UCBerkeley STAT150 Lecture Notes

Last update: 2026-01-30

Curated by Steven Huang (stvhuang.me@gmail.com)

Markov Chain

Lecture 3

Markov Chains (MC)

Most stochastic processes we'll study this semester have a special property called the **Markov Property**.

Informally, conditioned on the present state, the **past & future** behavior of a MC are **independent**.

Lecture 3 (ii)

Recall: A discrete stochastic process is a collection of RV's evolving in discrete time steps!

$$(X_n)_{n=0}^{\infty} = (X_0, X_1, X_2, \ldots)$$

Recall: Conditional probability of B given A

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

Lecture 3 (iii)

Def: (X_n) is a (discrete time) Markov Chain (MC) if

$$P(X_{n+1} = x_{n+1} \mid X_n = x_n, ..., X_0 = x_0) \\ = P(X_{n+1} = x_{n+1} \mid X_n = x_n)$$

for **all** $n \geq 1$ and $x_{n+1}, ..., x_0$

Lecture 3 (iv)

Discrete Time Markov Chains are discussed in:

- Durrett §1
- Pinsky-Karlin §3-4

Reading, looking at the examples, & doing exercises here is encouraged!

Lecture 3 (v)

In other words, when asking what the MC will do next, all that matters is its current state (not how it got there, etc.).

Def: $P\big(X_{n+1} = x_{n+1} \mid X_n = x_n\big)$ is called the **transition probability** from x_n to x_{n+1} at time $n + 1$.

Lecture 3 (vi)

In this course, we will almost only study MC's that have **time-homogenous** transition prob.'s.

This means that, for all i & j, $P\big(X_{n+1} = j \mid X_n = i\big)$ is the **same** for all times n.

So we write

$$\begin{aligned} p_{i,j} &= P(X_1 = j \mid X_0 = i) \\ &= P(X_2 = j \mid X_1 = i) \\ &= P(X_3 = j \mid X_2 = i) \\ &= \ldots \end{aligned}$$

Lecture 3 (vii)

- We write either $p_{i,j} = p(i, j)$.
- $p_{i,j}$ = prob. the MC transitions from i to j, supposing it is currently at i.

Def: The **state space** S is the collection of all possible states that the MC can visit.

$\rightarrow$ Sometimes Ω instead of S.

Lecture 3 (viii)

For a time-homogenous MC with finite state space $|S| < \infty$, we can put the transition probabilities in a matrix:

Def: Transition Matrix

$$\underline{P} = \left(p_{i,j}\right)_{i,j}$$

i.e. $\underline{P}$ is a matrix whose i, j'th[1] entry is

$$p_{i,j} = P\left(X_{n+1} = j \mid X_n = i\right)$$

(for all n)

[1] i'th row and j'th column

Lecture 3 (ix)

- If $|S| = \infty$ but countable, e.g. $S = \{0, 1, 2, ...\}$ then $\underline{P}$ is an infinite matrix.
- For now, we'll mainly focus on MC's with finite state spaces.

Note: in this case we can always encode S by $\{0, 1, ..., N\}$ or $\{1, 2, ..., N\}$ – often it is convenient to do so.

Lecture 3 (x)

Many nice examples of MC's in:

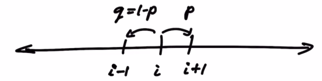
- [D] §1.1
- [PK] §3.3

Please take a look. We'll do some today — but not all of them.

Lecture 3 (xi)

Examples:

(1) Simple Random Walk (SRW) on integers $\mathbb{Z} = \{0, \pm 1, \pm 2, \ldots\}$, $p \in (0,1)$.



$$p_{i,j} = \begin{cases} p & j = i+1 \\ q & j = i-1 \\ 0 & \text{o/w} \end{cases} \quad (q = 1-p)$$

- If $p = \frac{1}{2}$, **symmetric** SRW.

Lecture 3 (xii)

I.e. each step of walk, we flip a coin that is heads w.p. p, tails w.p. $q = 1 - p$. If heads we step right, if tails we step left.

$$S_n = \sum_{k=0}^{n} \xi_k, \quad S_0 = 0$$

$$\xi_1, \xi_2, \ldots \text{ IID } \begin{cases} +1 & \text{w.p. } p \\ -1 & \text{w.p. } q \end{cases}$$

Lecture 3 (xiii)

For any n,

$$
\begin{aligned}
&P(S_{n+1} = j \mid S_n = i, S_{n-1} = s_{n-1}, \dots, S_0 = s_0) \\
&= P(S_{n+1} = j \mid S_n = i) \\
&= P(\xi_{n+1} = j - i) \\
&= \begin{cases} p & \text{for } j = i + 1 \\ q & \text{for } j = i - 1 \\ 0 & \text{o/w} \end{cases}
\end{aligned}
$$

$\therefore$ (X_n) is a time-hom. MC.

Lecture 3 (xiv)

State space $\mathbb{Z}$ is infinite:

$$\underline{P} = \begin{array}{cc} & \begin{array}{ccccc} \cdots & i-1 & i & i+1 & \cdots \end{array} \\ \begin{array}{c} \vdots \\ i-1 \\ i \\ i+1 \\ \vdots \end{array} & \begin{pmatrix} \ddots & & & & 0 \\ q & 0 & p & & \\ & q & 0 & p & \\ & & q & 0 & p \\ 0 & & & & \ddots \end{pmatrix} \end{array}$$

ALL 0's HERE

i, j'th entry = $p_{i,j}$ for $i, j \in \mathbb{Z}$.

Lecture 3 (xv)

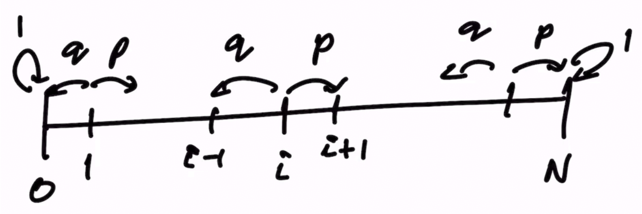
(2) Gambler's Ruin

SRW with **absorbing states** at 0 and N.

Def: A state $i \in S$ is absorbing if $p(i,i) = 1$.

I.e. once in state i, MC will stay there forever.

Lecture 3 (xvi)



$$p_{0,0} = p_{N,N} = 1$$

$$p_{i,i+1} = p, \quad 1 \le i < N$$

$$p_{i,i-1} = q, \quad 1 \le i < N$$

$$p_{i,j} = 0, \quad \text{o/w.}$$

Lecture 3 (xvii)

Called Gambler's Ruin because:

- Start with $\$X_0$.
- Each step bet \$1. Win \$1 w.p. p, lose \$1 w.p. q.
- Until either win jackpot $\$N$ or go bust \$0 — and then stop playing either way.

Lecture 3 (xviii)

Interesting Questions:

$$T = \min\{n \geq 1 : X_n \in \{0, N\}\}$$
$$= \# \text{ bets until game over}$$

For a given $1 \leq x_0 < N$,

$$\text{Question: } E_{x_0}(T)^{[2]} = E(T|X_0 = x_0)$$
$$\text{Question: } P_{x_0}(T = N) = P(T = N|X_0 = x_0)$$

I.e. expected # bets & prob. of jackpot, starting with $\$x_0$.

[2]Expected value conditioned on starting point being x_0.

Lecture 3 (xix)

- These questions will be investigated in the workshops & later this semester.
- SRW & Gambler's Ruin will be important examples which we'll come back to often.

Lecture 3 (xx)

$$\underline{P} = \begin{array}{c|cccccc} & 0 & 1 & 2 & \cdots & N-1 & N \\ \hline 0 & 1 & & & & & \\ 1 & q & 0 & p & & O & \\ 2 & & q & 0 & p & & \\ \vdots & & & & \ddots & & \\ N-1 & & O & & q & 0 & p \\ N & & & & & & 1 \end{array}$$

i, j'th entry = $p_{i,j}$ for $0 \leq i, j \leq N$.

Lecture 4

Recall: $(X_n)_{n=0}^{\infty}$ is a (time-homogenous) MC if

$$P(X_{n+1} = j \mid X_n = i, X_{n-1} = x_{n-1}, \ldots X_0 = x_0)$$
$$= P(X_{n+1} = j \mid X_n = i)$$
$$= p_{ij} \quad \text{(i.e., doesn't depend on } n\text{)}$$

for **all** n and states $i, j, x_{n-1}, \ldots, x_0$.

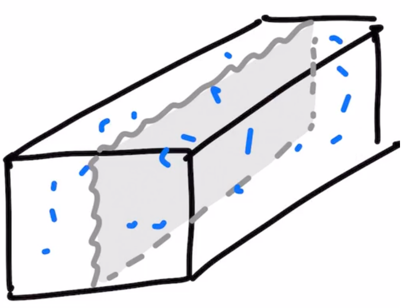
- $\underline{P} = \underbrace{p_{ij}}_{\text{tr. prob. from } i \text{ to } j} = \text{transition matrix}$

Lecture 4 (ii)

We did a couple examples ((1) SRW & (2) Gambler's ruin).
Let's see one more:

(3) Ehrenfest model:
Physicists (wife & husband):
Tatiana & Paul Ehrenfest proposed this model to explain the 2nd law of thermodynamics:
In particular, an isolated system will naturally evolve toward max entropy (least ordered).

Lecture 4 (iii)



N molecules in a chamber divided by a thin membrane. Sometimes a molecule slips through.

Let $X_n = \#$ molecules on the left side.

Lecture 4 (iv)

To get a mathematical model:

- Each step, select a molecule uniformly at random (amongst all N) and then move it to the other side.

$$\begin{aligned} p(i, i-1) &= P(\text{pick mol. from left}) \\ &= \frac{i}{N} \\ p(i, i+1) &= P(\text{pick mol. from right}) \\ &= \frac{N-i}{N} = 1 - \frac{i}{N} \end{aligned}$$

(& all other $p_{ij} = 0$)

Lecture 4 (v)

- We are often interested in the long run (LR) behavior of MC's.
- E.g., for the Ehrenfest chain, does $P(X_n = i) \overset{n\to\infty}{\to} ???$

By the 2nd law of thermodynamics, you might expect this **equilibrium distribution** to be:

$$\lim_{n\to\infty} P(X_n = i) = \frac{\overbrace{\binom{N}{i}}^{\#\text{ subsets size } i}}{\underbrace{2^N}_{\#\text{ subsets}}}, \quad 0 \leq i \leq N^{[3]}$$

[3] $\sum_{i=0}^{N} \binom{N}{i} = 2^n$ Binomial Theorem

Lecture 4 (vi)

We'll come back to this problem and LR behavior of MC's in depth, starting in [D] §1.4 and [PK] §4.

Homework:

Take a look at the other examples in [D] §1.1.

Lecture 4 (vii)

[D] §1.2: Multi-step transition prob.'s

$$P = \left(p_{ij}\right), \quad p_{ij} = P\left(X_{n+1} = j \mid X_n = i\right)$$

E.g., $X_n = \begin{cases} 0 & \text{Rainy} \\ 1 & \text{Sunny} \end{cases}$ on day n.

Suppose $\underline{P} = \begin{pmatrix} .4 & .6 \\ .7 & .3 \end{pmatrix}$

I.e., given rainy today, the probability of sunny tomorrow is 60%.

Lecture 4 (viii)

$$0 \equiv \text{Rainy}, \quad 1 \equiv \text{Sunny}$$

$$\underline{P} = \begin{pmatrix} .4 & .6 \\ .7 & .3 \end{pmatrix}$$

$$\begin{aligned} & P(\text{Sunny 2 days from now} \mid \text{Rainy today}) \\ & = P(X_{n+2} = 1 \mid X_n = 0) \\ & = P(X_{n+2} = 1, X_{n+1} = 0 \mid X_n = 0) \\ & \qquad + P(X_{n+2} = 1, X_{n+1} = 1 \mid X_n = 0)^{[4]} \\ & = p_{00}p_{01} + p_{01}p_{11} = .4(.6) + .6(.3) \end{aligned}$$

[4] LoTP

Lecture 4 (ix)

But what is $p_{00}p_{01} + p_{01}p_{11}$?

$$\underline{P} = \begin{pmatrix} .4 & .6 \\ .7 & .3 \end{pmatrix} = \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix}$$

$$\underline{P}^2 = \begin{pmatrix} p_{00}p_{00} + p_{01}p_{10} & p_{00}p_{01} + p_{01}p_{11} \\ p_{10}p_{00} + p_{11}p_{10} & p_{10}p_{01} + p_{11}p_{11} \end{pmatrix}$$

$$\left(\underline{P}^2\right)_{ij} = \sum_k p_{ik}p_{kj}$$

$$p_{00}p_{01} + p_{01}p_{11} = \left(\underline{P}^2\right)_{01}$$

Lecture 4 (x)

This holds in general:

To get m step transition prob.'s

$$p_{ij}^m = p^m(i,j) = P\big(X_{n+m} = j \mid X_n = i\big)$$

We simply take P to the mth power P^m.

Lecture 4 (xi)

Theorem

Let (X_n) be a (time-hom.) MC on state space S.

Then, for any m and $i, j \in S$,

$$\begin{aligned} p_{ij}^m &= P\big(X_{n+m} = j \mid X_n = i\big) \\ &= (P^m)_{ij} \end{aligned}$$

Lecture 4 (xii)

Useful if P can be diagonalized:

$$\underline{P} = U^{-1} D U$$

- D = Diagonal matrix of eigenvalues
- U = Rows are corr. left eigenvectors
- U^{-1} = Columns are corr. right eigenvectors

Then $\underline{P}^m = U^{-1} D^m U$

$$D = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_N \end{pmatrix}, \quad D^m = \begin{pmatrix} \lambda_1^m & & \\ & \ddots & \\ & & \lambda_N^m \end{pmatrix}$$

Lecture 4 (xiii)

Eg: 2-State MC

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

$$= \begin{pmatrix} 1 & \frac{p}{p+q} \\ 1 & -\frac{q}{p+q} \end{pmatrix} \begin{pmatrix} 1^{[5]} & 0 \\ 0 & 1-p-q \end{pmatrix} \begin{pmatrix} \frac{q}{p+q} & \frac{p}{p+q} \\ 1 & -1 \end{pmatrix}$$

Note: $|1-p-q| < 1$ unless MC is trivial:

[5]Leading eigenvalue

Lecture 4 (xiv)

Otherwise, $(1-p-q)^m \to 0$ as $m \to \infty$.

$$P^m = \begin{pmatrix} 1 & \frac{p}{p+q} \\ 1 & -\frac{q}{p+q} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & (1-p-q)^m \end{pmatrix} \begin{pmatrix} \frac{q}{p+q} & \frac{p}{p+q} \\ 1 & -1 \end{pmatrix}$$

$$\overset{m\to\infty}{\to} \begin{pmatrix} 1 & \frac{p}{p+q} \\ 1 & -\frac{q}{p+q} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{q}{p+q} & \frac{p}{p+q} \\ 1 & -1 \end{pmatrix}$$

$$= \begin{pmatrix} \frac{q}{p+q} & \frac{p}{p+q} \\ \frac{q}{p+q} & \frac{p}{p+q} \end{pmatrix} = \begin{pmatrix} \pi \\ \pi \end{pmatrix}$$

Where $\pi = (\pi_0, \pi_1)$ is the left eigenvector for the leading eigenvalue $\lambda = 1$.

Lecture 4 (xv)

This suggests that — regardless of the initial state X_0 — the limiting distribution of this 2-state MC is

$$\lim_{n \to \infty} P(X_n = i) = \pi_i$$

- We'll come back to this later (Main MC Theorem)

Lecture 4 (xvi)

Eg: $0 \equiv$ Rainy, $\quad 1 \equiv$ Sunny

$$P = \begin{pmatrix} .4 & .6 \\ .7 & .3 \end{pmatrix}, \quad p = 0.6, q = 0.7$$

$$\pi = (\pi_0, \pi_1) = \left(\frac{q}{p+q}, \frac{p}{p+q}\right)$$
$$= \left(\frac{.7}{1.3}, \frac{.6}{1.3}\right)$$

$\therefore$ About 53.8% of days rainy 😄

46.2% of days sunny 😔

Lecture 5

From last class:

Theorem. Let (X_n) be a (time-hom.) MC on state space S. Then, for any m and $i, j \in S$,

$$\begin{aligned} p_{ij}^m &= P(X_{n+m} = j \mid X_n = i) \\ &= (\underline{P}^m)_{ij} \end{aligned}$$

Proof:

Lecture 5 (ii)

This essentially follows by the Law of Total Probability (LoTP).

In this context, we call it the **Chapman-Kolmogorov equations**

$$p_{ij}^{m+n} = \sum_k p_{ik}^m p_{kj}^n$$

For all $n, m \geq 0$ & $i, j \in S$.

Lecture 5 (iii)

CK Equations

$$p_{ij}^{m+n} = \sum_k p_{ik}^m p_{kj}^n$$

(1) Why is it true?

(2) How does it prove the theorem?

Lecture 5 (iv)

(1) If MC transitions from i to j in $n + m$ steps, then it is at some state k at time n on its way there. By LoTP

$$
\begin{aligned}
&P\big(X_{n+m} = j \mid X_0 = i\big) \\
&= \sum_k P\big(X_{n+m} = j, X_n = k \mid X_0 = i\big)
\end{aligned}
$$

Lecture 5 (v)

$$= \sum_k P(X_{n+m} = j, X_n = k \mid X_0 = i)$$

$$= \sum_k P(X_{n+m} = j \mid X_n = k, X_0 = i) P(X_n = k \mid X_0 = i)^{[6]}$$

$$\overset{\text{MC}}{=} \sum_k P(X_{n+m} = j \mid X_n = k) P(X_n = k \mid X_0 = i)$$

$$\overset{\text{time hom.}}{=} \sum_k p_{kj}^m p_{ik}^n$$

[6] $P(A, B|C) = P(A|B, C)P(B|C)$

Lecture 5 (vi)

(2) Theorem follows by induction:

$$\begin{aligned}\underline{P}^m &= \underline{P}^{m-1}\underline{P} \\ &= \left(p_{ij}^{m-1}\right)\left(p_{ij}\right)\end{aligned}$$

$$\begin{aligned}\therefore \left(\underline{P}^m\right)_{ij} &= \sum_k p_{ik}^{m-1} p_{kj} \\ &= p_{ij}^m \quad \text{by CK-Eqn.}\end{aligned}$$

(Recall: $(AB)_{ij} = \sum_k A_{ik}B_{kj}$)

Lecture 5 (vii)

[D] §1.3: Classification of states

Many interesting questions about MC's involve their long-run (LR) behavior. We saw some examples last class.

Another such question is:
Which states are visited ∞ many times, which only $< \infty$ times?

Lecture 5 (viii)

First, we set up some notation:

(X_n) a MC on state space S.

- For $x \in S$ and an event A,

$$P_x(A) = P(A \mid X_0 = x)$$

  the prob. of A starting at x.

- For $y \in S$, $T_y = \min\{n \geq 1 : X_n = y\}$, time of first visit (**after** time 0) to state y.

Lecture 5 (ix)

We let

$$\rho_{yy} = P_y\big(T_y < \infty\big)$$

Be prob. the MC ever returns to y, having started at y.

$$\begin{aligned} \text{Def } y \text{ is transient if } \rho_{yy} &< 1 \\ \text{recurrent if } \rho_{yy} &= 1 \end{aligned}$$

Lecture 5 (x)

Intuitively, by Markov property, each time (X_n) visits y it starts afresh.

