of this matrix is a *probability distribution*
  - It describes the probability distribution of the next observed value *given* some value for the current state
  - Therefore, the elements in a row of the matrix must be nonnegative ($p_{ij} \geq 0$) and sum to 1:

$$\sum_{j=1}^{s} p_{ij} = \sum_{j=1}^{s} P(X_{t+1} = j \mid X_t = i) = 1$$

$$\text{for all } i = 1, \ldots, s$$

Example 2

- What are the transition probabilities for the gambling problem?
- Suppose your current wealth is $X_t=i$
  - What are the possible values for X_{t+1}?
    - If you play and win, your wealth will be $X_{t+1}=i+1$
    - If you play and loose, your wealth will be $X_{t+1}=i-1$
    - If you do not play, your wealth will be $X_{t+1}=i$
      - This only happens if $i=0$ or $i=2x_0$, since at that point you stop playing!

Example 2

- Suppose $0<i<2x_0$, so you keep playing

$$P(X_{t+1} = i+1 \mid X_t = i) = p$$

$$P(X_{t+1} = i-1 \mid X_t = i) = 1-p$$

$$P(X_{t+1} = j \mid X_t = i) = 0 \qquad \text{if } j \neq i-1, i+1$$

Example 2

- Suppose i=0

$$P(X_{t+1} = 0 \mid X_t = 0) = 1$$
$$P(X_{t+1} = j \mid X_t = 0) = 0 \quad \text{if } j \neq 0$$

- Suppose $i=2x_0$

$$P(X_{t+1} = 2x_0 \mid X_t = 2x_0) = 1$$
$$P(X_{t+1} = j \mid X_t = 2x_0) = 0 \quad \text{if } j \neq 2x_0$$

Example 2

- For simplicity, let us assume that x_0=2
  - The state space is then {0,1,2,3,4}
  - The transition probability matrix is then:

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1-p & 0 & p & 0 & 0 \\ 0 & 1-p & 0 & p & 0 \\ 0 & 0 & 1-p & 0 & p \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Example 3

- In Smalltown
  - 90% of all sunny days are followed by a sunny day
  - 80% of all cloudy days are followed by a cloudy day
- Model Smalltown's weather as a Markov chain

Example 3

- X_t = weather on day t
- State space: $\{\text{sunny,cloudy}\} = \{1,2\}$
- Transition probability matrix:

$$P = \begin{pmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{pmatrix}$$

Example 3'

- In Smalltown
  - If the last two days have been sunny, then tomorrow will be sunny with probability 95%
  - If yesterday was cloudy and today was sunny, then tomorrow will be sunny with probability 70%
  - If yesterday was sunny and today was cloudy, then tomorrow will be sunny with probability 60%
  - If the last two days have been cloudy, then tomorrow will be sunny with probability 80%
- Can Smalltown's weather be modeled as a Markov chain?

Example 3'

- X_t = weather on days *t-1* and *t*
- State space: {(sunny,sunny), (cloudy,sunny), (sunny,cloudy), (cloudy,cloudy)} = {1,2,3,4}
- Transition probability matrix:

$$P = \begin{pmatrix} 0.95 & 0 & 0.05 & 0 \\ 0.70 & 0 & 0.30 & 0 \\ 0 & 0.60 & 0 & 0.40 \\ 0 & 0.20 & 0 & 0.80 \end{pmatrix}$$

n-Step transition probabilities

- Recall the transition probabilities

$$P(X_{t+1} = j \mid X_t = i) = p_{ij}$$

- It is sometimes also interesting to answer questions like:
  - If the Markov chain is currently in state i, what is the probability that n periods from now the Markov chain is in state j ?

n-Step transition probabilities

- In general, we would like to know

$$P(X_{t+n} = j \mid X_t = i) = P_{ij}(n)$$

- These are called the *n-step transition probabilities*
- Clearly, $P_{ij}(1)=p_{ij}$
- Let's look at $P_{ij}(2)$

