Applied Probability Notes

Andrew Martin

June 27, 2018

Lectured by Lewis Mitchell < 3

Contents

1 Probability Review **2**

2 Intro **2**

3 Probability and Measure **2**
3.1 Laws of large numbers . . . 2
3.2 Smallest σ-algebra . . . 4
3.3 Euclidean Borel field . . . 4

4 Discrete Time Markov Chains **7**
4.1 Basic Definitions . . . 7
4.1.1 Analysis of the two gamblers . . . 11
4.2 Hitting probabilities and hitting times . . . 12
4.3 Expected Hitting time . . . 14
4.4 Classification of states . . . 15

5 Branching processes **21**
5.1 Periodicity . . . 25
5.2 Limiting behaviours . . . 26
5.3 Positive and null recurrence . . . 27
5.4 Solidarity Summary . . . 28
5.5 Global balance equations . . . 31
5.6 Long term behaviour . . . 33
5.6.1 Partial Balance Equations . . . 33
5.7 Time Reversed Markov Chains . . . 35

6 Martingales **39**
6.1 Definition . . . 39
6.2 Almost sure events and events of probability 0 . . . 39
6.3 Stopping times and the Optimal stopping theorem . . . 43

7 Brownian Motion **49**

8 Guest Lecture **50**
8.1 Newsvendor model . . . 50

1 Probability Review

Assumed knowledge
Ω is the sample space $\cap$ intersect $\cup$ union Law of total probability:

$$\begin{aligned} P(A) &= \sum_{i=1}^{n} P(A \cap B) \\ &= \sum_{i=1}^{n} P(A|B_i)P(B_i) \end{aligned}$$

Hint for proving: use the fact that the set $\{B_1, B_2, \ldots, B_n\}$ forms a partition of Ω then

$$P(A) = P(\cup_{i=1}^{n}(A \cap B_i))$$

For events A B and C

2 Intro

3 Probability and Measure

3.1 Laws of large numbers

Weak Law of large numbers

If $X_1, X_2 \ldots$ are i i d RVs with mean $E[X_i] = \mu$ then for all $\varepsilon > 0$:

$$\lim_{n\to\infty} P\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) = 0$$

I.e. if n is large, $\frac{S_n}{n}$ will approach μ

Strong Law of Large Numbers

Same setup

$$P\left(\lim_{n\to\infty} \frac{S_n}{n} = \mu\right) = 1$$

I.e. with probability 1, $\frac{S_n}{n} = \frac{\sum_{i=1}^{n} x_n}{n}$ converges to μ.
These statements refer to the long run but don't say much about the short term.

Central Limit Theorem

X_i iid RVs with mean $E[X_i] = \mu$ and variance $E[(X_i - \mu)^2] = \sigma^2$ then

$$P\left(\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \le y\right) \to \Phi(y)$$

Where $\Phi(y) = \frac{1}{\sqrt{2\pi}\int_{-\infty}^{y} e^{-x^2/2}dx}$ is the standard normal distribution $N(0,1)$. I.e. the probability tends to the standard normal. Informally:

$$S_n \approx n\mu + \sqrt{n\sigma^2}\mathcal{Z}$$

Measure and Probability Theory

Simple notation:

- Ω is a set of outcomes
- $\mathcal{F}$ is a set of events defined over Ω
- $P : \mathcal{F} \to [0, 1]$ is a function assigning probabilities to events in $\mathcal{F}$

The three form a triple in measure theory, describing the probability space.
If this is true then $\mathcal{F}$ is called a σ field or σ algebra, meaning that $\mathcal{F}$ is a non-empty collection of subsets in $\mathcal{R}$ such that:

1. if $A \in \mathcal{F}$ then $A^C \in \mathcal{F}$
2. If $\{A_i : A_i \in \mathcal{F}\}$ is a countable sequence of sets, then

$$\cup_i A_i \in \mathcal{F}$$

Note that $\cap_i A_i = (\cup_i A_i^C)^C$, implying that a$\sigma$-field is also closed under countable intersections

Borel σ-field

A most important type of σ-field is the Borel σ-field, which contains sets known as Borel sets:
The smallest σ-field $\mathcal{B}$ that contains all open intervals (a, b) such that $(-\infty \leq a < b \leq \infty)$ is called a Borel σ-field. Any set B$\in \mathcal{B}$
A number of different intervals can be used to generate the same Borel σ-field.
But why use this stuff?
Some measures can be undefined, meaning we need them sometimes. The probability for a particular occurrence of a continuous RV is 0, even if it is not an impossible event. But what about an infinite collection of individual events like this. So this is where those fields come in.
E.g. A simplified version of X Lotto by SA Lotteries, where you pick 6 numbers between 1 and 45 and SA Lotteries draws 6 numbers and if yours match, you win. There are

$$\binom{45}{6} = 145060$$

Ways of choosing 6 numbers. I.e. 145060 distinct 6 tuples $\omega_j = \{\omega_{j1}, \omega_{j2}, \ldots, \omega_{j6}\}$ Each of these is different i.e. $\omega_j \cap \omega_k \notin \Omega$ for $j \neq k$ But what are the possible *events* in this game? You could bet on any subset of Ω. So any subset of Ω is a potential event. Also $\mathcal{F}$ can be known as the *power set* of Ω. Note that $\mathcal{F}$ contains Ω and $\emptyset$. Because you could either not play $\emptyset$ or play every single 6-tuple Ω. Since $\mathcal{F}$ contains all subsets of Ω, it satisfies the following:

$$\text{If } A \in \mathcal{F} \text{ then } A^c \in \mathcal{F}$$

$$\text{If } A, B \in \mathcal{F} \text{ then } A \cup B \in \mathcal{F}$$

$$\text{If } A_i \in \mathcal{F} \text{ then } \cup_{i=1}^{n} \in \mathcal{F}$$

If something holds these two conditions, it is called an **algebra**.
If Ω is finite, then you can have an infinite union, which gives: A σ- algebra satisfies both, plus:

$$\text{If } A_i \in \mathcal{F} \text{ for } i = 1, \ldots \text{ then } \cup_{i=1}^{\infty} A_i \in \mathcal{F}$$

Using the X-lotto example. There are $N = 8,145,060$ possible ω_j, and so any particular 6-tuble has a $1/N$ chance of occurrence. Let A be the set of 6-tuples you have chosen. Then the probability $P(A)$ for $A \in \mathcal{F}$ is given by the number n of elements in the set A divide by the total number of elements in Ω. In particular $P(\Omega) = 1$ (if you choose every 6-tuple) and $P(\emptyset) = 0$ (if you do not choose any 6-tuples). $P(A)$ for $A \in \mathcal{F}$ is a function called a probability measure. Definition for probability measures:
A function $P : \mathcal{F} \to [0, 1]$ from a σ algebra $\mathcal{F}$ of subsets of a set Ω into the unit interval is a probability measure on the measurable space $\{\Omega, \mathcal{F}\}$ if it satisfies the following:

$$P(A) \geq 0 \quad \text{for all } A \in \mathcal{F}$$
$$P(\Omega) = 1$$
$$P\left(\cup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) \text{ for all disjoint sets } A_i \in \mathcal{F}$$

This no longer applies to the X-lotto example as there is only a finite number of disjoint sets, but the bottom equation still works since $\emptyset \cap \emptyset = \emptyset$. Then in general if $\mathcal{F}$ is finite, it is sufficient to check that:

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) \quad \text{for all disjoint sets } A_1, A_2 \in \mathcal{F}$$

In the X-lotto example, if disjoint sets A_1 and A_2 have n_1 and n_2 elements respectively then $P(A_1 \cup A_2) = \frac{n_1+n_2}{N} = \frac{n_1}{N} + \frac{n_2}{N} = P(A_1) + P(A_2)$

Our X-lotto example is therefore completely described by the triple $\{\Omega, \mathcal{F}, P\}$ (called the probability space). In general we may state for a probability space that it consists of:

- The Sample space Ω
- A σ-algebra
- A probability measure $P : \mathcal{F} \to [0.1]$ satisfying the equations shown above.

In the X-Lotto game, the collection $\mathcal{F}$ of events is in fact an algebra, but since it is finite, it is automatically a σ algebra as well.

Theorem 2.5 If an algebra contains only a finite number of sets, then it is a σ-algebra.

But why a σ-algebra?

Example 2.3 Toss a fair coin infinitely many times and receive 1 unit reward for heads, and 0 for tails. Given an event $A_{k,n}$ which consists of all the outcomes where the winnings after n tosses is k units. The probability $P(A_{k,n})$ is given by:

$$P(A_{k,n}) = \begin{cases} \binom{n}{k}\frac{1}{2^n} & \text{if } n \geq k \\ 0 & \text{if } k > n \end{cases}$$

For $q = 1, 2, \ldots$, consider now the events that after n tosses, the average win per coin toss, k/n is contained in the interval $\left[\frac{1}{2} - \frac{1}{q}, \frac{1}{2} + \frac{1}{q}\right]$. These events correspond to the sets $B_{q,n}$, where:

$$B_{q,n} = \bigcup_{k=\lceil n/2-n/q \rceil}^{\lfloor n/2+n/q \rfloor} A_{k,n}$$

Then the set $\cap_{m=n}^{\infty} B_{q,m}$ corresponds to the event that from the n^{th} tossing onwards, the average win per coin toss will stay in the interval $\left[\frac{1}{2} - \frac{1}{q}, \frac{1}{2} + \frac{1}{q}\right]$. The set $\cup_{n=1}^{\infty} \cap_{m=n}^{\infty} B_{q,m}$ corresponds to the event that there exists an n such that from the n^{th} tossing onwards, the average win per coin toss will stay in the interval.

Finally the set $\cap_{q=1}^{\infty} \cup_{n=1}^{\infty} \cap_{m=n}^{\infty} B_{q,m}$ corresponds to the event that the average win per coin toss converges to $1/2$ as $n \to \infty$. Now the strong Law of Large Numbers therefore states that the latter event has probability 1. That is:

$$P(\cap_{q=1}^{\infty} \cup_{n=1}^{\infty} \cap_{m=n}^{\infty} B_{q,m}) = 1$$

which is only defined if:

$$\cap_{q=1}^{\infty} \cup_{n=1}^{\infty} \cap_{m=n}^{\infty} B_{q,m} \in \mathcal{F}$$

To guarantee this, $\mathcal{F}$ must be a σ-algebra.

3.2 Smallest σ-algebra

The smallest σ-algebra containing a given collection $\mathcal{C}$ of sets is called the σ-algebra generated by $\mathcal{C}$ and is usually denoted by $\sigma(\mathcal{C})$.

The idea of a smallest σ-algebra of subsets of Ω is always relative to a collection of subsets of Ω, because without reference to such a collection of sets, the smallest σ-algebra of subsets of Ω is just $\{\phi, \Omega\}$. An important special case of a smallest σ-algebra is where $\Omega = R$ and $\mathcal{C}$ is the collection of all open intervals.

3.3 Euclidean Borel field

The σ-algebra generated by the collection of all open intervals $\{(a, b) : a < b \text{ where } a, b \in R\}$ in R is called the Euclidean Borel field, denoted by $\mathcal{B}$ and its members are called the Borel sets.

$\mathcal{B}$ works for closed intervals and half-open intervals too, and works for higher dimensions.

Theorem 2.6 Let $\{\Omega, \mathcal{F}, P\}$ be a probability space, then the following hold for sets in $\mathcal{F}$

1. $P(\emptyset) = 0$
2. $P(A^C) = 1 - P(A)$
3. $A \subset B$ implies that $P(A) \leq P(B)$
4. $P(A \cup B) + P(A \cap B) = P(A) + P(B)$
5. If $A_n \subset A_{n+1}$ then $P(A_n) \uparrow P(\cup_{n=1}^{\infty} A_n)$ If $A_n \supset A_{n+1}$ then $P(A_n) \downarrow P(\cup_{n=1}^{\infty} A_n)$
6. $P(\cup_{n=1}^{\infty} A_n) \leq \sum_{n=1}^{\infty} P(A_n)$

Where the $\uparrow$ means increasing sequence of numbers, and $\downarrow$ means a decreasing sequence of numbers.

Example 2.4: The uniform probability measure Place ten numbered balls in a bowl. Randomly draw a ball and write down the corresponding number in the tenths place of a decimal number between 0 and 1. Replace the ball and repeat, except in the hundredths place, etc.
E.g. picking 5 and then 7 gives the number 0.57. Repeating this infinitely many times gives a random number between 0 and 1 such that the sample space is in the unit interval $[0, 1)$. In this example for a given number $x \in [0, 1)$ the probability that a random number chosen as above is less than or equal to x is in fact x i.e. $P(X \leq x) = x$.
Suppose we only draw 2 balls and end up with 0.57. Then to get a number less than 0.57:

1. There are 5 ways of choosing the first number and 10 ways of choosing the second number.
2. There is only 1 way of choosing a 5 first, and 8 ways of choosing the second number less than or equal to 7
3. There are 100 ways to pick 2 balls with replacement.

Hence there are 58 ways of choosing a number less than or equal to 0.57 and so in this experiment where we only choose 2 balls, the probability we obtain a number less than or equal to 0.57 is $58/100 = 0.58$. This holds for longer chains. We see that for $x \in [0, 1]$ we have that $P([0, x]) = x$.
It follows that $P([x]) = P([x, x]) = 0$ and $P([0, 1]) = 1$.
Since this is continuous, half open intervals are the same as open is the same as closed.
Any finite union of intervals can be written as a finite union of disjoint intervals but cutting out the overlap. Therefore this probability measure extends to finite unions of intervals by adding up the lengths of the associated disjoint intervals. Moreover the collection of all finite unions of subintervals in $[0, 1]$ including $[0, 1]$ and the empty set, is closed under complements and finite unions. Hence we have derived the probability measure P for an algebra $\mathcal{F}_0$ of subsets of $[0, 1]$ given by:

$$\mathcal{F}_0 = (a, b), [a, b], (a, b], [a, b), \text{ where } a, b \in [0, 1] \text{ and } a \leq b$$

Along with their finite unions.
This is a special case of Lebesgue measure, which assigns the length of an interval. The algebra $\mathcal{F}_0$ does not however contain all of the Borel sets in $[0, 1]$, and therefore not all probability measures are defined. In particular, for a countable sequence of sets $A_j \in \mathcal{F}_0$, the probability $P(\bigcup_{j=1}^{\infty} A_j)$ is not always defined because their is no guarantee that: $\bigcup_{j=1}^{\infty} A_j \in \mathcal{F}_0$.
We therefore need to extend the probability measure p on $\mathcal{F}_0$ to a probability measure defined on the Borel sets in $[0, 1]$, which is normally achieved using the concept of "outer measure".
Any subset A of $[0, 1]$ can always be completely covered by a finite OR countably infinite union of sets in the algebra $\mathcal{F}_0 : A \subset \bigcup_{j=1}^{\infty} A_j$, for $A_j \in \mathcal{F}_0$.
The "probability" of A is therefore bounded above by $\sum_{j=1}^{\infty} P(A_j)$ and the infimum of this bound taken across all sequences of sets $A_j \in \mathcal{F}_0$ where $A \subset \bigcup_{j=1}^{\infty} A_j$ yields the "outer measure"

Definition 2.19 Outer measure:
IF $\mathcal{F}_0$ is an algebra of subsets of Ω then the outer measure P^* of an arbitrary subset A of Ω is:

$$P^*(A) = \inf_{A \subset (\bigcup_{j=1}^{\infty} A_j), A_j \in \mathcal{F}_0} \sum_{j=1}^{\infty} P(A_j)$$

Note that $\bigcup_{j=1}^{\infty} A_j \in \mathcal{F}_0$ is not a requirement in this definition.
Without loss of generality we may also assume that the infimum of Definition 2.19 is taken over all disjoint sets $A_j \in \mathcal{F}_0$, in which case if it happens that $A \in \mathcal{F}_0$ then $P^*(A) = P(A)$.

It seems reasonable to now also consider for what other sets in Ω does the outer measure correspond to a probability measure.
Note that the outer measure P^* satisfies the first two equations of definition 2.15 of probability measure, but in general the third equation does not hold for arbitrary sets. It is however possible to extend the outer measure to a probability measure on a σ-algebra $\mathcal{F}$ containing $\mathcal{F}_0$

Theorem 2.7 If P is a probability measure on $\{\Omega, \mathcal{F}_0\}$, where $\mathcal{F}_0$ is an algebra, let $\mathcal{F} = \sigma(\mathcal{F}_0)$ (the smallest σ algebra containing $\mathcal{F}_0$). Then the outer measure P^* is the unique probability measure on $\{\Omega, \mathcal{F}\}$ which coincides with P on $\mathcal{F}_0$.
Hence for the experiment earlier, there exists a σ algebra $\mathcal{F}$ of subsets of $\Omega = [0, 1]]$ containing the algebra $\mathcal{F}_0$, for which the outer measure is the unique probability measure that coincides with P. This probability measure assigns to each interval in $[0, 1]$ its length as its probability and is called the *Uniform* probability measure.
A probability space becomes decidedly more interesting when we define a random variable on it, as below:

In broad terms, a RV is just a numerical translation of the outcomes of the sample space.

Example 2.5 Toss a coin: Here $\Omega = \{H, T\}$ and the σ algebra is $\mathcal{F} = \{\Omega, \emptyset, \{H\}, \{T\}\}$ with corresponding probability measure $P(\{H\}) = P(\{T\}) = 1/2$.

$$\text{Define:} \quad X(\omega) = \begin{cases} 1 & if \quad \omega = H \\ 0 & if \quad \omega = T \end{cases}$$

For an arbitrary Borel set B we have $P(\{\omega \in \Omega : X(\omega) \in B\})$

$$= \begin{cases} P(\{H\}) = 1/2 \quad if \quad 1 \in B \quad and \quad 0 \notin B \\ P(\{T\}) = 1/2 \quad if \quad 1 \notin B \quad and \quad 0 \in B \\ P(\{H, T\}) = 1 \quad if \quad 1 \in B \quad and \quad 0 \in B \\ P(\emptyset) = 0 \quad if 1, 0 \notin B \end{cases}$$

In this example, the set is automatically equal to one of the elements of $\mathcal{F}$ and so $P(X(\omega) \in B) = P(\{\omega \in \Omega : X(\omega) \in B\})$ is well defined. In general however, we need to confine our RVs $X : \Omega \to R$ to those for which we can make probability statements about events of the type ,where B is an arbitrary Borel set, which is only possible if these sets are members of $\mathcal{F}$.

Definition 2.20: Random variable: If $\{\Omega, \mathcal{F}, P\}$ is a probability space, then a mapping $X : \Omega \to R$ is called a random variable defined on $\{\Omega, \mathcal{F}, P\}$, if X is $\mathcal{F}$-measurable. This means that for every Borel set B, $\{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}$.

Theorem 2.8 A mapping $X : \Omega \to R$ is $\mathcal{F}$ measurable if and aonly if for all $x \in R$ the sets $\{\omega \in \Omega : X(\omega) \leq x\}$ are members of $\mathcal{F}$. In example 2.4, consider that I set $X(\omega) = 1$ if $\omega \in \Omega = [0, 1]$ is rational, and $X(\omega) = 0$ otherwise.
Would such a RV be $\mathcal{F}$ measurable? In short - no.