So

$$\begin{aligned}
\rho_{yy} &= P(\text{ever visit } y) \\
&= P(\text{visit } y \text{ at least once}) \\
\rho_{yy}^2 &= P(\text{visit } y \text{ at least twice}) \\
&\vdots \\
\rho_{yy}^k &= P(\text{at least } k \text{ visits})
\end{aligned}$$

Lecture 5 (xi)

To formalize this reasoning, we introduce the idea of a **stopping time**, which will play an important role in this course.

Def. T is a **stopping time (ST)** if only the values of $X_0, X_1, ..., X_n$ are needed to determine if the event $\{T = n\}$ has occurred.

Lecture 5 (xii)

Eg SRW. $X_0 = 0$.

- $T = T_1 = 1^{\text{st}}$ visit to 1
  Is a ST.
- $T = \min\{n \geq 1 : X_{n+1} = 1\} =$ time just before first visit to 1.
  Is **not** a ST, since eg.
  $(X_0, X_1, X_2) = (0, -1, 0)$ is not enough to determine if $T = 2$.

Lecture 5 (xiii)

The **strong Markov property** is the fact that the Markov property also holds at ST's.

Note this is not trivial, since ST's can be **random**.

Theorem $(X_n)_{n=0}^{\infty}$ a MC, T a ST. Given $T = n$ and $X_T = y$, $(X_{T+k})_{k=0}^{\infty}$ is a MC started at y.

Lecture 5 (xiv)

This is quite intuitive, so we won't prove it. — See [D] p11-12, where it is shown

$$P\big(X_{T+1} = z \mid X_T = y, T = n\big) = p_{yz}$$

Lecture 5 (xv)

By the strong MP, indeed

$$P_y\left(T_y^k < \infty\right) = P(\text{at least } k \text{ visits})$$
$$= \rho_{yy}^k$$

If $\rho_{yy} < 1$, $P_y\left(T_y^k < \infty\right) \to 0$ as $k \to \infty$.
So total # visits to y is $< \infty$ (w.p. 1).

Lecture 5 (xvi)

Why? # visits to y after time 0 is Geometric$\big(1 - \rho_{yy}\big)$:

$$P(\#\text{visits} = k) = \rho_{yy}^k\big(1 - \rho_{yy}\big), k = 0, 1, \ldots$$

$$\therefore E(\#\text{visits}) = \frac{\rho_{yy}}{1 - \rho_{yy}} < \infty$$

If $P(\#\text{visits} = \infty) > 0$, then $E(\#\text{visits}) = \infty$ (contradiction).

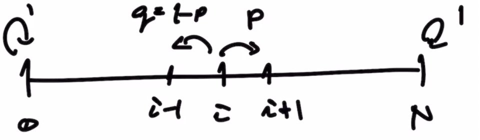
Lecture 5 (xvii)

On the other hand,

If $\rho_{yy} = 1$, then if started from y the MC will visit y ∞ many times (wp 1).

Lecture 5 (xviii)

E.g., Gambler's Ruin



0, N are clearly recurrent — any absorbing state is: starting from there, not only do you return, you can never leave!

Lecture 5 (xix)

All other states $0 < i < N$ are transient.

Note: This means (wp 1) the game will eventually end (either 0 = bankrupt or N = jackpot).[7]

Why?

[7] All states between 0 and N will only be visited finitely many times.

Lecture 5 (xx)

First note if $p = 0$ or $p = 1$ the result is obvious since starting at $0 < i < N$, the MC will never return to i.

If $p \in (0, 1)$, then

$$
\begin{aligned}
P(T_1 = \infty) &\geq q > 0 \Rightarrow 1 \text{ trans.} \\
P(T_2 = \infty) &\geq q^2 > 0 \Rightarrow 2 \text{ trans.} \\
&\vdots \\
P(T_{N-1} = \infty) &\geq q^{N-1}\,[8] > 0 \Rightarrow N-1 \text{ trans.}
\end{aligned}
$$

[8] Also $> p$

Lecture 5 (xxi)

A more surprising fact — which we'll prove later on — is that for SRW on $\mathbb{Z}$, all states are recurrent if $p = \frac{1}{2}$.

- Also true on $\mathbb{Z}^2$.
- But all states transient for SRW on $\mathbb{Z}^d$, $d \geq 3$.

"A wandering walker finds their way home, but not so for a bird".

Lecture 6

Continuing [D] §1.3 — Classification of States

Main Idea

- State space S can be split into "communication classes".
- x, y in the same class if possible to get from x to y and back again.
- States in the same class have many of the same properties — we call these "class properties".

Lecture 6 (ii)

Def: We say that y is **accessible** from x, and we write $x \to y$, if $\rho_{xy} = P_x(T_y < \infty) > 0$[9].

In other words,

$$p_{xy}^m > 0, \quad \text{for some } m > 0$$

- [D]urrett instead say "x comunicates with y" – we wouldn't do this, because the following definition is standard in probability:

[9] Time of first return to y.

Lecture 6 (iii)

Note:

$$\rho_{xy} = P_x\big(T_y < \infty\big)$$

ρ: Greek letter Rho

This is what is used in [D]urrett. It can look like a p in handwriting. I'll do my best — usually clear by context (ask me if not).

Lecture 6 (iv)

Def: We say that $x \neq y$ **communicate**, $x \leftrightarrow y$, if $x \to y$ and $y \to x$. We trivially define all $x \leftrightarrow x$.

"$\leftrightarrow$" is an **equivalence relation** on S. This means:

- a. $x \leftrightarrow x$ for all x **(Reflexive)**
- b. $x \leftrightarrow y \Rightarrow y \leftrightarrow x$ for all x, y **(Symmetric)**
- c. $x \leftrightarrow y$ & $y \leftrightarrow z \Rightarrow x \leftrightarrow z$ for all x, y, z **(Transitive)**

Lecture 6 (v)

Consequently, the state space S of any MC (X_n) is the disjoint union of "communication classes".

$$\begin{aligned} C_x &= \text{Communication class of } x^{[10]} \\ &= \{y \in S : x \leftrightarrow y\} \end{aligned}$$

Note: $x \leftrightarrow x$ and $x \leftrightarrow y \Rightarrow y \leftrightarrow x$ are trivial.

[10]It is possible some $C_x = C_y$. In fact, $C_x = C_y$ if $x \leftrightarrow y$.

Lecture 6 (vi)

We need only check that: $\leftrightarrow$ is transitive:

Lemma: Suppose $x \to y$ & $y \to z$. Then $x \to z$.[11]

Proof:

$$x \to y \Rightarrow p_{xy}^m > 0 \text{ for some } m$$

$$y \to z \Rightarrow p_{yz}^n > 0 \text{ for some } n$$

$$\therefore p_{xz}^{n+m} \geq^{[12]} p_{yz}^n p_{xy}^m > 0 \quad \square$$

[11]Same holds for $\leftrightarrow$.

[12]There may be many way from x to z in $n+m$ steps. But one ways is to first go to y in n steps and then go to z in m steps.

Lecture 6 (vii)

Eg: Gambler's Ruin

3 communication classes.

$$C_0 = \{0\}, C_N = \{N\}$$
$$C_1 = \dots = C_{N-1} = \{1, 2, \dots, N-1\}$$

Note: $x \in C_x$ for any x.
If x is absorbing then $C_x = \{x\}$.

Lecture 6 (viii)

We will see in the lectures to come that many important properties of states in a MC are "class properties" – meaning if x has this property, then so do **all** $y \in C_x$.

Lecture 6 (ix)

The following result is useful, especially when $|S| = \infty$.

Theorem. If $x \to y$ but $\rho_{yx} < 1$ then x is transient.

Intuitively, there is a chance that the MC will visit y before returning to x. And once at y, there is a chance it may never visit x in the future.

$$\therefore \rho_{xx} < 1 \Rightarrow x \text{ is transient.}$$

Lecture 6 (x)

Proof

Let $l = \min\{m : p_{xy}^m > 0\}$ be the length of the shortest path from x to y. $l < \infty$ since $x \to y$.

Note: by minimality, x is not visited on this path.

$$\therefore 1 - \rho_{xx} \geq \underbrace{p_{xy}^l}_{>0} \left(1 - \underbrace{\rho_{yx}}_{<1} \right) > 0$$

$$\Rightarrow \rho_{xx} < 1 \Rightarrow x \text{ is transient.} \square$$

Lecture 7

Def: We say that y is **accessible** from x, and we write $x \to y$, if $\rho_{xy} = P_x(T_y{}^{[13]} < \infty) > 0$.

In other words,

$$p_{xy}^m > 0 \quad \text{for some} \quad m > 0.$$

- Durrett instead say "x communicates with y" – we won't do this, because the following definition is standard in probability:

[13] Tiem of the fist return to y

Lecture 7 (ii)

Def: We say that x and y **communicate**, $x \leftrightarrow y$, if $x \to y$ and $y \to x$. We trivially define all $x \leftrightarrow x$.

"$\leftrightarrow$" is an **equivalence relation** on S. This means:

- a. $x \leftrightarrow x$ for all x **(Reflexive)**
- b. $x \leftrightarrow y \Rightarrow y \leftrightarrow x$ for all x, y **(Symmetric)**
- c. $x \leftrightarrow y$ and $y \leftrightarrow z \Rightarrow x \leftrightarrow z$ for all x, y, z **(Transitive)**

Lecture 7 (iii)

Consequently, the state space S of any MC (X_n) is the disjoint union of "communication classes".

$$\begin{aligned} C_x &= \text{Communication class of } \ x \\ &= \{y \in S : x \leftrightarrow y\}. \end{aligned}$$

Note: $x \leftrightarrow x$ and $x \leftrightarrow y \Rightarrow y \leftrightarrow x$ are trivial.

Lecture 7 (iv)

The following result is useful, especially when $|S| = \infty$.

Theorem. If $x \to y$ but $\rho_{yx} < 1$, then x is transient.

Intuitively, there is a chance that the MC will visit y before returning to x. And once at y, there is a chance it may never visit x in the future.

$$\therefore \rho_{xx} < 1 \Rightarrow x \text{ is transient.}$$

Lecture 7 (v)

From now on:

Transient $\equiv$ Trans

Recurrent $\equiv$ Rec.

This theorem immediately implies:

Corollary. If x is Rec and $x \to y$, then $\rho_{yx} = 1$.

(By theorem $\rho_{yx} = 1$, or else x is Trans.)

Lecture 7 (vi)

As you might expect:

Theorem. Trans/Rec is a class property i.e. all states in the same class are trans or rec.

We begin now working towards a proof of this.

Lecture 7 (vii)

Do example 1.14, p14 on your own.

- Draw tr. diagram.
- Find comm. classes.
- Determine rec/trans for each state (& notice that it is the same within classes).

Lecture 7 (viii)

In order to prove the theorem, we show:

Lemma. If x is rec and $x \leftrightarrow y$ then y is rec.

Note: Lemma $\Rightarrow$ Theorem.

Why?

Lecture 7 (ix)

This lemma follows by a useful characterization of recurrence:

A state y is Rec $\Leftrightarrow$

$$E_y N_y \overset{(1)}{=} \sum_{n=1}^{\infty} p_{yy}^n \overset{(2)}{=} \infty$$

$$\text{where } N_y = \#\text{ visits to } y \text{ after time } 0$$

$$= \#\{n \geq 1 : X_n = y\}.$$

Lecture 7 (x)

To see why: (1)

$$N_y = \sum_{n=1}^{\infty} \mathbf{1}_{X_n = y}, \quad \mathbf{1}_{X_n = y} = \begin{cases} 1 & \text{if} \quad X_n = y \\ 0 & \text{if} \quad X_n \neq y \end{cases}$$

By LoE, $E_x N_y = \sum_{n=1}^{\infty} P_x(X_n = y) = \sum_{n=1}^{\infty} p_{xy}^n$

In particular, if $x = y$,

$$E_y N_y = \sum_{n=1}^{\infty} p_{yy}^n$$

Lecture 7 (xi)

(2) For this, we use the SMP.

$$E_x N_y = \sum_{k=1}^{\infty} P_x\left(N_y \geq k\right)^{[14]}$$

$$= {\rho_{xy}}^{[15]} \sum_{k=1}^{\infty} \rho_{yy}^{k-1\,[16]}$$

$$= \frac{\rho_{xy}}{1 - \rho_{yy}}^{[17]}$$

[14]Review: Tail formula for a RV with values in 0, 1, 2, ...

[15]Visit 1st time

[16]Return k-1 more times (at least)

[17]Geometric sum

Lecture 7 (xii)

If $x \leftrightarrow y$:

$$E_y N_y = \frac{\rho_{yy}}{1 - \rho_{yy}}$$

Lecture 7 (xiii)

Recall $\rho_{yy} = 1 \Leftrightarrow$ Rec.

$$\therefore E_y N_y = \sum_{n=1}^{\infty} p_{yy}^n = \frac{\rho_{yy}}{1-\rho_{yy}} = \infty$$

$$\Leftrightarrow \text{Rec.}$$

Lecture 7 (xiv)

With this, we can finally prove our lemma:

Lemma. If x is Rec and $x \leftrightarrow y$ then y is Rec.

If x is Rec and $x \leftrightarrow y$, we already know from earlier that $\rho_{yx} = 1$.

$$\therefore \text{some } k, l \text{ such that}$$

$$p_{xy}^k > 0$$

$$p_{yx}^l > 0$$

Lecture 7 (xv)

To show y is Rec, we show $\sum_{n=1}^{\infty} p_{yy}^n = \infty$.

$$\begin{aligned}
\sum_{n=1}^{\infty} p_{yy}^n &\geq \sum_{n=1}^{\infty} p_{yy}^{l+n+k} \\
&\geq \sum_{n=1}^{\infty} \underbrace{p_{yx}^l}_{>0} \quad p_{xx}^n \underbrace{p_{xy}^k}_{>0} \\
&= p_{yx}^l p_{xy}^k \underbrace{\sum_{n=1}^{\infty} p_{xx}^n}_{=\infty \text{ since x is rec}} \\
&= \infty
\end{aligned}$$

$\therefore$ y is Rec. $\square$

Lecture 7 (xvi)

[D]urrett p15-17 prove these previous results + more:

Def: A subset $C \subset S$ of the state space is **closed** if you can't get out once in:

$$i \in C, j \notin C \Rightarrow p_{ij} = 0$$

Def: A set $A \subset S$ is **irreducible** if for all $i, j \in A \Rightarrow i \leftrightarrow j$.

Lecture 7 (xvii)

Durrett (see Theorem 1.7) shows:

If C is **finite**, **closed** and **irreducible**, then all states in C are Rec.

To prove this now, using what we already know, we just need to make the following observation:

Lecture 7 (xviii)

Obs: Suppose $C \subset S$ is **finite** and **closed**. Then at least one state is recurrent.

This is sort of obvious. Once in C you are stuck there forever. There are only $|C| < \infty$ states. So if the MC runs forever (∞ many steps), at least one state must be visited ∞ many times. (See p17 for formal proof.)

Lecture 7 (xix)

Another class property

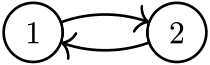
Def: The **period** of a state x is

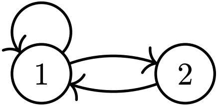
$$\text{g.c.d.}^{[18]}\ \{m \geq 1 : p_{xx}^m > 0\}$$

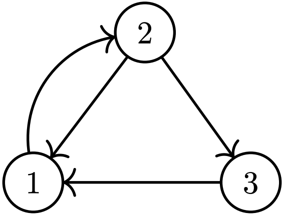
- A state x is **aperiodic** if its period is 1.
- An MC is **aperiodic** if all of its states are aperiodic.

[18] greatest common divisor

Lecture 7 (xx)

- 

1, 2 have period 2
- 

1, 2 aperiodic
- 

1, 2, 3 aperiodic, $p_{33}^3 > 0, p_{33}^5 > 0, \gcd\{3, 5, ...\} = 1$

Lecture 7 (xxi)

Theorem Aperiodicity is a class property.

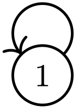
I.e. all states in the same comm. class have the same period.

Proof. Think about it yourself. Might be on HW or in workshop.

Lecture 7 (xxii)

A useful observation

If in the tr. diagram there is a “loop” at x:



[i.e. $p_{xx} > 0$][19]

Then $C_x =$ class of x is aperiodic.

Why?

[19]Pos. entry on diagonal of P.

Lecture 7 (xxiii)

[D] §1.4 – Stationary distributions

The main result here is that if (X_n) on $|S| < \infty$:

1. Aperiodic
2. Irreducible (only 1 class)

Then (X_n) has a SD π:

$$\lim_{n \to \infty} p_{ij}^n = \pi_j$$ [20][21]

We'll prove this later. For now we just assume it.

---

[20] $\pi = (\pi_1, \pi_2, ..., \pi_m), |S| = m$

[21] = LR prop. of time spent in state j, does not depend on X_0.

Lecture 8

[D] §1.4-1.6: Stationary Distributions

Main result here is that if

1. $|S| < \infty$
2. Aperiodic (all states period 1)
3. Irreducible ($i \leftrightarrow j$ for all $i, j \in S$)

Then (X_n) has a stationary distribution $\pi = \pi \underline{P}$ and

$$\lim_{n\to\infty} p_{ij}^n = \pi_j$$

Moreover,

Lecture 8 (ii)

Let $N_n(j)$ = # visits to j by time n

$E_j T_j$ = mean return time to j

$$\lim_{n\to\infty} \frac{N_n(j)}{n} = \frac{1}{E_j T_j} = \pi_j$$

$$\begin{aligned} \text{Therefore, } \pi_j &= \text{long-run proportion of time in } j \\ &= \text{inverse mean return time to } j. \end{aligned}$$

So π tells us about the long-run behavior of a MC.

Lecture 8 (iii)

- We will prove these results later (§1.7)
- For now we study their consequences through some examples
- Markov chains with $|S| = \infty$ can have stationary distributions but the situation is more complicated (§1.10)

Lecture 8 (iv)

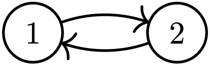
Important to note:

$$\lim_{n\to\infty} \underbrace{p_{ij}^n}_{\text{depends on } i \ \& \ j} = \underbrace{\pi_j}_{\text{does not depend on } i}$$

I.e. dependence on initial state X_0 decreases as $n \to \infty$.

Lecture 8 (v)

May **not** be true if (X_n) is periodic or reducible:

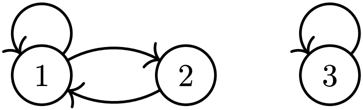


(periodic)

$$p_{11}^n = \begin{cases} 1 & \text{if } n \text{ is even} \\ 0 & \text{if } n \text{ is odd} \end{cases}$$

$$\therefore \lim_{n \to \infty} p_{11}^n \text{ does not exist!}$$

Lecture 8 (vi)



(reducible)

$\lim_{n\to\infty} p_{ij}^n$ depends on i.

$$p_{33}^n = 1 \text{ for all } n$$
$$p_{13}^n = p_{23}^n = 0 \text{ for all } n$$

$\therefore \lim_{n\to\infty} p_{ij}^n$ depends on which class i is from.

Lecture 8 (vii)

[22]Suppose we start (X_n) randomly,
i.e. X_0 is a RV on $S = \{1, 2, ..., m\}$
some distribution:

$$q_i = P(X_0 = i), \quad 1 \leq i \leq m$$

Then what is $P(X_n = j)$? $\quad 1 \leq j \leq m$

---

[22]Why $\pi = \pi \underline{P}$?