2-Step transition probabilities

- We go from state i at time t to state j at time $t+2$ by passing through some state k
- However, we do not know (or care!!) what the intermediate state k is
- So

$$
\begin{aligned}
P_{ij}(2) &= P(X_{t+2} = j \mid X_t = i) \\
&= \sum_{k=1}^{s} P(X_{t+1} = k \mid X_t = i) \cdot P(X_{t+2} = j \mid X_{t+1} = k) \\
&= \sum_{k=1}^{s} p_{ik} \cdot p_{kj}
\end{aligned}
$$

2-Step transition probabilities

- If we arrange the probabilities $P_{ij}(2)$ in a matrix, we have

$$
P(2) = \begin{pmatrix} p_{11}(2) & p_{12}(2) & \cdots & p_{1s}(2) \\ p_{21}(2) & p_{22}(2) & \cdots & p_{2s}(2) \\ \vdots & \vdots & & \vdots \\ p_{s1}(2) & p_{s2}(2) & \cdots & p_{ss}(2) \end{pmatrix} = P \cdot P = P^2
$$

n-Step transition probabilities

- We go from state i at time t to state j at time $t+n+1$ by passing through some state k at time $t+n$
- However, we do not know (or care!!) what the intermediate state k is
- So

$$
\begin{aligned}
P_{ij}(n+1) &= P(X_{t+n+1} = j \mid X_t = i) \\
&= \sum_{k=1}^{s} P(X_{t+n} = k \mid X_t = i) \cdot P(X_{t+n+1} = j \mid X_{t+n} = k) \\
&= \sum_{k=1}^{s} P_{ik}(n) \cdot p_{kj}
\end{aligned}
$$

n-Step transition probabilities

- If we arrange the probabilities $P_{ij}(n+1)$ in a matrix, we have

$$
P(n+1) = \begin{pmatrix} p_{11}(n+1) & p_{12}(n+1) & \cdots & p_{1s}(n+1) \\ p_{21}(n+1) & p_{22}(n+1) & \cdots & p_{2s}(n+1) \\ \vdots & \vdots & & \vdots \\ p_{s1}(n+1) & p_{s2}(n+1) & \cdots & p_{ss}(n+1) \end{pmatrix} = P(n) \cdot P
$$

- We can conclude that $P(n) = P^n$ for all $n=1,2,\ldots$

n-Step transition probabilities

- What about $n=0$?

$$P_{ij}(0) = P(X_t = j \mid X_t = i) = \begin{cases} 1 & \text{if } j = i \\ 0 & \text{if } j \neq i \end{cases}$$

- So $P(0) = I = P^0$

- In matrix notation, this is equivalent to $\mathbf{P}^{(n)} = \mathbf{P}^{(n-1)}\mathbf{P}$

where $\mathbf{P}^{(n)}$ is the matrix whose elements are the *n-step transition probabilities.*

- This fundamental relationship specifying the *n-step transition probabilities* in terms of those of lower order is called the **Chapman-Kolmogorov** equation.
- A more general form of the **Chapman-Kolmogorov** equation is the following $\mathbf{P}^{(n)} = \mathbf{P}^{(n-m)}\mathbf{P}^{(m)}$.
- Applying Chapman-Kolmogorov equation iteratively yields that $\mathbf{P}^{(n)} = \mathbf{P}^{n}$. That is, the transition matrix of the n-step probabilities is just the matrix of the one-step transition probabilities raised to the power n.

Example 3

- Recall the 1st Smalltown example:
  - 90% of all sunny days are followed by a sunny day
  - 80% of all cloudy days are followed by a cloudy day

$$P = \begin{pmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{pmatrix}$$

$$P^2 = \begin{pmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{pmatrix} \cdot \begin{pmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{pmatrix} = \begin{pmatrix} 0.83 & 0.17 \\ 0.34 & 0.66 \end{pmatrix}$$