Definition 2.21: Generated σ algebra: If X is a RV, then define $X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\}$. Then the σ algebra $\mathcal{F}_x = \{X^{-1}(B)$, for all $B \in \mathcal{B}\}$, is called the σ algebra generated by X.
In the coin tossing example, the mapping X is one-to-one, and therefore $\mathcal{F}_X = \mathcal{F}$, but in general $\mathcal{F}_X$ is much smaller than $\mathcal{F}$

Example 2.6 Roll a Die: Consider the roll of a die where we set X to be 1 if the die shows an even face or 0 if it shows an odd face, then: $\mathcal{F}_X = \{\{1, 2, 3, 4, 5, 6\}, \emptyset, \{2, 4, 6\}, \{1, 3, 5\}\}$ You now decide to conduct a further investigation where you need to use the RV Y which has value 1 if the die shows a number less than 4 and value 0 otherwise. Thus:

$$\mathcal{F}_Y = \{\Omega, \emptyset, \{1, 2, 3\}, \{4, 5, 6\}\}$$

Of course, Y is $\mathcal{F}_Y$ measurable, but Y is not $\mathcal{F}_X$ measurable. Therefore, if you only record X you cannot say anything about Y , and vice-versa. To be able to answer questions about both X and Y, you need to work with the σ-algebra generated by:

$$\mathcal{F} = \{\Omega, \emptyset, \{1, 3, 5\}, \{2, 4, 6\}, \{1, 2, 3\}, \{4, 5, 6\}\}$$

Moral: You need to know what investigation you need to do *before* conducting the experiment, so that you know what information you need to record. Otherwise you will not have the information necessary to make meaningful statements.

Definition 2.22: Induced Probability measure: Given a RV X define for arbitrary $B \in \mathcal{B}$

$$\mu_X(B) = P(X^{-1}(B)) = P(\{\omega \in \Omega : X(\omega) \in B\})$$

Then $\mu_X(B)$ is a probability measure on $\{R, \mathcal{B}\}$ is induced by the RV X, where

1. for all $B \in \mathcal{B}, \mu_X(B) \geq 0$
2. $\mu_X(R) = 1$ and,
3. for all disjoint $B_j \in \mathcal{B}$, $\mu_X(\bigcup_{j=1}^{\infty} B_j) = \sum_{j=1}^{\infty} \mu_X(B_j)$

Therefore, the RV X maps the probability space $\{\Omega, \mathcal{F}, P\}$ into a new probability space $\{R, \mathcal{B}, \mu_X\}$, which in turn is mapped back by X^{-1} into (the possibly smaller) probability space: $\{\Omega, \mathcal{F}_X, P\}$

Definition 2.23: Distribution function If X is a random variable, which induces probability measure μ_X, the function $F(x) = \mu_X((-\infty, x])$, for $x \in R$ is called the distribution function of X.

Theorem 2.9 A distribution function of a RV is always right continuous and monotonic non-decreasing. That is, for all $x \in R$:

$$\lim_{\delta \downarrow 0} F(x+\delta) = F(x) \quad and$$
$$F(x_1) \leq F(x_2) \quad if \quad x_1 < x_2, \quad with$$
$$\lim_{x \downarrow -\infty} F(x) = 0 \quad and \quad lim_{x \uparrow \infty} F(x) = 1$$

Yay no more measure theory!

4 Discrete Time Markov Chains

4.1 Basic Definitions

Particular class of random processes. Markov Chains are memoryless - the next state only depends on the present state.

Definition 3.24 Random processes (stochastic processes) are a family, or sequence, of random variables X_n where n is a parameter running over a suitable index set T. Each $X_n, n \in T$ is a (measurable) function, which maps Ω into a state space S. For any fixed

This family is called a **realisation** or **sample path** of X at ω.

Let n correspond to discrete units of time, such that $T = \{0, 1, 2, \ldots\}$. This process is said to occur in **discrete** time. Furthermore, we restrict to discrete RVs, resulting in a countable state space. Let $X_n, n \in N$ be a sequence of RVs with a countable state space S.

Definition 3.25 Discrete time Markov Chain DTMC The discrete time random process $X_n, n \in N$ is a DTMC if it satisfies

$$P(X_n = s | X_0 = x_0, X_1 = x_1, \ldots, X_{n-1} = x_{n-1}) = P(X_n = s | X_{n-1} = x_{n-1})$$

for all $n \geq 1$ and all $s, x_0, \ldots x_{n-1} \in S$. This is the memoryless property.

Since S is countable, it has a one-to-one relation to some subset of the natural numbers. Without loss of generality, we can assume $S \subset N$. This simplifies notation considerably, as we can essentially give up the physical interpretation of the states for a numerical one, and can write more general expressions like

$$P(X_{n+1} = j | X_n = i), \text{ for } i, j \in S$$

This is a transition probability. These govern the evolution of the Markov chain - the evolution of the sequence $X_1, X_2, \ldots$ from some starting state X_0. Note that the "starting state" X_0 can be specified deterministically, or it can be chosen randomly from some distribution across S.

Definition 3.26 Time homogeneous Markov Chain A Markov chain X_n is time-homogeneous if we have

$$P(X_{n+1} = j | X_n = i) = P(X_1 = j | X_0 = i) \text{ for all } n \text{ and } i, j \in S$$

In which case we write:

$$P_{i,j} = P(X_{n+1} = i | X_n = i) \; \forall n \text{ and } i, j \in S$$

In contrast if the transition probabilities depend on the current time n as well as on i and j, then the Markov chain is called time-inhomogeneous (we won't do much of that yet).

Definition 3.27 Transition matrix The transition matrix P of a time-homogeneous Markov chain is the $|S| \times |S|$ matrix of transition probabilities

$$p_{i,j} = P(X_{n+1} = j | X_n = i) \quad \rightarrow \quad P = [p_{i,j}]$$

Transition probabilities satisfy the following

1. $0 \leq p_{i,j} \leq 1$
2. $\sum_{j \in S} p_{i,j} = 1$ Note this means the row sums to 1 (this is convention, can be done with columns!!)

Definition 3.28 m-step transition matrix The m-step transition matrix $P^{(m)} = [p_{i,j}^{(m)}]$ of a time-homogeneous Markov chain is the $|S| \times |S|$ matrix of transition probabilities:

$$p_{i,j}^{(m)} = P(X_{n+m} = j | X_n = i)$$

The m-step transition probability $p_{i,j}^{(m)}$ is the probability that the process starting in state i at time n, finds itself in state j at time $n + m$. In general there are multiple possible "paths" that result in this outcome. The probability $p_{i,j}^{(m)}$ is the sum of these possible paths.

Theorem 3.10 Calculating the m-step transition matrix

$$P^{(m)} = P^m$$

Proof: use induction on m, which is valid since it is a **discrete** time markov chain. Trivially true for $m = 1$. Assume true for $m - 1$, i.e. $P^{(m-1)} = P^{m-1}$. Condition $p_{i,j}^{(m)}$ on the state after $m - 1$ steps so that:

$$\begin{aligned}
p_{i,j}^{(m)} &= P(X_m = j, X_0 = i) \\
&= \sum_{k \in S} P(X_m = j | X_{m-1} = k, X_0 = i) P(X_{m-1} = k | X_0 = i) \text{ using equation 0.5 (law of total probability)} \\
&= \sum_k P(X_m = j | X_{m-1} = k) P(X_{m-1} = k | X_0 = i) \text{ markov memoryless property} \\
&= \sum_k P(X_m = j | X_{m-1} = k) P_{i,k}^{(m-1)} \text{ by assumption} \\
&= \sum_k p_{k,j} p_{i,k}^{(m-1)} \\
&= \sum_k p_{i,k}^{(m-1)} p_{k,j} \\
&= \sum_k [P^{(m-1)}]_{ik} \times p_{kj} \\
&= [P^m]_{ij} \text{ by matrix multiplication}
\end{aligned}$$

As required.

Example 3.10 Numerical example using theorem 3.10 if

$$P = \begin{bmatrix} 0 & \frac{1}{4} & \frac{3}{4} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}, \quad \text{what is } p_{1,2}^{(3)}$$

By theorem 3.10, $P^{(3)} = P^3$, so:

$$p_{1,2}^{(}3) = [P]_{1,2}^3 = \begin{bmatrix} \frac{10}{64} & \frac{25}{64} & \frac{29}{64} \\ \frac{12}{64} & \frac{28}{64} & \frac{14}{64} \\ \frac{11}{32} & \frac{15}{32} & \frac{6}{32} \end{bmatrix} = \frac{25}{64}$$

Example 3.11 two gamblers:

Player A and player B are engaged in a contest comprised of a series of games in which A has probability p of winning any given game, and player B has probability $q = 1 - p$ of winning any given game.

Each time a player wins, the player's fortune is increased by one dollar, and conversely, the player's fortune decreases by one dollar if they lose.

The players' combined fortune is N dollars. A starts with a fortune of k and B starts with a fortune of $N - k$.

The players stop playing when one of them has lost all of their money.

There is a definite advantage to A if $p > q$ and a definite disadvantage if $p < q$. If $p = q$ it is considered "*fair*".

Consider the Markov chain X_n representing A's fortune after n games for $n \in N$.
The state space of the process X_n is clearly: $S = \{0, 1, 2, \ldots, N\}$
A is ruined when the state of the process is 0, and B is ruined when the state of the process is N
The transition probabilities for A's fortune for $0 < i < N$ (note not including 0 or N) are given by:

$$p_{ij} = \begin{cases} p, & j = i + 1 \\ 1 - p, & j = i - 1 \\ 0, & \text{otherwise} \end{cases}$$

i.e. you can only move by 1 at any time. States 0 and N are absorbing states, as once the process enters one of those states, it cannot leave. The transition probability matrix for the process X_n is the $(N + 1) \times (N + 1)$ matrix:

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

The future value of a player's fortune generally does depend on the player's history. E.g. if we know that $X_5 = 1$, then the probability that A is ruined before B will be higher than if we know that $X_5 = N - 1$. However, the fortune of A is still a Markov chain because the future fortune does not depend on the history given the current fortune

That is, they do not depend on the history of wins and losses, only on the player's current fortune. Note that the fortune of A (or B) is a time-homogeneous Markov chain, because the transition probabilities do not depend on the time index $n \in N$.

What is the probability that A eventually loses all their money?
We can answer this by letting W be the event that A eventually loses all of their money, and then let

$$\begin{aligned} u_i &= P(W|X_0 = i) \\ &= P(X_n \to 0 \text{ before } X_n \to N|X_0 = i) \end{aligned}$$

We can immediately identify the "boundary conditions" that are relevant for this problem:

$$u_0 = 1 \quad u_N = 0$$

In order to obtain u_i for general values of i, we employ a **first step analysis** (this is basically a PDEs question with a sparse matrix).

Here, we make use of the Law of total probability to simplify the calculation. To do so, we need to choose a suitable set of partitioning events. In this case, we use the events defined by the state at the end of the first transition. In our problem, for $0 < i < n$ we have:

$$\begin{aligned} u_i &= P(W|X_0 = i) \\ &= \sum_{j \in S} P(W|X_1 = j, X_0 = i) \\ &= \sum_{j \in S} P(W|X_1 = j) P(X_1 = j|X_0 = i) \text{ Markov property.} \\ &= \sum_{j \in S} P(W|X_0 = j) P(X_1 = j|X_0 = i) \text{ Time homogeneity.} \\ &= \sum_{j \in S} u_j P_{ij} \text{ Definition and matrix} \\ &= pu_{i+1} + qu_{i-1} \end{aligned}$$

So now we have the system of difference equations

$$u_i = pu_{i+1} + qu_{i+1}, \quad 0 < i < N$$

With boundary conditions $u_0 = 1$, $u_N = 0$. There are several possible approaches we can use to solve this system, but we shall apply a very systematic, standard method for solving them. Note that The difference equations are:

1. **Second-order** because the associated characteristic (or auxiliary) equation is of second order as we shall see shortly (think about newtons law for the equation)
2. **Homogeneous** (there's no term not associated with u)
3. **Linear -** because all u_i's are raised to the first power, and all products only involve one u_i.

Difference equations can be thought of as the discrete analogue of differential equations.

To solve this you find the characteristic equation - by 'guessing' a form of solution, then check if it satisfies the DE. It turns out we can apply a similar technique with difference equations. Lets try a solution of the form:

$$u_i = w^i$$

Substituting this into the difference equation gives:

$$w^i = pw^{i+1} + qw^{i-1}, \quad 0 < i < N$$

Divide through by w^{i-1} and rearrange to get the characteristic equation:

$$pw^2 - w + q = 0$$

The two roots of the characteristic equation are (using completing the square):

$$pw^2 - (p+q)w + q = (w-1)(pw-q) = 0$$

Which gives:

$$w_1 = 1 \quad w_2 = \frac{q}{p}$$

Since the $w_1 = 1$ appears, this means things are working as they should. Note that if $p = q$ we have a repeated root, so we must study the cases of $p \neq q$ and $p = q$ separately.

1. $p \neq q$ Recall the principle of superposition of solutions of linear DEs. This suggests a general solution of form:

   $$u_i = A_1(1)^i + A_2 \left(\frac{q}{p}\right)^i = A_1 + A_2 \frac{q^i}{p^i}$$

   Now using the boundary conditions to determine that values of A_1 and A_2:

   $$u_0 = 1 \implies A_1 + A_2 = 1$$

   $$u_N = 0 \implies A_1 + A_2 \left(\frac{q}{p}\right)^N = 0$$

$$\implies A_2 = \frac{1}{1 - \left(\frac{q}{p}\right)^N}$$

$$A_1 = 1 - A_2$$

This gives:

$$u_i = \frac{\left(\frac{q}{p}\right)^N - \left(\frac{q}{p}\right)^i}{\left(\frac{q}{p}\right)^N - 1} \quad \text{for } p \neq q$$

2. $p = q$ We have repeated roots $w_1 = w_2 = 1$ and so in a manner that is analogous to the solution of DEs, we guess:

$$u_i = (A_1 + A_2 i)\, w^i = A_1 + A_2 i$$

Now using the boundary conditions to find A_1 and A_2:

$$u_0 = 1 \implies A_1 = 1 \quad u_N = 0 \implies 0 = 1 + A_2 N \implies A_2 = \frac{-1}{N}$$

$$u_i = \frac{N-i}{N} \quad \text{for } p = q$$

So the solution is for $0 \leq i \leq N$ is:

$$u_i = \begin{cases} \frac{\left(\frac{q}{p}\right)^N - \left(\frac{q}{p}\right)^i}{\left(\frac{q}{p}\right)^N - 1} & \text{for } p \neq q \\ \frac{N-i}{N} & \text{for } p = q \end{cases}$$

When player A has a definite advantage, $p > q$, then u_i's probability that A loses all of her money will drop off dramatically for small i. Whereas if $q > p$ this probability becomes small for large i (close to N). Recall that A losing the contest is equivalent to B winning the contest. This gives the symmetry.

Consider the probability that A eventually *wins* all the money: That is: $v_i = P(X_n \to N before 0 | X_0 = i)$ We could solve this using the method above, with the appropriate boundary conditions, or we could use a simple argument exploiting the problem's symmetry. This is solved in class exercise 2. The answer of course is, for $0 \leq i \leq N$.

$$v_i = \begin{cases} \frac{\left(\frac{q}{p}\right)^i - 1}{\left(\frac{q}{p}\right)^N - 1} & \text{for } p \neq q \\ \frac{i}{N} & \text{for } p = q \end{cases}$$

4.1.1 Analysis of the two gamblers

We can now ask the question: **What is the probability that the contest eventually comes to an end?** This is equivalent to the question **What is the probability that A eventually wins or loses all of their money?** The answer is:

$$u_i + v_i = 1$$

Note that while it is a logical possibility that the players can play forever (win,lose,win,lose,...), the probability is still equal to 0. The set of all games that go on forever is a set of measure 0. It has no size with respect to Lebague measure.

Gamblers again What is the expected duration of the contest, given that A has initial fortune i? Let H denote the (random) time until A wins or loses the contest (that is, the process reaches 0 or N), and let:

$$k_i = E\left(H | X_0 = i\right)$$

Note that$H$ is a RV taking values on the countably infinite state space $\{0, 1, 2\ldots\} \cup \{\infty\}$. Instead k_i is its (conditional) expectation and is thus a constant value, not a RV.

Note hitting times can (unfortunately) be infinite In particular, the expectations k_i satisfy the system of difference equations:

$$k_i = 1 + pk_{i+1} + qk_{i-1}, \quad 0 < i < N$$

$$k_0 = 0 \quad and \quad k_n = 0$$

The 1 in the first equation is the unit of time spent in playing one game, (which is also the minimum amount of time which must be spent playing if the contest has not yet ended). This is a system of second order non-homogeneous linear difference equations. By writing:

$$pk_{i+1} - k_i + qk_{i-1} = -1$$

We see that this would be identical to the homogeneous equation (only with the -1)

We construct a solution as follows (similar to in DEs), by solving the general solution, and adding the particular solution.

So for $p \neq q$, the solution to the homogeneous equation is:

$$k_i = A_1 + A_2 \left(\frac{q}{p}\right)^i, \quad 0 \leq i \leq N$$

As from DEs, to add the -1 we try polynomial solutions. Starting with a constant C

$$pC - C + qC = 0$$

So that doesn't work. Try $C_1 + iC_2$

$$p(c_1 + C_2(i+1)) - C_1 - C_2 i + q(C_1 + C_2(i-1)) = -1$$

$$C_1[P - 1 + q] + C_2[i(p - 1 + q) + p - q] = -1$$

$$C_1 = \frac{1}{q-p}$$

Note that the constant C_1 appears in the homogeneous form of the solution, so it cancels again as expected. So the general form is:

$$k_i = A_1 + A_2 \left(\frac{q}{p}\right)^i + \frac{i}{q-p}$$

Using the boundary conditions

$$k_0 = 0 \quad gives \quad A_1 = -A_2$$

$$k_N = 0 \quad gives \quad A_2 = \frac{N}{(p-q)\left(\left(\frac{q}{p}\right)^i - 1\right)}$$

This gives:

$$k_i = \frac{1}{p-q}\left[N\frac{1-\left(\frac{q}{p}\right)^i}{1-\left(\frac{q}{p}\right)^N} - i\right], \quad p \neq q$$

For $p =$

4.2 Hitting probabilities and hitting times

This is a special case of a markov chain known as a random walk.

The probabilities u_i that we obtained before were examples of hitting probabilities. In particular, u_i corresponds to the probability of hitting state 0 (an absorbing state), before hitting state N (another absorbing state) given that the starting state is i. Also the random time H above was an example of a hitting time, i.e. the time at which the process first hits an absorbing state.

We now define the notion of hitting probability and hitting time in more general terms. Let $X_n, n \in N$ be a markov chain with transition matrix P, and state space $\mathcal{S}$. Let $\mathcal{A}$ be a subset of $\mathcal{S}$, which contains the state, or states, for which we want to know the hitting probability.

Definition 3.29 Hitting time The hitting time of a subset $\mathcal{A} \subset \mathcal{S}$ is the RV:

$$H^{\mathcal{A}} : \Omega \to \{0, 1, 2, \ldots\} \cup \{\infty\}$$

$$H^{\mathcal{A}} = \inf\{n \geq 0 : X_n \in \mathcal{A}\}$$

Note in this case Ω corresponds to the set of all possible sample paths. We use the infimum instead of the minimum, because we cannot be sure for all simple paths that the minimum value of n exists. We are therefore forced to use the convention that the infimum of the empty set is ∞

Note in the previous problem we had $\mathcal{A} = \{\prime, \mathcal{N}\}$. and we knew that for sure there exists a finite value of n such that $X_n \in \mathcal{A}$ because we showed that with probability one, the contest ends after a finite number of time-steps.