Lecture 8 (viii)

$$\begin{aligned} P(X_n = j) &= \sum_{i=1}^{m} P(X_n = j, X_0 = i) \quad \text{(LoTP)} \\ &= \sum_{i=1}^{m} P(X_0 = i) P(X_n = j \mid X_0 = i) \\ &= \sum_{i=1}^{m} q_i p_{ij}^{n} \\ &= \left(q \underline{P}^{n}\right)_j \end{aligned}$$

$$\text{Where } q = (q_1, q_2, \ldots, q_m)$$

$$\underline{P} = \left(p_{ij}\right)$$

Lecture 8 (ix)

Recall that stationary distribution $\pi = (\pi_1, \ldots, \pi_m)$ has property that

$$\lim_{n\to\infty} p_{ij}^n = \pi_j, \quad \text{for any } i$$

It follows that $\lim_{n\to\infty} P(X_n = j) = \pi_j$ no matter what q is:

$$P(X_n = j) = \sum_i q_i p_{ij}^n$$

$$\therefore \lim_{n\to\infty} P(X_n = j) = \sum_i q_i \pi_j = \pi_j \overbrace{\sum_i q_i}^{=1} = \pi_j$$

Lecture 8 (x)

Q:

Now, what happens if we start (X_n) in from its stationary distribution: $X_0 \sim \pi$

I.e: $\pi_i = P(X_0 = i)$?

Lecture 8 (xi)

$$P(X_n = j) = \sum_i P(X_{n-1} = i) p_{ij} \quad \text{(LoTP+Cond. prob.)}$$

Take $n \to \infty$:

$$\pi_j = \sum_i \pi_i p_{ij}$$ [23]

That is,

$$\pi = \pi \underline{P}$$ [24]

---

[23] $= (\pi \underline{P})_j$

[24] Probably the most important formula in §1.

Lecture 8 (xii)

Note $\pi = \pi \underline{P} \Rightarrow$

$$\Rightarrow \pi \underline{P}^2 = (\pi \underline{P})\underline{P} = \pi \underline{P} = \pi$$

etc.

$$\Rightarrow \pi = \pi \underline{P}^n$$

for any n.

$$\Rightarrow \text{if } X_0 \sim \pi, \text{then } P(X_n = i) = \pi_i$$

Lecture 8 (xiii)

In other words:

If we start (X_n) from π, i.e. $X_0 \sim \pi$, then (X_n) stays in π forever:

$$X_0 \sim \pi \Rightarrow X_n \sim \pi$$

for all n.

For this reason, we also sometimes call π the **equilibrium**.

Lecture 8 (xiv)

There are many simple examples in §1.4 for you to look at.

- Can always try to find stationary distribution π by solving the system of equations $\pi = \pi \underline{P}$ directly. Sometimes this is doable.
- However, often this is complicated. Several important special cases where π can be found are in §1.6.

Lecture 8 (xv)

A special case where it is easy to find π:

For any $\underline{P}$, rows sum to 1:

$$\sum_j p_{ij} = 1, \quad \text{for any } i$$

Def: $\underline{P}$ is doubly stochastic if also columns sum to 1:

$$\sum_i p_{ij} = 1, \quad \text{for any } j$$

Lecture 8 (xvi)

If $|S| = m$ and $\underline{P}$ doubly stochastic then $\pi = \left(\frac{1}{m}, \ldots, \frac{1}{m}\right)$.

I.e. stationary distribution is uniform on S:

$$(\pi \underline{P})_j = \sum_i \frac{1}{m} p_{ij} = \frac{1}{m} \sum_i p_{ij}{}^{[25]} = \frac{1}{m} = \pi_j$$

$$\therefore \pi = \pi \underline{P} \quad \square$$

- Other important special cases where π can be found are in §1.6.

[25] $\sum_i p_{ij} = 1$ since $\underline{P}$ is doubly stochastic.

Lecture 8 (xvii)

Detailed Balance Equations

Suppose that for some $\pi = (\pi_1, \ldots, \pi_m)$

$$\pi_i p_{ij} = \pi_j p_{ji} \quad \text{for all } i, j \quad \text{(DB)}$$

Then $\pi = \pi \underline{P}$.

$$(\pi \underline{P})_j = \sum_i \pi_i p_{ij} \stackrel{\text{DB}}{=} \sum_i \pi_j p_{ji}$$

$$= \pi_j \overbrace{\sum_i p_{ji}}^{=1} = \pi_j$$

$$\Rightarrow \pi \underline{P} = \pi$$

Lecture 8 (xviii)

For what MC's (X_n) should we expect DB to hold?[26]

$$\pi_i^{[27]}\, p_{ij} = \pi_j^{[28]}\, p_{ji}$$

∴ DB says LR prop. of transitions $i \to j$ = LR prop. of transitions $j \to i$.

---

[26] DB⇒SD, but not SD⇒DB

[27] $= \lim_{n\to\infty} P(X_n = i) =$ LR proportion of time in i

[28] LR proportion of time in j

Lecture 8 (xix)

Eg: Recall the Ehrenfest chain:

N molecules separated by a thin membrane into 2 chambers. Sometimes a molecule slips through.

X_n = # molecules in 1st chamber.

Each step pick a random molecule and move through.

$$p_{ij} = \begin{cases} \frac{i}{N} & \text{for } j = i-1 \\ 1 - \frac{i}{N} & \text{for } j = i+1 \\ 0 & \text{o/w} \end{cases}$$

Lecture 9

Eg: Recall the Ehrenfest chain:

N molecules separated by a thin membrane into 2 chambers. Sometimes a molecule slips through.

X_n = # molecules in 1st chamber.

Each step pick a random molecule and move through.

$$p_{ij} = \begin{cases} \frac{i}{N} & \text{for } j = i - 1 \\ 1 - \frac{i}{N} & \text{for } j = i + 1 \\ 0 & \text{o/w} \end{cases}$$

Lecture 9 (ii)

Intuition: In the long run as $n \to \infty$, any given particle should be approximately equally likely to be in either chamber.

$$\therefore X_n = \#\text{ in 1st chamber} \approx \text{Bin}\left(N, \frac{1}{2}\right)$$

$$\therefore P(X_n = i) \approx \binom{N}{i}\left(\frac{1}{2}\right)^i\left(\frac{1}{2}\right)^{N-i} = \frac{\binom{N}{i}}{2^N}$$

$$\therefore \text{Expect} \quad \pi_i = \lim_{n\to\infty} P(X_n = i) = \frac{\binom{N}{i}}{2^N}$$ [29]

[29] Prob. a uniform random subset is size i.

Lecture 9 (iii)

It is somewhat involved to solve $\pi = \pi \underline{P}$ for this Markov chain. It is easier to use the DBE.

But why should we Expect

$$\pi_i p_{ij} = \pi_j p_{ji} \quad \forall i, j$$

to hold for this MC?

Lecture 9 (iv)

π_i = long-run prop. of time spent in state i

$\pi_i p_{ij}$ = long-run prop. of transitions made by MC from $i \to j$

$\pi_j p_{ji}$ = long-run prop. of transitions made by MC from $j \to i$

Note: only possible transitions are $i \to i \pm 1$ (all other $p_{ij} = 0$)

Lecture 9 (v)

Also: every time the MC transitions $i \to i+1$, it must first make a transition $i+1 \to i$ before it can make another $i \to i+1$ again.

$\therefore$ At all times n, the number of transitions from $i \to j$ and the number of transitions from $j \to i$ made by time n are equal up to ± 1.

Lecture 9 (vi)

That is, for any n,

$$\frac{\#(i \to j \text{ by time } n)}{n} = \frac{\#(j \to i \text{ by time } n) \pm 1}{n}$$ [30]

Taking $n \to \infty$,

$$\pi_i p_{ij} = \pi_j p_{ji} \quad \text{(DBE)}$$

[30] $\frac{\pm 1}{n} \to 0$

Lecture 9 (vii)

Next we use the DBE to verify that $\pi = (\pi_1, \ldots, \pi_N)$ is the SD ($\pi = \pi \underline{P}$) where[31]

$$\pi_i = \frac{\binom{N}{i}}{2^N}$$

Check:[32]

$$\frac{\binom{N}{i}}{2^N} p_{i,i+1} \stackrel{?}{=} \frac{\binom{N}{i+1}}{2^N} p_{i+1,i}$$

[31]Last class: DB $\Rightarrow$ SD

[32]Can assume $|i - j| = 1$, since otherwise $p_{ij} = p_{ji} = 0$. So can assume $j = i + 1$

Lecture 9 (viii)

$$\frac{\binom{N}{i}}{\cancel{2^N}} p_{i,i+1} \stackrel{?}{=} \frac{\binom{N}{i+1}}{\cancel{2^N}} p_{i+1,i}$$

$$\binom{N}{i}\left(1-\frac{i}{N}\right) \stackrel{?}{=} \binom{N}{i+1}\frac{i+1}{N}$$

$$\frac{\cancel{N}!}{i!(N-i)!}\frac{N-i}{\cancel{N}} \stackrel{?}{=} \frac{\cancel{N}!}{(i+1)!(N-i-1)!}\frac{i+1}{\cancel{N}}$$

$$\frac{1}{i!(N-i-1)!} = \frac{1}{i!(N-i-1)!} \quad \checkmark$$

Lecture 9 (ix)

$\therefore$ Since DB $\Rightarrow$ SD, we know therefore

$$\pi = \pi \underline{P}$$

where

$$\pi = \left(\frac{\binom{N}{1}}{2^N}, \frac{\binom{N}{2}}{2^N}, \ldots, \frac{\binom{N}{N}}{2^N} \right)$$

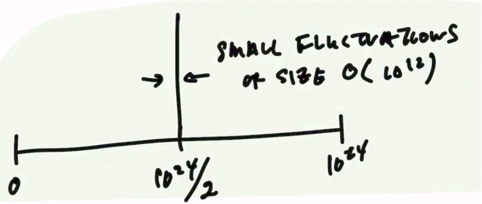
Lecture 9 (x)

$$
\begin{aligned}
\pi_i &= \frac{\binom{N}{i}}{2^N} \\
&= \lim_{n\to\infty} P(X_n = i) \\
&= \text{long-run proportion of time spent in state } i \\
&\quad (i \text{ molecules in 1st chamber})
\end{aligned}
$$

Lecture 9 (xi)

For a gas, $N \approx 10^{24}$

PMF of $\pi \sim \text{Bin}\left(10^{24}, \frac{1}{2}\right)$:



molecules in 1st chamber (large n)

$$= 500,000,000,000, \ldots, \ldots, \ldots, \ldots^{[33]}$$

[33] As $n \to \infty$, $\frac{X_n}{N} \approx \frac{1}{2}$. Small fluctuations in here.

Lecture 9 (xii)

Reversible MC's

The MP is symmetric in time: "Given current state, past & future are independent."

However, convergence to a stationary distribution is not. Typically, as $n \to \infty$, organized $\to$ disorganized state.

Eg. Ehrenfest chain: $X_0 = N$ (all in 1st chamber). As $n \to \infty$, molecules in 1st or 2nd chambers wp $\frac{1}{2}, \frac{1}{2}$.

Lecture 9 (xiii)

This suggests that if we want to get time symmetry, we must start from the stationary distribution: $X_0 \sim \pi$.

Def: Let $(X_n)_{n=0}^N$ be a MC run up to time N. Then $Y_n = X_{N-n}$ is called the **backwards chain**.

$$\begin{aligned} (Y_n)_{n=0}^N = (&Y_0, && Y_1, \ldots\ldots\ldots, && Y_N) \\ &= X_N && = X_{N-1} && = X_0 \end{aligned}$$

Lecture 9 (xiv)

Theorem Suppose (X_n) has SD π. $X_n \sim \pi$. Then the backwards chain

$$(Y_n = X_{N-n})_{n=0}^{N}$$

1. Is a MC
2. Has SD π
3. Has transition probabilities

$$q_{ij} = p_{ji}\frac{\pi_j}{\pi_i}$$

Lecture 9 (xv)

Proof: Will be on HW #2.

Note: If (X_n) satisfies the DBE,

$$q_{ij} = p_{ji}\frac{\pi_j}{\pi_i} = p_{ij}$$

In this case, we call (X_n) **time reversible**.

If you watch a movie of (X_n) started from equilibrium $X_0 \sim \pi$, you won't be able to tell if time is moving forward or backward.

Lecture 10

Eg Prof has 3 umbrellas. Each day

$$\begin{tikzcd} \text{Home} \arrow[r, bend left, "\text{PM}"] & \text{Office} \arrow[l, bend left, "\text{AM}"] \end{tikzcd}$$

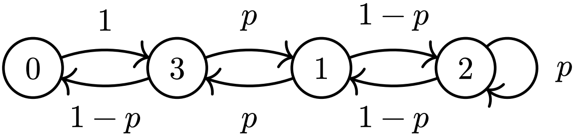
Rains w.p. p on each trip.

Takes umbrella if raining & has one at current location (o/w gets wet).

Q: What long-run fraction of time does prof get wet?

Lecture 10 (ii)

X_n = # umbrellas at current location



Can try to solve $\pi = \pi \underline{P}$ directly.

But easier to use DB.

(DB because for any i, j such that $p_{ij} > 0$, if we go $i \to j$ we must $j \to i$ before $i \to j$ again.)

Lecture 10 (iii)

Normalizing method

1. First find $x_i p_{ij} = x_j p_{ji} \forall i, j$.
2. Then $\pi_i = \frac{x_i}{\sum_j x_j}$ is SD.

We are "normalizing" in step 2. so that $\sum_i \pi_i = 1$.

Lecture 10 (iv)

$$\begin{aligned} x_0 &= (1-p)x_3 \\ x_3 p &= x_1 p \\ x_1(1-p) &= x_2(1-p) \\ &\Rightarrow \begin{cases} x_1 = x_2 = x_3 \\ x_0 = (1-p)x_3 \end{cases} \end{aligned}$$

Since we will normalize after, it doesn't matter which x_i we take, as long as $\neq 0$. $x_1 = 1$ is easy:

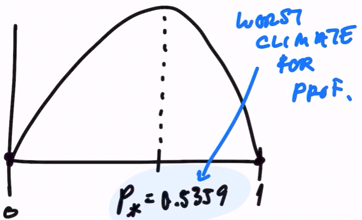
$$x = (1-p, 1, 1, 1) \Rightarrow \pi = \left(\frac{1-p}{4-p}, \frac{1}{4-p}, \frac{1}{4-p}, \frac{1}{4-p}\right)$$

Lecture 10 (v)

Q: What LR fraction of time does prof get wet?

This is $\pi_0 p$ = LR prop. of times it is raining in his current location before the next trip.

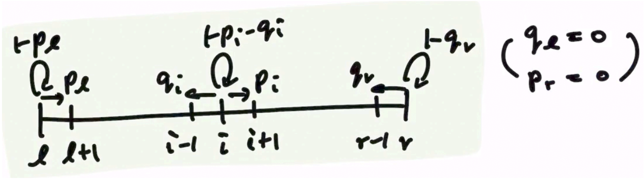
$$\pi_0 p = \frac{p(1-p)}{4-p}$$



Lecture 10 (vi)

The previous MC is a special case of "brith & death" (B&D) MC.

In such a MC, $|S| = \{l, l+1, ..., r\}$ for some $l < r$ and



All other $p_{ij} = 0$.

Lecture 10 (vii)

A B&D MC always satisfies DB:

($\#\{i \to j$ by time $n\} = \#\{i \leftarrow j$ by time $n\} \pm 1$ for any i, j and time n.)

You can use this to find π.

$$\pi_{l+i} = \pi_l \frac{p_l p_{l+1} \cdots p_{l+i-1}}{q_{l+1} q_{l+2} \cdots q_{l+i}}$$

Lecture 10 (viii)

DB:

$$\pi_l p_l = \pi_{l+1} q_{l+1}$$

$$\Rightarrow \pi_{l+1} = \pi_l \frac{p_l}{q_{l+1}}$$

$$\pi_{l+1} p_{l+1} = \pi_{l+2} q_{l+2}$$

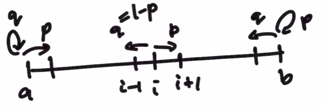
$$\Rightarrow \pi_{l+2} = \pi_{l+1} \frac{p_{l+1}}{q_{l+2}} = \pi_l \frac{p_l p_{l+1}}{q_{l+1} q_{l+2}}$$

$\vdots$ etc.

$$\pi_{l+i} = \pi_l \frac{p_l p_{l+1} \cdots p_{l+i-1}}{q_{l+1} q_{l+2} \cdots q_{l+i}}$$

Lecture 10 (ix)

Eg SRW on $\{a, a+1, \ldots, b\}$ with "reflecting" boundary points a, b:



$$\pi_{a+i} = \pi_a \frac{p_a p_{a+1} \cdots p_{a+i-1}}{q_{a+1} q_{a+2} \cdots q_{a+i}} = \pi_a \left(\frac{p}{q}\right)^i$$

Case 1. If $p = q = \frac{1}{2}$, $\pi_a = \pi_{a+1} = \ldots = \pi_b$

$$\Rightarrow \pi = \left(\frac{1}{b-a+1}, \ldots \frac{1}{b-a+1}\right) \quad \text{(Uniform)}$$

Lecture 10 (x)

Case 2. If $p \neq q$, then $\pi_{a+i} = \pi_a \left(\frac{p}{q}\right)^i$

We need $\sum_{i=0}^{b-a} \pi_{a+i} = 1$, so

$$\frac{1}{\pi_a} = \sum_{i=0}^{b-a} \left(\frac{p}{q}\right)^i = \frac{\left(\frac{p}{q}\right)^{b-a+1} - 1}{\frac{p}{q} - 1}$$

Therefore,

$$\pi_{a+i} = \frac{\frac{p}{q} - 1}{\left(\frac{p}{q}\right)^{b-a+1} - 1} \left(\frac{p}{q}\right)^i$$

Lecture 10 (xi)

Metropolis-Hastings Algorithm

Used in Bayesian statistics, image reconstruction, for studying complicated statistical physics models, etc...

Suppose we want to sample from some distribution π on a set S.

Lecture 10 (xii)

We can construct a MC (X_n) with SD π. If we run (X_n) for a long time, $P(X_n = i) \approx \pi_i$, so the position of (X_n) after a long time is close to a sample from the distribution π.

Metropolis-Hastings gives us a method of constructing such a MC (X_n):

Lecture 10 (xiii)

Let (Y_n) be a MC on S with tr. prob. q_{ij}.

Then let (X_n) be the MC on S with tr. prob.

$$p_{ij} = q_{ij} r_{ij}$$

$$\text{where } r_{ij} = \min\left\{ \frac{\pi_j q_{ji}}{\pi_i q_{ij}}, 1 \right\}$$

Lecture 10 (xiv)

In other words, if (X_n) is currently at i, we select a possible next transition according to what (Y_n) would do ($i \to j$ w.p. q_{ij}), **however** we accept this transition only with probability

$$r_{ij} = \min\left\{ \frac{\pi_j q_{ji}}{\pi_i q_{ij}}, 1 \right\}^{[34]}$$

Otherwise we stay at i in this step.

[34] w.p $1 - r_{ij}$, we reject and staty at i

Lecture 11

Metropolis-Hastings Algorithm

Used in Bayesian statistics, image reconstruction, for studying complicated statistical physics models, etc...

Suppose we want to sample from some distribution π on a set S.

Lecture 11 (ii)

We can construct a MC (X_n) with SD π. If we run (X_n) for a long time, $P(X_n = i) \approx \pi_i$, so the position of (X_n) after a long time is close to a sample from the distribution π.