Example 3

- Some *n*-step transition probability matrices are:

$$P^3 = \begin{pmatrix} 0.83 & 0.17 \\ 0.34 & 0.66 \end{pmatrix} \cdot \begin{pmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{pmatrix} \approx \begin{pmatrix} 0.78 & 0.22 \\ 0.44 & 0.56 \end{pmatrix}$$

$$P^{10} \approx \begin{pmatrix} 0.68 & 0.32 \\ 0.65 & 0.35 \end{pmatrix} \qquad P^{20} \approx \begin{pmatrix} 0.67 & 0.33 \\ 0.67 & 0.33 \end{pmatrix}$$

Long-run behavior

- Note that, as n grows large,
  - the matrix P^n hardly changes
  - the rows of the matrix are (approximately) the same
- Let's see what happens to P^n for the gambling example

Example 2

- For simplicity, let us assume that $x_0=2$
  - The state space is then $\{0,1,2,3,4\}$
- Let's also assume that $p=½$
  - The transition probability matrix is then:

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ ½ & 0 & ½ & 0 & 0 \\ 0 & ½ & 0 & ½ & 0 \\ 0 & 0 & ½ & 0 & ½ \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Example 2

- Some n-step transition probability matrices are:

$$P^2 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{4} & 0 & \frac{1}{4} & 0 \\ \frac{1}{4} & 0 & \frac{1}{2} & 0 & \frac{1}{4} \\ 0 & \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad P^3 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \frac{5}{8} & 0 & \frac{1}{4} & 0 & \frac{1}{8} \\ \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{8} & 0 & \frac{1}{4} & 0 & \frac{5}{8} \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Example 2

- Some more n-step transition probability matrices are:

$$P^{10} \approx \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ .73 & .02 & 0 & .02 & .23 \\ .485 & 0 & .03 & 0 & .485 \\ .23 & .02 & 0 & .02 & .73 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad P^{15} \approx \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ .75 & 0 & 0 & 0 & .25 \\ .50 & 0 & 0 & 0 & .50 \\ .25 & 0 & 0 & 0 & .75 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Long-run behavior

- Again, it seems that, as n grows large,
  - the matrix P^n hardly changes
- However, in this case
  - the rows of the matrix do *not* approach the same values

Example 2

- Next, let us look at the case $p=\frac{1}{4}$
  - The transition probability matrix is then:

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \frac{3}{4} & 0 & \frac{1}{4} & 0 & 0 \\ 0 & \frac{3}{4} & 0 & \frac{1}{4} & 0 \\ 0 & 0 & \frac{3}{4} & 0 & \frac{1}{4} \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Example 2

- Some n-step transition probability matrices are:

$$P^2 \approx \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ .75 & .19 & 0 & .06 & 0 \\ .56 & 0 & .38 & 0 & .06 \\ 0 & .56 & 0 & .19 & .25 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad P^3 \approx \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ .89 & 0 & .09 & 0 & .02 \\ .56 & .28 & 0 & .10 & .06 \\ .42 & 0 & .28 & 0 & .30 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Example 2

- Some more n-step transition probability matrices are:

$$P^{10} \approx \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ .97 & .005 & 0 & 0 & .025 \\ .89 & 0 & .01 & 0 & .10 \\ .66 & .01 & 0 & .01 & .32 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad P^{15} \approx \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ .975 & 0 & 0 & 0 & .025 \\ .90 & 0 & 0 & 0 & .10 \\ .675 & 0 & 0 & 0 & .325 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

- Now, if the unconditional probability $P\{X_n = j\}$ (called the state probabilities) are desired, the initial state must be specified. It is however possible that the initial conditions are given by a distribution over the possible states at time zero. Let , and let $\mathbf{a}^{(n)}$ denote the row vector whose elements are the $p_i^{(n)}$. Then, the initial conditions are given by $\mathbf{a}^{(0)}$. To get $\mathbf{p}^{(n)}$, we proceed as follows.

$$p_j^{(n)} = P\{X_n = j\} = \sum_i P\{X_n = j | X_0 = i\} P\{X_0 = i\} = \sum_i p_{ij}^{(n)} p_i^{(0)}$$

Therefore, $p_j^{(n)} = \sum_i p_i^{(0)} p_{ij}^{(n)}$

- Recognizing the right hand side as multiplication of vectors, we conclude that

$$\mathbf{a}^{(n)} = \mathbf{a}^{(0)}\mathbf{P}^{(n)}$$

- In other words, the vector of state probabilities for time *n* is given by the vector of the initial state probabilities multiplied by the *n-step transition* matrix. Thus, it is clear that the Markov chain is completely specified when its transition matrix **P** and the vector of initial conditions $\mathbf{a}^{(0)}$ are known.