Definition 3.30: Hitting probability The hitting probability of a subset $\mathcal{A}$ of $\mathcal{S}$ is defined as:

$$u_i^{\mathcal{A}} = P(H^{\mathcal{A}} < \infty | X_0 = i)$$

I.e. $u_i^{\mathcal{A}}$ is the probability that the process ever hits some state in $\mathcal{A}$ given it starts in i. When all of the states in the set $\mathcal{A}$ are absorbing states (or when $\mathcal{A}$ is an absorbing class [explained later]), then $u_i^{\mathcal{A}}$ is a special case of a hitting probability, referred to as an **absorption probability**.

Theorem 3.11: Minimal non-negative solution for hitting probabilities The hitting probabilities $u_i^{\mathcal{A}}$ are given by the minimal non-negative solution to the system of equations:

$$u_i^{\mathcal{A}} = \begin{cases} 1 & for\ i \in \mathcal{A} \\ \sum_{j \in \mathcal{S}} p_{i,j} u_j^{\mathcal{A}} & for\ i \notin \mathcal{A} \end{cases}$$

In general terms, this means that if: x_i, $i \in \mathcal{S}$ is another solution to this, with $x_i \geq 0$ for all i, then $x_i \geq u_i$ for all i. I.e. this is the inf of the set.

Proof: First the probability satisfies the equations:
If $X_0 = i \in A$, then $H^{\mathcal{A}} = 0$, so $u_i^{\mathcal{A}} = 1$
If $X_0 = i \notin A$, then $H^{\mathcal{A}} \geq 1$, and we have:

$$\begin{aligned} u_i^{\mathcal{A}} &= P(H^{\mathcal{A}} < \infty | X_0 = i) \quad \text{which by the Law of total probability, gives:} \\ &= \sum_{j \in \mathcal{S}} P(H^{\mathcal{A}} < \infty | X_1 = j, X_0 = i) P(X_1 = j | X_0 = i) \\ &= \sum_{j \in \mathcal{S}} P(H^{\mathcal{A}} < \infty | X_1 = j) P(X_1 = j | X_0 = i) \text{ markov property} \\ &= \sum_{j \in \mathcal{S}} P(H^{\mathcal{A}} < \infty | X_0 = j) P(X_1 = j | X_0 = i) \text{ time homogeneous} \\ &= \sum_{j \in \mathcal{S}} u_j^{\mathcal{A}} p_{ij} \end{aligned}$$

Note that all these proofs seem follow the same order of steps. Suppose now that x_i, $i \in \mathcal{S}$ is any solution to this. Then $u_i^{\mathcal{A}} = x_i = 1$ for $i \in \mathcal{A}$. For $i \notin \mathcal{A}$, we have:

$$x_i = \sum_{j \in \mathcal{S}} p_{ij} x_j = \sum_{j \in \mathcal{A}} p_{ij}(1) + \sum_{j \notin \mathcal{A}} p_{ij} x_j$$

Since $\mathcal{A}$ and $\mathcal{A}^c$ form a **partition** of $\mathcal{S}$.
Substitute for x_j (recursion) to obtain:

$$\begin{aligned} x_i &= \sum_{j \in \mathcal{A}} p_{ij}(1) + \sum_{j \notin \mathcal{A}} p_{ij} \left(\sum_{k \in A} p_{jk} + \sum_{k \notin A} p_{jk} x_k \right) \\ &= P(X_1 \in \mathcal{A} | X_0 = i) + P(X_1 \notin \mathcal{A}, X_2 \in \mathcal{A} | X_0 = i) + \sum_{j \notin A} \sum_{k \notin A} p_{ij} p_{jk} x_k \\ &= P(\text{absorved in 2 or less steps}) + Junkterm \end{aligned}$$

Use recursion here, the junk term doesn't disappear, but you end up with $P($ absorbed in N steps$) + Junkterm$
Now since x is non-negative, the junk term has to be non-negative.
All of the other terms sum to $P(H^{\mathcal{A}} \leq n | X_0 = i)$ We can therefore write:

$$x_i \geq P(H^{\mathcal{A}} \leq n | X_0 = i)$$

But this is for arbitrary n

$$\begin{aligned} \implies x_i &\geq \lim_{n \to \infty} P(H^{\mathcal{A}} \leq n | X_0 = i) \\ \implies x_i &\geq P(H^{\mathcal{A}} < \infty | X_0 = i) \\ \implies x_i &\geq u_i^{\mathcal{A}} \end{aligned}$$

Which means that this is the minimal non-negative solution.

4.3 Expected Hitting time

A similar result holds for the expected hitting times $k_i^{\dashv}$ As before, let $\mathcal{A}$ be a subset of $\mathcal{S}$.

Theorem 3.12 Minimal non-negative solution for hitting times. The mean hitting times $k_i^{\dashv}$, $i \in \mathcal{S}$ are given by the minimal non-negative solution to the system of equations:

$$k_i^{\mathcal{A}} = \begin{cases} 0 & i \in \mathcal{A} \\ 1 + \sum_{j \notin \mathcal{A}} p_{ij} k_j^{\mathcal{A}} & i \notin \mathcal{A} \end{cases}$$

If a problem with a finite state space, it can be visualised - so this stuff doesn't seem super relevant. If we get a system of linear equations in an infinite state space, the minimality property can be useful. Where there are multiple possible solutions from which we must choose, as is demonstrated in the following:

Example 3.13 Infinitely Rich player B

Consider a modified version of the gamblers ruin contest. Only in this case, B is infinitely rich.
The state space then becomes $\mathcal{S} = \{0, 1, 2, \ldots\}$. As before, player A has lost all of their money, when the state of the process reaches 0.

Player B can never be bankrupt, and there is no bound on how much money can potentially be won by player A.

This could be seen as if B is a casino, and A is playing at the casino.

Let $u_i = u_i^{\mathcal{A}} = u_i^{\{0\}}$ be the probability of absorption in state 0, given that the process starts in state i. Given the new state space, we have the system of difference equations:

$$u_i = p u_{i+1} + q u_{i-1}, \quad i \geq 1$$

With the single BC, $u_0 = 1$. Recall the general solution of the system of equations involved two coefficients A_1, A_2 , and the two boundary conditions were necessary to completely determine these coefficients. In this problem we have no second boundary condition. The result from theorem 3.11 can help in this situation.

Consider $p \neq q$ General solution as in the previous case was:

$$u_i = A_1 + A_2 \left(\frac{q}{p}\right)^i, \quad i \geq 0$$

The boundary condition gives:

$$u_0 = 1 = A_1 + A_2$$

$$\implies u_i = 1 + A_2 \left(\left(\frac{q}{p}\right)^i - 1 \right)$$

How can we determine A_2?
If $p > q$, then $\lim_{i \to \infty} \left(\frac{q}{p}\right)^i = 0$, so $\lim_{i \to \infty} u_i = 1 - A_2$
Constraints $0 \leq u_i \leq 1$ for all $i \geq 0$ implies:

$$0 \leq A_2 \leq 1$$

Using the theorem, we find the minimal non-negative solution by selecting a value of A_2 from the infinite number of possible values on the interval $[0, 1]$. In particular, since $-1 \leq \left(\left(\frac{q}{p}\right)^i - 1 \right) \leq 0$, when $q < p$, then u_i is minimised (non-negatively) over all possible choices of $A_2 \in [0, 1]$ when $A_2 = 1$ This gives:

$$u_i = \left(\frac{q}{p}\right)^i$$

However, If $q > p$, then for all $i \geq 1$

$$\left(\frac{q}{p}\right)^i - 1 > 0$$

Then the minimal, non-negative solution is obtained by setting $A_2 = 0$, so

$$u_i = 1, \quad i \geq 0$$

Finally consider $q = p$

$$u_i = A_1 + A_2 i$$

The boundary condition $u_0 = 1$ gives $A_1 = 1$ so

$$u_i = 1 + A_2 i$$

The minimal is obtained by setting $A_2 = 0$, which gives:

$$u_i = 1, \quad i \geq 0$$

Putting these all together, for $i \geq 0$:

$$u_i = \begin{cases} \left(\frac{q}{p}\right)^i & when\ q < p \\ 1 \quad when\ q > p \end{cases}$$

This means that if the casino has an advantage, or if the game is fair, player A will eventually lose with probability 1.

4.4 Classification of states

Classifying different states of a DTMC
Consider a DTMC with transition matrix P

Definition 3.31: Accessible states State j is said to be accessible from state i if $p_{ij}^{(m)} > 0$ for some integer $m > 0$. In words: state j is accessible from i if there is a probability that j can be reached from i in a finite number of steps. The notation $i \to j$ is used to signify that j is accessible from i

Definition 3.32: Ephemeral states If state i does not have access to itself, i is ephemeral. In general we shall disregard such states. This is shown by a column of zeros in the transition matrix

Definition 3.33: Communicating states States i and j communicate if each is accessible from the other. The notation $i \leftrightarrow j$ is used to signify that i and j communicate. $p_{ii}^{(m)} = 0$ for all $m > 0$ If they do not communicate, then either $p_{ij}^{(m)} = 0\ \forall m > 0$ or $p_{ji}^{(m)} = 0\ \forall m > 0$ or both. Communication is an equivalence relation (if you ignore ephemeral states)- so it is reflexive, symmetric, and transitive.

Proof of transitivity Given $i \leftrightarrow j$ and $j \leftrightarrow k$, this implies there exist n and m such that:

$$p_{ij}^{(n)} > 0 \quad and \quad p_{jk}^{(m)} > 0$$

Conditioning on the n^{th} state, we can then write

$$p_{ik}^{(n+m)} = \sum_{l \in \mathcal{S}} P_{il}^{(n)} P_{lk}^{(m)} \geq P_{ij}^{(n)} P_{jk}^{(m)} > 0$$

Similarly, show that $p_{ki}^{(r)} > 0$ for some r
Since $\leftrightarrow$ is an equivalence relation, it partitions the set of (non-ephemeral) states of the process into mutually disjoint classes, and thus:

1. Every non-ephemeral state has to be in **exactly** one class
2. i and j are in the same class iff $i \leftrightarrow j$

It is really easy to prove that this is a partition

Example 3.14 Consider the 6-state markov chain with transition matrix

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & \frac{1}{6} & \frac{1}{2} & 0 & \frac{1}{6} & \frac{1}{6} \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

State 1 is ephemeral. State $1 \to 2$, $2 \to 5$, $3 \to 4$, $5 \to 6$ $6 \to 2$ all with prob 1 But state $4 \to \{2, 3, 5, 6\}$ So $\{3, 4\}$ form a communicating class, and $\{2, 5, 6\}$ forms another communicating class. Once the Markov chain enters $\{2, 5, 6\}$ it stays

there forever - we call this a recurrent class
This notion motivates the concepts of recurrence and transience.

Transient states are revisited infinitely often, whereas transient states are only revisited a finite number of times, and is never revisited thereafter. This is important because every state of a Markov chain is one or the other.

If in the gamblers ruin case where B is infinitely rich, we let A gain a dollar every time A hit \$0, i.e. 0 becomes a reflecting barrier, rather than an absorbing state. We want to know: would state 0 be a recurrent or transient state?

Let $f_{ii}^{(n)}$ be the probability that starting from i, the first return to state i occurs at the n^{th} transition, such that for a fixed state i and for each $n \geq 1$:

$$f_{ii}^{(n)} = P(X_n = i, X_l \neq i, l = 1, \ldots, n-1 | X_0 = i)$$

Physically it makes sense that we need $f_{ii}^{(0)} = 0$ for all i, because if no transitions have occurred, then the probability of a revisit is trivially zero. Define for an arbitrary fixed state i:

$$f_{ii} = \sum_{n=1}^{\infty} f_{ii}^{(n)} \quad 0 \leq f_{ii} \leq 1$$

We see that f_{ii} is the probability that when the process starts in state i, it returns to state i after a finite number of transitions. I.e. f_{ii} is the probability of ever revisiting state i. Equivalently, f_{ii} is the probability that the process revisits i *at least once.*

We can now formally define recurrence and transience:

Definition 3.34: Recurrence and transience A state i is

$$recurrent \quad if \quad f_{ii} = 1$$

$$transient \quad if \quad f_{ii} < 1$$

$$ephemeral \quad if \quad f_{ii} = 0$$

In the following theorem, we show that if a state is recurrent, it will be occupied infinitely often with probability 1

Theorem 3.13 For a Markov process X_n, $n \geq 0$

$$P(X_n = i \text{ for infinitely many } n | X_0 = i) = \begin{cases} 1 & \text{if i is recurrent} \\ 0 & \text{if i is transient} \end{cases}$$

Proof of this theorem Let A_i^{mj} be the event that the process visits i at least m times after time $j \geq 0$ and define:

$$\begin{aligned} Q_i^m &= P(A_i^{m,0} | X_0 = i) \\ &= \text{probability that the process visits i at least m times after time zero given we started in state i} \end{aligned}$$

Let E^k be the event that the first return to state i occurs at time $k > 0$, that is:

$$E^k = \{X_k = i, X_l \neq i, l = 1, \ldots, k-1\}, \quad k > 0$$

And so, by 3.30,

$$P(E^k | X_0 = i) = f_{i,i}^{(k)}$$

Clearly the events are disjoint
By considering all possible realisations of the process for which $X_0 = i$ and the first return to state i occurs at time $k > 0$. Partitioning on the time k of first return to i we have (using the law of total prob) that

$$Q_i^m = \sum_{k=1}^{\infty} P(A_i^{m,0} | E^k, X_0 = i) P(E^k | X_0 = i) = \sum_{k=1}^{\infty} P(A_i^{m,0} | E^k, X_0 = i) f_{ii}^{(k)}$$

Note this is a bit different to the first step analysis seen in other proofs, where we partitioned on the state of the process after *one* transition. Here we have partitioned on the first return time k, but we can still use the Law of total probability

since the events E^k are disjoint.
Consider the first conditional probability $P(A_i^{m,0}|E^k, X_0 = i)$

$$\begin{aligned}
P(A_i^{m,0}|E^k, X_0 = i) &= P(A_i^{m,0}|X_k = i, X_l \neq i, l = 1, \ldots, k-1, X_0 = i) \\
&= P(A_i^{m-1,k}|X_k = i) \text{ by the markov property} \\
&= P(A_i^{m-1,0}|X_0 = i) \text{ time homogeneous} \\
&= Q_i^{m-1}
\end{aligned}$$

using (3.31) and proceeding with recursion, we get:

$$\begin{aligned}
Q_i^m = \sum_{k=1}^{\infty} Q_i^{m-1} f_{ii}^{(k)} &= Q_i^{m-1} \sum_{k=1}^{\infty} f_{ii}^{(k)} \\
&= Q_i^{m-2} f_{ii}^2 \\
&\vdots \\
&= Q_i^1 f_{ii}^{m-1}
\end{aligned}$$

By definition $Q_i^1 = P(A_i^{1,0}|X_0 = i) = f_{ii}$ and recall that f_{ii} is the probability that the process revisits state i at least once after time 0. Hence:

$$Q_i^m = f_{ii}^m$$

Taking limits then gives:

$$\lim_{m\to\infty} Q_i^m = \begin{cases} 1, & f_{ii} = 1 \text{ (recurrent)} \\ 0, & f_{ii} < 1 \text{ (transient)} \end{cases}$$

Proving the result.
Sometimes f_{ii} can be difficult to calculate, and we need a characterisation in terms of the one-step transition matrix P

Theorem 3.14 If state i is transient, then

$$\sum_{n=1}^{\infty} p_{ii}^{(n)} < \infty \text{ (it converges)}$$

If state i is recurrent then

$$\sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty \text{ (it diverges)}$$

Proof The proof involves MGFs and a few preliminaries
For $n \geq 1$,

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

This is left as an exercise (use the law of total probability and consider all possible realisations of the process for $X_0 = i$ and $X_n = i$ and the first return to state i occurs at the k^{th} transition). This expression is a convolution. These are best handled using generating functions, and so we define:

$$P_{ii}(z) = \sum_{n=0}^{\infty} p_{ii}^{(n)} z^n, \quad |z| < 1$$

$$F_{ii}(z) = \sum_{n=0}^{\infty} f_{ii}^{(n)} z^n \quad |z| < 1$$

Need to know these are well-defined.

$$\begin{aligned}
|P_{ii}(z)| \leq \sum_{n=0}^{\infty} |P_{ii}^{(n)} z^n| &= \sum_{n=0}^{\infty} P_{ii}^{(n)} |z|^n \\
&\leq \sum_{n=0}^{\infty} 1|z|^n = \frac{1}{1-|z|} \\
&< \infty \text{ for } |z| < 1
\end{aligned}$$

Therefore it is bounded. Likewise for F_{ii}:

$$\begin{aligned}
|F_{ii}(z)| &\leq \sum_{n=0}^{\infty} |f_{ii}^{(n)} z^n| \\
&\leq \sum_{n=0}^{\infty} |f_{ii}^{(n)} \quad \text{for } |z| \leq 1 \\
&= \sum_{n=1}^{\infty} |f_{ii}^{(n)} \text{ since } f_{ii}^{(0)} = 0 \text{ by def} \\
&= f_{ii} \leq 1
\end{aligned}$$

So they are well defined.
Proof of theorem cont.
Multiply (3.32) by z^n and sum from $n = 1 \to \infty$.

$$\begin{aligned}
\sum_{n=1}^{\infty} p_{ii}^{(n)} z^n &= \sum_{n=1}^{\infty} z^n \sum_{k=0}^{n} f_{ii}^{(k)} p_{ii}^{(n-k)} \\
&= \sum_{n=0}^{\infty} z^n \sum_{k=0}^{n} f_{ii}^{(k)} p_{ii}^{(n-k)} \text{ since } f_{ii}^{(0)} = 0 \\
&= (\sum_{k=0}^{\infty} f_{ii}^{(k)} z^k)(\sum_{n=0}^{\infty} p_{ii}^{(n)} z^n)
\end{aligned}$$

For the LHS, note that $p_{ii}^{(0)} = 1$ so that

$$P_{ii}(z) - 1 = F_{ii}(z) P_{ii}(z)$$

Giving:

$$P_{ii}(z) = \frac{1}{1 - F_{ii}(z)}$$

We can now prove the theorem itself. We exploit (we will justify this later)

$$\lim_{\to 1^-} F_{ii}(z) = \sum_{n=0}^{\infty} f_{ii}^{(n)} = f_{ii}$$

and $f_{i,i} < 1$ **if** it is transient

$$\begin{aligned}
f_{ii} < 1 \Leftrightarrow \sum_{n=0}^{\infty} f_{ii}^{(n)} < 1 \Leftrightarrow \lim_{\to 1^-} F_{ii}(z) < 1 \\
\Leftrightarrow \lim_{z \to 1^-} \frac{1}{1 - F_{ii}(z)} < \infty \Leftrightarrow \lim_{z \to 1^-} P_{ii}(z) < \infty \\
\Leftrightarrow \sum_{n=0}^{\infty} p_{ii}^{(n)} < \infty \Leftrightarrow \sum_{n=1}^{\infty} p_{ii}^{(n)} < \infty \\
\text{since } p_{ii}^{(0)} = 1 < \infty
\end{aligned}$$

Similarly it follows that if i is **recurrent** (i.e. $f_{ii} = 1$) then

$$\sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty$$

Theorem 3.15 In a DTMC, state i is transient if and only if

$$\sum_{n=1}^{\infty} p_{j,i}^{(n)} < \infty \quad \text{for all j}$$

Proof is left as an exercise - similar to 3.14

Example 3.15 Consider a DTMC with state space Z and transition probabilities

$$p_{ij} = \begin{cases} p \text{ for } j = i+1 \\ q \text{ for } j = i-1 \\ 0 \text{ otherwise} \end{cases}$$

So $q = 1-p$, $p > 0$, $q > 0$. **How many communicating classes are there?**
The entire state space is a single communicating class Is state 0 recurrent or transient. By theorem 3.14, state 0 is recurrent if:

$$\sum_{n=1}^{\infty} p_{00}^{(n)} = \infty$$

Transient if this is $<$ instead. So in this case: $p_{00}^{(n)} = 0$ for odd n as it cannot step back to zero in an odd number of steps. Looking at the $2n$ transitions, we require a total of n transitions to the "right" and n transitions to the "left" to get back to where we started (0). The probability of each instance of this is $p^n q^n$. We must then multiply this probability by the number of ways that these moves can occur. Therefore:

$$p_{00}^{(2n)} = \binom{2n}{n} p^n q^n$$

Hence, state 0 is transient if

$$\sum_{n=1}^{\infty} \binom{2n}{n} p^n q^n < \infty$$

and is recurrent otherwise.