Metropolis-Hastings gives us a method of constructing such a MC (X_n):

Lecture 11 (iii)

Let (Y_n) be a MC on S with tr. prob. q_{ij}.

Then let (X_n) be the MC on S with tr. prob.

$$p_{ij} = q_{ij} r_{ij}$$

$$\text{where } r_{ij} = \min\left\{ \frac{\pi_j q_{ji}}{\pi_i q_{ij}}, 1 \right\}$$

Lecture 11 (iv)

In other words, if (X_n) is currently at i, we select a possible next transition according to what (Y_n) would do ($i \to j$ w.p. q_{ij}), **however** we accept this transition only with probability

$$r_{ij} = \min\left\{\frac{\pi_j q_{ji}}{\pi_i q_{ij}}, 1\right\}$$[35]

Otherwise we stay at i in this step.

[35] w.p $1 - r_{ij}$, we reject and staty at i

Lecture 11 (v)

Then X_n has SD π.

To see this we check DB:

$$\pi_i p_{ij} = \pi_j p_{ji} \quad \forall i, j$$

Recall $p_{ij} = q_{ij} r_{ij} = q_{ij} \min\left\{\frac{\pi_j q_{ji}}{\pi_i q_{ij}}, 1\right\}$.

Case 1:

$$\text{Suppose } \frac{\pi_j q_{ji}}{\pi_i q_{ij}} \leq 1$$

$$\text{Then } r_{ij} = \frac{\pi_j q_{ji}}{\pi_i q_{ij}} \text{ and } r_{ji} = 1$$

Lecture 11 (vi)

$$\begin{aligned}
\therefore \pi_i p_{ji} &= \pi_i q_{ji} r_{ji} \\
&= \pi_i q_{ji} \frac{\pi_j q_{ji}}{\pi_i q_{ji}} \\
&= \pi_j q_{ji}
\end{aligned}$$

$$\begin{aligned}
\pi_j p_{ji} &= \pi_j q_{ji} r_{ji} \\
&= \pi_j q_{ji}
\end{aligned}$$

$$\Rightarrow \pi_i p_{ji} = \pi_j p_{ji}$$

Lecture 11 (vii)

Case 2: $\frac{\pi_j q_{ji}}{\pi_i q_{ji}} \geq 1$ is symmetric, and follows similarly.

Therefore (X_n) satisfies DB with π. Hence $\pi = \pi \underline{P}$.

$$\therefore \lim_{n\to\infty} P(X_n = i) = \pi_i, \text{as desired}$$

Lecture 11 (viii)

Example: Generating samples from Geometric(λ), $\lambda \in (0,1)$, using SRW.

Then

$$\begin{aligned}\pi_i &= P(\text{Geometric}(\lambda) = i) \\ &= \lambda^i(1-\lambda), i = 0, 1, \ldots\end{aligned}$$

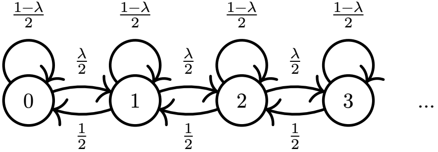
(X_n) SRW: $q_{ji} = \begin{cases} \frac{1}{2} & j = i \pm 1 \\ 0 & \text{o/w} \end{cases}$

$$r_{ji} = \min\left\{1, \frac{\pi_j \cancel{q_{ji}}}{\pi_i \cancel{q_{ji}}}\right\} = \min\{1, \lambda^{j-i}\}$$

Lecture 11 (ix)

Since $\lambda < 1, r_{i,i+1} = \lambda$ & $r_{i,i-1} = 1$[36]

(Y_n) has $p_{i,j}$:



[37]

Running (Y_n) for a long time gives an approx sample from Geometric(λ) distribution.

[36] $= \min\{1, \frac{1}{\lambda}\}$

[37] irreducible, aperiodic, $|S| = \infty$

Lecture 11 (x)

[D] §1.7 – Proof of main theorems

Recall from last week:

Main MC Theorem If (X_n) is aperiodic, irreducible and $|S| < \infty$, then there is a SD π (i.e. $\pi = \pi \underline{P}$ & $\sum \pi_i = 1$) and

1. $\lim\limits_{n\to\infty} p_{ij}^n = \pi_j \quad$ for any i
2. $\lim\limits_{n\to\infty} \frac{N_n(j)}{n}$[38] $= \frac{1}{E_j[T_j]}$[39] $= \pi_j$

[38] LR prop. of time in j.

[39] Inverse mean return time to j.

Lecture 11 (xi)

We'll prove this theorem this week. Along the way we'll prove some more general results:[40]

Theorem 1.19 Suppose (X_n) is irreducible, aperiodic and has a SD $\pi = \pi \underline{P}$. Then $\lim_{n\to\infty} p_{ij}^n = \pi_j \quad$ for all i, j

Theorem 1.20 Suppose (X_n) is irreducible and Rec. Then it has a stationary measure $\mu \geq 0 : \sum_i \mu_i p_{ji} = \mu_j$ for all i.

---

[40]Holds for $|S| = \infty$.

Lecture 11 (xii)

Note: We don't assume $|S| < \infty$ above.

However, if $|S| < \infty$ then the stationary measure μ can be normalized to get a SD:

$$\pi_i = \frac{\mu_i}{\sum_j \mu_j}$$ [41]

Then $\pi \geq 0$, $\sum_i \pi_i = 1$, & $\pi = \pi \underline{P}$.

$\therefore$ Theorems 1.19 & 1. 20 $\Rightarrow$ Part (1.) in the main theorem.

[41] $\sum < \infty$ since $|S| < \infty$

Lecture 11 (xiii)

However, if $|S| = \infty$, we may not be able to normalize.

E.g. (X_n) symmetric SRW on Z.

$\mu_i \equiv 1 \quad \sum_j \mu_j = \infty$

$\therefore$ π would have to be $\pi = 0$ which is **not** a prob. dist. (does not add to 1)

Lecture 11 (xiv)

To prove theorem 1.19, we need:

Lemma If (X_n) has a SD π (i.e. $\pi = \pi \underline{P}$ & $\sum_i \pi_i = 1$) then all states with $\pi_j > 0$ are Rec.

Proof. From previous lectures,

$$E_i N_j = \sum_{n=1}^{\infty} p_{ij}^n$$

$$\therefore \sum_i \pi_i E_i N_j = \sum_{i=1}^{\infty} \pi_i \sum_{n=1}^{\infty} p_{ij}^n$$

Lecture 12

[D] §1.7 – Proof of main theorems

Recall from last week:

Main MC Theorem If (X_n) is aperiodic, irreducible and $|S| < \infty$, then there is a s.d. π (i.e. $\pi = \pi \underline{P}$ & $\sum \pi_i = 1$) and

1. $\lim\limits_{n\to\infty} p_{ij}^n = \pi_j \quad$ for any i
2. $\lim\limits_{n\to\infty} \frac{N_n(j)}{n}$[42] $= \frac{1}{E_j[T_j]}$[43] $= \pi_j$

---

[42]LR prop. of time in j.

[43]Inverse mean return time to j.

Lecture 12 (ii)

We'll prove this theorem this week. Along the way we'll prove some more general results:[44]

Theorem 1.19 Suppose (X_n) is irreducible, aperiodic and has a SD $\pi = \pi \underline{P}$. Then $\lim_{n\to\infty} p_{ij}^n = \pi_j \quad$ for all i, j

Theorem 1.20 Suppose (X_n) is irreducible and Rec. Then it has a stationary measure $\mu \geq 0 : \sum_i \mu_i p_{ji} = \mu_j$ for all i.

[44]Holds for $|S| = \infty$.

Lecture 12 (iii)

To prove theorem 1.19, we need:

Lemma If (X_n) has a SD π (i.e. $\pi = \pi \underline{P}$ & $\sum_i \pi_i = 1$) then all states with $\pi_j > 0$ are Rec.

Proof. From previous lectures,

$$E_i N_j = \sum_{n=1}^{\infty} p_{ij}^n$$

$$\therefore \sum_i \pi_i E_i N_j = \sum_{i=1}^{\infty} \pi_i \sum_{n=1}^{\infty} p_{ij}^n$$

Lecture 12 (iv)

$$\sum_i \pi_i E_i N_j = \sum_i \pi_i \sum_{n=1}^{\infty} p_{ij}^n = \sum_{n=1}^{\infty} \sum_i \pi_i p_{ij}^n$$

$$= \sum_{n=1}^{\infty} \left(\pi \underline{P}^n\right)_j = \sum_{n=1}^{\infty} \pi_j = \infty$$

(since $\pi = \pi P \Rightarrow \pi = \pi P^n$ and $\pi_j > 0$)

Recall: $E_i N_j = \frac{\rho_{ij}}{1-\rho_{jj}}$ where $\rho_{ij} = P_i\left(T_j < \infty\right)$

Lecture 12 (v)

$$\therefore \sum_i \pi_i \frac{\rho_{ij}}{1-\rho_{jj}} = \frac{1}{1-\rho_{jj}} \sum_i \pi_i \rho_{ij} = \infty$$

However,

$$\sum_i \pi_i \rho_{ij} \leq \sum_i \pi_i = 1$$

$$\therefore \frac{1}{1-\rho_{jj}} = \infty \Rightarrow \rho_{jj} = 1$$

$$\Rightarrow j \text{ is Rec.} \quad \square$$

Lecture 12 (vi)

We are ready to prove:

Theorem 1.19 Suppose (X_n) is irreducible, aperiodic and has a stationary distribution $\pi = \pi \underline{P}$. Then $\lim_{n\to\infty} p_{ij}^n = \pi_j$ for all i, j.

[Proof is long but beautiful. We'll split it into several parts]

Proof. (X_n) a MC on S with tr. prob. p_{ij}. We define $(Y_n) = (X_n, X_n')$ on $S^2 = S \times S$ with tr. prob q as follows:

Lecture 12 (vii)

$$q_{(i_1,j_1),(i_2,j_2)} = p_{i_1 i_2} p_{j_1 j_2}$$

I.e. the 2 co-ordinates transition independently according to $\underline{P}$.

Step 1. (Y_n) is irreducible.

Lemma 1.16 in [D] shows that if state x is aperiodic, $p^n_{xx} > 0$ for all sufficiently large n. We'll use this without proof.

Lecture 12 (viii)

Since (X_n) irreducible,

$$p^k_{i_1 i_2} > 0 \text{ and } p^l_{j_1 j_2} > 0 \text{ for some } k, l$$

Since (X_n) aperiodic, so i_2, j_2 are aperiodic,

$$p^n_{i_2 i_2}, p^n_{j_2 j_2} > 0 \text{ for all large } n$$

$$\begin{aligned}
\therefore\ q^{k+l+n}_{(i_1,j_1),(i_2,j_2)} &= p^{k+l+n}_{i_1,i_2} p^{k+l+n}_{j_1,j_2} \\
&\geq \left(p^k_{i_1 i_2} p^{l+n}_{i_2 i_2}\right)\left(p^l_{j_1 j_2} p^{k+n}_{j_2 j_2}\right) > 0 \text{ for all large } n
\end{aligned}$$

Lecture 12 (ix)

$\therefore$ (Y_n) is irreducible.

Step 2 $\hat{\pi}_{(i,j)} = \pi_i \pi_j$ is a SD for (Y_n).

This is because, recall, the coordinates move independently like (X_n).

So by lemma, all states $(i,j) \in S^2$ are recurrent for $(Y_n) = (X_n, X'_n)$.

Let $T = \min\{n \geq 0 : X_n = X'_n\} = 1^{\text{st}}$ time co-ordinates equal.

Note: $P(T < \infty) = 1$. (all states (x,x) is recurrent.)

Lecture 12 (x)

Step 3 $P(X_n = x, T \leq n) = P(X'_n = x, T \leq n)$

$$\begin{aligned}
&P(X_n = x, T \leq n) \\
&= \sum_{m=1}^{n} \sum_{y} P(T = m, X_m = y, X_n = x) \\
&= \sum_{m=1}^{n} \sum_{y} P(T = m, X_m = y) P(X_n = x \mid X_m = y) \\
&= \sum_{m=1}^{n} \sum_{y} P(T = m, X'_m = y)^{[45]} \; P(X'_n = x \mid X'_m = y)^{[46]} \\
&= P(X'_n = x, T \leq n)^{[47]}
\end{aligned}$$

[45] by def of T

[46] X_n, X'_n iid.

[47] by 1st two lines of reasoning instead applied to (X'_n).

Lecture 12 (xi)

Finally,

Step 4 By previous step, the distributions of X_n and X'_n agree on $\{T \leq n\}$. So,

$$\begin{aligned} &|P(X_n = x) - P(X'_n = x)| \\ &= |P(X_n = x, T > n) - P(X'_n = x, T > n)| \\ &\leq P(X_n = x, T > n) + P(X'_n = x, T > n) \end{aligned}$$ [48]

[48] By triangle inequality.

Lecture 12 (xii)

$$\therefore \sum_x |P(X_n = x) - P(X'_n = x)|$$
$$\leq 2P(T > n)$$
$$\to 0 \quad \text{as} \quad n \to \infty$$

All above holds regardless of what X_0 is.

Now, let $X_0 = i$ & $X'_0 \sim \pi$. Then,

$$\sum_j |p^n_{ij} - \pi_j| \to 0 \quad \text{as} \quad n \to \infty$$
$$\Rightarrow \lim_{n\to\infty} p^n_{ij} = \pi_j \quad \forall i, j \quad \square$$

Lecture 12 (xiii)

Next, we prove:

Theorem 1.20 Suppose (X_n) is irreducible and recurrent. Then it has a **stationary measure** $\mu \geq 0 : \sum_i \mu_i p_{ij} = \mu_j$ for all j.

Recall: if $|S| < \infty$ we can use μ to find SD: $\pi_i = \frac{\mu_i}{\sum_j \mu_j}$.

Lecture 12 (xiv)

Proof. Fix $i \in S$. Let

$$T_i = \min\{n \geq 1 : X_n = i\}$$

We show

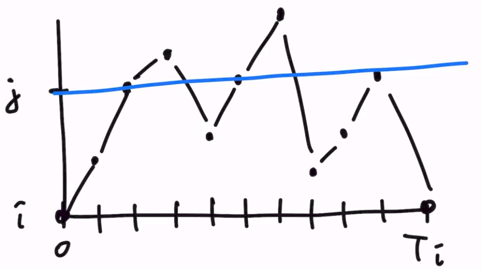
$$\mu_i(j) = \sum_{n=0}^{\infty} P_i(X_n = j, T_i > n)$$

$$= \text{expected \# visits to j before time } T_i \text{, starting from i}$$

is a stationary measure.

Lecture 12 (xv)

This is called the "cycle trick".



$$\begin{aligned}\mu_i(j) &= E_i[\#\text{ visits to j during } \{0, 1, ..., T_i - 1\}] \\ &= E_i[\#\text{ visits to j during } \{1, 2, ..., T_i\}] \\ &= (\mu_i \underline{P})(j)\end{aligned}$$

Lecture 13

Next, we aim to prove part (2) of the main MC theorem. This follows by:

Theorem 1.21 If (X_n) is Irr and Rec, then

$$\lim_{n\to\infty} \frac{N_n(j)}{n} = \frac{1}{E_j T_j}$$

And

Lecture 13 (ii)

Theorem 1.22 If (X_n) is Irr and has a SD π (i.e., $\pi = \pi \underline{P}$ and $\sum_i \pi_i = 1$), then

$$\pi_j = \frac{1}{E_j T_j}$$

Therefore, if (X_n) is Irr, Aper, and $|S| < \infty$, we know (X_n) is Rec and so a SD π exists. So,

$$\lim_{n \to \infty} \frac{N_n(j)}{n} \overset{1.21}{=} \frac{1}{E_j T_j} \overset{1.22}{=} \pi_j$$

Lecture 13 (iii)

- Also note that Theorem 1.22 implies that if (X_n) is Irr and has a SD π, then the SD is **unique**. (i.e., if $\pi' = \pi' \underline{P}$, then $\pi' = \pi$)
- If an Irr MC has an equilibrium, it is unique!

Lecture 13 (iv)

Note that Theorem 1.21 does not require aperiodicity.

Eg. $0 \leftrightarrow 1$

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

Irreducible, recurrent and has a stationary distribution, although period = 2. $\pi = \left(\frac{1}{2}, \frac{1}{2}\right)$ (doubly stochastic.)

$\frac{1}{2} = E_1 T_1 =$ long-run proportion of time spent in state 1.

Lecture 13 (v)

First, we note that Thm 1.21 $\to$ Thm 1.22:

Proof of Thm 1.22

Suppose (X_n) is irreducible and has a stationary distribution π.

By a previous lemma, if there is a stationary distribution π, then all i with $\pi_i > 0$ are recurrent. Therefore, since (X_n) is irreducible, all states are recurrent (recurrence is a class property). Hence, by Thm 1.21,

$$\lim_{n\to\infty} \frac{N_n(j)}{n} = \frac{1}{E_j T_j}$$

Lecture 13 (vi)

This holds no matter what X_0 is.

Assuming $X_0 \sim \pi$, and taking expectations on both sides,

$$E_\pi\left[\frac{N_n(j)}{n}\right] \to \frac{1}{E_j T_j} \quad \text{as} \quad n \to \infty$$

But since π is a stationary distribution,

$$E_\pi[N_n(j)] = \sum_{m=1}^{n} \overbrace{P_\pi(X_m = j)}^{=\pi_j} = n\pi_j$$

Therefore, $\frac{1}{E_j T_j} = \pi_j \quad \square$

Lecture 13 (vii)

Finally,

Proof of Thm 1.21 Suppose (X_n) is irreducible & recurrent. We show

$$\frac{N_n(j)}{n} \to \frac{1}{E_j T_j}$$

(no matter what X_0 is).

First, if $X_0 = j$, then the times $t_1, t_2, \ldots$ between visits to j are iid.

Lecture 13 (viii)

Let

$$\begin{aligned} T_k &= \min\{n \geq 1 : N_n(j) = k\} \\ &= \text{time of } k^{\text{th}} \text{ return.} \\ &= \sum_{l=1}^{k} t_l \end{aligned}$$

By LLN,

$$\frac{T_k}{k} \to E_j T_j \quad \text{as} \quad k \to \infty$$

since T_k = sum of k iid rvs with mean $E_j T_j$.

Lecture 13 (ix)

If $X_0 \neq j$, it doesn't make a difference in the limit: $t_1 < \infty$ since (X_n) is irreducible & recurrent. Then $t_2, t_3, \ldots$ are iid with mean $E_j T_j$.

$\therefore$, no matter what X_0 is,

$$\lim_{k \to \infty} \frac{T_k}{k} = E_j T_j$$

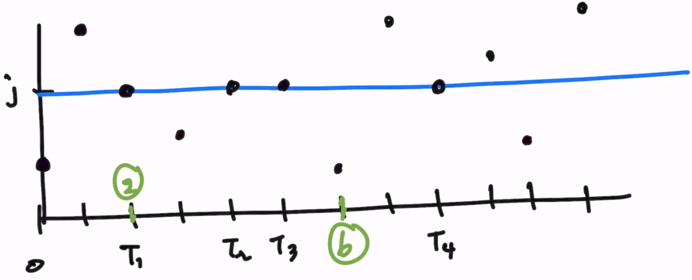
Lecture 13 (x)

To finish the proof, note

$$T_{N_n(j)} \leq n < T_{N_n(j)+1}$$

- $N_n(j) =$ # visits by time n.
- $T_{N_n(j)} = n$ if $X_n = j$, or else $T_{N_n(j)} < n$.
- At time n, have visited $N_n(j)$ times. So $n <$ time of $(N_n(j) + 1)^{\text{th}}$ visit $= T_{N_n(j)+1}$.

