Chapter 1 - Definitions & Examples

Definition: Stochastic process

A family of random variables \( \{ X_i \}_{i \in I} \) where \( i \) is an index set, usually denoting time (so \( I = \mathbb{N} \cup \{0\} \) or \( \mathbb{R}_+ \))

First Part of the Course - Discrete time

A s.p. is now \( \{ X_n \}_{n \in \mathbb{N}} \)

- each \( n \in \mathbb{N} \), \( X_n \) is a random variable, i.e., a function from probability space \( \Omega \) into a state space \( S \)

- \( S \) can be: Finite / Infinite \( \leq \) Countable / Uncountable

Examples of Stochastic Process

- Random Walk, sequence of i.i.d.s.
- Example (Closing Value of Stock)

Example

- Stock Awesome.com.cy has a value after the end of each day, when the market closes.
- What are the questions of interest?

- Long term behavior?
- Is day-to-day information enough?
- Does information we possess matters?

Answer to last question? YES! History affects probabilities!
Conditional probability: We use information we already have

- Event $B$, $P(B) > 0$ (necessary)
- We have knowledge of event $B$ - we know it happened
- Q: What is the prob. of something else happening?

Then

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{P(A|B)}{P(B)} = \frac{P(A, B)}{P(B)}$$

Disconnected events: (= Mutually exclusive)

(i.e. Sunny vs Rainy) $P(AB) = 0 \Rightarrow P(A|B) = 0$

Independent events: $P(AB) = P(A)P(B)$

(information about A is useless in terms of info about B)

$$\Rightarrow P(A|B) = P(A)$$

Lecture 23/08 End

The three tools of conditioning

1. Multiplication Rule:

$$P(A_1 \cdots A_n) = P(A_1)P(A_2|A_1)P(A_3|A_2A_1) \cdots P(A_n|A_{n-1}A_1)$$

(This is part of the first hw)

Example:

There are $N$ keys in a basket and only one opens a door.
What is the probability that the door will open at the $n$-th trial? ($1 \leq n \leq N$)

Define $A_i = \{i$-th key did not open door$\}$

We want $P(A_1 \cdots A_{n-1} A_n^c) = \frac{1}{N}$ (why?)

Then, by the multiplication rule,

$$P(A_1 \cdots A_{n-1} A_n^c) = \frac{1}{N} \cdot \frac{N-1}{N-1} \cdot \frac{N-2}{N-2} \cdots \frac{N-n}{N-n} = \frac{1}{N}$$ (Wow!)
y? Use permutations of the N keys!

\[ J \ (N-1)! \quad \text{that place the correct key at the n-th position} \]

\[ N! \quad \text{all perms.} \]

\[ \Rightarrow P\{A_i \ldots A_{n-1} A_n\} = \frac{(N-1)!}{N!} = \frac{1}{N} \]

3). Law of total probability.

\[ \{B_i\}_{i \in N}, P\{B_i\} > 0 \quad \forall i \ (\text{finite or countably many}) \]

\[ \text{and so that} \quad P\{B_i \cap B_j\} = 0 \quad \text{if} \ i \neq j \quad \text{and} \quad \sum_{i=1}^{\infty} P\{B_i\} = 1 \]

They partition \( \Omega \).

\[ P\{A\} = \sum_{i=1}^{\infty} P\{A \cap B_i\} \]

\[ \text{why?} \quad \sum_{i=1}^{\infty} P\{A \mid B_i\} P\{B_i\} \]

C) Bayes Formula (aposteriori probability)

We know an event happened. What is the prob. that something else happened before it??

\[ P(A \mid B) = \frac{P\{A \cap B\}}{P(B)} = \frac{P\{A \cap B\}}{P(B)} \cdot \frac{P\{A\}}{P\{B\}} = \frac{P\{B \mid A\} \cdot P(A)}{P(B)} \]

longer version

\[ P(A \mid B) = \frac{\sum_{j} P\{B \mid A_j\} \cdot P\{A_j\}}{\sum_{j} P\{B \mid A_j\} \cdot P\{A_j\}} \]

What is this? What properties satisfy \( A_i \)?
Chapter 1 - Finite Markov Chains

Definitions & Examples

- Discrete time stochastic process, \( \{X_n\}_{n \in \mathbb{N}} \). Index \( n \) denotes time (or time step).
- State space \( S \) is finite for this chapter!