Recall the ratio test from first year: The ratio test says that if

$$\lim_{n\to\infty} \left|\frac{U_{n+1}}{U_n}\right| \begin{cases} < 1 \text{ then it is convergent} \\ = 1 \text{ the test is inconclusive} \\ > 1 \text{ then it is divergent} \end{cases}$$

In this case, let $U_n = \binom{2n}{n} p^n q^n$

$$\begin{aligned} \left|\frac{U_{n+1}}{U_n}\right| &= \frac{\binom{2(n+1)}{n+1} p^{n+1} q^{n+1}}{\binom{2n}{n} p^n q^n} \\ &= \frac{\frac{2(n+1)!}{(n+1)!(n+1)!}(pq)^{(n+1)}}{\frac{(2n)!}{n!n!}(pq)^n} \\ &= \frac{(2n+2)(2n+1)}{(n+1)^2} pq \\ &= \frac{(n+1)(n+\frac{1}{2})}{(n+1)^2} 4pq \\ &= \frac{(n+\frac{1}{2})}{n+1} 4pq \end{aligned}$$

$$\lim_{n\to\infty} \left|\frac{U_{n+1}}{U_n}\right| = 4pq = 4p(1-p) \leq 1$$

So if $p \neq \frac{1}{2}$ U_n converges, and state 0 is transient. For $p = \frac{1}{2}$ we get $4pq = 1$ so the ratio test doesn't tell us anything about the convergence of the series. So we must try another test or comparison with a series for which we do know the convergence properties.

$$\begin{aligned}
\sum_{n=1}^{\infty}\binom{2n}{n}p^nq^n &= \sum_{n=1}^{\infty}\binom{2n}{n}\left(\frac{1}{2}\right)^n\left(\frac{1}{2}\right)^n \\
&= \sum_{n=1}^{\infty}\frac{2n(2n-1)(2n-2)(2n-3)\dots(1)}{n(n-1)(n-2)(n-3)\dots(1)n!}\left(\frac{1}{2}\right)^n\left(\frac{1}{2}\right)^n \\
&= \sum_{n=1}^{\infty}\frac{(2n-1)(2n-3)\dots(1)}{n!}2^n\left(\frac{1}{2}\right)^n\left(\frac{1}{2}\right)^n \\
&= \sum_{n=1}^{\infty}\frac{(n-1/2)(n-3/2)\dots(1/2)}{n(n-1)(n-2)} \\
&> \sum_{n=1}^{\infty}\frac{1}{2n} \text{ which diverges} \\
&= \infty
\end{aligned}$$

We have used $n-\frac{1}{2} > n-1$ etc. in the last step. So we have concluded that U_n is divergent if $p = q = \frac{1}{2}$. and hence 0 is recurrent.

What was special about 0 though? With little thought it is easy to see that any state $i \in Z$ could have been chosen, and the same conditions could be derived. It turns out that this is true for all **communicating classes** of any markov chain.

Theorem 3.16 Solidarity theorem In a communicating class either all states are recurrent together, or are transient together. (Solidarity property is a property shared by all states of a communicating class) **Proof** Let i, j be in the same communicating class C. We just have to prove that if i is recurrent then j is also recurrent.

$$i, j \in C \Leftrightarrow i \leftrightarrow j \text{ (i communicates with j)}$$

which implies there exists $m, n > 0$ such that

$$p_{ij}^{(m)} > 0, \quad p_{ji}^{(n)} > 0$$

Just have to prove they converge or diverge

This then implies that

$$p_{jj}^{(n+l+m)} = \sum_{h\in S}\sum_{k\in S}p_{jh}^{(n)}p_{hk}^{(l)}p_{kj}^{(m)} \geq p_{ji}^{(n)}p_{ii}^{(l)}p_{ij}^{(m)}$$

By picking the specific term in the double sum corresponding to $h = k = i$, because any term $p_{rs}^{(t)} \geq 0$
Thus:

$$\sum_{l=1}^{\infty}p_{jj}^{(n+l+m)} \geq p_{ji}^{(n)}\left(\sum_{l=1}^{\infty}p_{ii}^{(l)}\right)p_{ij}^{(m)} = \infty$$

Because the middle term on the RHS diverges (since i is assumed to be recurrent) Therefore:

$$\sum_{l=1}^{\infty}p_{jj}^{(n+l+m)} = \sum_{t=n+m+1}^{\infty}p_{jj}^{(t)} = \infty$$

$$\implies \sum_{t=1}^{\infty}p_{jj}^{(t)} = \infty$$

I.e. j is recurrent!
Comment: $\sum_{n=1}^{\infty}p_{ii}^{(n)}$ being finite/infinite is a useful theoretical way of expressing the fact that i is transient/recurrrent. However, for testing in practical situations, this will have limitations.

Theorem 3.17: A markov chain with a finite number of states has at least one recurrent class/state. **Proof:** Suppose there are N states, $j = 1, \dots, N$, then we have

$$\sum_{j=1}^{N}p_{ij}^{(n)} = 1$$

since the DTMC must be in *some* state j after n steps.
Now sum from $n=1$ to k and let $k \to \infty$

$$\lim_{k\to\infty} \sum_{n=1}^{k} \sum_{j=1}^{N} p_{ij}^{(n)} = \lim_{k\to\infty} \sum_{n=1}^{k} 1 = \infty$$

Since the RHS diverges, the LHS must diverge also. So:

$$\sum_{n=1}^{\infty} \sum_{j=1}^{N} p_{ij}^{(n)} = \infty$$

Assume now that all states are transient, such that the sum converges for all i, j, and:

$$\sum_{j=1}^{N} \sum_{n=1}^{\infty} p_{ij}^{(n)} < \infty$$

Since this is an absolutely convergent double series, the order of summation can be changed, i.e.:

$$\sum_{n=1}^{\infty} \sum_{j=1}^{N} p_{ij}^{(n)} < \infty$$

Which contradicts the previous statement Therefore the DTMC has at least one recurrent state

Definition 3.35 Irreducible DTMC A DTMC is **irreducible** if all states communicate with each other.
In other words, a Markov chain consisting of a single communicating class is called irreducible.

Example 3.16 A three state markov chain in which all possible transitions have a positive probability. It is possible to reach any state from any other. Since the markov chain is irreducible, we cannot simplify the problem by focusing on a smaller subset of recurrent states. If we added an absorbing 4th state, then it is no longer irreducible.

Theorem 3.18 An irreducible finite-state DTMC is recurrent. This is effectively a restatement of the last two theorems.

Definition 3.36: Closed communicating class A communicating class C is **closed** if

$$\sum_{j\in C} p_{ij} = 1 \text{ for all } i \in C$$

That is, no state in C has access to any state outside of C.
Absorbing states are a special case of this, consisting of just one state.

Theorem 3.19 If C is a recurrent class, then it is closed. If C is not closed, then $p_{ij} > 0$ for some $i \in C$, and $j \notin C$, so $i \to j$. We cannot have $j \to k$ for any $k \in C$, otherwise, j would be in C. (since $j \to k$, $k \to i \implies j \to i \implies j \Leftrightarrow i$) So if C is not closed, this implies there is a positive probability of never returning to i if the process leaves C via the state $j \notin C$. This means that i is transient.
Since transience is a solidarity property, then C must be transient - which is a contradiction.
The converse of this theorem is not true in general. However for a finite state space, it will hold.

5 Branching processes

A particular class of DTMC. Wide applications in biology, sociology and engineering.

Consider a population consisting of individuals able to reproduce. Suppose each individual *will*, by the end of its lifetime, have produced j new offspring with probability $p_{1,j}$ for $j \geq 0$, independently of all other individuals.
This means the lines of descent from different individuals of the same generation are independent.

Let X_0 be the number of individuals initially present (the $zero^{th}$ generation). All of their offspring form the first generation, denoted X_1. In general, let X_n denote the size of the n^{th} generation

Suppose that $p_{1,0} > 0$, or that the probability an individual producing zero offspring is greater than zero, and hence that a line of descent of an individual (and hence a population of individuals) may become extinct.

To find the probability that the line of descent from an individual eventually becomes extinct, we my model the population process on a state space:

$$\mathcal{S} = \{0, 1, 2, 3, \ldots\} \text{ where } X_n = \text{ size of the nth generation, and, } X_0 = 1$$

$$p_{ij} = P(\text{i individuals produce j offspring})$$

This process has a countable collection of states, and changes only occur at a countable collection of time points corresponding to the i^{th} generation. The process is memoryless and time homogeneous.
Hence:

- this process is a DTMC
- the probability of extinction is a hitting probability on state 0
- the time to extinction is a hitting time on state 0

Assuming that $X_0 > 0$, then the following cases arise:

1. $0 < p_{1,0} < 1$: State 0 is a class on its own, it is a recurrent and absorbing state, because if there are no individuals, there can be no offspring. In the population sence - state zero represents extinction. The remaining states form a single transient class such that there is a positive probability that the population will leave this set at the end of the n^{th} generation. That is, using the independence of each individual in the n^{th} generation: $p_{1,0}^{X_n} > 0$.
2. $p_{1,0} = 1$, the trivial case where state 0 will immediately be occupied in the first generation. All states other than 0 are ephemeral
3. $p_{1,0} = 0$: State 0 is unreachable. The states $\{1, 2, 3, \ldots\}$ are transient if $p_{11} < 1$, else they are all individual absorbing classes. But either way, there is no possibility of extinction.

We only care about case 1. If we let $U_i^{(0)}$ be the probability of eventually hitting state 0 given that we start in state $i \in \mathcal{S}$, it is also an absorption probability that represents the extinction probability of a population of size i. Independence of distinct lines of descent implies that since $U_1^{(0)}$ is the probability of any individual's line of descent becoming extinct, that:

$$U_i^{(0)} = \left(U_1^{(0)}\right)^i \text{ for all } i \geq 1$$

Clearly we also have that $U_0^{(0)} = 1$, i.e.:

$$U_i^{(0)} = \left(U_1^{(0)}\right)^i \text{ for all } i \geq 0$$

This is good because it can apply for every state i now, meaning law of total probability is on the table!
We can now use a first step analysis to consider $U_{X_0}^{(0)}$ where $X_0 = 1$. That is,

$$U_1^{(0)} = \sum_{j=0}^{\infty} p_{1j} U_j^{(0)} = \sum_{j=0}^{\infty} P_{1j} \left(U_1^{(0)}\right)^j$$

As before, look for the minimal non-negative solution to this system of equations. Now if $X_0 = 1$:

$$\lim_{n\to\infty} P(X_n = 0) = U_1^{(0)}$$

But what values can $U_1^{(0)}$ take? We know that for $0 < p_{1,0} < 1$, that the states are transient, and so these states must eventually be vacated. In fact we have that:

$$\lim_{n\to\infty} P(X_n > N) = 1 - U_1^{(0)}, \text{ for any finite value of } N$$

That is, there is a probability of $1 - U_1^{(0)}$ that the population grows without bound. Now, the state of the process at generation n, is just the sum of the number of offspring Z_i from each individual of the $(n-1)^{st}$ generation. This gives:

$$X_n = \sum_{i=1}^{X_{n-1}} Z_i$$

By conditioning on X_{n-1}, we can evaluate the expected number in the n^{th} generation.

$$E[X_n] = E\left[E[X_n|X_{n-1}]\right]$$

(This is the law of total expectation)
To make further progress with this, we need an expression for the expected number of offspring, μ, that any individual may produce. That is:

$$\mu = \sum_{j=0}^{\infty} j p_{1j}$$

Then we may write:

$$E[X_n] = E\left[E[X_n|X_{n-1}]\right] = E\left[E\left[\sum_{i=1}^{X_{n-1}} Z_i|X_{n-1}\right]\right]$$

Since this condition implies that we know X_{n-1}:

$$E[X_n] = E[\mu X_{n-1}] = \mu E[X_{n-1}]$$

Using this result we can see:

$$\begin{aligned} E[X_1] &= \mu E[X_0] = \mu \\ E[X_2] &= \mu E[X_1] = \mu^2 \\ &\vdots \\ E[X_n] &= \mu E[X_{n-1}] = \mu^n \end{aligned}$$

So what does the value of the expected number of offspring for an individual μ imply for the survival or extinction of the population?

1. $\mu < 1$ - will go extinct $U_1^{(0)} = 1$
2. $\mu > 1$ $U_1^{(0)}$ should be less than 1 (it is possible not to go extinct)
3. $\mu = 1$ Is hard to explain but will be 1 in the end

Consider:

$$\begin{aligned} E[X_n] &= \sum_{j=1}^{\infty} jP(X_n = j) \\ &\geq \sum_{j=1}^{\infty} 1P(X_n = j) = P(X_n \geq 1) \text{ (probability of not going extinct)} \end{aligned}$$

Then if $\mu < 1$ we get:

$$\lim_{n\to\infty} E[X_n] = \lim_{n\to\infty} (\mu^n) \to 0 \implies P(X_n \geq 1) \to 0$$
$$\text{Therefore } P(X_n = 0) \to 0 \text{ as } n \to \infty, \text{ so } U_1^{(0)} = 1$$

For the other two cases, where $\mu \geq 1$

Example 3.17 Binary branching process Let X_n be the number of individuals in a population at time n, where the population evolves according to the following transition probabilities $p_{1,0} = p$,
$p_{1,1} = r$,
$p_{1,2} = q$,
I.e. someone can have 0, 1, or 2 children such that $p + r + q = 1$ and $p, q > 0$ and $r \geq 0$
Using first step analysis:

$$U_1^{(0)} = p + rU_1^{(0)} + q\left(U_1^{(0)}\right)^2$$

which is quadratic in $U_1^{(0)}$ - we seek the minimal non-negative solution (as usual) Consider general quaratics $y(x) = p + rx + qx^2$ and find points for which $y(x) = x$. Clearly:

- $y(0) = p > 0$ and
- $y(1) = p + r + q = 1$.

Considering the derivative:

$$\frac{dy}{dx} = r + 2qx \left(= \sum_{j=0}^{\infty} jp_{1,j}x^{j-1} \text{ in general} \right)$$

which is the mean number of offspring μ from an individual at $x = 1$ (using the definition of the mean number of births $= r + 2q$) Summarising: we have $y(1) = 1$, $y(0) = p > 0$ and the derivative at $x = 1$ is μ. The curve is also convex.

1. $\mu > 1$: $y(x)$ crosses the line $y = x = 1$ from below to above, so $\exists x^* \in (0, 1)$ such that $y(x^*) = x^*$, which means $U_1^{(0)} < 1$.
2. $\mu = 1$: $y(x)$ is a tangent to the line $y = x = 1$ (it touches) and so $U_1^{(0)} = 1$
3. $\mu < 1$: $y(x)$ crosses the line from above to below, and so $U_1^{(0)} = 1$

In general, For $p_{1,0} < 1$

$$y(x) = \sum_{j=0}^{\infty} p_{1,j}x^j \text{ for all } x \in [0, 1]$$

As all the $p_{1,j} \geq 0$, $y(x)$ is monotonic non-decreasing on $[0, 1]$, and the derivative:

$$\frac{dy}{dx} = \sum_{j=0}^{\infty} jp_{1,j}x^{j-1} > 0 \quad \forall x \in [0, 1]$$

Hence all of the arguments employed in the binary branching process example hold for all in general where $p_{1,0} < 1$ and so:

1. $\mu > 1 \implies U_1^{(0)} < 1$
2. $\mu = 1 \implies U_1^{(0)} = 1$
3. $\mu < 1 \implies U_1^{(0)} = 1$

Example 3.19 Ternary branching processes with $0, 1, 2, 3$ offspring with probs $1/3, 1/3, 1/6, 1/6$ respectively. We wish to fin dthe prob of a line of descent becoming extinct. I.e. minimal non-negative solution to the system:

$$x = \frac{1}{3} + \frac{1}{3}x + \frac{1}{6}x^2 + \frac{1}{6}x^3$$

$x = 1$ is clearly a solution, giving us that:

$$0 = \frac{1}{6}(x - 1)(x^2 + 2x - 2)$$
$$\text{so } x = 1 \quad \textit{or} \quad x = -1 \pm \sqrt{3}$$

And so the minimal non-negative solution to this is: $U_1^{(0)} = -1 + \sqrt{3} \approx 0.732 < 1$ Note that $\mu = \frac{1}{3} + 2\frac{1}{6} + 3\frac{1}{6} = \frac{7}{6} > 1$ Which means this solution was correct.

Example 3.18 Same scenario with different probabilities $7/18, 1/3, 1/6, 1/9$ respectively. Same question:

$$x = \frac{7}{18} + \frac{1}{3}x + \frac{1}{6}x^2 + \frac{1}{9}x^3$$

$x = 1$ is clearly a solution again, giving us that:

$$0 = \frac{1}{18}(x - 1)^2(2x + 7)$$

So $x = 1$ or $-7/2$ - giving the minimal non-negative solution is $U_1^{(0)} = 1$ (check μ for confirmation)

Example 3.20 Same different probs $4/9, 1/3, 1/9, 1/9$ respectively.

$$x = \frac{4}{9} + \frac{1}{3}x + \frac{1}{9}x^2 + \frac{1}{9}x^3$$

Once again $x = 1$ is clearly a solution giving:

$$0 = \frac{1}{9}(x-1)(x^2 + 2x - 4)$$

So $x = 1$ or $x = -1 \pm \sqrt{5}$ So minimal non-negative solution to this is $U_1^{(0)} = 1$ Note that $\mu = \frac{1}{3} + \frac{2}{9} + \frac{3}{9} = \frac{8}{9} < 1$.

5.1 Periodicity

We have seen that recurrence/transience are solidarity properties of a communicating class. Another solidarity property is that *period* of a state. Have to use the following result:

Theorem 3.20: Kolmogorov Suppose W is a set of integers with the property that

$$m, n \in W \implies m + n \in W$$

Define d to be the greatest common divisor (gcd) of all elements of W. Then:

1. $nd \in W$ for all sufficiently large n
2. $d \not| m \implies m \notin W$ (i.e. if m is not exactly divisible by d, then m is not in W)

Proof omitted bc number theory

Example 3.21 $1 \to 2$, $2 \leftrightarrow 3$, $3 \leftrightarrow 4$ $4 \to 1$ Let $W = \{n : p_{1,1}^{(n)} > 0\}$ in the Markov Chain given in lectures. Then $W = \{4, 6, 8, \ldots\}$ (Observe the possible paths)
Using Kolmogorov:
The GCD of the elements of W is $d = 2$, $nd \in W$ for $n \geq 2$, and if $2 \not| m \implies m \notin W$

Now if we add a path: $2 \leftrightarrow 4$ In this case, $W = \{3, 4, 5, 6, 7, \ldots\}$
So using Kolmogorov again: The GCD is 1, $nd \in W$ for $n \geq 3$, and $1 \not| m \implies m \notin W$. Which is trivial in this case.