Lecture 13 (xi)



$N_2(j) = 1$

$$T_1 \le 2 < T_2, \quad T_1 = 2, \quad T_2 = 4$$

$N_6(j) = 3$

$$T_3 \le 6 < T_4, \quad T_3 = 5, \quad T_4 = 8$$

Lecture 13 (xii)

Hence,

$$\underbrace{\frac{T_{N_n(j)}}{N_n(j)}}_{\to E_j T_j} \le \frac{n}{N_n(j)} < \frac{T_{N_n(j)+1}}{N_n(j)}$$

$$= \underbrace{\frac{T_{N_n(j)+1}}{N_n(j)+1}}_{\to E_j T_j} \underbrace{\frac{N_n(j)+1}{N_n(j)}}_{\to 1}$$

Since j is recurrent, $N_n(j) \to \infty$ as $n \to \infty$. (Will visit ∞ many times.)

Therefore, $\lim_{n\to\infty} \frac{n}{N_n(j)} = E_j T_j \quad \square$

Lecture 13 (xiii)

The following generalization is proved in Durrett:

Theorem 1.23 Suppose (X_n) is irreducible, has a stationary distribution π, & $\sum_x |f(x)|\ \pi_x < \infty$. Then

$$\frac{1}{n}\sum_{m=1}^{n} f(X_m) \to \sum_x f(x)\pi_x$$

Taking $f(x) = \mathbf{1}_{x=j}$, we recover

$$\frac{N_n(j)}{n} \to \pi_j \quad \text{as} \quad n \to \infty.$$

Lecture 13 (xiv)

See §1.5 for some applications of Theorem 1.23.

Eg Repair chain

Machine has 3 parts: 1, 2, 3. Working as long as ≥ 2 parts working. Otherwise, parts are replaced, then machine working next day.

$$S = \{0, 1, 2, 3, 12, 13, 23\}$$

$$\begin{cases} 0 & \text{None broken} \\ 1,\ 2,\ 3 & \text{Part i broken} \\ 12,\ 13,\ 23 & \text{Parts i,j broken} \end{cases}$$

Lecture 13 (xv)

$$
\underline{P} = \begin{array}{c|ccccccc}
 & 0 & 1 & 2 & 3 & 12 & 13 & 23 \\
\hline
0 & .93 & .01 & .02 & .04 & & & \\
1 & & .94 & & & .02 & .04 & \\
2 & & & .95 & & .01 & & .04 \\
3 & & & & .97 & & .01 & .02 \\
12 & 1 & & & & & & \\
13 & 1 & & & & & & \\
23 & 1 & & & & & &
\end{array}
$$

IF 2 PARTS BROKEN, WILL HAVE NONE BROKEN NEXT DAY, SO $\to 0$.

Lecture 13 (xvi)

Q: Operate system for 5 years ($\approx 1,800$ days). About how many parts 1,2,3 will we use?

A: Can show

$$\pi = \frac{1}{8910}\left(\underbrace{3000}_{0}, \underbrace{500}_{1}, \underbrace{1200}_{2}, \underbrace{4000}_{3}, \underbrace{22}_{12}, \underbrace{60}_{13}, \underbrace{128}_{23}\right)^{[49]}$$

[49]by solving $\pi = \pi \underline{P}$

Lecture 13 (xvii)

Therefore, over 1800 days, will use about

$$1800(\pi_{12}+\pi_{13}) = 1800 * \frac{82}{8910} \approx 16.5$$

parts of type 1.

- Here we are using Thm 1.23 where

$$f_x = \mathbb{1}_{x\in\{12,13\}} = \begin{cases} 1 & x \in \{12,13\} \\ 0 & x \notin \{12,13\} \end{cases}$$

Lecture 14

Last week we proved the main MC theorem for finite state space MC's:

Theorem. (X_n) an IRR, APER MC on state space $|S| < \infty$. Then

1. There is a SD π: $\pi = \pi \underline{P}$ & $\sum \pi_j = 1$
2. $\lim_{n\to\infty} p_{ij}^n = \pi_j$ for all i, j.
3. $\pi_j = \lim_{n\to\infty} \frac{N_n(j)}{n}$ = LR prop. of steps in state $j = \frac{1}{E_j T_j}$ = inverse mean return time to j

Lecture 14 (ii)

To prove this, we proved several general results:

Theorem 1.19 Suppose (X_n) is Irr, aperiodic and has a SD $\pi = \pi \underline{P}$. Then $\lim_{n\to\infty} p_{ij}^n = \pi_j$ for all i, j

Theorem 1.20 Suppose (X_n) is Irr and Rec. Then it has a stationary measure $\mu \geq 0$: $\sum_i \mu_i p_{ij} = \mu_j$ for all i.

Lecture 14 (iii)

Theorem 1.21 (X_n) Irr & Rec. Then

$$\lim_{n\to\infty}\frac{N_n(j)}{n}=\frac{1}{E_jT_j}$$

Theorem 1.22 If (X_n) IRR & has SD π (i.e. $\pi=\pi\underline{P}$ & $\sum_i \pi_i=1$) then

$$\pi_j=\frac{1}{E_jT_j}$$

Theorem 1.23 Suppose (X_n) is IRR has SD π, & $\sum_x |f(x)|\,\pi_x<\infty$. Then

$$\frac{1}{n}\sum_{m=1}^{n} f(X_m)\to\sum_x f(x)\pi_x$$

Lecture 14 (iv)

Notice that some of the previous results can hold even when $|S| = \infty$.

[D] §1.11: Infinite State Spaces

In this lecture we'll discuss the a generalized main MC theorem that can hold even when $|S| = \infty$.

- The main issue is recurrence when $|S| = \infty$. It is not enough **just** to be Rec:

Lecture 14 (v)

Def Suppose j is Rec: $P_j\left(T_j < \infty\right) = 1$. Then we call j

- Positive Recurrent (Pos Rec) if $E_j T_j < \infty$
- Null Recurrent (Null Rec) if $E_j T_j = \infty$.
- Pos Rec: will return & mean wait time is finite.
- Null Rec: will return, but mean wait time is infinite.

Lecture 14 (vi)

- Boltzmann (1887) defined a state that is aperiodic & Pos Rec to be **ergodic**.

Def We call a MC ergodic if all states are ergodic.

Lecture 14 (vii)

Facts:

1. Pos/Null Rec is a class property. Therefore, so is ergodicity.
2. If a communication class is finite & Rec, then it is Pos Rec. Therefore for finite state space MC's Rec $\equiv$ Pos Rec.

Lecture 14 (viii)

Generalized Main MC Theorem.

(X_n) an irreducible, ergodic MC. Then

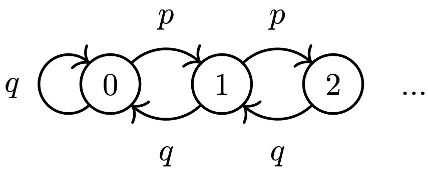
1. There is a SD π: $\pi = \pi \underline{P}$ & $\sum \pi_i = 1$
2. $\lim_{n\to\infty} p_{ij}^n = \pi_j$ for all i, j.
3. $\pi_j = \lim_{n\to\infty} \frac{N_n(j)}{n}$ = LR prop. of steps in state $j = \frac{1}{E_j T_j}$ = inverse mean return time to j

Lecture 14 (ix)

- We need Pos Rec so that we can normalize the stationary measure in theorem 1.20 above to get a stationary distribution: $\pi = \pi \underline{P}$ and $\sum \pi_i = 1$.

Lecture 14 (x)

EG SRW Reflected at 0.



[50]

What is SD π? Does it exist?

---

[50] $q = 1 - p$, $S = \{0, 1, ...\}$, $|S| = \infty$

Lecture 14 (xi)

This is a B&D MC (recall this means $p_{ij} = 0$ unless $j \in \{i, i \pm 1\}$). So to find SD, we try to solve DB equations.

Recall normalizing method:

1. Find $x_i p_{ij} = x_j p_{ji}$ (usually start by setting some $x_i = 1$)
2. Then put $\pi_i = \frac{x_i}{\sum_j x_j}$.

This works provided $\sum_j x_j < \infty$.

Lecture 14 (xii)

Set $x_0 = 1$.

$$\begin{aligned}
x_0 p_{01} = x_1 p_{10} &\Rightarrow p = x_1 q \\
&\Rightarrow x_1 = \frac{p}{q} \\
x_1 p_{12} = x_2 p_{21} &\Rightarrow \frac{p}{q} \cdot p = x_2 q \\
&\Rightarrow x_2 = \left(\frac{p}{q}\right)^2 \\
&\vdots \\
&\Rightarrow x_j = \left(\frac{p}{q}\right)^j
\end{aligned}$$

Lecture 14 (xiii)

In order to normalize, we need $\sum_j x_j = \sum_{j=0}^{\infty} \left(\frac{p}{q}\right)^j < \infty$

If $p \geq \frac{1}{2}$ then $\frac{p}{q} \geq 1$

$$\therefore \sum_j x_j = \begin{cases} \infty & p \geq \frac{1}{2} \\ \frac{1}{1-\frac{p}{q}} & p < \frac{1}{2} \end{cases}$$

So for $p < \frac{1}{2}$

$$\pi_i = \left(1 - \frac{p}{q}\right)\left(\frac{p}{q}\right)^i = \frac{1-2p}{1-p}\left(\frac{p}{q}\right)^i$$ [51]

---

[51] > 0

Lecture 14 (xiv)

(X_n) is clearly irreducible. $\therefore$ For $p < \frac{1}{2}$, (X_n) IRR & has SD.

By Theorem 1.22 above,

$$E_j T_j = \frac{1}{\pi_j} = \left(\frac{q}{p}\right)^j \frac{1-p}{1-2p} < \infty$$

$\therefore$ For $p < \frac{1}{2}$ all states are Pos Rec.

Lecture 14 (xv)

What happens for $p \geq \frac{1}{2}$?

- In this case, RW is getting pull away from the reflecting boundary. It seems that RW can drift off to ∞ and not have an equilibrium, at least when $p > \frac{1}{2}$. $p = \frac{1}{2}$ seems less clear.

Lecture 15

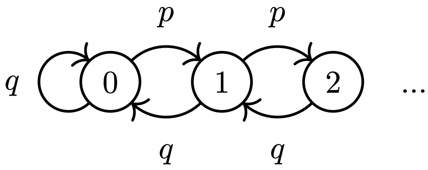
Generalized Main MC Theorem (X_n) an **irreducible, ergodic** MC. Then

1. There is a SD $\pi : \pi = \pi \underline{P}$ & $\sum \pi_i = 1$
2. $\lim_{n\to\infty} p_{ij}^n = \pi_j$ for all i, j.
3. $\pi_j = \lim_{n\to\infty} \frac{N_n(j)}{n} = \text{LR prop. of steps in state } j$

$= \frac{1}{E_j T_j} = \text{inverse mean return time to } j$

Lecture 15 (ii)

Eg: SRW Reflected at 0



What is SD π? Does it exist?

Lecture 15 (iii)

This is a B&D MC (recall this means $p_{ij} = 0$ unless $j \in \{i, \pm 1\}$). So to find SD, we try to solve DB equations.

Recall normalizing method:

1. Find $x_i p_{ij} = x_j p_{ji}$ (usually start by setting some $x_i = 1$)
2. Then put $\pi_i = \frac{x_i}{\sum_j x_j}$.

This works provided $\sum_i x_i < \infty$.

Lecture 15 (iv)

In order to normalize, we need $\sum_j x_j = \sum_{j=0}^{\infty} \left(\frac{p}{q}\right)^j < \infty$

If $p \geq \frac{1}{2}$ then $\frac{p}{q} \geq 1$

$$\therefore \sum_j x_j = \begin{cases} \infty & p \geq \frac{1}{2} \\ \frac{1}{1-\frac{p}{q}} & p < \frac{1}{2} \end{cases}$$

So for $p < \frac{1}{2}$

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

Lecture 15 (v)

(X_n) is clearly irreducible.

$\therefore$ For $p < \frac{1}{2}$, (X_n) is Irr & has SD.

By theorem 1.22 above,

$$E_j T_j = \frac{1}{\pi_j} = \left(\frac{q}{p}\right)^j \frac{1-p}{1-2p} < \infty$$

$\therefore$ For $p < \frac{1}{2}$ all states are Pos Rec.

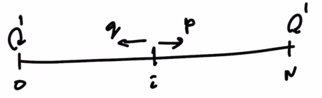
Lecture 15 (vi)

What happens for $p \geq \frac{1}{2}$?

- In this case, RW is getting pulled away from the reflecting boundary. It seems that RW can drift off to ∞ and not have an equilibrium, at least when $p > \frac{1}{2}$. $p = \frac{1}{2}$ seems less clear.

Lecture 15 (vii)

Recall from HW 1



$$P_x(\text{hit N before 0}) = \frac{\left(\frac{q}{p}\right)^x - 1}{\left(\frac{q}{p}\right)^N - 1}$$

$$\to 1 - \left(\frac{q}{p}\right)^x$$

As $N \to \infty$, provided $\frac{q}{p} < 1$ ($p > \frac{1}{2}$)

Lecture 15 (viii)

$\therefore$ For SRW reflected at 0:

$$P_x{}^{[52]}\left(T_0 < \infty\right) = \left(\frac{q}{p}\right)^x < 1$$

$\therefore$ For $p > \frac{1}{2}$, all states are Trans.

What about the borderline case $p = \frac{1}{2}$?

[52] $x \geq 1$

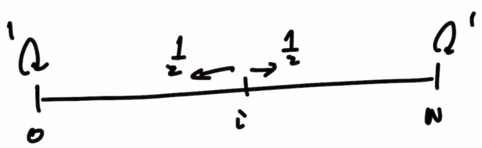
Lecture 15 (ix)

Symmetric ($p = \frac{1}{2}$) SRW on $\mathbb{Z}$ is Rec. Therefore SD is symmetric SRW reflected at 0. (Why? Start at 0. Either $X_1 = 0$, or $X_1 = 1$. If $X_1 = 1$, since SRW on $\mathbb{Z}$ is Rec., we will visit 0 again w.p. 1.)

However, it is Null Rec, and so doesn't have a SD.

Lecture 15 (x)

By FSA can show:



$$T = \min\{n : X_n \in \{0, N\}\}$$
$$= \text{time at end of game.}$$

$$E_x T = x(N - x)$$

Lecture 15 (xi)

$\therefore$ Taking $N \to \infty$: for sym. SRW reflected at 0, $E_1 T_0 = \infty$.

$\therefore E_0 T_0 = \frac{1}{2} + \frac{1}{2}(1 + E_1 T_0)$

$\Rightarrow E_0 T_0 = \infty$

$\therefore$ 0 is Null Rec

$\therefore$ All states are Null Rec ((X_n) Irr.)

Lecture 15 (xii)

Branching Processes (BP)

Start with 1 particle $X_0 = 1$.

Suppose that each particle in any generation gives birth to an IID number of particles before it dies, according to some **distribution**[53] with mean μ.

---

[53] "offspring distribution" often denoted by ξ.

Lecture 15 (xiii)

Let X_n = # particles in nth generation.
(X_n) is a MC.

$$\begin{aligned} & X_0 = 1 \\ \& \quad & X_n \mid (X_{n-1} = m) = \sum_{i=1}^{m} \xi_i^n \end{aligned}$$

where ξ_i^n IID, $E\xi = \mu$.

Q: Will MC ever visit (the absorbing state) 0 – i.e. will the population die out eventually?

Lecture 15 (xiv)

We will ignore the trivial case $P(\xi = 1) = 1$. In this case each particle gives birth to 1 particle, $\mu = 1$, and the population survives.

Lecture 15 (xv)

In all other cases, survival only depends on whether $\mu > 1$ or $\mu \leq 1$.

Theorem (X_n) a BP with offspring distribution ξ, $\mu = E\xi$. Suppose $P(\xi = 1) \neq 1$. Then

$$\begin{aligned} \rho = P(\text{extinct}) &= P(X_n = 0 \ \text{ eventually}) \\ &= \begin{cases} 1 & \mu \leq 1 \\ < 1 & \mu > 1 \end{cases} \end{aligned}$$

Moreover, $\rho = P(\text{extinct})$ is smallest positive solution to $x = \sum_k P(\xi = k) x^k$.

Lecture 15 (xvi)

- The borderline case $\mu = 1$ is the most surprising.
- Also very useful that this result also gives us a formula for $P(\text{survive}) = 1 - \rho$.

Lecture 15 (xvii)

Recall

$$X_n \mid (X_{n-1} = m) = \sum_{i=1}^{m} \xi_i$$

$$\Rightarrow E(X_n \mid X_{n-1} = m) = m\mu$$

$$\Rightarrow E(X_n \mid X_{n-1}) = \mu X_{n-1}$$

$$\Rightarrow E(X_n) = \mu E(X_{n-1})$$

Since $X_0 = 1$, this implies

$$E(X_n) = \mu^n \quad \text{for all } n \geq 0$$

Lecture 15 (xviii)

Hence

$$E(X_n) \overset{n\to\infty}{\to} \begin{cases} 0 & \mu < 1 \\ 1 & \mu = 1 \\ \infty & \mu > 1 \end{cases}$$

By Markov's Inequality,[54]

$$P(X_n > 0) \leq E(X_n) \to 0 \text{ if } \mu < 1$$
$$\therefore P(\text{extinct}) \to 1 \text{ for } \mu \leq 1$$

[54] $P(X \geq a) \leq \frac{E(X)}{a}$ for $a > 0$.

Lecture 16

Start with 1 particle $X_0 = 1$.

Suppose that each particle in any generation gives birth to an IID number of particles before it dies, according to some **distribution**[55] with mean μ.

[55] "offspring distribution" often denoted by ξ.

Lecture 16 (ii)

Let X_n = # particles in nth generation. (X_n) is a MC.

$$X_0 = 1$$

$$X_n \mid (X_{n-1} = m) = \sum_{i=1}^{m} \xi_i^n$$

where ξ_i^n iid, $E\xi = \mu$.

Q: Will MC ever visit (the absorbing state) 0 – i.e. will the population die out eventually?

Lecture 16 (iii)

We will ignore the trivial case $P(\xi = 1) = 1$. In this case each particle gives birth to 1 particle, $\mu = 1$, and the population survives.

Lecture 16 (iv)

In all other cases, survival only depends on whether $\mu > 1$ or $\mu \leq 1$.

Theorem (X_n) a BP with offspring distribution ξ, $\mu = E\xi$. Suppose $P(\xi = 1) \neq 1$. Then

$$\begin{aligned}\rho = P(\text{extinct}) &= P(X_n = 0 \ \text{ eventually}) \\ &= \begin{cases} 1 & \mu \leq 1 \\ < 1 & \mu > 1 \end{cases}\end{aligned}$$

Moreover, $\rho = P(\text{extinct})$ is smallest positive solution to $x = \sum_k P(\xi = k)x^k$.

Lecture 16 (v)

- The borderline case $\mu = 1$ is the most surprising.
- Also very useful that this result also gives us a formula for $P(\text{survive}) = 1 - \rho$.