The most natural thing to evaluate is the probabilities of the process.

\[
(\dagger) \quad P\{X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n\}
\]

\( \forall n \in \mathbb{N}, \forall (i_0, \ldots, i_n) \in S^{n+1} \)

Probability (\( \dagger \)) is an intersection of events, so, by the multiplication rule,

\[
(\dagger) = P\{X_0 = i_0\} \cdot P\{X_1 = i_1 | X_0 = i_0\} \cdots P\{X_n = i_n | X_{n-1} = i_{n-1}, \ldots, X_0 = i_0\} \quad (\star\star)
\]

Initial distribution \quad Transition probabilities

Can we simplify (\( \star\star \))? \textbf{NOT ALWAYS}

the process is \underline{Markov} then it simplifies greatly:

\underline{Markov Property} (Restarting!)

- In normal words: In order to make predictions for the future of the system, all you need to know is its present state (not its past history).

- In mathy words: Conditional on the present, past and future are independent.

- In mathy symbols:

\[
P\{X_n = i_n | X_0 = i_0, \ldots, X_{n-1} = i_{n-1}\} = P\{X_n = i_n | X_{n-1} = i_{n-1}\}
\]
A **Markov chain** (MC) is a stochastic process that satisfies the Markov Property.

A **time-homogeneous** MC satisfies

\[ P \{ X_n = i_n \mid X_0 = i_0, \ldots, X_{n-1} = i_{n-1} \} = P(i_{n-1}, i_n) \]

for some function \( P : \mathbb{S} \times \mathbb{S} \rightarrow [0,1] \) that does not depend on \( n \).

This implies:

\[ P \{ X_{n+k} = i \mid X_{n+k-1} = j \} = P(j,i) \]

\( \forall k \geq 1 \)

\( \forall n \geq 0 \).

The **initial distribution** \( \phi \) gives the starting positions:

\[ P \{ X_0 = i \} = \phi(i) \]

which can be simplified as:

\[ P \{ X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n \} = \phi(i_0) P(i_0,i_1) P(i_1,i_2) \ldots P(i_{n-1},i_n) \]

**Particularization**

Three-state MC:

\[ \begin{array}{c}
\frac{1}{3} \\
A \\
\frac{1}{3} \\
B \\
\frac{2}{3} \\
C \\
\frac{1}{2} \\
\end{array} \]

- **Directed graph**
- **Is it a MC?** (Yes, why?)
- \( P(B,C) = \frac{1}{2} \)

**How to organise?** From each state, you go to a different state.

- **Two coordinates!**

\[ \begin{bmatrix}
\frac{1}{3} & 0 & \frac{1}{3} \\
\frac{2}{3} & \frac{1}{2} & 0 \\
0 & \frac{1}{4} & \frac{3}{4} \\
\end{bmatrix} = P \]

This is called a **transition matrix** (i.e., matrix of transition probabilities). It's a **stochastic matrix**. The rows sum up to 1 (why?).
In general, any directed graph, where the sum of outbounds probability sum to 1, can be turned into a MC. (Why?) Because a MC is uniquely completely defined by the transition probabilities, which in turn can be defined by the transition matrix.

$$P = \begin{bmatrix} p(x_1,x_1) & \cdots & p(x_1,x_N) \\ \vdots & \ddots & \vdots \\ p(x_N,x_1) & \cdots & p(x_N,x_N) \end{bmatrix}$$

**Stochastic matrix** $P$

Entries $P_{ij} = p(i,j)$, $0 \leq P_{ij} \leq 1$ (why?)

**Stochastic** $\sum_{j=1}^{N} P_{ij} = 1$ (row sum) (why?)

Anything satisfying $1 \times 2$ is a transition matrix for a MC.

**Example**

1. Two-state MC - A (highly unrealistic) weather model.

Let $X_n = \begin{cases} 0 \text{ rainy} \\ 1 \text{ not rainy} \end{cases}$ during the $n$-th day of the measurement.