Definition 3.37: Period of a state The **period** $d(i)$ of a state i in an arbitrary DTMC is the GCD of the set $W = \{n : p_{i,i}^{(n)} > 0\}$ From theorem 3.20 we have:

1. $p_{i,i}^{(nd(i))} > 0$ for suff large n
2. $d(i) \not| m \implies p_{i,i}^{(m)} = 0$
   Ie the probability of being in state i after m steps is zero if $d(i) \not| m$

Solidarity is covered in the following theorem:

Theorem 3.21 If i and j belong to the same communicating class, then $d(i) = d(j)$.
Proof: Suppose i and j are both members of communicating class $\mathcal{C}$.
Then there exist $m, n > 0$ such that $p_{i,j}^{(m)} > 0$ and $p_{j,i}^{(n)} > 0$, so that: Want to show that $d(i) = d(j)$

$$\begin{aligned} p_{i,i}^{(m+ld(j)+n)} &= \sum_{h \in S} \sum_{k \in S} p_{i,h}^{(m)} p_{h,k}^{(ld(j))} p_{k,i}^{(n)} \\ &\geq p_{i,j}^{(m)} p_{j,j}^{(ld(j))} p_{j,i}^{(n)} \end{aligned}$$

The middle term on the RHS is strictly positive for all l sufficiently large, say $l \geq l_0$. and the other two are positive by assumption.
Therefore:

$$p_{i,i}^{(m+ld(j)+n)} > 0 \quad \text{for } l \geq l_0$$

Hence:

$$\begin{aligned}
&d(i) \mid (m + l_0 d(j) + n) \quad \text{and} \quad d(i) \mid (m + (l_0+1)d(j) + n) \\
\implies &d(i) \mid (m + (l_0+1)d(j) + n) - (m + l_0 d(j) + n) \\
\implies &d(i) \mid d(j)
\end{aligned}$$

Similarly we can show that $d(j) \mid d(i)$. If $d(i)$ divides $d(j)$ exactly, and vice versa, then $d(i) = d(j)$. Since this is a solidarity property, it makes sense to talk about the period of a communicating class, and hence of an irreducible Markov Chain.

Definition 3.38 : Aperiodic state State i is called **aperiodic** if it has period $d(i) = 1$. The communicating class containing i is also called **aperiodic**. Limiting behaviour of some markov chains relies on this periodicity, e.g. for**aperiodic** markov chains $\lim_{n\to\infty} p_{ii}^{(n)}$ exists. Whereas a periodic markov chain will not approach a limit.

5.2 Limiting behaviours

In general we are often interested in the behaviour of $p_{ij}^{(n)}$ as $n \to \infty$. If state i is transient, from theorem 3.14, we know that

$$\sum_{n=1}^{\infty} p_{i,i} < \infty$$

Therefore $p_{i,i}^{(n)} \to 0$ as $n \to \infty$. Otherwise the sum could not be finite. On the other hand, if i is recurrent,:

$$\sum_{n=1}^{\infty} p_{i,i} = \infty$$

But this is compatible with either:

$$\lim_{n\to\infty} p_{i,i}^{(n)} > 0$$

or

$$\lim_{n\to\infty} p_{i,i}^{(n)} = 0$$

E.g. the harmonic series $\sum_{n=1}^{\infty} \frac{1}{n} = \infty$ but $\lim_{n\to\infty} \frac{1}{n} = 0$.

So we are only interested in recurrent states for this section.

Theorem 3.22 For an irreducible recurrent DTMC of period d:

$$p_{ii}^{(n)} = 0 \text{ if } d \not| \, n$$

And

$$\lim_{n\to\infty} p_{i,i}^{(nd)} = \lambda_i = \frac{d}{\sum_{n=1}^{\infty} n f_{i,i}^{(n)}}$$

Proof omitted The first expression is a trivial consequence of periodicity, and the latter is quite 'involved'.

Instead we will illustrate it with an example:

Example 3.23 Consider the irreducible Markov chain with transition probability matrix:

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

This Markov chain has no limiting distribution, since;

$$P^2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad P^3 = P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \text{ etc.}$$

That is, $\lim_{n\to\infty} P_{i,j}^{(n)}$ DNE for any i, j. This is due to the fact that $d = 2$.

Similarly, due to the deterministic nature of this Markov chain, it is easy to see (without any calculation) that:

$$P_{i,i}^{(2n)} = 1$$

And,

$$\lim_{n\to\infty} P_{11}^{(2n)} = 1 \quad \lim_{n\to\infty} P_{22}^{(2n)} = 1$$

I.e. $\lambda_1 = \lambda_2 = 1$. So we know $d = 2$, $\sum_{n=1}^{\infty} nf_{ii}^{(n)}$

$$\sum_{n=1}^{\infty} nf_{ii}^{(n)} = \mu_i$$

Where μ_i is the mean number of steps required for the first return to i. And in this case we know $\mu_i = 2$.

This example was simple, normally we can't just calculate the λ by inspection. Most of this section will deal with:

- Interpret the physical meaning of λ_i
- Find a general method for calculating the λ_i.

Interpretation of Theorem 3.22
Imagine the process starts in state i, and a very long time later is observed (we have forgotten how long). Intuitively, suppose we define $\pi_i = P(\text{this observation is in state } i)$
We don't know whether or not a multiple of d steps has occurred - so we need to determine the *probability* that the number of transitions since time 0 is a multiple of d. To address this, we apply the principle of "maximum entropy" - where we make the assumption which maximises the uncertainty - this minimising our assumed prior knowledge of the number of transitions since 0, which we shall denote by n. That is, in the absence of any knowledge about n, we assume that all values of n (below) are equally likely:

$$n \mod d \in [0, d-1]$$

Hence, we can state:

$$P(\text{number of transitions since 0 is not a multiple of } d) = \frac{d-1}{d}$$
$$P(\text{number of transitions since 0 is a multiple of d}) = \frac{1}{d}$$

Using theorem 3.22 we have:

$$\pi_i = \frac{d-1}{d} \times 0 + \frac{1}{d} \times \lambda_i = \lambda_i/d$$

so that:

$$\pi_i = \frac{1}{\sum_{n=1}^{\infty} f_{i,i}^{(n)}} = \frac{1}{\mu_i}$$

Where μ_i = mean number of steps required for first return to i.

5.3 Positive and null recurrence

So $\pi_i = \frac{1}{\mu_i}$ can be interpreted as the mean proportion of time that the process spends in i, given it starts in i. Note, it is also the long-term probability of finding the process in state i, given it started in state i. Also for an irreducible, aperiodic ($d = 1$) Markov chain, the informal argument given above suggests that $\lambda_i = \pi_i$. There are two cases:

- $\lambda_i > 0 \implies \pi_i > 0$
  - The state is called 'positive recurrent'
  - $\mu_i < \infty$, i.e. expected return time is finite.
  - $\lim_{n\to\infty} p_{i,i}^{(nd)} > 0$
- $\lambda_i = 0 \implies \pi_i = 0$
  - The state is called 'null recurrent'
  - $\mu_i = \infty$, i.e. expected return time is infinite
  - $\lim_{n\to\infty} p_{i,i}^{(nd)} = 0$

So given a recurrent state i, we have a couple of ways to determine whether the state is positive or null recurrent.

Theorem 3.23 Null recurrence and positive recurrence are solidarity properties
Proof is left for a tutorial or CE
Because of solidarity, in order to determine whether a communicating class is transient, positive recurrent, or null recurrent, it is possible to determine the nature of just one state in the communicating class. An irreducible DTMC can therefore be called transient, positive recurrent, or null recurrent.

5.4 Solidarity Summary

- Period
- Transience
- Positive recurrence
- Null recurrence

We have investigated some properties of $\lim_{n\to\infty} p_{i,i}^{(n)}$ for a state i. Now we shall investigate some properties of $\lim_{n\to\infty} p_{i,j}^{(n)}$ for states $i \neq j$. we saw earlier that if j is transient, then $\sum_{n=1}^{\infty} p_{i,j}^{(n)} < \infty$ for any state i. In order for this series to be finite, it must be true that $\lim_{n\to\infty} p_{i,j}^{(n)} = 0$.
Now looking at the limit where j is recurrent, by first defining:

$$f_{i,j}^{(n)} = P(X_n = j, X_l \neq j, l = 1, \ldots, n-1 | X_0 = i)$$

I.e. probability of first entry into j at time n, starting from state i

This is just a generalisation of $f_{i,i}^{(n)}$. Define:

$$f_{i,j} = \sum_{n=1}^{\infty} f_{i,j}^{(n)}$$

which is the hitting probability for reaching state j, starting in state i.

Theorem 3.24 Let j be an aperiodic recurrent state in a DTMC, and let i be any other state. Then:

$$p_{i,j}^{(n)} \to f_{i,j}\lambda_j \quad as\ n \to \infty$$

Proof: A simple variant of an earlier expression (3.32) gives:

$$p_{ij}^{(n)} = \sum_{k=0}^{n} f_{ij}^{(k)} p_{jj}^{(n-k)} \text{ (law of total probability)}$$

For any i, j. We know that for any $\epsilon > 0$, there exists N such that:

$$\sum_{k=N}^{\infty} f_{ij}^{(k)} < \epsilon$$

We know this because $\sum_{k=1}^{\infty} f_{ij}^{(k)} \leq 1$ by definition, which implies $\lim_{k\to\infty} f_{ij}^{(k)} = 0$ and therefore such an N must exist.
Then for $n > N$, we have:

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

$$\text{now:} \quad p_{ij}^{(n)} = \sum_{k=0}^{N-1} f_{ij}^{(k)} p_{jj}^{(n-k)} + \sum_{k=N}^{n} f_{ij}^{(k)} p_{jj}^{(n-k)}$$

$$\sum_{k=0}^{N-1} f_{ij}^{(k)} p_{jj}^{(n-k)} \leq p_{ij}^{(n)} < \sum_{k=0}^{N-1} f_{ij}^{(k)} p_{jj}^{(n-k)} + \epsilon$$

Aside: it is possible that $\lim_{n\to\infty} p_{ij}^{(n)}$ does not exist (e.g. it could oscillate). However, $\inf_{k\geq n} p_{ij}^{(k)}$ is an increasing sequence (an increasing function of n), and the limit of this sequence does exist, which is known as the $\liminf$ of the sequence $p_{ij}^{(n)}$. Similarly, $\sup_{k\geq n} p_{ij}^{(k)}$ is a decreasing sequence (a decreasing function of n), and the limit of it exists, and is known as the $\limsup$ of the sequence.

Back to the problem:
Now take the limit $n \to \infty$ in the previous expression.
Theorem 3.22 tells us that:

$$\lim_{n\to\infty} p_{jj}^{(n-k)} = \lambda_j$$

so:

$$\begin{aligned}\sum_{k=0}^{N} f_{ij}^{(k)}\lambda_j &\le \lim_{n\to\infty} \inf p_{ij}^{(n)} \\ &\le \lim_{n\to\infty} \sup p_{ij}^{(n)} \\ &\le \sum_{k=0}^{N-1} f_{ij}^{(k)}\lambda_j + \epsilon\end{aligned}$$

So:

$$\left| \lim_{n\to\infty} \sup p_{ij}^{(n)} - \lim_{n\to\infty} \inf p_{ij}^{(n)} \right| \le \epsilon$$

For All $\epsilon > 0$ so they are equal. Hence the limit $\lim_{n\to\infty} p_{ij}^{(n)}$ exists and lets call it $\lambda_{i,j}$ So:

$$\sum_{k=0}^{N-1} f_{ij}^{(k)}\lambda_j \le \lambda_{ij} \le \sum_{k=0}^{N-1} f_{ij}^{(k)}\lambda_j + \epsilon$$

Now let $N \to \infty$

$$\sum_{k=0}^{\infty} f_{ij}^{(k)}\lambda_j \le \lambda_{ij} \le \sum_{k=0}^{\infty} f_{ij}^{(k)}\lambda_j + \epsilon$$

$$f_{ij}\lambda_j \le \lambda_{ij} \le f_{ij}\lambda_j + \epsilon$$

Where $f_{ij} = \sum_{k=1}^{\infty} f_{ij}^{(k)}$.
Therefore since ϵ is arbitrary we get:

$$\lambda_{ij} = f_{ij}\lambda_j$$

A more complicated analysis proves the similar result when j is periodic.

Corollary 3.1 For an irreducible, aperiodic, recurrent DTMC:

$$\lim_{n\to\infty} p_{ij}^{(n)} = \lambda_j$$

Proof: The DTMC is irreducible and recurrent, so $f_{ij} = 1$ for all i. Therefore it is true (quality proof).

$$\lim_{n\to\infty} p_{ij}^{(n)} = \lambda_{ij} = f_{ij}\lambda_j = \lambda_j$$

Application: Consider a K state, aperiodic, irreducible DTMC, which we have seen must be recurrent. Further we know that:

$$\sum_{j=1}^{K} p_{ij}^{(n)} = 1, \qquad i \in \mathcal{S}\ \forall n \ge 1$$

Let $n \to \infty$, to show that:

$$\sum_{j=1}^{K} \lambda_j = 1$$

which implies there is at least one j such that $\lambda_j > 0$, meaning that j is positive recurrent.

Theorem 3.25 For an aperiodic, recurrent DTMC, the λ_j satisfy:

$$\lambda_j = \sum_{i \in \mathcal{S}} \lambda_j p_{ij}, \quad j \in \mathcal{S}$$

Proof: Since it is memoryless, we can't use first step analysis. We have to do the opposite ('last step analysis' in Lewis' words).
Condition on the state of the process i after n transitions,

$$p_{jj}^{(n+1)} = \sum_{i \in \mathcal{S}} p_{ji}^{(n)} p_{ij}$$

Want to take the limit of both sides as $n \to \infty$, thus giving

$$\lambda_j = \sum_{i \in \mathcal{S}} \lambda_i p_{ij}$$

But like before, we have to be able to interchange the order of limit and summation. How can we justify it?
Aside: we have a set of numbers $f_i^{(n)}$ with the property that

$$\lim_{n \to \infty} f_i^{(n)} = f_i \quad \forall i$$

When is $\lim_{n\to\infty} \sum f_i^{(n)} = \sum f_i$

1. If the sum is over a finite number of terms
2. If $f_i^{(n)} \to f_i$ monotonically
3. If $|f_i^{(n)}| < g_i$ for all i, where $\sum_i g_i$ converges

In our case: none of those hold.
We can however use Fatou's Lemma, which states that if the $f_i^{(n)}$ are positive:

$$\lim_{n \to \infty} \inf \sum_i f_i^{(n)} \geq \sum_i f_i$$

We know that our limit exists.
In what follows we always know that our limit exists (see Corollary 3.1). So we can use Fatou's Lemma with $\liminf$ replaced by just lim.

$$\begin{aligned} 1 = \sum_{j \in \mathcal{S}} p_{ij}^{(n)} &= \lim_{n \to \infty} \sum_{j \in \mathcal{S}} p_{ij}^{(n)} \\ &\geq \sum_{j \in \mathcal{S}} \lim_{n \to \infty} p_{ij}^{(n)} \text{ using Fatou} \\ &= \sum_{j \in \mathcal{S}} \lambda_j \end{aligned}$$

Recall:

$$p_{jj}^{(n+1)} = \sum_{i \in \mathcal{S}} p_{ji}^{(n)} p_{ij}$$

Let $n \to \infty$, and use Fatou's Lemma to give:

$$\lambda_j \geq \sum_{i \in \mathcal{S}} \lim_{n \to \infty} p_{ji}^{(n)} p_{ij} = \sum_{i \in \mathcal{S}} \lambda_i p_{ij}$$

Aside (generalisation):
Now multiply by $p_{j,k}$ and sum over j to get:

$$\lambda_j \geq \sum_{j \in \mathcal{S}} \lambda_j p_{jk} \geq \sum_{j \in \mathcal{S}} \sum_{i \in \mathcal{S}} \lambda_j p_{ij} p_{jk}$$

The RHS is an absolutely convergent double series, we can change the order of summation:

$$\lambda_k \geq \sum_{i\in\mathcal{S}} \lambda_i \sum_{j\in\mathcal{S}} p_{ij}p_{jk} = \sum_{i\in\mathcal{S}} \lambda_i p_{ij}^{(2)}$$

By induction we have:

$$\lambda_k \geq \sum_{i\in\mathcal{S}} \lambda_i p_{ik}^{(n)} \ \forall n$$

end aside
We're going to try and use a proof by contradiction to force an equality.
Suppose this inequality is strict for some $n \geq 1$, and $k \in \mathcal{S}$, so that summing over k we have;

$$\sum_{k\in\mathcal{S}} \lambda_k > \sum_{k\in\mathcal{S}} \sum_{i\in\mathcal{S}} \lambda_j p_{ik}^{(n)}$$

We already know that the left hand side is ≤ 1 from earlier statements. So the series on the RHS is absolutely convergent, and we can change the order of summation to yield:

$$\lambda_k > \sum_{i\in\mathcal{S}} \lambda_j \sum_{k\in\mathcal{S}} p_{ik}^{(n)} = \sum_{i\in\mathcal{S}} \lambda_i$$

which is a contradiction. So the inequality cannot be strict, such that:

$$\lambda_k = \sum_{i\in\mathcal{S}} \lambda_i p_{ik}^{(n)} \ \forall n \geq 1, \ \forall k$$

Theorem 3.26 For an irreducible, aperiodic, **positive** recurrent DTMC:

$$\sum_{k\in\mathcal{S}} \lambda_k = 1$$

Proof: In the proof of theorem 3.25, we found that:

1. $\sum_{k\in\mathcal{S}} \lambda_k \leq 1$
2. $\lambda_k = \sum_{i\in\mathcal{S}} \lambda_i p_{ik}^{(n)} \ \forall n$

We can get the result if we take the limit $n \to \infty$ in the above expression, and justify the swap of limit and summation (we won't justify the swap because we're lazy)

$$\begin{aligned}
\lambda_k &= \lim_{n\to\infty} \sum_{i\in\mathcal{S}} \lambda_i p_{ik}^{(n)} \\
&= \sum_{i\in\mathcal{S}} \lambda_i \lim_{n\to\infty} p_{ik}^{(n)} \\
&= \sum_{i\in\mathcal{S}} \lambda_i \lambda_k
\end{aligned}$$

Therefore rearranging we see that:

$$\lambda_k(1 - \sum_{i\in\mathcal{S}} \lambda_i) = 0$$

Now for a positive recurrent state k, $\lambda_k \neq 0$. So by construction, it must be true that $\sum_{i\in\mathcal{S}} \lambda_i = 1$

5.5 Global balance equations

For the aperiodic, positive recurrent case, this result ties down to the interpretation of the λ_i as *steady state probabilities*, a.k.a *stationary probabilities* or *equilibrium probabilities*. The usual notation for these is $\lambda_i = \pi_i$. We now have a systematic way of calculating these probabilities, given by the solution to the system of linear equations known as the **global balance equations**

$$\pi_k = \sum_{i\in\mathcal{S}} \pi_i p_{ik}, \ k \in \mathcal{S}, \quad \text{subject to } \sum_i \pi_i = 1$$

Note that equilibrium in a Markov chain does not imply things stop moving. **It means certain statistical quantities settle down to constant values**.