- let **P** take the following numerical values.

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

Also, assume that the vector of initial conditions is given by $\mathbf{a}^{(0)}$ $=[1/3, 1/3, 1/3]$.

- It is easy to verify that

$$P^{(2)} = P^2 = \begin{bmatrix} 0.500 & 0.333 & 0.167 \\ 0.333 & 0.389 & 0.278 \\ 0.111 & 0.444 & 0.444 \end{bmatrix}$$

- From the relationship $\mathbf{a}^{(2)} = \mathbf{a}^{(0)}\mathbf{P}^{(2)}$, one can verify that

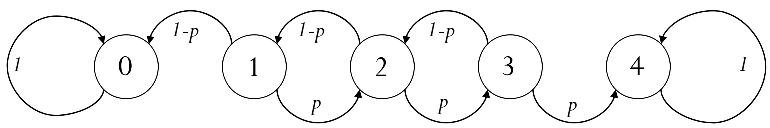
$\mathbf{a}^{(2)} = [0.315, 0.389, 0.296]$.

Classification of states and Markov chains

- In terms of their limiting behavior, there seem to be different types of Markov chains
- We will next discuss a useful way to classify different states and Markov chains

Communicating states

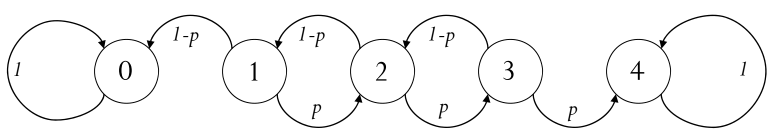
- We say that two states *communicate* if they are reachable from each other



  - States 1, 2, and 3 communicate with each other
  - States 0 and 4 do not communicate with any other state

Closed sets

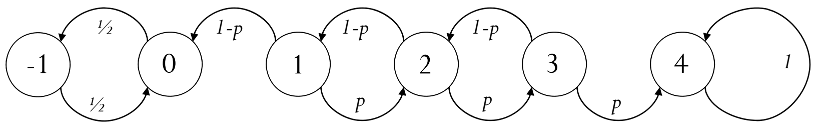
- We say that a set of states is *closed* if the Markov chain cannot leave that set of states



  - The sets {0} and {4} are closed sets

Closed sets – another example

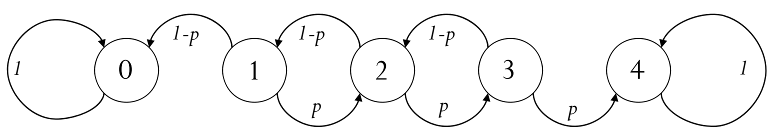
- A closed set can contain more than one state
  - Example:



  - The sets {-1,0} and {4} are closed sets

Absorbing states

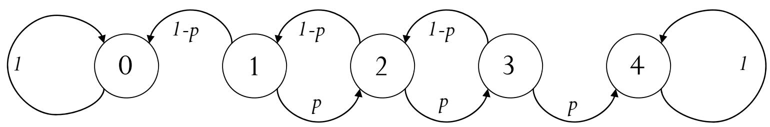
- Any state that we can never leave is called an *absorbing state*