Suppose that at the end of each day, I prob. $p$ to switch from rainy to not rainy, and a prob. $q$ to switch from not rainy to rainy.
The chain can be represented by

\[
\begin{pmatrix}
1-p & p \\
q & 1-q
\end{pmatrix}
\]

Finite Capacity queueing system

If you are in a bank, where the teller serves at most one person in a given time interval with prob. \( p \). If the queue length in front of a teller is two, no more people may enter the queue. If there is room, at most one client attempts to join the queue with prob. \( q \) during each time interval. Is this a NC?

Find the stochastic matrix. (the transition matrix)

\[
\begin{pmatrix}
0 & 1 & 2 \\
1 & 1-(1-p)q-(1-q)p & p \\
2 & (1-p)q & 1-p
\end{pmatrix}
\]

Gambler’s Ruin Problem (Random walk with absorbing boundaries)

\( p + p + q + 1 = 1 \). For \( 1 \leq x \leq N \), \( p(x, x) = 1 \), \( p(x, x+1) = p \), \( p(x, x-1) = q \).
Why are the transition matrices useful?

We need to find \[ P \{ X_n = j \mid X_0 = i \} \] (look deep into the future!

The only information we have are \( \phi \) (initial distribution) and \( P \) (the tr.m.)

\((**)\) is called the \( n \)-step transition.

\[ p_n(i,j) = P \{ X_n = j \mid X_0 = i \} = P \{ X_{n+k} = j \mid X_k = i \} \]

Assume \( \bar{\phi} = (\phi(0), \ldots, \phi(N)) \) i.d.

Then, by law of total probability

\[ p_n(i,j) = \sum_{i \in S} \phi(i) \cdot \left[ P \{ X_n = j \mid X_0 = i \} \right] \]

Need to know this!

\[ \text{aim}: p_n(i,j) = (P^n)_{i,j} \quad \text{the} \ (i,j)\ \text{entry of matrix} \ P \ \\text{r} \ \\
\text{roof}: \text{We use induction.} \ \\
\text{For} \ n=1 \ \text{true!} \ \\
\text{Assume} \ P_{n \otimes} (i,j) = (P^{n \otimes})_{i,j} \ \forall (i,j). \ \text{We'll show it for} \ n+1. \\
\text{Then} \ P \{ X_{n+1} = j \mid X_0 = i \} \stackrel{\text{why}}{=} \sum_{k \in S} \left[ P \{ X_n = k \mid X_0 = i \} \cdot P \{ X_{n+1} = j \mid X_n = k \} \right] \\
= \sum_{k \in S} p_n(i,k) \cdot p(k,j) \\
\text{why} = (P^n \cdot P)_{i,j} \ \\
\text{so, if} \ \phi_0 = (\phi_0(1), \ldots, \phi_0(N)) \Rightarrow \bar{\phi}_n = \phi_0 \cdot P^n.
The $n$-th step transition $P\{X_n = j \mid X_0 = i\} = p_n(i,j)$ is the $ij$-th coordinate of the matrix $P^n$. $P$ is transition matrix.

$$P\{X_n = j\} = \sum_{k=1}^{\infty} p_n(k,j) \cdot P\{X_0 = k\}$$

Let $\Phi_0$ be the vector $[\phi_0(1), \ldots, \phi_0(N)]$ that gives the initial distribution of the chain. Then (4) becomes (assuming $\Phi_n = [P\{X_n = 1\}, P\{X_n = 2\}, \ldots, P\{X_n = N\}]$)

$$\Phi_n = \Phi_0 \cdot P^n$$

**Example (Two-state MC)**

Let $P = \begin{bmatrix} 3/4 & 1/4 \\ 1/6 & 5/6 \end{bmatrix}$, Let $\Phi_0 = [1, 0]$ (⇒ Meaning?)

What is the probability that during the 6-th day it is sunny?

Answer

$$P^6 = \begin{bmatrix} 3/4 & 1/4 \\ 1/6 & 5/6 \end{bmatrix}^6 = \begin{bmatrix} 0.424 & 0.576 \\ 0.384 & 0.616 \end{bmatrix}$$

Then $\Phi_6 = \Phi_0 \cdot P^6 = \begin{bmatrix} 1 \ 0 \end{bmatrix} \begin{bmatrix} 0.424 & 0.576 \\ 0.384 & 0.616 \end{bmatrix} = \begin{bmatrix} 0.424 \\ 0.576 \end{bmatrix}$.