Lecture 16 (vi)

Recall

$$X_n \mid (X_{n-1} = m) = \sum_{i=1}^{m} \xi_i$$

$$\Rightarrow E(X_n \mid X_{n-1} = m) = m\mu$$

$$\Rightarrow E(X_n \mid X_{n-1}) = \mu X_{n-1}$$

$$\Rightarrow E(X_n) = \mu E(X_{n-1})$$

Since $X_0 = 1$, this implies

$$E(X_n) = \mu^n \quad \text{for all } n \geq 0$$

Lecture 16 (vii)

Hence

$$E(X_n) \overset{n\to\infty}{\rightarrow} \begin{cases} 0 & \mu < 1 \\ 1 & \mu = 1 \\ \infty & \mu > 1 \end{cases}$$

By Markov's Inequality,

$$P(X_n > 0) \leq E(X_n) \to 0 \text{ if } \mu < 1$$

$$\therefore P(\text{extinct}) \to 1 \text{ for } \mu \leq 1$$

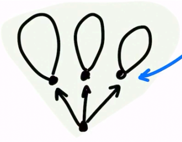
Lecture 16 (viii)

For $\mu > 1$, we need to be more careful:

Recall, $\rho = P(\text{extinct})$. By FSA,

$$\rho = \sum_{k=0}^{\infty} P(\xi = k)\rho^k$$

Why?



CHILDREN OF 1st PARTICLE.
THEIR "FAMILY TREES"
ARE INDEP, AND
EACH DIE OUT EVENTUALLY
W.P. ρ.

Lecture 16 (ix)

$$\varphi(t) = \sum_{k=0}^{\infty} P(\xi = k)t^k$$

is called the **generating function** of the RV ξ.

From the previous slide,

$$\rho = \varphi(\rho)$$

I.e. ρ is a root of φ. Recall that the theorem claims that ρ is the **smallest** pos. root.

Lecture 16 (x)

Also note that $\theta = 1$ is always a root:

$$\varphi(\theta) = \sum_{k=0}^{\infty} P(\xi = k)\theta^k$$

Set $\theta = 1$:

$$\varphi(1) = \sum_{k=0}^{\infty} P(\xi = k) = 1$$

Lecture 16 (xi)

Lemma. ρ is smallest pos. solution of $\varphi(\theta) = \theta$, $\theta \in [0, 1]$.

Proof.

$$\underbrace{P(X_n = 0)}_{\text{Die out by time } n} = \sum_{k=0}^{\infty} P(\xi = k) \underbrace{P(X_{n-1} = 0)^k}_{[56]}$$

[56] All family trees of all k children of 1st particle must die out (independently) by time $n - 1$.

Lecture 16 (xii)

Let $\rho_n = P(X_n = 0)$
By previous slide $\rho_n = \varphi(\rho_{n-1})$.

Note $\rho_0 \leq \rho_1 \leq \rho_2 \leq \ldots$
This is because $X_{n-1} = 0 \Rightarrow X_n = 0$ for any n.
$(A \supset B \Rightarrow P(A) \geq P(B))$

All $\rho_n \leq 1$. So by calculus (Monotone Convergence Theorem) the sequence converges:

$$\lim_{n\to\infty} \rho_n = \rho_\infty \left(= \sup_n \rho_n \right)$$

Lecture 16 (xiii)

$$\rho_n{}^{[57]} = \varphi(\rho_{n-1}{}^{[58]}) \Rightarrow \rho_\infty = \varphi(\rho_\infty)$$

So ρ_∞ is a solution to $\theta = \varphi(\theta)$.

To finish proof of lemma, we show ρ_∞ = smallest sol. in $[0, 1]$.

---

[57] $\longrightarrow \rho_\infty$

[58] $\longrightarrow \rho_\infty$

Lecture 16 (xiv)

Let p = smallest pos. sol. to $\varphi(\theta) = \theta$ in $[0, 1]$.

We show $p = \rho_\infty$.

Note: $\varphi(\theta) = \sum_{k=0}^{\infty} P(\xi = k)\theta^k$ is increasing in θ, since all $P(\xi = k) \geq 0$.

$\rho_0 = P(X_0 = 0) = 0 \leq p$, since $X_0 = 1$.

Since φ increasing, $\varphi(\rho_0) \leq \varphi(\rho)$
$\Rightarrow \rho_1 \leq \rho$ since $\varphi(\rho_0) = \rho_1$, and $\varphi(\rho) = \rho$.

Lecture 16 (xv)

Repeating argument, all $\rho_n \leq \rho$.

Take $n \to \infty$, $\rho_\infty \leq \rho \quad \square$

Using lemma, we can now study cases $\mu > 1$ and $\mu = 1$:

Lecture 16 (xvi)

$\mu > 1$ If $P(\xi = 0) = 0$, then clearly $\rho = 0$. Also, $\varphi(\theta) = \sum_{k=1}^{\infty} P(\xi = k)\theta^k$ (sum starts at $k = 1$), so $\varphi(0) = 0$.

If $P(\xi = 0) > 0$, then

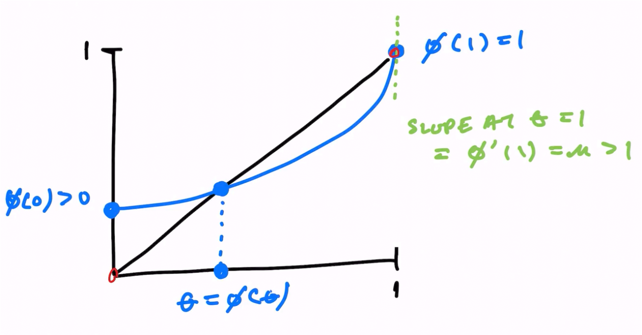
$$\varphi(\theta) = \sum_{k=0}^{\infty} P(\xi = k)\theta^k$$

$$\varphi'(\theta) = \sum_{k=1}^{\infty} P(\xi = k) \cdot k\theta^{k-1}$$

$$\varphi'(1) = \sum_{k=1}^{\infty} P(\xi = k) \cdot k = \mu.$$

If $\mu > 1$, slope of φ at $\theta = 1$ is larger than 1 = slope of diagonal

Lecture 16 (xvii)



Blue: generating function φ

Black: diagonal $y = \theta$

$\varphi(0) = P(\xi = 0)$

Lecture 16 (xviii)

$\mu = 1$ Recall, we exclude trivial case where $P(\xi = 1) = 1$. Then $\mu = 1$ and $\rho = 0$.

Suppose $\mu = 1$ & $P(\xi = 1) < 1$. We show φ has no root $\theta < 1$.

Note

$$\begin{aligned}\varphi'(\theta) &= \sum_{k=1}^{\infty} P(\xi = k) k \theta^{k-1} \\ &< \sum_{k=1}^{\infty} P(\xi = k) k = \mu = 1\end{aligned}$$

if $\theta < 1$.

Lecture 16 (xix)

$\therefore$ If $\theta < 1$ then

$$\begin{aligned} \int_\theta^1 \varphi'(u)du &= \varphi(1) - \varphi(\theta) \\ &= 1 - \varphi(\theta) \\ &< \int_\theta^1 1du = 1 - \theta \end{aligned}$$

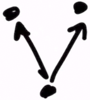
$\Rightarrow \varphi(\theta) > 1 - (1 - \theta) = \theta.$

$\therefore \varphi(\theta) > \theta$ for $\theta < 1.$

$\therefore \theta = 1$ only root in $[0, 1]$ $\quad\square$

Lecture 16 (xx)

Eg Binary Branching:



$\mu = 2\alpha$

$\mathbb{P}(\xi = 2) = \alpha$ $\quad$ $\mathbb{P}(\xi = 0) = 1 - \alpha$

$$\varphi(\theta) = \sum_{k=0}^{\infty} P(\xi = k)\theta^k = 1 - \alpha + \alpha\theta^2$$

$$\varphi(\theta) = \theta \Rightarrow 0 = (\theta - 1)(\alpha\theta - (1 - \alpha)).$$

Roots are $\theta = 1$ & $\theta = \frac{1-\alpha}{\alpha} < 1$ for $\alpha > \frac{1}{2}$ when $\mu > 1$.

Lecture 16 (xxi)

We turn now to: [D] §1.9 & 1.10 on **Exit Distributions & Times**. (Basically just FSA).

- We have already seen some of this in homework & workshops.
- Basis for this is FSA (First Step Analysis)

Eg: Gambler's ruin, $P_x(\text{Jackpot})$ & $P_x(\text{Ruin})$ is the exit distrib. from $\{1, 2, ..., N-1\}$ started at x.

Lecture 17

Short Lecture Today:

- We'll finish chapter 1
- Will spend the rest of week on ch. 2: Poisson processes.

Lecture 17 (ii)

§1.9 & 1.10 on **Exit Distributions** & **Times**

- We have already seen some of this in homework & workshops.
- Basis for this is FSA (First Step Analysis)

Eg: Gambler's Ruin, $P_x(\text{Jackpot})$ & $P_x(\text{Ruin})$ is the exit distribution from

$$\{1, 2, ..., N-1\}$$

started at x.

Lecture 17 (iii)

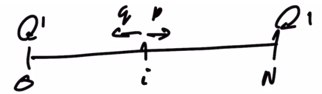
Def: For $A \subset S$, let

$$V_A = \min\{n \geq 0 : X_n \in A\}$$

be 1st time MC hits set A.

Lecture 17 (iv)

Eg: Gambler's Ruin



$$h(i) = P_i(V_N < V_0) = P_i(\text{Jackpot})$$

$$h(N) = P_N(\text{Jackpot}) = 1$$

$$h(0) = P_0(\text{Jackpot}) = 0$$

For $i \in \{1, 2, \dots, N-1\}$,
$h(i) = \sum_j p_{ij} h(j) = qh(i-1) + ph(i+1)$ is what we got by FSA on HW#1.

Lecture 17 (v)

Theorem Let (X_n) be a MC on S. Suppose $A, B \subset S$ such that $C = S \setminus (A \cup B)$ is finite.

Suppose

$$\begin{aligned} h(a) &= 1 && a \in A \\ h(b) &= 0 && b \in B \\ h(i) &= \sum_j{}^{[59]} p_{ij} h(j) && i \in C \end{aligned}$$

Then, $h(i) = P_i(V_A < V_B)$[60] if $P_i(\min\{V_A, V_B\} < \infty$[61]$) > 0$ for all $i \in C$.

Proof: Essentially FSA, but more technical, see [D].

---

[59] FSA

[60] Exit distribution

[61] Exit in finite time from C

Lecture 17 (vi)

- For Gambler's Ruin, take

$$A = \{N\}$$
$$B = \{0\}$$

If we can solve

$$h(N) = 1$$
$$h(0) = 0$$
$$h(i) = qh(i-1) + ph(i+1)$$

The theorem tells us $h(i) = P_i(\text{Jackpot})$

Lecture 17 (vii)

Look at Ex 1.42 [D] "Matching Pennies" on your own.

Wikipedia: "Genetic Drift"

Ex 1.43 "Wright-Fisher model with no mutation"

$$S = \{0, 1, ..., N\}$$

$$p_{ij} = \binom{N}{j}\left(\frac{i}{N}\right)^{j}\left(\frac{N-i}{N}\right)^{N-j}$$

$$= P\left(\text{Binomial}\left(N, \frac{i}{N}\right) = j\right)$$

X_n = # of "type A" genes. (Haploid model: A or B)

Lecture 17 (viii)

0 and N are absorbing states.

$$\text{Binomial}(N,0) \equiv 0$$
$$\text{Binomial}(N,1) \equiv N$$

Notice for $h(i) = \frac{i}{N}$,

$$h(i) = \sum_j p_{ij} h(j)$$

Indeed,

$$\sum_j \underbrace{\binom{N}{j}\left(\frac{i}{N}\right)^j \left(\frac{N-i}{N}\right)^{N-j}}_{p_{ij}} \underbrace{\frac{j}{N}}_{h_j}$$
$$= \frac{1}{N} E\left(\text{Bin}\left(N, \frac{i}{N}\right)\right) = \frac{i}{N}$$

Lecture 17 (ix)

$\therefore$ by Lemma, taking $A = \{N\}$, $B = \{0\}$,

$$\begin{aligned} &P_i(V_A < V_B) \\ &= P_i(\text{Fixation to 'type A' genes}) \\ &= \frac{i}{N} \\ &= \text{initial prop. of type A genes.} \end{aligned}$$[62]

[62] This is where our guess initially came from.

Lecture 17 (x)

Reducible Chains

Exit distributions are important for determining long run behavior of reducible MC's.

See ex 1.46 [D] for a nice example.

Lecture 17 (xi)

Idea Find closed communicating classes. Assuming, for example, $|S| < \infty$, the MC will eventually arrive in one such class, and stay there forever. So if $i \in S$ and $j \in C$, a closed class, we can find $\lim_{n\to\infty} p_{ij}^n$.

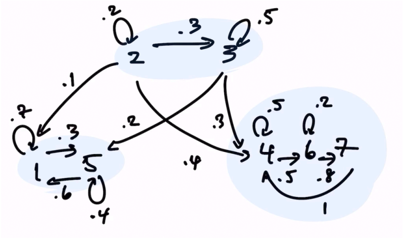
Lecture 17 (xii)

By finding

1. P_i(Hit C before any other closed class)
2. SD π for $\underline{P}$ restricted to C.

Then $\lim_{n\to\infty} p_{ij}^n = \pi_j P(\text{hit C 1st})$

Lecture 17 (xiii)



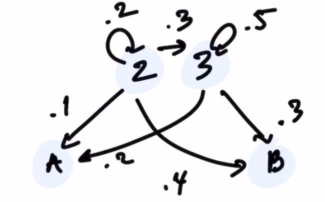
MC eventually ends up in $A = \{1, 5\}$ or $B = \{4, 6, 7\}$.

$$\pi_A = \frac{1}{3}(2, 1)$$

$$\pi_B = \frac{1}{9}(8, 5, 4)$$

Lecture 17 (xiv)

To find $P_i(V_A < V_B)$, it is easier to study:



$$h(A) = 1, \quad h(B) = 0$$

$$h(2) = 0.1 + 0.2h(2) + 0.3h(3)$$

$$h(3) = 0.2 + 0.5h(3)$$

By theorem, solution to this is exit distribution.

Lecture 17 (xv)

$$\Rightarrow h(2) = \frac{11}{40}$$
$$h(3) = \frac{2}{5}$$

Therefore, for example,

$$\lim_{n\to\infty} p_{21}^n = \underbrace{P_2(V_A < V_B)}_{h(2)} (\pi_A{}^{[63]})_1 = \frac{11}{40} \cdot \frac{2}{3}$$

$$\lim_{n\to\infty} p_{37}^n = (1 - P_3(V_A < V_B))(\pi_B)_7$$
$$= \left(1 - \frac{2}{5}\right) \cdot \frac{4}{17}$$

---

[63] Equil. dist. for class A

Lecture 18

Last lecture, we discussed exit distributions (§1.9).

Today, we'll quickly discuss exit times (§1.10).

Lecture 18 (ii)

Def For $A \subset S$, let

$$V_A = \min\{n \geq 0 : X_n \in A\}$$

be the first time the MC hits set A.

Lecture 18 (iii)

Eg Gambler's Ruin

$$
\begin{aligned}
g(i) &= E_i(\min\{V_0, V_N\}) \\
&= \text{Expected number of bets until game over,} \\
&\quad\ \text{starting with } i
\end{aligned}
$$

$$g(0) = g(N) = 0 \text{ for } i \in \{1, 2, ..., N-1\}$$

$$g(i) = 1^{[64]} + qg(i-1) + pg(i+1)$$

[64] 1 step, then start from $i-1$ or $i+1$.

Lecture 18 (iv)

Theorem. (X_n) MC on S. Suppose $A \subset S$ such that $C = S \setminus A$ is finite. Suppose

$$\begin{aligned} g(a) &= 0 && a \in A \\ g(i) &= 1 + \sum_j p_{ij} g(j) && i \in C \end{aligned}$$

Then $g(i) = E_i(V_A)$ if $P_i(V_A < \infty) > 0$ for all $i \in C$.

- For gambler's ruin, take $A = \{0, N\}$.

Lecture 18 (v)

Eg Waiting time for TT[65] while flipping a fair coin

$$S = \{0, 1, 2\}$$

MC in state i if last i flips are T. The first time we see TT is the first time we visit 2, started at 0: $E_0(V_2)$.

[65] 2 tails in a row

Lecture 18 (vi)

To find $g_i = E_i(V_2)$, we solve

$$\begin{cases} g_2 = 0 \\ g_0 = 1 + \frac{1}{2}(g_0 + g_1) \\ g_1 = 1 + \frac{1}{2}g_0 + \cancel{\frac{1}{2}g_2} \end{cases}$$

& then apply the theorem.

Lecture 18 (vii)

$$\begin{cases} g_0 = 1 + \frac{1}{2}(g_0 + g_1) \\ g_1 = 1 + \frac{1}{2} g_0 \end{cases}$$

$$\Rightarrow \begin{cases} g_0 = 2 + g_1 \\ 2g_1 = 2 + g_0 \end{cases}$$

$$\Rightarrow \begin{cases} g_0 = 6 \\ g_1 = 4 \end{cases}$$

$$\therefore E_0(V_2) = g_0 = 6$$

Poisson Processes

Lecture 18 (PP)

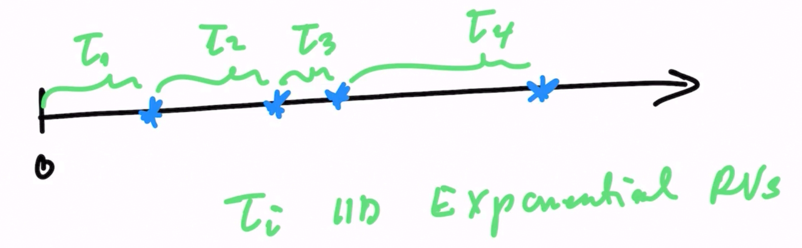
New chapter, §2 in [D]

Lecture 18 (PP) (ii)

- In Stat 134 (or equivalent course) you have already seen Poisson Processes.
- In Stat 150, we'll study this fundamental process more rigorously.

Lecture 18 (PP) (iii)

- The (1-dimensional) Poisson Process is a point process on the line $[0, \infty)$ with IID exponential "inter-arrival times" between points.



Lecture 18 (PP) (iv)

The Poisson Process is important because:

- Many real-world situations can be modeled using them:
  - Cars arriving at a toll booth
  - Galaxies in some region of the universe (3-D[66] Poisson Process)

[66] R^3

Lecture 18 (PP) (v)

- But also, many calculations work out nicely for the Poisson Process because exponential RVs have the “lack of memory” property.
- The exponential RV is, in fact, the **only** continuous RV with this property.

Lecture 18 (PP) (vi)

- In §3, we'll study "Renewal Processes" — which have IID inter-arrival times, which need not be exponential.
- These processes are more complicated, so it is crucial we understand the Poisson Process well first.

Lecture 18 (PP) (vii)

§2.1 — Exponential RVs

- This section is Stat 134 review. **Please** read on your own to refresh your memory.
- We'll just quickly summarize here:

Lecture 18 (PP) (viii)

Recall: "Survival function" of a RV X is

$$P(X > x) = 1 - F(x)^{[67]}$$

Def $X \sim \mathrm{Exp}(\lambda)$, Exponential RV with **rate** λ if

$$P(X > x) = e^{-\lambda x}, \quad x \geq 0.$$

[67] CDF

Lecture 18 (PP) (ix)

To get PDF,

$$\begin{aligned} f(x) = F'(x) &= \frac{d}{dx}\left(1 - e^{-\lambda x}\right) \\ &= \lambda e^{-\lambda x}. \end{aligned}$$

Can show:

$$\mathrm{E}\ X = \frac{1}{\lambda}$$

$$\mathrm{Var}\ X = \frac{1}{\lambda^2}$$

Lecture 18 (PP) (x)

Other useful properties:

1. Lack of Memory (LoM):

$$P(X > t + s \mid X > t) = P(X > s)$$

$$\left[\text{Proof: LHS} = \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} = e^{-\lambda s} = \text{RHS}\right]$$

$\therefore$ Conditional on $X > t$, RV starts afresh at time t, as though it were a brand new $\text{Exp}(\lambda)$.