  - The states 0 and 4 are absorbing states

Transient states

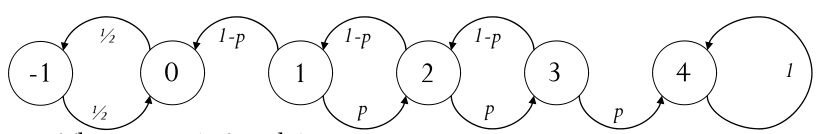
- If it is possible to leave a state and never return to it, that state is called a *transient state*



  - The states 1, 2, and 3 are transient states

Recurrent states

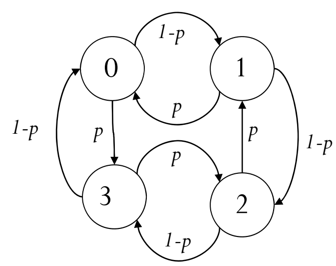
- Any state that is not transient is called a *recurrent state*



  - The states -1, 0 and 4 are recurrent states

(A)Periodic states

- If all paths leading from a state back to that state have a length that is a multiple of some number $k>1$ that state is called a *periodic state*
  - The largest such k is called the *period* of the state.



All states are periodic with period 2.

Ergodic Markov chains

- If all states in a Markov chain communicate with each other *and* are
  - Recurrent
  - Aperiodic

  then the chain is called *ergodic*
- Example:
  - Smallville's weather Markov chain

  but not:
  - Gambling Markov chain

Limiting behavior

- If a Markov chain is *ergodic*, then there exists some vector

such that $$\pi = \begin{pmatrix} \pi_1 & \pi_2 & \cdots & \pi_s \end{pmatrix}$$

$$\lim_{n\to\infty} P^n = \begin{pmatrix} \pi_1 & \pi_2 & \cdots & \pi_s \\ \pi_1 & \pi_2 & \cdots & \pi_s \\ \vdots & \vdots & & \vdots \\ \pi_1 & \pi_2 & \cdots & \pi_s \end{pmatrix}$$

Limiting behavior

- Interpretation:
  - After many transitions, the probability that we are in state i is approximately equal to π_i, *independent* of the starting state
  - The values π_i are called the *steady-state probabilities*, and the vector π is called the *steady-state (probability) distribution* or *equilibrium distribution*

Limiting behavior

- Can you explain why a steady-state distribution exists for ergodic Markov chains, but not for other chains?
  - Consider the case of
    - Periodic chains
    - non-communicating pairs of states
- In the following, we will assume we have an ergodic Markov chain

Limiting behavior

- How do we compute the steady-state probabilities π?
- Recall the following relationship between the multi-step transition probabilities:

$$P_{ij}(n+1) = \sum_{k=1}^{s} P_{ik}(n) \cdot p_{kj}$$

- The theoretical result says that, for large n

$$P_{ij}(n+1) \approx \pi_j$$
$$P_{ik}(n) \approx \pi_k$$

Limiting behavior

- So, for large n,

$$P_{ij}(n+1) = \sum_{k=1}^{s} P_{ik}(n) \cdot p_{kj}$$

becomes

$$\pi_j = \sum_{k=1}^{s} \pi_k p_{kj}$$

- In matrix notation,

$$\pi = \pi P$$

Limiting behavior

- But… $\pi = \pi P$
  - Has $\pi = 0$ as a solution…
  - In fact, it has an infinite number of solutions (because P is singular – why?)
  - Fortunately, we know that π should be a probability distribution, so not all solutions are meaningful!
    - We should ensure that

$$\sum_{i=1}^{s} \pi_i = 1$$

Example 3

- Recall the 1st Smalltown example:
  - 90% of all sunny days are followed by a sunny day
  - 80% of all cloudy days are followed by a cloudy day
  - Find the steady-state probabilities

$$P = \begin{pmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{pmatrix}$$

Example 3

- We need to solve the system

$$\begin{pmatrix} \pi_1 & \pi_2 \end{pmatrix} = \begin{pmatrix} \pi_1 & \pi_2 \end{pmatrix} \begin{pmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{pmatrix}$$
$$\pi_1 + \pi_2 = 1$$

or

$$\pi_1 = 0.9\pi_1 + 0.2\pi_2$$
$$\pi_2 = 0.1\pi_1 + 0.8\pi_2$$
$$\pi_1 + \pi_2 = 1$$

- We can ignore one of the first 2 equations! (Why?)