Example 3.25 Markov Chain with transition probability matrix is

$$P = \begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix}$$

Is aperiodic when $0 < a, b < 1$. If $= 0$ then you don't move, and 1 you oscillate. The global balance equations are:

$$\pi_1 = (1-a)\pi_1 + b\pi_2 \quad and \quad \pi_2 = a\pi_1 + (1-b)\pi_2$$

Subject to $\pi_1 + \pi_2 = 1$ One of the first two equations is redundant (always true when there are a finite number of states this is shown in CE4) We can arbitrarily strike out the first equation, leaving us with two equations in two unknowns. So the solution is $\vec{\pi} = (\pi_1, \pi_2) = (\frac{b}{a+b}, \frac{a}{a+b})$

Note for a general finite state DTMC with N states, the global balance equations yield $N+1$ equations in N unknowns. Removing one of the redundant equations (you can't remove the normalisation constraint!!!!), leaves N equations in N unknowns, which can be solved with normal linear algebra.
In the last few results and the last example, we have restricted our attention to aperiodic Markov chains ($d = 1$). However with a little more work, the same results can be derived for the periodic case.
Most importantly: for an irreducible, positive recurrent DTMC with period $d > 1$, the system still has a solution. However the interpretation should be that:

$$\lim_{n\to\infty} p_{jj}^{(nd)} = \lambda_j = d\pi_j$$

$$p_{jj}^{(m)} = 0 \quad if\ d \not| m$$

Example 3.26 Recall example 3.23 where

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

We saw that this Markov chain has no limiting distribution because $d = 2$. However it does have a stationary distribution, which is given by the solution to:

$$\pi_1 = \pi_1 p_{11} + \pi_2 p_{21} = \pi_2 \quad and \quad \pi_2 = \pi_1 p_{12} + \pi_2 p_{22} = \pi_1$$

And $\pi_1 + \pi_2 = 1$ Which trivially gives $\vec{\pi} = (\pi_1, \pi_2) = (1/2, 1/2)$
Recall for this example we have $\lambda_1 = \lambda_2 = 1$, and so $\lambda_1 = d\pi_1 = 2 * \frac{1}{2}$. So it is consistent.

This makes sense, because even if a given $p_{ii}^{(n)}$ oscillates indefinitely, you can still ask 'after a long time, what is the probability π_i of finding the process in state i.
Similarly we can ask what the mean proportion of time spent in state i is

Theorem 3.27 For an irreducible DTMC, if the equations:

$$x_k = \sum_{i\in\mathcal{S}} x_i p_{ik} \quad \text{such that} \quad \sum_{i\in\mathcal{S}} x_i = 1 \quad and \quad x_i > 0\ \forall i \in \mathcal{S}$$

have a solution, then the DTMC is positive recurrent, and furthermore, the stationary distribution is given by $\pi_i = x_i$
Proof: Will only prove aperiodic since we have pretty much only cared about aperiodic for proofs. From earlier:

$$x_k = \sum_i x_i p_{ik}^{(n)} \forall n$$

$$\text{Hence: } x_k = \lim_{n\to\infty} \sum_i x_i p_{ik}^{(n)}$$

Now suppose the DTMC is either transient or null recurrent and we have already seen that $\lim_{n\to\infty} p_{ik}^{(n)} = 0$ for all $k \in S$. Therefore:

$$\begin{aligned} x_k &= \lim_{n\to\infty} \sum_i x_i p_{ik}^{(n)} \\ &= \sum_{i\in S} x_i \lim_{n\to\infty} p_{ik}^{(n)} \text{ (lebasgue dominated convergence)} \\ &= 0 \forall k \in S \end{aligned}$$

But the $LHS > 0$ since $x_k > 0$ for at least some k - this means we have a contradiction. Therefore the DTMC is positive recurrent. Since all states are positive recurrent, we know that:

$$\lim_{n\to\infty} p_{ik}^{(n)} = \pi_i > 0$$

Therefore $\forall k \in S$:

$$\begin{aligned} x_k &= \lim_{n\to\infty} \sum_i x_i p_{ik}^{(n)} \\ &= \sum_i x_i \lim_{n\to\infty} p_{ik}^{(n)} \\ &= \sum_i x_i \pi_k \\ &= \pi_k \sum_i x_i \\ &= \pi_k \end{aligned}$$

As required.
This is important because to test if an irreducible DTMC is positive recurrent, we just have to see if there is a positive solution to the global balance equations. At the same time, we get the stationary distribution.

Example 3.27 Infinite number of states: Consider the DTMC with state space Z^+ with transition probabilities $p_{i,i+1} = p$ $p_{i,i-1} = q$, $p_{0,1} = 1$ and $p_{i,j} = 0$ otherwise.

This is two gamblers with infinitely rich player B, and 0 is not an absorbing state.
The cases where the communicating class Z^+ is positive recurrent, null recurrent or transient, each have a different implication for the benefactor (what are they?). A solution to the global balance equations for this problem exists only for the case $p < q$ (positive recurrent):

$$\pi_i = \frac{1}{2p}(1 - \frac{p}{q})(\frac{p}{q})^i$$

and there is no solution otherwise.
Hint for CE - the global balance equations yield a system of difference equations, with special equations for $i = 0, 1$
try a solution of the form $\pi_i = Aw^i$.

5.6 Long term behaviour

Solving the global balance equations
There is no all-purpose method for solving the global balance equations. However some guidelines:

- **Finite number of states**, we have $N + 1$ linear equations in N unknowns (just discard one of the redundant ones!) and then use linear algebra to solve
- **Countably infinite**
  - If the transition probabilities p_{ij} **do not** depend on i for $i \geq I$, for some finite I, then we can employ difference equation methods. Another is to employ *generating functions*
  - If the probabilities p_{ij} **do** depend on i, then it is often **not possible** to solve them analytically. There are however some exceptions (characterised in the next section)

5.6.1 Partial Balance Equations

Consider an irreducible, positive recurrent Markov chain.
Partition the state space S into two sets B and B^C.

Then the global balance equations for state j (flux out = flux in)

$$\begin{aligned}
\pi_j &= \sum_{i\in S}\pi_i p_{ij} \\
\pi_j = \pi_j \sum_{i\in S} p_{ji} &= \sum_{i\in S}\pi_i p_{ij} \\
\sum_{j\in B}\pi_j \sum_{i\in S} p_{ji} &= \sum_{j\in B}\sum_{i\in S}\pi_i p_{ij} \\
\sum_{j\in B}\pi_j \left(\sum_{i\in B} p_{ji} + \sum_{i\in B^C} p_{ji}\right) &= \sum_{j\in B}\left(\sum_{i\in B}\pi_i p_{ij} + \sum_{i\in B^C}\pi_i p_{ij}\right) \\
\sum_{j\in B}\sum_{i\in B}\pi_j p_{ji} + \sum_{j\in B}\sum_{i\in B^C}\pi_j p_{ji} &= \sum_{j\in B}\sum_{i\in B}\pi_i p_{ij} + \sum_{j\in B}\sum_{i\in B^C}\pi_i p_{ij} \\
\sum_{j\in B}\sum_{i\in B^C}\pi_j p_{ji} &= \sum_{j\in B}\sum_{i\in B^C}\pi_i p_{ij} \\
\sum_{j\in B}\pi_j \sum_{i\in B^C} p_{ji} &= \sum_{j\in B}\sum_{i\in B^C}\pi_i p_{ij}
\end{aligned}$$

At this point, this is just a cosmetic change to global balance. These are known as the **partial balance equations** and have the following physical interpretation:
total flux from B to B^C = total flux from B^C to B
Global balance and partial balance equations have the same 'information' about the markov chain. But sometimes the partial balance equations are easier to solve than the global balance equations.

Example 3.28 Skipfree irreducible chain on Z The process moves up or down by one unit per transition. Partition into sets B and B^C by cutting between $j-1$ and j. Since the chain is skipfree, the only way to pass between B and B^C is by going from $j-1$ to j and vice versa
The partial balance equations yield:

$$\pi_{j-1}p_{j-1,j} = \pi_j p_{j,j-1}$$

Since j can be chosen arbitrarily, this holds for all j. These equations are known as 'cut equations' in some texts. And in particular since they only refer to two states in S they are often known as the "local balance" equations.
Compare with the global balance equations:

$$\pi_j = \pi_{j-1}p_{j-1,j} + \pi_j p_{jj} + \pi_{j+1}p_{j+1,j} \quad \forall j \in S$$

Which is a much grottier equation.

Example 3.29: Ehrenfest Diffusion Model Markov chain in physics, Ehrenfest Diffusion Model for diffusion of gas through a membrance separating two chambers. In this model, we essentually represent each chamber as an 'urn', and the gas molecules represented as balls (numbered $1,\ldots,N$) distributed between the two 'urns'.
Let X_n be the number of molecules in the right-hand chamber after the n^{th} passage through the membrane. $0 \le X_n \le N$
The physical assumption is that for some constant C:

$$\begin{aligned}
&P(\text{transition right to left}) = Ci \\
&P(\text{transition left to right}) = C(N-i)
\end{aligned}$$

Where there are i in the right, and $N-i$ in the left. That is, the rate of transitions is proportional to the "*concentration gradient*". Normalisation gives:

$$\begin{aligned}
Ci + C(N-i) &= 1 \\
\text{so that } C &= \frac{1}{N} \text{ and hence} \\
p_{i,i+1} &= \frac{N-i}{N} \\
p_{i,i-1} &= \frac{i}{N} \\
p_{i,k} &= 0 \text{ otherwise}
\end{aligned}$$

We could write down and solve the global balance equations, but the partial balance equations provide a slightly easier way:

$$
\begin{aligned}
\pi_{j-1}p_{j-1,j} &= \pi_j p_{j,j-1} \\
\pi_{j-1}\frac{N-(j-1)}{N} &= \pi_j \frac{j}{N} \\
\pi_j &= \frac{N-j+1}{j}\pi_{j-1} \\
&= \frac{N-j+1}{j}\frac{N-j+2}{j-1}\pi_{j-2} \\
&= \frac{N!}{j!(N-j)!}\pi_0 \\
&= \binom{N}{j}\pi_0
\end{aligned}
$$

To find π_0 we use the normalisation condition:

$$
\begin{aligned}
1 = \sum_{j=1}^{N}\pi_j &= \sum_{j=1}^{N}\binom{N}{j}\pi_0 \\
&= \pi_0\sum_{j=0}^{N}\binom{N}{j}1^j 1^{N-j} \\
&= \pi_0 2^N \text{ using the binomial theorem} \\
&= 1 \implies \pi_0 = \frac{1}{2^N}
\end{aligned}
$$

The solution shows that π_j has the binomial distribution:

$$
\pi_j = \frac{1}{2^N}\binom{N}{j} = \binom{N}{j}\left(\frac{1}{2}\right)^j\left(\frac{1}{2}\right)^{N-j}
$$

An interesting thing is that is N is reasonably large (even if $N = 30$), then $\binom{N}{N/2}$ is much larger than the ends $\binom{N}{0}$ and $\binom{N}{N}$. I.e. we will almost never see the system in a state with all molecules in one chamber.
The fact that this happens is a combinatorial phenomenon, and provides insight into the combinatorial aspect of statistical mechanics (thermodynamics). The above technique of 'cutting' works whenever the state diagram forms a tree, i.e. a connected graph with no closed loops. On the other hand, if the Markov chain is not skipfree, then of course the partial balance equations still hold, but they are no simpler than the global balance equations - so its a waste of time.
Now look at the concept of running a Markov chain in reverse. From this idea we will then consider a special form where we consider when a DTMC is reversible, in the sense that it appears to be the same in forward time as in reverse time. Reversibility is a useful characteristic of DTMCs, where detailed balance holds for all i, j.

5.7 Time Reversed Markov Chains

Consider an irreducible positive recurrent DTMC $\{X_n : n \in N\}$ and its associated reverse time DTMC $\{X_n^* : n \in N\}$, where the DTMC is considered in reverse time. $\implies X_n^* = X_{\tau-n}$ Time reversal is essentially mirroring with respect to a time τ, which in itself is unimportant as it only defines where the origin is in reversed time. Why do we case:

1. With the aid of reversed time DTMCs, we can often learn more about regular DTMCs
2. We can often very simply and elegantly derive results, where direct derivation might be complicated
3. Balance equations of a complex DTMC can be derived by "guessing" the reversed time DTMC.

Is the time-reversed chain Markov? Apparently yes!

$$\begin{aligned}
&P(X_m = j|X_{m+1} = i, X_{m+2} = i_2, \ldots, X_{m+k} = i_k) \\
&= \frac{P(X_m = j, X_{m+1} = i, X_{m+2} = i_2, \ldots, X_{m+k} = i_k)}{P(X_{m+1} = i, X_{m+2} = i_2, \ldots, X_{m+k} = i_k)} \\
&= \frac{P(X_m = j, X_{m+1} = i)P(X_{m+2} = i_2, \ldots, X_{m+k} = i_k|X_m = j, X_{m+1} = i)}{P(X_{m+1} = i)P(X_{m+2} = i_2, \ldots, X_{m+k} = i_k|X_{m+1} = i)} \\
&= \frac{P(X_m = j, X_{m+1} = i)P(X_{m+2} = i_2, \ldots, X_{m+k} = i_k|X_{m+1} = i)}{P(X_{m+1} = i)P(X_{m+2} = i_2, \ldots, X_{m+k} = i_k|X_{m+1} = i)} \\
&= \frac{P(X_m = j, X_{m+1} = i)}{P(X_{m+1} = i)} \\
&= P(X_m = j|X_{m+1} = i) = P^*_{ij}
\end{aligned}$$

In general, a DTMC run in reverse time will be very different from that which is run in forward time.

Consider now the reversed time probabilities p^*_{ij} of a general stationary, irreducible, positive recurrent DTMC.

$$\begin{aligned}
p^*_{ij} &= P(x_m = j|x_{m+1} = i) \\
&= \frac{P(X_m = j, X_{m+1} = i)}{P(X_{m+1} = i)} \\
&= \frac{P(x_m = j)P(x_{m+1} = i|x_m = j)}{P(x_{m+1} = i} \text{ (Bayes' rule)} \\
&= \frac{\pi_j p_{ji}}{\pi_i}
\end{aligned}$$

We know that an irreducible DTMC may be considered in reverse time, and we also have expressions for the transition probabilities in reverse time. This brings us to consider a case where the forward time DTMC is statistically indistinguishable from the revered time one. If this is this case, we say the DTMC is time reversible.

Definition 3.39 Reversibility An irreducible DTMC $X_n : n \in N$ is reversible if

$$P(X_n = i, X_{n+1} = j) = P(X_n = j, X_{n+1} = i),\ \forall i, j \in S \textit{ and } \forall n$$

$$P(X_n = j) > 0\ \forall j$$

I.e. it appears identical in reverse time

Theorem 3.28 An irreducible MC X_n is reversible iff there exists a collection of positive numbers x_j such that the detailed balance equations hold, i.e.:

$$\begin{aligned}
&x_i p_{ij} = x_j p_{ji},\ \forall i, j \in S \\
&\sum_j x_j = 1
\end{aligned}$$

Proof:
Forward: If the DTMC is reversible, by the first condition of 3.39:

$$\begin{aligned}
P(X_n = i, X_{n+1} = j) &= P(X_n = j, X_{n+1} = i) \\
P(X_{n+1} = j|X_n = i)P(X_n = i) &= P(X_{n+1} = i|X_n = j)P(X_n = j) \\
&= p_{ij}P(X_n = i) = p_{ji}P(X_n = j)
\end{aligned}$$

Now for some fixed n let $x_j = P(X_n = j)$ for all $j \in S$. which is a set of positive numbers by the second condition of Def. 3.39

$$x_i p_{ij} = x_j p_{ji} \quad \forall i, j \in s$$

$$\sum_j x_j = 1$$

The last step is to show, under the first condition of Def. 3.39, that $P(X_n = j)$ are independent of n. We thus have:

$$P(X_n = i, X_{n+1} = j) = P(X_n = j, X_{n+1} = i)$$

Sum this over all j to get

$$\sum_j P(X_n = i, X_{n+1} = j) = \sum_j P(X_n = j, X_{n+1} = i)$$
$$P(X_n = i, X_{n+1} \in S) = P(X_n \in S, X_{n+1} = i)$$
$$\implies P(X_n = i) = P(X_{n+1} = i)$$

for all i in S

Backward: Since the x_j are positive, and both

$$x_i p_{ij} = x_j p_{ji} \quad \forall i, j \in S$$
$$\sum_j x_j = 1$$

hold, then from 3.27 we see that the DTMC is positive recurrent with $x_j = \pi_j$

In the steady state $P(X_n = j) = \pi_j > 0$ so that reversing the algebra on the previous part of the proof we have that:

$$P(X_n = i, X_{n+1} = j) = P(X_n = j, X_{n+1} = i)$$

Theorem 3.29: Kolmogorov's Criterion An irreducible positive recurrent DTMC is reversible iff

$$p_{j_1,j_2} p_{j_2,j_3} \dots p_{j_{n-1},j_n} p_{j_n,j_1} = p_{j_1,j_n} p_{j_n,j_{n-1}} \dots p_{j_3,j_2} p_{j_2,j_1}$$

for all sequences of disjoint states $j_1, \dots, j_n$. **Proof:**

Forwards: If reversibility holds then we have $\pi_i p_{ij} = \pi_j p_{ji}$, $\forall i, j \in S$, And hence:

$$\pi_{j_1} p_{j_1,j_2} \pi_{j_2} p_{j_2,j_3} \dots \pi_{j_{n-1}} p_{j_{n-1},j_n} \pi_{j_n} p_{j_n,j_1} = \pi_{j_1} p_{j_1,j_n} \pi_{j_n} p_{j_n,j_{n-1}} \dots \pi_{j_3} p_{j_3,j_2} \pi_{j_2} p_{j_2,j_1}$$

Dividing both sides by the positive constant $\pi_{j_1} \pi_{j_2} \dots \pi_{j_{n-1}} \pi_{j_n}$ yields:

$$p_{j_1,j_2} p_{j_2,j_3} \dots p_{j_{n-1},j_n} p_{j_n,j_1} = p_{j_1,j_n} p_{j_n,j_{n-1}} \dots p_{j_3,j_2} p_{j_2,j_1}$$

Backwards: Consider states j, k If $p_{jk} = p_{kj} = 0$, then regardless of the choice of x_j and x_k, we have that $x_j p_{jk} = x_k p_{kj}$ as required.

Suppose the cycle condition holds:

$$p_{j_1,j_2} p_{j_2,j_3} \dots p_{j_{n-1},j_n} p_{j_n,j_1} = p_{j_1,j_n} p_{j_n,j_{n-1}} \dots p_{j_3,j_2} p_{j_2,j_1}$$

for all sequences of disjoint state $j_1, \dots j_n$. Assume now that $p_{jk} > 0$. Then irreducibility implies that there must be a disjoint cycle, with all edges market by positive probabilities, containing the edge $\{j, k\}$, as k must be able to access j. The cycle condition then implies that $p_{kj} > 0$.