Now, assume you want to settle a city. You are the first settler and you observe this weather pattern! What is the probability that it will be sunny or rainy in two years time?
Two ways to do it.
1) Using Linear algebra
2) Using probabilistic method + finite sums!

Method 2. (Becomes really difficult when state space becomes bigger)

\[
\begin{bmatrix}
0 & 1 \\
1-a & a \\
1 & b & 1-b \\
\end{bmatrix}
\]

We are interested in finding \( \mathbb{P}\{X_n = 0\} \)

law of total probability:

\[
\mathbb{P}\{X_{n+1} = 0\} = \mathbb{P}\{X_{n+1} = 0, X_n = 0\} + \mathbb{P}\{X_{n+1} = 0, X_n = 1\}
\]

\[
= \mathbb{P}\{X_{n+1} = 0 \mid X_n = 0\} \mathbb{P}\{X_n = 0\} + \mathbb{P}\{X_{n+1} = 0 \mid X_n = 1\} \mathbb{P}\{X_n = 1\}
\]

\[
= (1-a) \mathbb{P}\{X_n = 0\} + b \mathbb{P}\{X_n = 0\} \cdot (1-a) + b \mathbb{P}\{X_n = 1\}
\]

\[
\mathbb{P}\{X_{n+1} = 0\} = b + (1-a\cdot b) \mathbb{P}\{X_n = 0\}
\]

Index lowered by 1 \( \implies \) Re-iterate!

\[
= b + (1-a\cdot b) \left[ b + (1-a\cdot b) \mathbb{P}\{X_{n-1} = 0\} \right]
\]

\[
= b + b(1-a\cdot b) + (1-a\cdot b)^2 \mathbb{P}\{X_{n-1} = 0\}
\]

\[
= b + b(1-a\cdot b) + \cdots + b(1-a\cdot b)^n + (1-a\cdot b)^{n+1} \mathbb{P}\{X_0 = 0\}
\]

\[
= b \left[ \frac{1 - (1-a\cdot b)^{n+1}}{1-(1-a\cdot b)} \right] + (1-a\cdot b)^{n+1} \mathbb{P}\{X_0 = 0\}
\]

\[
\mathbb{P}\{X_{n+1} = 0\} = \frac{b}{a+b} + (1-a\cdot b)^{n+1} \left[ \mathbb{P}\{X_0 = 0\} - \frac{b}{a+b} \right]
\]
Assume further that \( x + y = 1 \iff x = \frac{b}{a+b}, \quad y = \frac{a}{a+b}. \)

So now, \( 0 \leq x \leq 1, \quad 0 \leq y \leq 1, \quad x + y = 1 \) and \((x, y) = \left[ \frac{b}{a+b}, \frac{a}{a+b} \right] \).

Moreover,

\[
\begin{bmatrix}
\frac{b}{a+b} & \frac{a}{a+b}
\end{bmatrix} P = \begin{bmatrix}
\frac{b}{a+b} & \frac{a}{a+b}
\end{bmatrix}
\begin{bmatrix}
\frac{b}{x+b} & \frac{a}{x+b}
\end{bmatrix}
= \begin{bmatrix}
\frac{b}{x+b} & \frac{a}{x+b}
\end{bmatrix}
\begin{bmatrix}
\frac{b}{a+b} & \frac{a}{a+b}
\end{bmatrix}
\]

What happens if \( \Phi_0 = \begin{bmatrix}
\frac{b}{x+b} & \frac{a}{x+b}
\end{bmatrix} \)? Nothing! That's why it's called invariant distribution.

**Natural questions**

Do we always have an invariant distribution? No, what if 1-eigenvector has negative entries?

If we have one, is it unique? Not necessarily, what if 1 is not a simple eigenvalue?

Do we care about other eigenvalues? Yes! Here is why...