Lecture 18 (PP) (xi)

2. Sum of n IID Exp(λ) RVs is called a Gamma(n, λ).

$$T_n = \sum_{i=1}^{n} T_i, \quad T_i \quad \text{IID Exp}(\lambda).$$

$$f^{[68]}{}_{T_n}(t) = \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!}, \quad t \geq 0$$

- Proof: Induction (see [D]).
- Note: If $n = 1$, $f(t) = \lambda e^{-\lambda t}$, so Gamma$(1, \lambda) = \text{Exp}(\lambda)$.

[68] PDF

Lecture 18 (PP) (xii)

3. Minimum of independent exponential RVs is exponential. Its rate is the sum of rates:

Theorem. $X_i \sim \mathrm{Exp}(\lambda_i)$, and all X_i independent. Then

i) $M = \min\{X_1, \ldots, X_n\} \stackrel{d}{=} \mathrm{Exp}\left(\sum_{i=1}^{n} \lambda_i\right)$

Moreover, let $I =$ index of smallest X_i, $M = X_I$. Then,

Lecture 18 (PP) (xiii)

ii) $P(I = i) = \frac{\lambda_i}{\sum_{j=1}^{n} \lambda_j}$

iii) I, M[69] are independent.

Part (iii) is perhaps the most surprising.

[69] $= X_I$

Lecture 18 (PP) (xiv)

Proof:

i)

$$\begin{aligned} P(M > t) &= P(\text{all } X_i > t) \\ &= \prod_{i=1}^{n} P(X_i > t) \text{ (indep.)} \\ &= \prod_{i=1}^{n} e^{-\lambda_i t} \\ &= e^{-t \sum_{i=1}^{n} \lambda_i} \end{aligned}$$

Lecture 18 (PP) (xv)

ii) Can be proved easily by induction — see [D].

iii)

$$
\begin{aligned}
f_{I,M}(i,t) &= \lambda_i e^{-\lambda_i t} \prod_{j \neq i}^{[70]} e^{-\lambda_j t} \\
&= \frac{\lambda_i}{\sum_j \lambda_j} \cdot \underbrace{\left(\sum_j \lambda_j\right) e^{-\left(\sum_j \lambda_j\right)t}}_{\text{PDF of a rate} \sum \lambda_i \text{ Exp. RV.}} \\
&= P(I = i)^{[71]} \, f_M(t)
\end{aligned}
$$

---

[70] All other RVs must be bigger than t, i.e., survival function.

[71] By part ii).

Lecture 18 (PP) (xvi)

§2.2 – Defining the Poisson Process

Recall: $X \sim \text{Poisson}(\mu)$ if

$$P(X = k) = e^{-\mu} \frac{\mu^k}{k!}, \quad k = 0, 1, 2, \ldots$$

Can show:

$$\begin{aligned} EX &= \mu \\ \text{Var } X &= \mu \end{aligned}$$

See [D].

Lecture 18 (PP) (xvii)

Also easy to show:

Theorem. If $X_i \sim \text{Poisson}(\lambda_i)$ and all X_i indep. Then

$$\sum_{i=1}^{n} X_i \sim \text{Poisson}\left(\sum_{i=1}^{n} \lambda_i\right).$$

Proof: Stat 134. See [D], by induction. $\square$

Lecture 19

The Poisson Process

Def $(N_t : 0 \le t$[72] $< \infty)$ is a rate λ Poisson Process – $\mathrm{PP}(\lambda)$ – if

1. $N_0 = 0$
2. $N_{t+s} - N_s \overset{d}{=} \text{Poisson}(t\lambda)$, **for any** $t, s \ge 0$
3. Indep. increments:

$$N_{t_1} - N_{t_0}, N_{t_3} - N_{t_2}, ..., N_{t_n} - N_{t_{n-1}}$$ [73]

indep. for **any** $t_0 < t_1 < ... < t_n$.

---

[72] continuously

[73] sequence of disjoint intervals

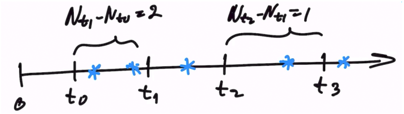
Lecture 19 (ii)

$$N_{t_1} - N_{t_0} = \#\text{ points in } (t_0, t_1]$$

$$(2.) \stackrel{d}{=} \text{Poisson}(\lambda(t_1 - t_0)^{[74]})$$

$$(3.)\ ^*\text{Indep.}^*\text{ of } N_{t_2} - N_{t_1} = \#\text{ points in } (t_1, t_2]$$

$$\stackrel{d}{=} \text{Poisson}(\lambda(t_2 - t_1)^{[75]})$$



(1.) $N_0 = 0$ because no points yet at time 0.

---

[74]length of $(t_0, t_1]$
[75]length of $(t_1, t_2]$

Lecture 19 (iii)

In particular, all

$$\begin{aligned} N_t{}^{[76]} &= \#\text{ points in } (0,t] \\ &= \#\text{ points by time } t \\ &\stackrel{d}{=} \text{Poisson}(\lambda t). \end{aligned}$$

This is why it is called a Poisson Process.

---

[76] $= N_t - N_0$

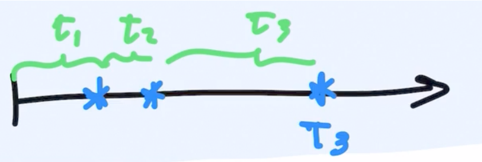
Lecture 19 (iv)

Constructing a rate λ Poisson Process:

$\tau_1, \tau_2, \ldots$ IID $\mathrm{Exp}(\lambda)$

Let $T_n = \sum_{i=1}^{n} \tau_i \overset{d}{=} \mathrm{Gamma}(n, \lambda)$.

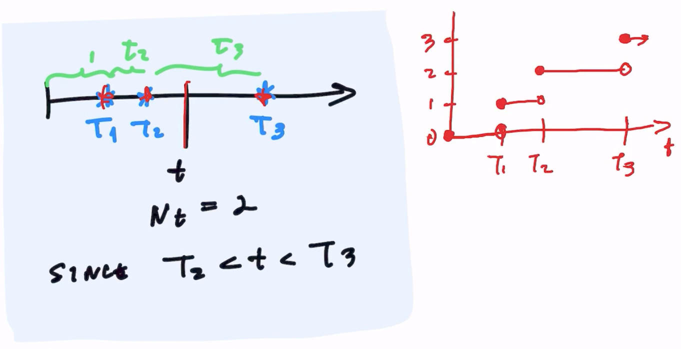
If we think of $\tau_i =$ lifetime of ith lightbulb, then T_n is the time at which we have gone through n lightbulbs in total.



$T_0 = 0, \quad T_1 = \tau_1, \quad T_2 = \tau_1 + \tau_2$

Lecture 19 (v)

Let $N_t = \max\{n : T_n \le t\} = \#$ of * to left of t



Poisson is discrete in values but continuous in time.

Lecture 19 (vi)

We will show that $(N_t)_{t \geq 0}$ is a $\mathrm{PP}(\lambda)$.

Clearly $N_0 = \max\{n \geq 0 : T_n \leq 0\} = 0$.

So we need to check (2.) & (3.) in definition above.

First step:

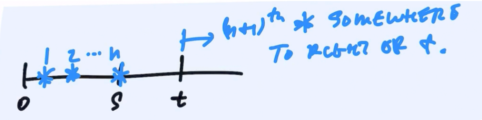
Lemma. $N_t \sim \mathrm{Poisson}(\lambda t)$

Lecture 19 (vii)

Proof. $\{N_t = n\} = \{T_n \leq t < T_{n+1}\}$.

I.e., to have exactly n points by time t, we need n^{th} point to arrive by time t & $(n+1)^{\text{th}}$ point to arrive at some later time $> t$.

$$\therefore P(N_t = n)^{[77]} = \int_0^t f_{T_n}(s)^{[78]} \; P(\tau_{n+1} > t - s) ds$$



---

[77] T_n, T_{n-1} indep, why?

[78] PDF for arrival time of n^{th} point, $T_n = \sum_{i=1}^{n} \tau_i$

Lecture 19 (viii)

$$T_n \sim \mathrm{Gamma}(n, \lambda), \quad f_{T_n}(s) = \lambda e^{-\lambda s} \frac{(\lambda s)^{n-1}}{(n-1)!}$$

$$\tau_{n+1} \sim \mathrm{Exp}(\lambda), \quad P\big(\tau_{n+1} > t - s\big) = e^{-\lambda(t-s)}$$

$$\therefore P(N_t = n) = \int_0^t \lambda e^{-\lambda s} \frac{(\lambda s)^{n-1}}{(n-1)!} e^{-\lambda(t-s)} ds$$

$$= e^{-\lambda t} \frac{\lambda^n}{(n-1)!} \int_0^t s^{n-1} ds = e^{-\lambda t} \frac{(\lambda t)^n}{n!}$$

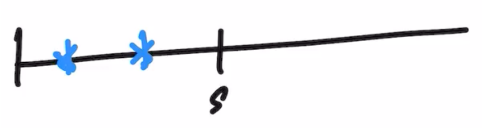
$$= P(\mathrm{Poisson}(\lambda t) = n).$$

Lecture 19 (ix)

We now check parts (2.) & (3.) in definition of $\mathrm{PP}(\lambda)$.

The key is LoM[79] & Lemma.

Proof of (2): $N_{t+s} - N_s \sim \mathrm{Poisson}(\lambda t)$.



By LoM the lightbulbs burning at time s starts afresh.

[79] lack of memory

Lecture 19 (x)

So if we ignore $*$ to left of s, & start counting $*$ starting at time s, we have another IID sequence of $\mathrm{Exp}(\lambda)$ inter-arrival times.

So by Lemma, the # of $*$ between s & $s+t$ is $\mathrm{Poisson}(\lambda t)$. Note the # of such $*$ is N_{s+t}[80] $- N_s$[81].

[80] # of $*$ to left of $t+s$

[81] # of $*$ to left of s

Lecture 19 (xi)

Moreover, observe that

$$\left(N_{t+s} - N_s\right)_{t \geq 0} \quad \& \quad \left(N_r\right)_{r \leq s}$$

are independent.[82]

This is because the length of time until a * after s is indep. of whatever happened before time s (again by LoM).

[82]This follows by 1. LoM, 2. τ_i are IID.

Lecture 19 (xii)

Finally, we check (3.)

Proof of indep. increments:

This follows by induction. The previous slide shows $\left(N_{t+t_{n-1}} - N_{t_{n-1}}\right)_{t\geq 0}$ & $(N_r)_{r\leq t_{n-1}}$ indep.

$\therefore$ In particular,

$$\underbrace{N_{t_n} - N_{t_{n-1}}}_{\substack{\text{depends on} \\ \left(N_{t+t_{n-1}} - N_{t_{n-1}}\right)_{t\geq 0}}}, \underbrace{N_{t_{n-2}} - N_{t_{n-3}}, \ldots, N_{t_1} - N_{t_0}}_{\substack{\text{depends on} \\ (N_r)_{r\leq t_{n-1}}}}$$

are indep.

Lecture 19 (xiii)

We'll skip §2.2.2 on "more realistic models" for now. $\mathrm{PP}(\lambda)$ pts arrive at rate λ, independent of time t.

The non-homogeneous[83] Poisson Process on p.105 is interesting. We may put something on HW #3 about this – take a look yourself.

[83] λ_t depends on t

Lecture 19 (xiv)

§2.3 – Compound Poisson Processes.

It is often useful to add one more layer of randomness.

Eg $(N_t)_{t \geq 0}$ is a $\mathrm{PP}(\lambda)$ modelling occurrences of earthquakes over time on Hayward fault.

Moreover, suppose each earthquake

Lecture 19 (xv)

Has an IID magnitude Y_i which is also indep. of (N_t).

Then

$$S_t = \sum_{i=1}^{N_t} Y_i$$ [84]

is the total magnitude felt along fault by time t.

[84] Compound Poisson Process.

Lecture 19 (xvi)

Note: For a regular $\mathrm{PP}(\lambda)$, all $Y_i = 1$.

- More examples of compound Poisson Processes in §2.2.

Lecture 19 (xvii)

Theorem. (N_t) a $\mathrm{PP}(\lambda)$. (Y_i) an IID sequence indep. of (N_t).

Let

$$S_t = \sum_{i=1}^{N_t} Y_i$$

denote the Compound Poisson Process.

Then

$$\mathrm{E}S_t = \lambda t \ \mathrm{E}Y$$
$$\mathrm{Var}S_t = \lambda t \ \mathrm{E}(Y^2).$$

Lecture 19 (xviii)

In a similar way, we can calculate 2nd moment:

$$
\begin{aligned}
\mathrm{E}\big(S_t^2\big) &= \sum_{n=0}^{\infty} P(N_t = n)\ \mathrm{E}\big(S_t^2 \mid N_t = n\big) \\
&= \sum_{n=0}^{\infty} P(N_t = n)\big[n\ \mathrm{Var}Y + (n\ \mathrm{E}Y)^{2\,[85]}\big] \\
&= \mathrm{Var}Y\ \mathrm{E}(N_t) + (\mathrm{E}Y)^2\ \mathrm{E}\big(N_t^2\big) \\
&= \lambda t\big(\mathrm{Var}Y + (\mathrm{E}Y)^2\big) \\
&= \lambda t\ \mathrm{E}\big(Y^2\big) \quad \square
\end{aligned}
$$

---

[85] $\mathrm{Var}X = \mathrm{E}\big(X^2\big) - (\mathrm{E}X)^2$

Lecture 19 (xix)

Proof

$$\begin{aligned}
\mathrm{E}(S_t) &= \sum_{n=0}^{\infty} P(N_t = n)\ \mathrm{E}(S_t \mid N_t = n) \\
&= \sum_{n=0}^{\infty} P(N_t = n) n\ \mathrm{E}Y^{[86]} \\
&= \mathrm{E}Y \sum_{n=0}^{\infty} n P(N_t = n) \\
&= \mathrm{E}Y \cdot \mathrm{E}(N_t) \\
&= \lambda t\ \mathrm{E}Y^{[87]}
\end{aligned}$$

---

[86] $S_t \mid (N_t = n) = \sum_{i=1}^{n} Y_i$

[87] $N_t \sim \mathrm{Poisson}(\lambda t)$

Lecture 19 (xx)

Eg Customers arrive at a store according to a rate $\lambda = 81$ Poisson Process.

Suppose each customer spends an IID amount of money with mean \$8 and SD \$6. Find the mean revenue after 1 day, i.e. at time $t = 1$. Also find its SD.

Lecture 19 (xxi)

$$R = \sum_{i=1}^{N_1} Y_i, \quad \mathrm{E}Y = 8, \quad \mathrm{Var}Y = 36, \quad N_1 \sim \mathrm{Poisson}(81)$$

By theorem, $\mathrm{E}R = 81^{[88]} \cdot 8^{[89]} = \$648$

$$\begin{aligned} \& \ \mathrm{Var}R &= 81\ \mathrm{E}(Y^2) \\ &= 81\big(\mathrm{Var}Y + (\mathrm{E}Y)^2\big) \\ &= 81(36 + 64) = 8,100 \end{aligned}$$

$\therefore \mathrm{SD}(R) = \sqrt{8100} = \$90$

---

[88] $\lambda t = \lambda \cdot 1 = 81$

[89] $\mathrm{E}Y$

Lecture 20

Last week we defined & constructed the 1-dimensional, rate λ Poisson Process:

Definition $(N_t, t \geq 0)$ is a $\mathrm{PP}(\lambda)$ if

1. $N_0 = 0$
2. Poisson increments:
   $(N_{t+s} - N_s)$[90] $\sim \mathrm{Poisson}(\lambda t$[91]$)$, for any s, t
3. Indep. increments:
   $N_{t_1} - N_{t_0}, ..., N_{t_n} - N_{t_{n-1}}$ indep for any $t_0 < t_1 < ... < t_{n-1} < t_n$

---

[90]# pts in $(t, t+s]$

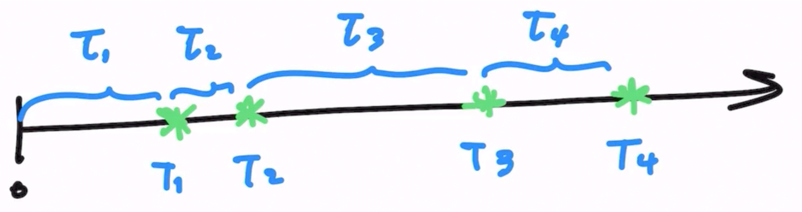
[91]length of $= (s, t+s]$

Lecture 20 (ii)

Construction:

"Point process" on $[0, \infty)$

with IID $\mathrm{Exp}(\lambda)$ inter-arrival times between points.



$$N_t = \# \text{ of } * \text{ in } [0, t]$$
$$= \max\{n : T_n \le t\}$$

Lecture 20 (iii)

Exponential RVs are very special: **only** continuous RV with Lack of Memory (LoM) property:

$$P(\tau > t + s \mid \tau > s) = P(\tau > t)$$

"Given it survives to s, it starts afresh at time s."

- This is key to many interesting/important properties of $\mathrm{PP}(\lambda)$ – the subject of §2.4 in [D].

Lecture 20 (iv)

We also studied Compound Poisson Processes, where each point is given an IID value Y_i. Then

$$S_t = \sum_{i=1}^{N_t} Y_i$$

Note: Standard $\mathrm{PP}(\lambda)$ is the case where $Y_i \equiv 1$ (as then $S_t = \sum_{i=1}^{N_t} 1 = N_t$).

Lecture 20 (v)

(§2.4 in [D]) Thinning of Poisson Processes[92]

Theorem. Let $(N_t, t \geq 0)$ be a $\mathrm{PP}(\lambda)$. Suppose Y is a RV on $\{1, 2, ..., m\}$, with $P(Y = j) = p_j$. Then the compound processes

$$N_j(t) = \sum_{i=1}^{N_t} \mathbb{1}_{\{Y_i = j\}}$$ [93]

are **independent** rate λp_j Poisson Processes.

---

[92] taking a PP, and spliting it up into indep. sub. PPs

[93] # of type j pts arrving by time t

Lecture 20 (vi)

Proof. For any s, t & $k_1, \ldots, k_m$[94]

$$P(N_1(t+s) - N_1(s) = k_1, \ldots, N_m(t+s) - N_m(s) = k_m)$$

$$= \binom{k_1 + \ldots + k_m}{k_1, \ldots, k_m} p_1^{k_1} \ldots p_m^{k_m} \cdot P\left(\underbrace{N_{t+s} - N_s}_{\text{Poisson}(\lambda t)} = k_1 + \ldots + k_m\right)$$

$$= \frac{\cancel{(k_1 + \ldots + k_m)!}}{k_1! \ldots k_m!} p_1^{k_1} \ldots p_m^{k_m} \cdot e^{-\lambda t} \frac{(\lambda t)^{k_1 + \ldots + k_m}}{\cancel{(k_1 + \ldots + k_m)!}}$$

$$= \prod_{j=1}^{m} e^{-\lambda p_j t} \frac{\left(\lambda p_j t\right)^{k_j}}{k_j!} = \underbrace{\prod_{j=1}^{m}}_{\text{indep.}} P\left(\text{Poisson}\left(\lambda p_j\right) = k_j\right)$$

[94] k_i: # of type i pts. arriving in $(s, t+s]$

Lecture 20 (vii)

This proves independence of the $N_j(t)$ processes & shows that they each satisfy condition (2.) in the definition of PP.