Example 3

- Solving

$$\pi_1 = 0.9\pi_1 + 0.2\pi_2$$
$$\pi_1 + \pi_2 = 1$$

gives $(\pi_1 \quad \pi_2) = (2/3 \quad 1/3)$

- Compare with

$$P^{20} \approx \begin{pmatrix} 0.67 & 0.33 \\ 0.67 & 0.33 \end{pmatrix}$$

First-Passage and First-Return probabilities

- By calculating $\mathbf{a}^{(n)}$, we answer the question of "what is the probability of being in a certain state at a certain time?" Another type of question which is frequently of interest is "how long will it take to reach a certain state?" This question has to do with the first passage (or return) to a given state. Specifically, let

$$f_{ij}^{(n)} = P\{X_n = j, X_{n-1} \neq j, X_{n-2} \neq j, \ldots, X_1 \neq j | X_0 = i\}$$

be the probability of the first passage from state i to state j in n steps. Then, it is clear that , and for $n > 1$,

$$f_{ij}^{(n)} = p_{ij}^{(n)} - \sum_{k=1}^{n-1} f_{ij}^{(k)} p_{jj}^{(n-k)}$$

- Thus, the $f_{ij}^{(n)}$ can be obtained iteratively if the $p_{ij}^{(n)}$ are known.
- Note that when $i=j$ we talk of first return instead of first passage.

- let **P** take the following numerical values.

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

Also, assume that the vector of initial conditions is given by $\mathbf{a}^{(0)}$ =[1/3, 1/3, 1/3].

- It is easy to verify that

$$P^{(2)} = P^2 = \begin{bmatrix} 0.500 & 0.333 & 0.167 \\ 0.333 & 0.389 & 0.278 \\ 0.111 & 0.444 & 0.444 \end{bmatrix}$$

- From the relationship $\mathbf{a}^{(2)} = \mathbf{a}^{(0)}\mathbf{P}^{(2)}$, one can verify that $\mathbf{a}^{(2)} = [0.315, 0.389, 0.296]$.

$$P^{(3)} = P^3 = \begin{bmatrix} 0.222 & 0.417 & 0.361 \\ 0.315 & 0.389 & 0.296 \\ 0.444 & 0.352 & 0.204 \end{bmatrix} \quad P^{(4)} = P^4 = \begin{bmatrix} 0.380 & 0.37 & 0.25 \\ 0.327 & 0.389 & 0.287 \\ 0.253 & 0.407 & 0.34 \end{bmatrix}$$

- Now, let $\mathbf{F}^{(n)}$ be the matrix whose elements are the . Then, from the recursive formula above, we have that
- $\mathbf{F}^{(2)} = \begin{bmatrix} 0.500 & 0.167 & 0.167 \\ 0.333 & 0.278 & 0.278 \\ 0.111 & 0.333 & 0.444 \end{bmatrix}$, $\mathbf{F}^{(3)} = \begin{bmatrix} 0.222 & 0.167 & 0.361 \\ 0.148 & 0.167 & 0.296 \\ 0.111 & 0.111 & 0.204 \end{bmatrix}$
- $\mathbf{F}^{(4)} = \begin{bmatrix} 0.1296 & 0.0556 & 0.0741 \\ 0.0864 & 0.0926 & 0.0957 \\ 0.0494 & 0.1111 & 0.1420 \end{bmatrix}$
- For instance, $f_{11}^{(2)} = p_{11}^{(2)} - p_{11}p_{11}$ =0.5,
- $f_{12}^{(2)} = p_{12}^{(2)} - p_{12}p_{22}$ =0.333 - 0.5*0.333= 0.167

Mean first passage time

- Consider an ergodic Markov chain
- Suppose we are currently in state *i*
  - What is the expected number of transitions until we reach state *j* ?
- This is called the *mean first passage time* from state *i* to state *j* and is denoted by m_{ij}
  - For example, in Smallville's weather example, m_{12} would be the expected number of days until the first cloudy day, given that it is currently sunny
- How can we compute these quantities?