**Long time behavior for stochastic matrix**

\[
P = \begin{bmatrix}
1 - R & a \\
b & 1 - b
\end{bmatrix}
\]

Can we say something about \( \lim_{n \to \infty} P^n \)?

The characteristic polynomial \( p(\lambda) = \det(P - \lambda I) \) is \( \begin{vmatrix}
1 - R - \lambda & R \\
b & 1 - b - \lambda
\end{vmatrix}
\)

\( = (1-R-\lambda)(1-b-\lambda) - ab \)

\( = \lambda^2 - (1-R+1-b)\lambda + (1-a)(1-b) \)

\( = \lambda^2 - (2-a-b)\lambda + 1 - (1-a-b) \)

\( = (\lambda-1)(\lambda-(1-a-b)) \)
\( n \lim_{n \to \infty} P\{X_{n+1} = 0\} = \frac{b}{a+b}. \) \hspace{1cm} \text{(no info about } \phi_0 \text{ required!)}

Repeat for \( P\{X_{n+1} = 1\} = 1 - P\{X_{n+1} = 0\} \Rightarrow \lim_{n \to \infty} P\{X_{n+1} = 1\} = \frac{a}{a+b}. \)

Method 1: Works in general and is computationally easy!

Reminder: An eigenvalue of a real matrix \( A \) is a complex number \( \lambda \) s.t.

\[ \text{If real vectors } \vec{v}, \vec{u}, \text{ so that } \vec{v} A = \lambda \vec{v} \text{ or } A \vec{u} = \lambda \vec{u} \]

(\( \forall \vec{v} \text{ is called left (or right) eigenvector).} \)

\[ \text{Eigenvalues are roots of the polynomial } \det(A-xI) = 0 \]

Any stochastic matrix has a real eigenvalue \( \lambda = 1 \).

Why?

\[
\begin{bmatrix}
    a_{11} & \cdots & a_{1n} \\
    \vdots & \ddots & \vdots \\
    a_{n1} & \cdots & a_{nn}
\end{bmatrix}
\begin{bmatrix}
    1 \\
    \vdots \\
    1
\end{bmatrix}
= \begin{bmatrix}
    \sum_{j=1}^{n} a_{1j} \\
    \vdots \\
    \sum_{j=1}^{n} a_{nj}
\end{bmatrix}
\begin{bmatrix}
    1 \\
    \vdots \\
    1
\end{bmatrix}
\]

\[ \begin{bmatrix}
    1 \\
    \vdots \\
    1
\end{bmatrix} \text{ is a right eigenvector.} \]

Let's find a left \( 1 \)-eigenvector for \[ \begin{bmatrix}
    1-a & a \\
    b & 1-b
\end{bmatrix} \]

Assume the eigenvector is \((x, y)\). Then:

\[
\begin{bmatrix}
    1-a & a \\
    b & 1-b
\end{bmatrix} \begin{bmatrix}
    x \\
    y
\end{bmatrix} = \begin{bmatrix}
    x \\
    y
\end{bmatrix} \iff \begin{cases}
    (1-a)x + by = x \\
    ax + (1-b)y = y
\end{cases}
\]

\[ \iff \begin{cases}
    -ax + by = 0 \\
    ax - by = 0
\end{cases} \iff \begin{cases}
    ax = by \\
    y = \frac{a}{b} x
\end{cases} \iff \frac{y}{x} = \frac{a}{b} \implies \begin{cases}
    (x, y) = x \left( 1, \frac{a}{b} \right), \forall x \in \mathbb{R} \text{ Which eigenvector to choose?}
\end{cases}
\]
eigenvalues $\lambda = 1$ (yay!) $\lambda = 1 - q - q b$. Note that $|1 - q - q b| < 1$.

raise $P$ to powers, first diagonalize.

$$D = \begin{bmatrix} 1 & 0 \\ 0 & 1 - q - q b \end{bmatrix} = Q^{-1} P Q,$$

where $Q$ needs to be specified.

The columns of the matrix $Q$ are right-eigenvectors, corresponding to the ordering of the eigenvalues.