So we are left with only considering the states j, k such that $p_{j,k} > 0$ and $p_{kj} > 0$ choose an arbitrary reference state i_0 and consider $j \in S$. Then irreducibility implies that there exist disjoint states $\{j_1, \dots, j_n, j_1^*, \dots, j_m^*\} \in S \backslash \{i_0, j\}$ such that:

$$p_{i_0 j_1} p_{j_1,j_2} \dots p_{j_{n-1},j_n} p_{j_n j} > 0$$
$$p_{j,j_m} p_{j_m^*,j_{m-1}^*} \dots p_{j_2^*,j_1^*} p_{j_1^*,i_0} > 0$$

The cycle condition then says that:

$$(p_{i_0,j_1} p_{j_1,j_2} \dots p_{j_{n-1},j_n} p_{j_n,j})(p_{j,j_m} p_{j_m^*,j_{m-1}^*} \dots p_{j_2^*,j_1^*} p_{j_1^*,i_0}) = (p_{i_0,j_1^*} p_{j_1^*,j_2^*} \dots p_{j_{m-1}^*,j_m^*} p_{j_m^*,j})(p_{j,j_n} p_{j_n,j_{m-1}} \dots p_{j_2,j_1} p_{j_1,i_0})$$

(can cross divide so star terms on one side and non-star on the other) Then for some positive constant B, write:

$$x_j = B \frac{p_{i_0,j_1} p_{j_1,j_2} \dots p_{j_{n-1},j_n} p_{j_n,j}}{p_{j,j_n} p_{j_n,j_{m-1}} \dots p_{j_2,j_1} p_{j_1,i_0}}$$

$$x_j = B \frac{p_{i_0,j_1^*} p_{j_1^*,j_2^*} \dots p_{j_{m-1}^*,j_m^*} p_{j_m^*,j}}{p_{j,j_m} p_{j_m^*,j_{m-1}^*} \dots p_{j_2^*,j_1^*} p_{j_1^*,i_0}}$$

By the fractions formed from the cycle condition. Hence we see that the value of $x_j, j \in S$ is independent of the path chosen to define it.
Therefore, since $p_{jk} > 0$ and $p_{kj} > 0$ we can also write:

$$x_k = B\frac{p_{i_0,j_1}p_{j_1,j_2}\cdots p_{j_{n-1},j_n}p_{j_n,j}p_{jk}}{p_{jk}p_{j,j_n}p_{j_n,j_{m-1}}\cdots p_{j_2,j_1}p_{j_1,i_0}}$$

$$x_k = \frac{1}{p_{kj}}\left(B\frac{p_{i_0,j_1}p_{j_1,j_2}\cdots p_{j_{n-1},j_n}p_{j_n,j}}{p_{j,j_n}p_{j_n,j_{m-1}}\cdots p_{j_2,j_1}p_{j_1,i_0}}\right)p_{jk}$$

$$x_k = x_j\frac{p_{jk}}{p_{kj}}$$

Hence for all $j, k \in S$

$$x_k p_{kj} = x_j p_{jk} \text{ detailed balance}$$

Now using the previous theorem, its reversible.

6 Martingales

6.1 Definition

A method of gambling in which one doubles the stakes after each loss. Stochastic processes are characterised by the dependence relationships which exist amongst the variables. In the previous section we talked about DT random processes, such that:

$$P(X_n = s|X_0 = x_0, \ldots, X_{n-1} = x_{n-1}) = P(X_n = s|X_{n-1} = x_{n-1})$$

The increments of a random process are the differences $X_s - X_t$ between its values at different times $t < s$.
The markovian description above implies that the random process has independent increments. This means that $X_s - X_t$ and $X_u - X_v$ for s, t, u, v are independent RVs whenever the two time intervals do not overlap, and, more generally, any finite number of increments assigned to pairwise non-overlapping time intervals are mutually (not just pairwise) independent.

6.2 Almost sure events and events of probability 0

In elementary examples, an RV X can only take on a finite number of distinct values from a finite state space Ω and:

$$P(A) = 0 \implies A = \phi \text{ (the impossible event)}$$

$$P(A) = 1 \implies A = \Omega \text{ (the "sure" event)}$$

More commonly however, for most applications, uncountability is the usual situation. Then, when X can take on uncountably many different values, the implications are different. E.g. if we choose an ω from $[0, 1]$ with a uniform dist, we get:

$$P(x = \omega) = 0 \text{ but } X = \omega \text{ is not impossible!}$$

$$P(x \neq \omega) = 1 \text{ but} X \neq \omega \text{ is not "sure"!}$$

In general, we therefore say that if $P(A) = 1$, then A is an almost sure event. (a.s.).
Conversely, if $P(A) = 0$ then event A is a null event
As we will see later, $X = Y$ a.s. or $P(X = Y) = 1$ implies that

$$E(X) = E(Y)$$

Lemma 3.1

$$E(T) = \sum_{k=1}^{\infty} P(T \geq k) \text{ discrete}$$

$$E(T) = \int_{t=0}^{\infty} P(T \geq t)dt \text{ continuous}$$

Proof: We proved this in tute 1.

Corollary 3.2

$$E(T) < \infty \implies P(T \geq k) \to 0 \text{ as } k \to \infty \implies P(T < \infty) = 1$$

$$E(T) < \infty \implies \sum_{k=n}^{\infty} kP(T = k) \to 0 \text{ as } n \to \infty \implies nP(T \geq n) \to 0 \text{ as } n \to \infty$$

Definition 3.40: Martingale (a) A random process $\{X_n : n \in N\}$ is a martingale if for all n

$$E(|X_n|) < \infty \text{ (absolutely cvgt)}$$

$$E(X_{n+1}|X_0, \ldots, X_n) = X_n \text{ a.s.}$$

Recall that $E(X_n)$ is a number, but notice that $E(X_n|X_0, \ldots X_{n-1})$ is a RV.
Since $E(X_n|X_n) = X_n$, trivially, we can rewrite the second as:

$$E\left[X_{n+1} - X_n|X_0, \ldots, X_n\right] = 0 \text{ a.s.}$$

There is a more general definition than this one, i.e.:

Definition 3.41 Martingale (b) A random process $\{X_n : n \in N\}$ is a martingale with respect to a random process $\{Y_n : n \in N\}$, if:

$$E[|X_n|] < \infty$$
$$E\left(X_{n+1}|Y_0, \ldots, Y_n\right) = X_n \text{ a.s.}$$

Of course in classical theory, this last equation means that:

$$E[X_{n+1}|\mathcal{F}_n] = X_n \text{ a.s.}$$

where $\mathcal{F}_n$ is the σ-algebra generated by $(Y_0, \ldots, Y_n)$ or $\sigma(Y_0, \ldots, Y_n)$. That is, X_n is *adapted* to $\mathcal{F}_n$, so X_n is measurable with respect to $\mathcal{F}_n$, for all n.
The random process $\{X_n : n \in N\}$ is said to be a martingale (with respect to $\mathcal{F}_n$)

What we need is that $\mathcal{F}_n$ is 'rich enough' (large enough) to contain all the information we want to answer questions about X.
Note that these equations imply that X_n is a measurable function of $Y_0, \ldots$. I.e.:

$$X_n = E(X_{n+1}|Y_0, \ldots, Y_n) = g(Y_0, \ldots, Y_n)$$

And $g(\cdot)$ must be measurable with respect to $Y_0, \ldots, Y_n$ by the above comments. Hence:

$$\begin{aligned} E(X_n|Y_0, \ldots, Y_n) &= E\left(g(Y_0, \ldots, Y_n)|Y_0, \ldots, Y_n\right) \\ &= g(Y_0, \ldots, Y_n) \\ &= X_n \text{ a.s.} \end{aligned}$$

Which is another way of writing the martingale property.

Definition 3.42: Expectation of a conditional expectation Let X and Y be RVs, then:

$$E\left[E[X|Y]\right] = E[Y]$$

assuming the expectations exist.
Proof: to see why this is true, let X and Y be RVs such that:

$$F_Y(y) = P(Y \leq y) \quad \textit{and} \quad f_Y(y) = \frac{dF_Y(y)}{dy}$$

Note that $P(X \leq x|Y = y)$ is well defined. Then this gives:

$$\begin{aligned} E\left[E[X|Y]\right] &= \int_y E[X|Y = y] f_Y(y) dy \\ &= \int_y \left(\int_x P(X \geq x|Y = y) dx \right) f_Y(y) dy \\ &= \int_x \int_y P(X \geq x|Y = y) f_Y(y) dy dx \\ &= \int_x P(X \geq x) dx \text{ LOTP} \\ &= E(X) \text{ undoing the step in the second line} \end{aligned}$$

Using a "Tricky trick" as Lewis put it...
Hence, we can (back to martingale stuff) say:

$$\begin{aligned} E(X_{n+1}) &= E(E(X_{n+1}|Y_0, \ldots, Y_n)) \\ &= E(X_n) \text{ using 3.42} \end{aligned}$$

Therefore, by applying this recursively, we get that:

$$E(X_n) = E(X_0) \quad \forall n$$

The martingale property, in terms of increments, is that the future increments of X_n have conditional mean zero, given the past and present values of the process.

Example 3.8 A fair game. Suppose a gambler has initial wealth X_0 and makes bets with various odds, such that the bets made are all for fair games, in the sense that the expected net gains are zero. Then the wealth of the gambler at time m, given by X_m, is a Martingale.

The random walk $\{X_n : n \in N\}$ arising in the gamblers ruin problem is an independent increment process, and if $p = \frac{1}{2}$ (which corresponds to a fair game), it is also a Martingale.
A Martingale is usually associated with the concept of fairness in gambling. **Identifying Martingales in processes is a very useful pursuit.**

Example 3.9: Sum of Indep RVs If X_j for $j \in N$ are independent RVs with $E(X_j) = 0$ and $E(|X_j|) < \infty$, then:

$$Z_n = \sum_{j=0}^{n} X_j \ for\ n \geq 0$$

is a Martingale. This follows because of the following:

$$E(|Z_n|) = E(\sum_{j=0}^{n} X_j) = E(|\sum_{j=0}^{n} X_j|) \leq \sum_{j=0}^{n} E(|X_j|) < \infty$$

$$\begin{aligned} E(Z_{n+1}|X_0, \ldots, X_n) &= E(Z_n + X_{n+1}|X_0, \ldots, X_n) \\ &= E(Z_n|X_0, \ldots, X_n) + E(X_{n+1}|X_0, \ldots, X_n) \\ &= Z_n + E(X_{n+1}|X_0, \ldots, X_n) \\ &= Z_n + E(X_{n+1}) \text{ using independence} \\ &= Z_n + 0 = Z_n \end{aligned}$$

So it is a Martingale.

Example 3.10: Product of indep RVs Same conditions except $E(X_j) = 1$, and

$$Z_n = \prod_{j=0}^{n} X_j \quad for \quad n \geq 0$$

Is a Martingale, this follows because:

$$E(|Z_n|) = E(|\prod_{j=0}^{n} X_j|) = E(\prod_{j=0}^{n} |X_j|) \leq \prod_{j=0}^{n} E(|X_j|) < \infty$$

This inequality is because expectation of products is the product of expectations plus covariances

$$\begin{aligned} E(Z_{n+1}|X_0, \ldots, X_n) &= E(Z_n X_{n+1}|X_0, \ldots, X_n) \\ &= E(Z_n|X_0, \ldots, X_n) E(X_{n+1}|X_0, \ldots, X_n) \text{ cov is 0 bc indep} \\ &= Z_n * E(X_{n+1}) = Z_n * 1 = Z_n \end{aligned}$$

Example 3.11: A marble game (Pólya's Urn) A bowl initially contains n red and m green marbles. A marble is drawn from the bowl at random, and the nreplaced along with another marble of the same colour. Hence after the i^{th} draw, there will be $n + m + i$ marbles in the bowl. If we let Y_i be the number of red marbles in the bowl at time i, then the conditional distribution of Y_{i+1} is:

$$Y_{i+1} = \begin{cases} Y_i + 1 & \text{with probability } \frac{Y_i}{n+m+i} \\ Y_i & \text{with probability } 1 - \frac{Y_i}{n+m+i} \end{cases}$$

The conditional expectation (by construction we don't have to condition on the whole past):

$$\begin{aligned} E[Y_{i+1}|Y_i] &= (Y_i + 1)\frac{Y_i}{n+m+i} + Y_i(1 - \frac{Y_i}{n+m+i} \\ &= Y_i \left(\frac{n+m+i+1}{n+m+i} \right) \end{aligned}$$

We wanted this to give $E[Y_{i+1}|Y_i] = Y_i$. The 'scaling factor' $\left(\frac{n+m+i+1}{n+m+i}\right)$ forces this not to be a martingale. So lets construct one from it! Let X_i be the proportion of red marbles in the bowl at time i, then $\{X_i i \in N\}$ is a Martingale wrt $\{Y_i|i \in N\}$. This follows because:

$$Y_0 = \frac{n}{n+m} \text{ and } X_i = \frac{Y_i}{n+m+i} \quad i \geq 1$$

And so $E[|X_i|]$ is finite for all i (in fact $0 \leq X_i \leq 1$ for all $i \geq 0$) and:

$$\begin{aligned}
E[X_{i+1}|Y_i, \dots, Y_0] &= E\left[\frac{Y_{i+1}}{n+m+i+1}|Y_i\right] \\
&= \frac{1}{n+m+i+1}E(Y_{i+1}|Y_i) \\
&= \frac{1}{n+m+i+1}Y_i\left(\frac{n+m+i+1}{n+m+i}\right) \\
&= \frac{Y_i}{n+m+i} = X_i
\end{aligned}$$

So X_i is a Martingale.

Example 3.12: Stock prices Let X_n be the closing price at the end of day n of a certain publicly traded security such as a share of stock. Now, the daily prices fluctuate, but many believe that in a **perfect** market, these price sequences $\{X_n : n \in N\}$ should be Martingales. That is, in a perfect market it should not be possible to predict with any degree of accuracy whether a future price of X_{n+1} will be higher or lower than the current price X_n. Otherwise there will be a no risk win.

Example 3.13: Branching processes Consider a population branching process as seen in the previous section. Let X_n be the size of the n^{th} generation, and let m be the mean number of offspring per individual. If we define $Z_n = \frac{X_n}{m^n}$ for $n \in N$ then Z_n is a martingale wrt X_n.

Example 3.14: Eponymous example At each toss of an unbiased coin, a bet is placed such that if a head is realised, the gambler gains as much as is bet, and loses that bet if a tail is seen. Let

$$\beta_k = \begin{cases} -1 & \text{if a tail is observed at the } k^{th} \; toss \\ +1 & \text{if a head is observed at the } k^{th} \; toss \end{cases}$$

Let the amount bet on the first game be $f_1 = 1$ unit and then define:

$$f_k = f_k(\beta_1, \dots, \beta_{k-1}), \quad for \; k \geq 2$$

As the k^{th} bet placed according to some prescribed function f_k. The net gain of the gambler after the n^{th} bet is the RV:

$$X_n = \sum_{k=1}^{n} \beta_k f_k(\beta_1, \dots, \beta_{k-1})$$

If we assume that $|f_k| < \infty$ for each k. Then $E[|X_n|]$ is finite for all $n \geq 1$ and

$$\begin{aligned}
E[X_{n+1}|\beta_1 \dots \beta_n] &= E\left[\sum_{k=1}^{n+1} \beta_k f_k(\beta_1, \dots, \beta_{k-1})|\beta_1, \dots, \beta_n\right] \\
&= \sum_{k=1}^{n} E\left[\beta_k f_k(\beta_1, \dots, \beta_{k-1})|\beta_1, \dots, \beta_n\right] + E[\beta_{n+1} f_{n+1}(\beta_1, \dots, \beta_k)|\beta_1, \dots, \beta_n)] \\
&= \sum_{k=1}^{n} \beta_k f_k(\beta_1, \dots, \beta_{k-1}) + f_{n+1}(\beta_1, \dots, \beta_k) E[\beta_{n+1}|\beta_1, \dots, \beta_n)] \\
&= \sum_{k=1}^{n} \beta_k f_k(\beta_1, \dots, \beta_{k-1}) + f_{n+1}(\beta_1, \dots, \beta_k) E[\beta_{n+1}] \\
&= X_n + f_{n+1}(\beta_1, \dots, \beta_k) E[\beta_{n+1}]
\end{aligned}$$

The second last step is since each flip is independent of the others. And if the coin is unbiased $E(\beta_{n+1}) = -1/2 + 1/2 = 0$ Which gives

$$E[X_{n+1}|\beta_1, \ldots, \beta_n] = X_n$$

And hence $\{X_n | n \geq 0\}$ is a Martingale with respect to $\{\beta_n \geq 0\}$

Consider the following strategy on the k^{th} round.

- IF you won $k-1$ bet 1 unit.
- If your last win was on hand $k-1-j$ then you bet 2^j units.

This is the eponymous martingale, and defines a messy $f_k(\beta_1, \ldots, \beta_{k-1})$ Then if the first head occurs on the T^{th} bet, the net gain X_T from the first T tosses is given by

$$X_T = \begin{cases} 1 & T = 0 \\ -(1 + 2 + 2^2 + \ldots + 2^{T-2}) + 2^{T-1} & T \geq 0 \end{cases} = 1 \forall T \geq 1$$

Note that if p is the probability that a head is realised at each toss (and the eponymous strategy is still played), then:

$$\begin{aligned} P(T = n) &= (1-p)^{n-1} p \quad p \neq 0, 1 \\ \implies P(T < \infty) &= \sum_{n=1}^{\infty} (1-p)^{n-1} p \\ &= 1 \end{aligned}$$

For every finite value of T:

$$X_T = -\sum_{j=0}^{T-2} 2^j + 2^{T-1} = 1$$

and so $E(X_T) = 1$ I.e. T is almost surely finite, and X_T is almost surely defined with $E(X_T) = 1$.
Note that in this strategy, we are assuming the gambler can lose an arbitrarily large amount before the first head occurs.
Now in contrast, for a fixed time n, in the case of an unbiased coin we have:

$$\begin{aligned} E[X_n] &= \sum_{k=1}^{n} E[\beta_k f_k(\beta_1, \ldots, \beta_{k-1})] \\ &= \sum_{k=1}^{n} E[\beta_k] E[f_k(\beta_1, \ldots, \beta_{k-1})] \text{ independence} \\ &= \sum_{k=1}^{n} 0 E[f_k(\beta_1, \ldots, \beta_{k-1})] \\ &= 0 \end{aligned}$$

6.3 Stopping times and the Optimal stopping theorem

Definition 3.43 Stopping time If $\{X_n, n \in N\}$ is a stochastic process, then a RV T is a Markov (stopping) time w.r.t $\{X_n\}$ if:

1. The range of $T \subset N$
2. For each $n \geq 0$, whether or not the event $\{T = n\}$ occurs is determined by $(X_0, X_1, \ldots, X_n)$
   If we also have:
3. $T < \infty$ a.s. ($\equiv P(T < \infty) = 1$) then T is an almost sure finite Markov Time or stopping time

The description is often given by the indicator function:

$$\begin{aligned} I_{\{T=n\}} &= I_{\{T=n\}}(X_0, \ldots, X_n) \text{ (function of X's)} \\ &= \begin{cases} 1 & \text{if event } T = n \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

Example 3.15

- The k^{th} time a process visits a collection of states K is a stopping time
- The last time a process visits a collection of states K is not in general a stopping time (needs to view future)

Elementary Properties: If S and T are stopping times for a giving process, then so are

- $S+T$
- $S \wedge T \equiv \min(S,T)$
- $S \vee T \equiv \max(S,T)$

In particular, for a given n, $T \wedge n$ is a stopping time.