Mean first passage time

- We are currently in state *i*
- In the next transition, we will go to some state *k*
  - If $k=j$, the first passage time from *i* to *j* is 1
  - If $k \neq j$, the mean first passage time from *i* to *j* is $1+m_{kj}$
  - So:

$$m_{ij} = p_{ij} \times 1 + \sum_{\substack{k=1 \\ k \neq j}}^{s} p_{ik}(1 + m_{kj})$$

$$= 1 + \sum_{\substack{k=1 \\ k \neq j}}^{s} p_{ik} m_{kj} = \sum_{k=1}^{s} p_{ik} + \sum_{\substack{k=1 \\ k \neq j}}^{s} p_{ik} m_{kj}$$

Mean first passage time

- We can thus find all mean first passage times by solving the following system of equations:

$$m_{ij} = \sum_{k=1}^{s} p_{ij} + \sum_{\substack{k=1 \\ k \neq j}}^{s} p_{ik} m_{kj} \qquad i = 1, \ldots, s;\ j = 1, \ldots, s$$

- $m_{ij} = \mu_{ij}$
- What is m_{ii}?
  - The mean number of transitions until we return to state i
  - This is equal to $1/\pi_i$!

- the *mean first passage time from state i to state j is computed as*

$$\mu_{ij} = \sum_{n=1}^{\infty} n f_{ij}^{(n)}$$

$$\|\mu_{ij}\| = (\mathbf{I} - \mathbf{N}_j)^{-1}\mathbf{1},\ j \neq i$$

where

$\mathbf{I} = (m - 1)$-identity matrix

$\mathbf{N}_j$ = transition matrix $\mathbf{P}$ less its jth row and jth column of target state j

$\mathbf{1} = (m - 1)$ column vector with all elements equal to 1

The matrix operation $(\mathbf{I} - \mathbf{N}_j)^{-1}\mathbf{1}$ essentially sums the columns of $(\mathbf{I} - \mathbf{N}_j)^{-1}$.

- the first passage time to a specific state from all others,
  consider the passage states 2 and 3 (fair and poor) to state 1 (good). Thus, $j = 1$

$$\mathbf{P} = \begin{pmatrix} .30 & .60 & .10 \\ .10 & .60 & .30 \\ .05 & .40 & .55 \end{pmatrix}$$

$$\mathbf{N}_1 = \begin{pmatrix} .60 & .30 \\ .40 & .55 \end{pmatrix}, (\mathbf{I} - \mathbf{N}_1)^{-1} = \begin{pmatrix} .4 & -.3 \\ -.4 & .45 \end{pmatrix}^{-1} = \begin{pmatrix} 7.50 & 5.00 \\ 6.67 & 6.67 \end{pmatrix}$$

- Thus

$$\begin{pmatrix} \mu_{21} \\ \mu_{31} \end{pmatrix} = \begin{pmatrix} 7.50 & 5.00 \\ 6.67 & 6.67 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 12.50 \\ 13.34 \end{pmatrix}$$

- This means that, on the average, it will take 12.5 seasons to pass from fair to good soil and 13.34 seasons to go from bad to good soil.

Example 3

- Recall the 1st Smalltown example:
  - 90% of all sunny days are followed by a sunny day
  - 80% of all cloudy days are followed by a cloudy day

$$P = \begin{pmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{pmatrix}$$

- The steady-state probabilities are $\pi_1 = 2/3$ and $\pi_2 = 1/3$

Example 3

- Thus:
  - $m_{11} = 1/\pi_1 = 1/(2/3) = 1½$
  - $m_{22} = 1/\pi_2 = 1/(1/3) = 3$
- And m_{12} and m_{21} satisfy:

$$\begin{aligned} m_{12} &= 1 + p_{11} m_{12} \qquad & m_{12} &= 1 + 0.9 m_{12} \\ m_{21} &= 1 + p_{22} m_{21} \qquad & m_{21} &= 1 + 0.8 m_{21} \end{aligned}$$

$$\begin{aligned} m_{12} &= 10 \\ m_{21} &= 5 \end{aligned}$$

Absorbing chains

- While many practical Markov chains are ergodic, another common type of Markov chain is one in which
  - some states are absorbing
  - the others are transient
- Examples:
  - Gambling Markov chain
  - Work-force planning