So $Q = \begin{bmatrix} 1 \\ -a \\ 1 \\ b \end{bmatrix}$ (why?) $\implies Q^{-1} = \begin{bmatrix} b & a \\ a + b & 1 \\ -1 & b \\ a + b & a \end{bmatrix}$

Then $P^n = (Q D Q^{-1})^n = Q D^n Q^{-1} = Q \begin{bmatrix} 1 & 0 \\ 0 & (1 - q - q b)^n \end{bmatrix} Q^{-1}

= \begin{bmatrix} b(1 - q - q b)^n \\ a + b \\ b - b(1 - q - q b)^n \\ a + b \end{bmatrix} \implies \lim_{n \to \infty} P^n = \begin{bmatrix} \frac{1}{n} \\ \frac{1}{n} \end{bmatrix}$

Note $\pi$ is invariant distribution!

Question:
Do all stochastic matrices satisfy $\lim_{n \to \infty} P^n = \begin{bmatrix} \frac{1}{n} \\ \frac{1}{n} \end{bmatrix}$?

Answer: No.

What can fail? Three possible problems!

1) Cycle $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ $\implies P^n = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} P^n = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$

2) $P^2 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} I$; $P^3 = P^2 P = P$, $P^4 = (P^2)^2 = I$...

$$\left\{ \begin{array}{c} P^{2n+1} = P \\ P^{2n} = I \end{array} \right.$$
What makes the "theorem" fail?

We can predict the position accurately according to steps.

Periodicity (periodicity implies a lot of zeros in high powers).

\[
P = \begin{bmatrix}
\frac{1}{2} & \frac{1}{2} & 0 \\
\frac{1}{2} & \frac{1}{2} & 0 \\
0 & \frac{1}{2} & \frac{1}{2}
\end{bmatrix} \quad P \cdot P \cdot P \rightarrow P^n = \begin{bmatrix}
\frac{1}{2} & \frac{1}{2} & 0 \\
\frac{1}{2} & \frac{1}{2} & 0 \\
0 & \frac{1}{2} & \frac{1}{2}
\end{bmatrix} \neq \begin{bmatrix}
\frac{\pi}{4} & \frac{\pi}{4} \\
\frac{\pi}{4} & \frac{\pi}{4}
\end{bmatrix}
\]

What makes the "theorem" fail?

\[
\begin{bmatrix}
A & 0 \\
0 & B
\end{bmatrix} \quad \text{st. matrix.}
\]

Chain:

\[
\begin{array}{c}
\circ \quad 2 \quad \circ \\
\circ \quad 3 \quad \circ
\end{array}
\]

\[\text{no communication!}\]

Chain can be broken down to smaller chains! (⇒ Reducibility)

Again, there are plenty of zeroes!

\[
\begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 \\
\frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 \\
0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\
0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\
0 & 0 & 0 & \frac{1}{2} & 0 & 1
\end{bmatrix} \Rightarrow P^k \approx \begin{bmatrix}
1 & 0 & 0 & 0 & 0 \\
\frac{3}{4} & \frac{1}{4} & 0 & \frac{1}{2} & 0 \\
\frac{1}{2} & \frac{1}{2} & 0 & \frac{1}{4} & \frac{1}{4} \\
\frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\
0 & 0 & 0 & 0 & 1
\end{bmatrix}
\]

Random walk with absorbing boundaries!

the random walk ever hits 1 or 5 it will be stuck!

there are useless states in the limit (⇒ Transience). Again, look all the zeroes "\_".

\textbf{Fact:} If \( P \) is a stochastic matrix, such that for some \( n \)

\( P^n \) has all positive entries then the left eigenvector \( \zeta \) can be chosen with non-negative entries and the eigenvalue \( 1 \) is simple and all other eigenvalues are in absolute value less than 1.

\textbf{END OF LECTURE}
Section 13: Classification of States

Frobenius Thm.

Suppose $P$ is a stochastic matrix with all entries strictly positive. Then 1 is a simple eigenvalue, there exists an invariant distribution $\pi$ (hence unique) and all other eigenvalues have absolute value strictly less than 1.

$$\Rightarrow \lim_{n \to \infty} P^n = [\pi \pi \ldots]$$

What conditions are sufficient so that $P^n$ has all positive entries or $n$ sufficiently large?