Proofs:

- $$\begin{aligned} I_{\{S+T=n\}} &= \sum_{k=0}^{n} I_{\{S=k,T=n-k\}} \\ &= \sum_{k=0}^{n} I_{\{S=k\}} I_{\{T=n-k\}} \end{aligned}$$

  Since both I's are functions of $(X_0, \ldots, X_n)$ the proof is done.

- $$\begin{aligned} I_{\{S\wedge T\leq n\}} &= 1 - I_{\{S\wedge T>n\}} \\ &= 1 - I_{\{S>n\}} I_{\{T>n\}} \end{aligned}$$

  Which again is a function of the X's

- $$I_{\{S\vee T\leq n\}} = I_{\{S\leq n\}} I_{\{T\leq n\}}$$

  Which again is a function of the X's

Relate these back to martingales: We know that (USE THIS IN THE ASSIGNMENT APPARENTLY)

$$E[X_n|Y_0, \ldots, Y_n] = X_n \quad a.s.$$

$$E[X_{n+1}|Y_0, \ldots, Y_n] = X_n \quad a.s.$$

Lets assume for some given value of $k \geq 0$ that

$$E[X_{n+k}|Y_0, \ldots, Y_n] = X_n \quad a.s$$

Then

$$\begin{aligned} E[X_{n+k+1}|Y_0, \ldots, Y_n] &= E\left(E[X_{n+k+1}|Y_0, \ldots, Y_{n+k}]|Y_0, \ldots, Y_n\right) \text{ (tower property)} \\ &= E(X_{n+k}|Y_0, \ldots, Y_n) \\ &= X_n \ a.s. \text{ using the inductive hypothesis} \end{aligned}$$

Lemma 3.2 Suppose that $\{X_n : n \in N\}$ is a martingale, and T is a stopping time wrt $\{Y_n : n \in N\}$ then for each $n \geq 0$:

$$E[X_n] = E[X_{T\wedge n}] = E[X_0]$$

Proof: For $n \geq k$ we have:

$$\begin{aligned} E[X_n I_{\{T=k\}}] &= E\left[E[I_{\{T=k\}} X_n | Y_0, \dots, Y_k]\right] \\ &= E\left[I_{\{T=k\}} E[X_n | Y_0, \dots, Y_k]\right] \\ &= E[X_k I_{\{T=k\}}] \end{aligned}$$

(Gary Glonek proof)

$$\begin{aligned} E[X_{T\wedge n}] &= E\left[X_{T\wedge n}\left(I_{\{T\geq n\}} + \sum_{k=0}^{n-1} I_{\{T=k\}}\right)\right] \\ &= E\left[X_{T\wedge n} I_{\{T\geq n\}}\right] + \sum_{k=0}^{n-1} E\left[I_{\{T=k\}} X_{T\wedge n}\right] \\ &= E[I_{\{T\geq n\}} X_n] + \sum_{k=0}^{n-1} E\left[I_{\{T=k\}} X_k\right] \\ &= E[I_{\{T\geq n\}} X_n] + \sum_{k=0}^{n-1} E\left[I_{\{T=k\}} X_n\right] \\ &= E\left[X_n\left(I_{\{T\geq n\}} \sum_{k=0}^{n-1} I_{\{T=k\}}\right)\right] \\ &= E[X_n] \end{aligned}$$

Theorem 3.30 (Lebesgue Dominated Convergence Theorem (for this context))

$$if \begin{cases} Z_n \text{ is a real valued RV for all } n \geq 0 \\ |Z_n| \leq V \ a.s. \ and \ E[V] < \infty \\ Z_n \to Z \ as \ n \to \infty \end{cases}$$

Then $E[Z]$ exists and $E[Z_n] \to E[Z]$ a.s. as $n \to \infty$ This tells us when $Z_n \to Z$ implies $E[Z_n] \to E[Z]$ which is non-trivial! We need to check the second condition to show if this is okay though (thats the one which usually breaks it)

Optional Stopping Theorem Suppose $\{X_n : n \in N\}$ is a martingale and T is an a.s. finite stopping time wrt $\{Y_n : n \in N\}$ then

$$if \begin{cases} E[|X_T| < \infty \\ E[X_n I_{\{T>n\}}] \to 0 \quad as \ n \to \infty \end{cases}$$

Then $E[X_T] = E[X_0]$ **Proof:** Set $Z_n = X_T I_{\{T>n\}}$. $Z = X_T$ and $V = |X_T|$ then

- Z_n is a real RV for all $n \geq 0$
- $|Z_n| \leq V$ a.s. and $E[V] < \infty$
- $I_{\{T\leq n\}} \to 1$ a.s. as $n \to \infty$

Such that $Z_n = X_T I_{\{T>n\}} \to X_T = Z$ a.s. as $n \to \infty$ By construction, the DCT are satisfied, so that $E[Z_n] \to E[Z]$. Now we have that:

$$\begin{aligned} E[X_n I_{\{T\leq n\}}] &\to E[X_T] \text{ as } n \to \infty \\ but \ E[X_T I_{\{T\leq n\}}] &= E[X_{T\wedge n} I_{\{T\leq n\}}] \\ &= E[X_{T\wedge n}(1 - I_{\{T>n\}})] \\ &= E[X_{T\wedge n}] - E[X_{T\wedge n} I_{\{T>n\}}] \\ &= E[X_{T\wedge n}] - E[X_n I_{\{T>n\}}] \\ &= E[X_{T\wedge n}] - 0 \text{ by assumption} \\ &= E[X_0] \text{ by lemma 3.2} \end{aligned}$$

Therefore $E[X_T] = E[X_0]$
What this means is we can derive a lot of results about general discrete time markov processes.

Example 3.17 Random walk on the Integers Let $\{X_n : n \in N\}$ be an IID seq of RVs that take the values $\{-1, 1\}$ for each $n \in N$ as follows:

$$P(X_n = 1) = \frac{1}{2} = P(X_n = -1)$$

Let $S_0 = 0$ and $S_n = \sum_{i=1}^{n} X_i$ describe the position of the walk at time $n \in N$ and for $a, b \in Z^+$, we bound the walk as between $(-a, b)$. Setting $-a, b$ as absorbing states then it is also a markov chain with all transient interior states. Therefore with prob 1, the process spends only a finite time on these interior states such that $T < \infty$ is an a.s finite stopping time. Further, since a, b are finite, it is clear that

$$E[|S_T|] < \infty$$

and

$$E[S_n I_{\{T>n\}}] \to 0 \quad as\ n \to \infty$$

Hence the conditions of the OST hold, and the OST implies that

$$E[S_T] = E[S_0] = 0$$

Let $V_a = P(s_n$ reaches $-a$ before $b)$.

$$Then\ S_T = \begin{cases} -a & \text{with probability } V_a \\ b & \text{with probability } 1 - V_a \end{cases}$$

Hence:

$$\begin{aligned} E[S_T] &= -aV_a + b(1 - V_a) \\ &= E[S_0] \\ &= 0 \end{aligned}$$

Therefore

$$V_a = \frac{b}{a+b}$$

I.e. the hitting probability for a is $\frac{b}{a+b}$ This is ludicrously easier than with Markov chains. Lets now consider the RV $Z_n = S_n^2 - n$ Note that

$$S_{n+1} = S_n + X_{n+1} = \begin{cases} S_n + 1 & \text{with prob } 1/2 \\ S_n - 1 & \text{with prob } 1/2 \end{cases}$$

Show that this is a martingale:

$$\begin{aligned} E[Z_{n+1}|S_0, \ldots, S_n &= \frac{1}{2}[(S_n + 1)^2 - (n + 1)] + \frac{1}{2}[(S_n - 1)^2 - (n + 1)] \\ &= S_n^2 + 1 - n + 1 \\ &= S_n^2 - n = Z_n \end{aligned}$$

And:

$$\begin{aligned} E[|Z_n|] &= E[|S_n^2 - n|] \\ &\leq E[|S_n^2|] + E[|n|] \text{ triangle ineq} \\ &= E[S_n^2] + n \\ &= var(S_n) + E[S_n]^2 + n \\ &= var(S_n) + 0 + n \\ &= nVar(X_n) + n \\ &= n + n < \infty \end{aligned}$$

Since n is a finite number. Hence, Z_n is a martingale wrt. S_n. As above, since a, b are finite, the OST is satisfied, thus

$$E[Z_T] = E[Z_0]$$

Where

$$Z_T = \begin{cases} a^2 - T & \text{with probability } V_a \\ b^2 - T & \text{with probability } 1 - V_a \end{cases}$$

and $V_a = \frac{b}{a+b}$. Therefore:

$$\begin{aligned} 0 &= E[Z_T] \\ &= (a^2 - E[T])V_a + (b^2 - E[T])(1 - V_a) \\ &= a^2 V_a + b^2(1 - V_a) - E[T] \\ E[T] &= a^2 \left(\frac{b}{a+b}\right) + b^2 \left(1 - \frac{b}{a+b}\right) \\ &= a^2 \left(\frac{b}{a+b}\right) + b^2 \left(\frac{a}{a+b}\right) \\ &= ab \end{aligned}$$

Example 3.18 Biased random walk on the integers Let $\{X_n : n \in N\}$ be IID seq of RVs $\in \{-1, 1\}$ for each $n \in N$ $S_0 = 0$ and $S_n = \sum_{i=1}^n S_i$ position at time n.
Let $\theta_n = \frac{q}{p}^{S_n}$ for all $n \geq 0$. This is a martingale (won't bother showing) The conditions of the OST are satisfied for T since $a, b < \infty$ Therefore:

$$\begin{aligned} E[\theta_T] &= E[\theta_0] \\ &= \frac{q}{p}^0 = 1 \\ 1 = E[\theta_T] &= E\left[\left(\frac{q}{p}\right)^{S_T}\right] \\ &= V_a \left(\frac{q}{p}\right)^{-a} + (1 - V_a)\left(\frac{q}{p}\right)^b \\ \implies V_a &= \frac{1 - \left(\frac{q}{p}\right)^b}{\left(\frac{q}{p}\right)^{-a} - \left(\frac{q}{p}\right)^b} \end{aligned}$$

Consider again the biased and bounded random walk but now take $Y_n = S_n - n\mu$ where $\mu = p - q \neq 0$. Then Y_n is a martingale wrt X_n and again the OST holds so that:

$$\begin{aligned} E[Y_T] &= E[Y_0] = 0 \\ 0 = E[Y_T] &= E[S_T] - \mu E[T] \\ &= -aV_a + b(1 - V_a) - \mu E[T] \\ \implies E[T] &= \frac{b - V_a(a+b)}{\mu} \end{aligned}$$

and substitution of V_a completes the argument

OST conditions We have claimed the OST was satisfied, but we should have actually verified them. I.e. if T is an a.s. finite st then we want to know that

$$\begin{aligned} &E[|Y_T|] < \infty \\ &E[Y_n I_{\{T \geq n\}}] \to 0 \quad as\ n \to \infty \end{aligned}$$

We note that $E[|Y_T|] < \infty$ is not an immediate consequence of $E[|Y_n|] < \infty$ for all n.

Definition 3.44: Submartingale For some reason a submartingale is one where you're winning.
A random process $\{X_n : n \in N\}$ is called a submartingale with respect to a random process $\{Y_n : n \in N\}$, if

$$E[|X_n|] < \infty \text{ for all } n$$
$$E[X_{n+1}|Y_0, \ldots, Y_n] \geq X_n \ a.s. \ and$$
$$X_n \text{ is a measurable function of } Y_0, \ldots Y_n \ a.s.$$

Definition 3.45: Supermartingale For some reason a supermartingale is one where you're losing.
A random process $\{X_n : n \in N\}$ is called a supermartingale with respect to a random process $\{Y_n : n \in N\}$, if

$$E[|X_n|] < \infty \text{ for all } n$$
$$E[X_{n+1}|Y_0, \ldots, Y_n] \leq X_n \ a.s. \ and$$
$$X_n \text{ is a measurable function of } Y_0, \ldots Y_n \ a.s.$$

- Every martingale is also a submartingale and a supermartingale
- Conversely, any stochastic process that is simultaneously super and sub martingale, it is a martingale.

For a coin toss

- $p = \frac{1}{2}$ martingale
- $p < \frac{1}{2}$ supermartingale
- $p > \frac{1}{2}$ submartingale

Elementary properties of super-martingales X_n a super martingale and T is a stopping time wrt Y_n.

1. $E[X_{n+1}] = E[E[X_{n+1}|Y_0, \ldots, Y_n]] \leq E[X_n]$
2. $E[X_{n+k}|Y_0, \ldots, Y_n] \leq X_n$ a.s. for all $k \geq 0$
3. $E[X_n I_{\{T=k\}}] \leq E[X_k I_{\{T=k\}}]$ for all $n \geq k$

And for a submartingale we swap all the inequalities around.

7 Brownian Motion

8 Guest Lecture

Calculating optimal limits for transacting credit card customers.

Abstract: Credit card users can roughly be divided into 'transactors', who pay off their balance each month, and 'revolvers', who maintain an outstanding balance, on which they pay substantial interest. In this talk, we focus on modelling the behaviour of an individual transactor customer. Our motivation is to calculate an optimal credit limit from the bank's point of view. This requires an expression for the expected outstanding balance at the end of a payment period. We establish a connection with the classical newsvendor model. Furthermore, we derive the Laplace transform of the outstanding balance, assuming that purchases are made according to a marked point process and that there is a simplified balance control policy which prevents all purchases in the rest of the payment period when the credit limit is exceeded. We then use the newsvendor model and our modified model to calculate bounds on the optimal credit limit for the more realistic balance control policy that accepts all purchases that do not exceed the limit. We illustrate our analysis using a compound Poisson process example and show that the optimal limit scales with the distribution of the purchasing process, while the probability of exceeding the optimal limit remains constant. Finally, we apply our model to some real credit card purchase data.
Transactor - transacting credit card customer. Normalise so that it is a range up to 1 being the limit the person can go over. Transactors do not need to pay much interest. Given that roughly 70% are transactors, the rest are revolvers.
Revolvers come close to their limit and then just keep paying off the interest and a a portion of this range but remain close to 1. There is a percentage who do not abide to these rules and act differently. What should a bank do to control the account of a transactor? The bank gives you a credit score based on your likelihood of defaulting (done through a questionnaire). The bank keeps your purchasing data. Should banks use individuals' data rather than demographic data? What sort of credit limit should be used pp A panks profit from a transactor is driven by interchange and the cost of capital. The expected profit for an individual transactor in a single period it

$$E[R(T)] = \gamma E[B_l(T)] - \nu l$$

Where

- γ is the interchange rate
- ν is a cost for funding, and;
- $B_l(T)$ is the balance of the credit card at time T with credit limit l - more later.

We want to find

$$\hat{l} = argmax_{l\in\cap}\gamma E[B_l(T)] - \nu l$$

8.1 Newsvendor model

- $A(T)$ denote the random demand for newspapers in a single period of length T and $F_A(.)$ its distribution
- l the stock level of newspapers and;
- γ and ν are the unit profit and cost respectively.

The problem is to determine the optimal stock level

$$l* := argmax_{l\in R^+} \{\gamma E[A(T) \cap l] - \nu l\}$$

If F_A has a density f_A we can write:

$$E[A(T) \Uparrow l] = \int_0^l y f_A(y)dy + \int_l^\infty l f_A(y)dy$$

So

$$\gamma E[A(T) \Uparrow l] - \nu l = \gamma \int_0^l y f_A(y)dy - \gamma l F_A l + (\gamma - \nu)l$$

Differentiating wrt l we find a maximum is attained at $l*$ defined by

$$F_A(l*) = \frac{\gamma - \nu}{\gamma}$$

The optimal stock-out probability is ν/γ, which depends only on the mark-up used by the newsvendor.
In general, when F_A might not be absolutely continuous the solution is

$$l* = \inf_{l \in R+} \left\{ F_A(l) \geq \frac{\gamma - \nu}{\gamma} \right\}$$

This is close, but $E[A(T) \Uparrow] \neq E[B_l(T)]$ A realistic balance control policy will permit any number of purchases as long as their total value doesn't exceed l. We have a functional equation for the Laplace-Stieltjes transform of $E[B_l(T)]$ in this case, but it is difficult to work with.
Out next idea was to create a simplified balance control policy that

- Rejects a purchase that takes the outstanding balance over the limit l, and;
- freezes the outstanding balance at the level prior to the rejected purchase until the end of the period T

The balance in the real process is bounded above by that of the newsvendor model and below by that of our model. True process lies between this model and the newsvendor model (it should allow further purchases so long as they remain under l). The card-holder attempts to make purchases according to a marked point process

$$A(t) = \sum_{i=1}^{N}(t)\xi_i$$

where

- ξ_i is a sequence of non-negative indep RVs with common distribution function F
- N(t) is a counting process indep of ξ describing the number of events of $(0, t]$ in a renewal process with inter-event time distribution G.

The distributions F and G are of exponential order and all moments exist. These conditions are sufficient to ensure the existence of the Laplace-Stieltjes transforms.

$$\tilde{f}(\theta) = \int_0^\infty e^{-\theta z} F(dz)$$

$$\tilde{g}(\phi) = \int_0^\infty e^{-\phi u} G(du)$$

For $y \in [0, l]$ we want to derive an expression for

$$P(B_l(t) \in y$$

Condition Theorem. Density of the sum of two random variables gives the product of the moment generating functions.
Conditioning on the time τ and value ξ of the first purchase in $[0, t]$ either:

- There is no purchase in $[0, t]$, in which case $B_l(t) = 0$
- The first purchase at $\tau \in [0, t]$ is bigger than l. In this case the purchase is rejected and $B_l(t) = 0$
- The first purchase at $\tau \in [0, t]$ is of size $\xi \in (y, l]$. In this case, we know that $P(B_l(t) \in (y, l]) = 1$ whatever happens subsequently.
- The first purchase at $\tau \in [0, t]$ is of size $\xi \in [0, y]$ in this case, the process starts over with a remaining time $t - \tau$ until the end of the payment period and a remaining limit $l - \xi$

By the law of total probability

$$P(B_l(t) \in (y, l]) = S_l(y, t)$$

Define the three dimensional Laplace transform

Applying a transform to the integral equation we obtain:

$$\tilde{S}(\theta, \phi, \psi)$$

Data! Transactor with a \$5000 limit. Had 732 transactions from 11/11/11 to 27/2/13. No cash, foreign, declined transactions. We focus on supermarket transactions - 306 purchases over 473 days which total \$11,469.44. In a 30-day

period this works out to an average of \$727.50 in purchases, far less than the account credit limit of \$5000. Intuitively the limit is too high (even though we're just talking about supermarkets)
Supermarket time series
A Γ distribution fit to the value. with paramaters $\hat{k} = 2.8946$ and $\hat{\mu} = 0.07$ Interchange rate of 0.0054, cost of capital 0.0007 and statement period of 30 days, we calculate, would profit \$0.44 With a revised limit of \$1000 gives expected profit between \$3.17 and \$3.19. with a chance for the transaction to decline of 8%.
By lowering the limit closely to the optimal level this gives a substantial gain in profit, a decrease in the expected balance, and an increase in the probability of the card being declined. Potential ways to deal with the probability of being declined are to increase limit away from the optimal value, or to give them a lower limit but reserve a pool of funds to be used for any of a group of cardholders who exceed their limit.