Example

- State College admissions office has modeled the path of a student through State College as a Markov chain
- States:
  - 1=Freshman, 2=Sophomore, 3=Junior, 4=Senior, 5=Quits, 6=Graduates
  - Based on past data, the transition probabilities have been estimated

Example

- Transition probability matrix:

$$\begin{array}{c c}
 & \begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \end{array} \\
\begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array} & \begin{pmatrix} .10 & .80 & 0 & 0 & .10 & 0 \\ 0 & .10 & .85 & 0 & .05 & 0 \\ 0 & 0 & .15 & .80 & .05 & 0 \\ 0 & 0 & 0 & .10 & .05 & .85 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}
\end{array}$$

Example

- Clearly, states 5 and 6 are absorbing states, and states 1-4 are transient states
- Given that a student enters State College as a freshman, how many years will be spent as freshman, sophomore, junior, senior before entering one of the absorbing states?

Example

- Consider the expected number of years spent as a freshman before absorption (quitting or graduating)
  - Definitely count the first year
  - If the student is still a freshman in the 2^{nd} year, count it also
  - If the student is still a freshman in the 3^{rd} year, count it also
  - Etc.

Example

- Recall the transition probability matrix:

matrix Q matrix R

$$P = \left(\begin{array}{cccc|cc} .10 & .80 & 0 & 0 & .10 & 0 \\ 0 & .10 & .85 & 0 & .05 & 0 \\ 0 & 0 & .15 & .80 & .05 & 0 \\ 0 & 0 & 0 & .10 & .05 & .85 \\ \hline 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right)$$

matrix I

Example

- The number of years spent in states 1-4 when entering State College as a freshman can now be found as the first row of the matrix $(I\text{-}Q)^{-1}$
- If a student enters State College as a freshman, how many years can the student expect to spend there?
  - Sum the elements in the first row of the matrix $(I\text{-}Q)^{-1}$

Example

- In this example:

$$(I - Q)^{-1} \approx \begin{pmatrix} 1.11 & 0.99 & 0.99 & 0.88 \\ 0 & 1.11 & 1.11 & 0.99 \\ 0 & 0 & 1.18 & 1.05 \\ 0 & 0 & 0 & 1.11 \end{pmatrix}$$

- The expected (average) amount of time spent at State College is: 1.11 + 0.99 + 0.99 + 0.88 = 3.97 years

Example

- What is the probability that a freshman eventually graduates?
- We know the probability that a starting freshman will have graduated after n years from the n-step transition probability matrix
  - It's the (1,6) element of P^n !
- The probability that the student will eventually graduate is thus
  - the (1,6) element of $\lim_{n\to\infty} P^n$!

Example

$$\lim_{n\to\infty} P^n = \begin{pmatrix} \lim_{n\to\infty} Q^n & \left(\lim_{n\to\infty} \sum_{\ell=0}^{n-1} Q^\ell\right) R \\ 0 & I \end{pmatrix}$$

$$= \begin{pmatrix} 0 & (I-Q)^{-1} R \\ 0 & I \end{pmatrix}$$

Example

- So the probability that a student that starts as a freshman will eventually graduate is
  - the (1,6-4=2) element of $(I\text{-}Q)^{-1}R$
- In the example:

$$(I-Q)^{-1}R \approx \begin{pmatrix} 0.25 & \boxed{0.75} \\ 0.16 & 0.84 \\ 0.11 & 0.89 \\ 0.06 & 0.94 \end{pmatrix}$$