- Anything that overcomes
  - (A) Reducibility
  - (B) Periodicity
  - (C) Transience

Reducibility:

Problem: The chain could split into several pieces.

The states don't "communicate".

Definition:

Two states $i$ and $j$ communicate with each other (written $i \rightarrow j$) if there exist $m, n > 0$ such that $P^m(i, j) > 0$ and $P^n(j, i) > 0$

i.e.

\[
\text{1 closed circuit that contains } i \text{ and } j \text{ and can travel from } i \text{ to } j
\]

The relation $\leftrightarrow$ is

- i) reflexive: $i \leftrightarrow i$ ($P_0(i, i) > 0$)
- ii) symmetric $i \leftrightarrow j \Rightarrow j \leftrightarrow i$
- iii) transitive $i \leftrightarrow k, k \leftrightarrow j \Rightarrow i \leftrightarrow j$
Proof of transitivity:

Assume \( P_{m_1}(i,j) > 0 \), \( P_{m_2}(j,k) > 0 \). Then

\[
P_{m_1 + m_2}(i,k) = P \left\{ X_{m_1 + m_2} = k \mid X_0 = i \right\} \geq P \left\{ X_{m_1 + m_2} = k, X_{m_2} = j \mid X_0 = i \right\} \quad \text{(why?)}
\]

\[
L.T.P \quad = \quad \frac{P \left\{ X_{m_1 + m_2} = k \mid X_{m_1} = j, X_0 = i \right\} \cdot P \left\{ X_{m_1} = j \mid X_0 = i \right\}}{P \left\{ X_{m_1} = j \mid X_0 = i \right\}} \quad \text{(why?)}
\]

\[
N.P \quad = \quad \frac{P \left\{ X_{m_1 + m_2} = k \mid X_{m_1} = j \right\} \cdot P \left\{ X_{m_1} = j \mid X_0 = i \right\}}{P \left\{ X_{m_1} = j \mid X_0 = i \right\}} \quad \text{(why?)}
\]

\[
T.H. \quad = \quad \frac{P \left\{ X_{m_2} = k \mid X_0 = j \right\} \cdot P \left\{ X_{m_1} = j \mid X_0 = i \right\}}{P \left\{ X_{m_1} = j \mid X_0 = i \right\}} \quad \text{(why?)}
\]

\[
P_{m_1}(i,j) \cdot P_{m_2}(j,k) > 0.
\]

The relation \( \leftrightarrow \) is an equivalence relation.

It partitions the state space into communication classes: \( S/\leftrightarrow \)

Any matrix satisfying the Perron-Frobenius theorem must have

only 1 communication class. Such a matrix is called

irreducible. (i.e. \( \forall (i,j) \exists \lambda \in \mathbb{R} \quad P_n(i,j) > 0, P_m(i,j) > 0 \).)

The communication classes are called

Transient, if the chain eventually leaves and never returns.

Recurrent, if the chain keeps visiting each state in the class \( \infty \) many times.

Example:

Berliner's ruin problem. (Fair game)

\[
P = \begin{bmatrix}
0 & 1/2 & 1/2 & 0 & 1/2 \\
1/2 & 0 & 1/2 & 0 & 1/2 \\
1/2 & 1/2 & 0 & 1/2 & 0 \\
0 & 1/2 & 1/2 & 0 & 1/2 \\
0 & 0 & 0 & 1/2 & 1/2
\end{bmatrix}
\]
Three communication classes!

\( \{0, 5\}, \{4, 3\}, \{1, 2, 3\} \)  

\[
\begin{bmatrix}
0 & 1/2 & 1/2 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 \\
2 & 1/4 & 1/4 & 1/4 & 0 & 0 \\
3 & 0 & 0 & 1/4 & 0 & 0 \\
4 & 0 & 0 & 0 & 1/4 & 0 \\
5 & 0 & 0 & 0 & 0 & 1/4
\end{bmatrix}
\]

Three communication classes.

\( \{0, 1, 5\}, \{2, 4\}, \{3\} \)

Rec. Rec. T.

Chain that starts in a recurrent class always stays in it!

\(15\) Periodicity:

Assume \( P \) is the matrix for an irreducible MC (otherwise work with each block).

Definition: A period of a state \( i \), \( d = d(i) \) is the largest common divisor of the integer set:

\[
\mathcal{J}_i = \left\{ n \geq 0 : p_n(i, i) > 0 \right\}
\]
Properties of $J_i$

If $m, n \in J_i$ then $m + n \in J_i$ and $P_{m+n}(i, i) \geq P_m(i, i) P_n(i, i) > 0$.

If $d$ is the greatest lower bound of $J_i$, then $d \in \{0, d, 2d, \ldots\}$.

Exercise 1.2.1 (Readers are expected to prove)

Define $D_i = \{x \in \mathbb{N}_0 \mid d \leq x \leq J_i\}$. Then $\gcd(D_i) = 1$.

Thus there exist $m, n \in D_i$ with $\gcd(m, n) = 1$.

Euclid's algorithm implies that there exist $\alpha_0, \beta_0 \in \mathbb{Z}$ such that $m\alpha_0 + n\beta_0 = 1$.

Now consider the set $A = \{mx + ny \mid x, y \in \mathbb{N}_0\}$.

Because $J_i$ (and by extension $D_i$) are closed under addition, $A \subseteq D_i$. We'll show (*) for the set $A$.

First note that there are two consecutive integers in $A$.

Let $x > |\alpha_0|, y > |\beta_0| \in \mathbb{N}$. Then

$N = mx + ny \in A$ and $mx + ny + 1 = m(x + \alpha_0) + n(y + \beta_0) = N + 1$.

We'll show that all numbers $> N^2$ are in $A$.

Let $m > N^2$ and write $m - N^2 = kN + r, 0 \leq r < N$.

Then $m = r + N^2 + kN = r(N + 1) + (N - r + k)N \in A \subseteq D_i \subseteq \mathbb{N}_0$.
The period is a class property.

Let \( P \) be irreducible (so only one communication class).

Let \( n,m \) such that \( P_n(i,j) > 0 \), \( P_m(j,i) > 0 \), and let \( d = \gcd \{ J_i \} \). We'll show that \( d = \gcd \{ J_i \} \).

\[ P_{m+n}(i,i) \geq P_n(i,j) P_m(j,i) > 0 \Rightarrow m+n \in J_i \Rightarrow d \mid m+n. \]

\[ P_{m+n}(j,j) \geq P_m(j,i) P_n(i,j) > 0 \Rightarrow m+n \in J_j \Rightarrow \]

or any other \( l \in J_j \) , \( m+n+l \in J_j \)

\[ P_{m+n+l}(i,i) \geq P_n(i,j) P_l(j,j) P_m(j,i) > 0 \Rightarrow m+n+l \in J_i \Rightarrow \]

\[ d \mid m+n+l = d \mid l . \Rightarrow d \leq \gcd \{ J_j \} \]

Similarly, interchanging the roles of \( i,j \) we have \( d \leq \gcd \{ J_i \} \). Therefore, \( d \leq \gcd \{ J_j, J_i \} \).

3.3 Irreducible, a periodic chains

\textit{A Warning:} \( P \) is aperiodic iff \( d=1 \). Then, \( \exists ! \) a unique invariant probability vector \( \pi \) satisfying \( \pi^T P = \pi^T \).

Moreover, if \( \pi \) is any initial probability vector

\[ \lim_{n \to \infty} \pi P^n = \pi \]

Also \( \pi(i) > 0 \) \( \forall i \)
Theorem (No proof)

If $P$ is irreducible with period $d$, $P$ will have $d$ eigenvalues with absolute value 1, the $d$ complex numbers $z$, s.t. $|z|^d = 1$

Each is simple.

In particular, 1 is simple and there exists a unique invariant prob. $\pi$.

Given any initial distribution $\bar{q}$, for large $n$ $\bar{q}P^n$ will cycle through $d$ different distributions but they will average to $\pi$

$$\lim_{n \to \infty} \frac{1}{d} \left[ \bar{q}P^{n+1} + \ldots + \bar{q}P^{n+d} \right] = \pi$$

$\pi$ does not represent the limit $\lim_{n \to \infty} P^n(j,i) / \pi(i)$

It still is the average proportion of time spent on each state!