Conditions (1.) & (3.) are immediately inherited from (N_t).

Lecture 20 (viii)

Example 2.8: "Poissonization"

- Some problems become easier if # of objects is Poisson rather than a fixed (non-random) number.

Eg: # of spectators at a game is Poisson ($\lambda = 2263$). What is the prob. that for all 365 days of year, at least one person in crowd has this B-day?

Lecture 20 (ix)

Ans Supposing uniformly random B-days (not so realistic):

$$N_j = \#\text{ born on day } j \text{ of } 365$$

are IID $\text{Poisson}\left(\lambda_j = \frac{2263}{365} = 6.2\right)$

$$P\left(\text{all } N_j > 0\right) = \left(1 - e^{-6.2}\right)^{365\,[95]}$$
$$\approx 47.6\%$$

---

[95] $P(\text{Poisson}(\lambda) = 0) = e^{-\lambda}$

Lecture 20 (x)

§2.2.2 [D]: Non-homogeneous Poisson Processes

We skipped this last week:

This is a point process on $[0, \infty)$ with Poisson increments — but rate of arrival of points changes with time.

- More realistic in many applications.

Lecture 20 (xi)

Def $(N_t, t \geq 0)$ is a Poisson Process with rate $(\lambda_r$[96]$, r \geq 0)$ if

1. $N_0 = 0$
2. $N_t - N_s \sim \text{Poisson}\left(\int_s^t \lambda_r dr\right)$
3. Indep. increments[97]

If $\lambda_r = \lambda$ then regular $\text{PP}(\lambda)$, since

$$\int_s^t \lambda dr = \lambda \int_s^t dr = \lambda(t-s)$$[98]

[96]rate at which points arrive at time r

[97]condition (1.) and (3.) are the same as for regular (hom.) PP

[98]length of $(s, t]$

Lecture 20 (xii)

Note: Inter-arrival times no longer independent or exponential.

See p.106 in [D].

Thinning property can be generalized to non-hom.

Case:

Lecture 20 (xiii)

Theorem. Suppose we have a Poisson Process with rate λ, however keep a point arriving at time r only with prob. $p(r)$ (and otherwise delete point). Then the resulting process is a non-hom. Poisson Process with rate $(\lambda p(r), r \geq 0)$.

Lecture 20 (xiv)

Example 2.9 (M/G/∞ queue[99])

Calls arrive at a call center with so many agents according to a $\mathrm{PP}(\lambda)$. The duration of any given call follows some distribution with CDF G with $G(0) = 0$ & mean μ.

$$G(y) = P(\text{call time} \le y)$$

[99]for future reference.

Lecture 20 (xv)

By the previous theorem, for any t, the # of ongoing calls at time t is Poisson with mean

$$\int_0^t \lambda[1 - G(t-s)]^{[100]}\, ds = \lambda \int_0^t [1 - G(t-s)] ds$$

Taking $t \to \infty$, this converges to

$$\lambda \int_0^\infty [1 - G(s)] ds =^{[101]} \lambda\mu$$

---

[100] prob. call placed at time s is still ongoing

[101] "Tail rule:" $\mathrm{E}[X] = \int_0^\infty P(X > x) dx$, for non-negative RV X

Lecture 20 (xvi)

∴ In the long run/equilibrium, the # of calls in system will be Poisson with mean $\lambda\mu$.

Lecture 21

Theorem Suppose we have a Poisson Process with rate λ, however we keep a point arriving at time τ only with probability $p(\tau)$ (and otherwise delete the point). Then the resulting process is a non-homogeneous Poisson Process with rate
$(\lambda p(r), r \geq 0)$.

Lecture 21 (ii)

Example 2.9 (M/G/∞ queue)

Calls arrive at a call center with infinitely many agents according to a $\mathrm{PP}(\lambda)$. The duration of any given call follows some distribution with CDF G with $G(0) = 0$ & mean μ.

$$G(y) = P(\text{call time} \leq y)$$

Lecture 21 (iii)

By the previous theorem, for any t, the number of ongoing calls at time t is Poisson with mean

$$\int_0^t \lambda[1 - G(t-s)^{[102]}]^{[103]}\, ds = \lambda \int_0^t [1 - G(t-s)]ds$$

$$=^{[104]} \lambda \int_0^t [1 - G(u)]du$$

[102]prob. call takes at most $t - s$ time

[103]prob. call placed at time s is ongoing

[104]change of variables $u = t - s$, $du = -ds$

Lecture 21 (iv)

Taking $t \to \infty$, this converges to

$$\lambda \int_0^\infty [1 - G(s)]ds =^{[105]} \lambda\mu$$

$\therefore$ In the long run/equilibrium, the number of calls in the system will be Poisson with mean $\lambda\mu$.

[105] Tail rule: $E[X] = \int_0^\infty P(X > x)dx$

Lecture 21 (v)

Example Customers arrive at a store at a rate of 10/hour. 60% men, 40% women. Men stay for an $\mathrm{Exp}(2)$ duration, women for a $\mathrm{Uniform}\left(0, \frac{1}{2}\right)$ duration. What is the probability that in equilibrium (i.e. after the store has been open for a long time) there are 4 men and 2 women in the store?

Lecture 21 (vi)

By thinning, men and women arrive according to independent Poisson Processes with rates 6 and 4 per hour. Since $\mu_m = \frac{1}{2}$ & $\mu_w = \frac{1}{4}$, by the previous result, in equilibrium the # men and # women are independent Poissons with means $3 = 6 \cdot \frac{1}{2}$ & $1 = 4 \cdot \frac{1}{4}$.

Lecture 21 (vii)

$$\therefore P(M = 4^{[106]}, W = 2^{[107]}) = e^{-3}\frac{3^4}{4!} \cdot e^{-1}\frac{1^2}{2!}$$

[106] $\approx \mathrm{Poi}(3)$

[107] $\approx \mathrm{Poi}(1)$

Lecture 21 (viii)

Superposition of Poisson Processes

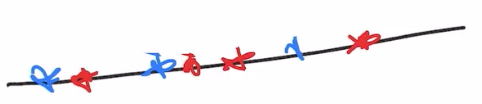
Instead of thinning, we can also add independent PPs to get another PP. Rates are added!

Theorem $N_1, ..., N_m$ independent $\mathrm{PP}(\lambda_i)$. Then $N_t = \sum_{i=1}^{m} N_i(t)$ is a Poisson Process, rate $\sum_{i=1}^{m} \lambda_i$.

Lecture 21 (ix)

Example “Poisson Race”

2 independent Poisson Processes: one red rate λ, one blue rate μ. What is the probability we see 6 reds before 4 blues?



Lecture 21 (x)

- Equivalently, we need at least 6 reds in the first 9 arrivals.
- Red + Blue is a $\mathrm{PP}(\lambda + \mu)$.
- Each arrival is Red w.p. $\frac{\lambda}{\lambda+\mu}$ and Blue w.p. $\frac{\mu}{\lambda+\mu}$.

Lecture 21 (xi)

$$\therefore P(6 \text{ reds before } 4 \text{ blues})$$

$$=^{[108]} \sum_{k=6}^{9} \binom{9}{k} \left(\frac{\lambda}{\lambda+\mu}\right)^{k} \left(\frac{\mu}{\lambda+\mu}\right)^{9-k}$$

[108] $P\left(\mathrm{Bin}\left(9, \frac{\lambda}{\lambda+\mu}\right) \geq 6\right)$

Lecture 21 (xii)

Last Section on Poisson Processes: **Conditioning**

Theorem Let (N_t) be a $\mathrm{PP}(\lambda)$. Let $s < t$ and $n \geq 0$. The conditional distribution of $N_s \mid N_t = n$ is $\text{Binomial}\big(n, \frac{s}{t}\big)$. That is,

$$P(N_s = k \mid N_t = n) = \binom{n}{k}\left(\frac{s}{t}\right)^k\left(1 - \frac{s}{t}\right)^{n-k}$$

Lecture 21 (xiii)

This follows by the following remarkable fact:

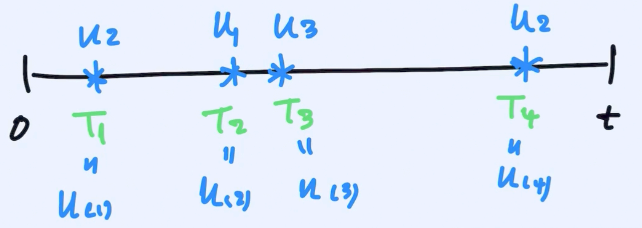
Theorem Let (N_t) be a $\mathrm{PP}(\lambda)$. Then conditional on $N_t = n$, the arrival times $T_k = \sum_{i=1}^{k} \tau_i$, $1 \leq k \leq n$, are distributed as the order statistics of n IID $\mathrm{Uniform}(0, t)$ random variables.

$u_1, u_2, \ldots, u_n$ IID $\mathrm{Uniform}(0, t)$
$u_{(1)}, u_{(2)}, \ldots, u_{(n)}$ order statistics

- $u_{(1)}$: smallest $u_i = T_1$ (1st arrival time)
- $u_{(2)}$: 2nd smallest $u_i = T_2$
- …
- $u_{(n)}$: largest $u_i = T_n$ (last arrival time)

Lecture 21 (xiv)

Eg Supposing $N_t = 4$, i.e. 4 points by time t:



$u_1, \ldots, u_4$ IID Uniform on $[0, t]$

$u_{(1)} =$ smallest, $u_{(2)} =$ 2nd smallest, ...

Lecture 21 (xv)

First, note that the 2nd theorem implies the 1st:

Conditional on $N_t = n$, n IID Uniform$(0, t)$ points are placed on $[0, t]$. Each has probability $\frac{s}{t}$ of landing in $[0, s]$. So,

$$P(N_s = k | N_t = n) = \binom{n}{k} \left(\frac{s}{t}\right)^k \left(1 - \frac{s}{t}\right)^{n-k}$$

Lecture 21 (xvi)

Proof of 2nd Theorem

First note that the joint PDF of $\left(u_{(1)}, u_{(2)}, \ldots, u_{(n)}\right)$ is

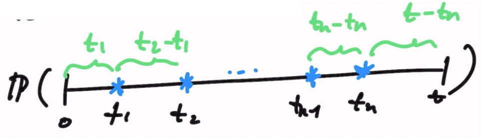
$$f(t_1, \ldots, t_n) = \begin{cases} \frac{n!}{t^n} & 0 < t_1 < \ldots < t_n < t \\ 0 & \text{otherwise} \end{cases}$$

This is a Stat 134 fact. Easy to see since $\frac{1}{t^n}$ is the joint PDF of $(u_1, u_2, \ldots, u_n)$ and there are $n!$ ways to order them.

Lecture 21 (xvii)

So to prove the theorem, we show that $\frac{n!}{t^n}$ is also the joint PDF of the arrival times $(T_1, \ldots, T_n)$ conditional on $N_t = n$. To see this: $P(N_t = n) = e^{-\lambda t}\frac{(\lambda t)^n}{n!} = P(\mathrm{Poi}(\lambda t) = n)$

To have $T_1 = t_1, \ldots, T_n = t_n$, the joint density of the inter-arrival times is:



$$= P(\mathrm{Exp}(\lambda) = t_1)P(\mathrm{Exp}(\lambda) = t_2 - t_1)$$
$$\cdots P(\mathrm{Exp}(\lambda) = t_n - t_{n-1})P(\mathrm{Exp}(\lambda) > t - t_n)$$

Lecture 21 (xviii)

$$= \lambda e^{-\lambda t_1} \quad \lambda e^{-\lambda(t_2 - t_1)} \quad \cdots \quad \lambda e^{-\lambda(t_n - t_{n-1})} \quad e^{-\lambda(t - t_n)}$$

$$= \lambda^n e^{-\lambda t} \text{ (telescoping)}$$

$\therefore$ the conditional joint density of $(T_1, \ldots, T_n)$ is

$$\frac{\lambda^n e^{-\lambda t}}{e^{-\lambda t} \frac{(\lambda t)^n}{n!}} = \frac{n!}{t^n}$$

Lecture 21 (xix)

The conditioning property is very useful for calculations. See examples and exercises in [D] & [PK]. We'll do one example from [PK]:

Lecture 21 (xx)

Eg Customers arrive at a party according to a $\mathrm{PP}(\lambda)$. The entrance fee depends on the time of arrival. At time t, the cover charge is $e^{-\beta t}$, $\beta > 0$. I.e. cover charge decays exponentially with rate β.

Find the mean revenue by time t.

Lecture 21 (xxi)

$$R_t = \sum_{k=1}^{N_t} e^{-\beta T_k}$$

$$E[R_t] \overset{\text{LoTP}}{=} \sum_{n=0}^{\infty} E\left[\sum_{k=1}^{n} e^{-\beta T_k} \mid N_t = n\right] P(N_t = n)$$

$$= \sum_{n=0}^{\infty} nE\left[e^{-\beta u}\right] P(N_t = n), \quad u \sim \text{Uniform}(0, t)$$

$$= E\left[e^{-\beta u}\right] \sum_{n=0}^{\infty} nP(N_t = n) = E\left[e^{-\beta u}\right] E[N_t]$$

$$= (\lambda t)\left(\frac{1}{t}\int_0^t e^{-\beta s} ds\right) = \frac{\lambda}{\beta}\left(1 - e^{-\beta t}\right)$$

Renewal Processes

Lecture 22

A renewal process is a generalization of the Poisson point processes.

Recall that the inter-arrival times between points in a $\mathrm{PP}(\lambda)$ are IID $\mathrm{Exp}(\lambda)$.

Lecture 22 (ii)

The Exp distribution is nice because it has the lack-of-memory property. This makes many exact calculations possible for the PP.

However, often Exp inter-arrival times are not present in real-world situations. We will see applications to **queues**.

Lecture 22 (iii)

Def: (N_t) is a renewal process (RP) if the inter-arrivals between points are IID.[109]

[109]not necessarily Exp.

Lecture 22 (iv)

- Times at which points arrive are called "renewal times." This is because the process starts afresh at such times with another IID seq of inter-arrival times.
- By LoM, this in fact holds at **any** time in a PP (not only at arrival times).

Lecture 22 (v)

Since no other distribution (other than Exp) has LoM, it is often difficult to do exact calculations about events at finite times of an RP.

But we can still study its LR behavior/equilibrium precisely, due to IID inter-arrival times.

Lecture 22 (vi)

The key to studying the LR behavior of an RP is the Law of Large Numbers (LLN).

Notation:

- $\tau_1, \tau_2, \ldots$ IID with common CDF $F(x) = P(\tau \leq x)$.
- $T_n = \sum_{i=1}^{n} \tau_i = n^{\text{th}}$ arrival time.
- $N_t = \max\{n : T_n \leq t\} = \#$ of arrivals by time t.

Lecture 22 (vii)

Eg

(X_n) a MC, $X_0 = x$. Let $T_n = \text{time of } n^{\text{th}} \text{ return to } x$.
By SMP[110], $N_t{}^{[111]} = \max\{n : T_n \leq t\}$ is a RP (times between visits are IID).

[110]strongg Markov Property

[111]# of returns to x by time t

Lecture 22 (viii)

The first important result:

Theorem. Let (N_t) be a RP with mean inter-arrival time $\mu = \mathrm{E}\tau$. ~~(Assume~~ $P(\tau > 0) > 0$~~, o/w process is degenerate.)~~
Then

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

Lecture 22 (ix)

Notice that, in particular, this proves that for a MC

$$\lim_{n\to\infty} \frac{N_n(j)^{[112]}}{n} = \frac{1}{\mathrm{E}_j \tau_j}$$

I.e. LR prop. of time spent in j is equivalent to inverse mean return time.

[112] # of visits to j by time n

Lecture 22 (x)

This theorem follows by the strong LLN (SLLN):

If $X_1, X_2, \ldots$ IID with $\mu = \mathrm{E}X$ and $S_n = \sum_{i=1}^{n} X_i$, then

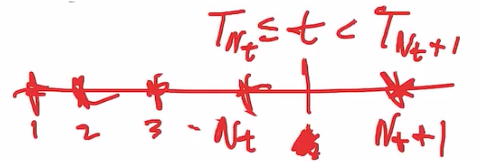
$$P\left(\frac{S_n}{n} \underset{n\to\infty}{\to} \mu\right) = 1$$

Lecture 22 (xi)

Proof of theorem using SLLN:

Let $\tau_i = X_i$. Then $S_n = T_n$.
So by SLLN, $\frac{T_n}{n} \to \mu$ as $n \to \infty$ (w.p. 1).

Next, observe $T_{N_t} \leq t < T_{N_t+1}$. (we discussed this during PP lectures.)



Lecture 22 (xii)

$$\therefore \frac{T_{N_t}}{N_t} \leq \frac{t}{N_t} \leq \frac{T_{N_t+1}}{N_t}$$

$$= \frac{T_{N_t+1}}{N_t+1} \frac{N_t+1}{N_t}$$

As $t \to \infty$, $N_t \to \infty$.

So $\frac{N_t+1}{N_t} \to 1$, $\frac{T_{N_t}}{N_t} \underset{\text{SLLN}}{\to} \mu$ & $\frac{T_{N_t+1}}{N_t+1} \underset{\text{SLLN}}{\to} \mu$.

Lecture 22 (xiii)

$$\therefore \text{ letting } t \to \infty$$

$$\mu \leq \lim_{t\to\infty} \frac{t}{N_t} \leq \mu$$

$$\Rightarrow \lim_{t\to\infty} \frac{N_t}{t} = \frac{1}{\mu}$$

Lecture 22 (xiv)

Most applications involve instead the following extension (cf. **Compound** PPs).

Def: (N_t) a RP with inter-arrival times (sometimes also called "holding times") $\tau_1, \tau_2, \ldots$. Let $r_1, r_2, \ldots$ be an IID seq of "rewards" indep. of holding times.

Lecture 22 (xv)

Then

$$R_t = \sum_{i=1}^{N_t} r_i$$

is called a **Renewal-Reward Process**.

- Each time a point arrives, a reward (possibly negative) is collected. $R_t = \text{total reward by time } t$.

Lecture 22 (xvi)

Theorem

$$P\left(\lim_{t\to\infty}\frac{R_t}{t}=\frac{\mathrm{E}r}{\mathrm{E}\tau}\right)=1$$

I.e.

$$\mathrm{LR}\ \frac{\mathrm{Reward}}{\mathrm{Time}}=\frac{\mathrm{E(Reward/Cycle)}}{\mathrm{E(Time/Cycle)}}$$

Lecture 22 (xvii)

Proof

$$\frac{R_t}{t} = \overbrace{\frac{N_t}{t}}^{\frac{1}{\mu} \text{ by prev. thm.}} \left(\underbrace{\frac{1}{N_t} \sum_{i=1}^{N_t} r_i}_{\mathrm{E}r \text{ by SLLN}} \right)$$

$$\therefore \lim_{t\to\infty} \frac{R_t}{t} = \frac{\mathrm{E}r}{\mathrm{E}\tau}$$

Lecture 22 (xviii)

Eg Long run car costs.

Lifetime of car is random with PDF h. Suppose we buy a new car when it breaks down or after T years, whichever comes first. Suppose new car costs $\$A$ and if a car breaks down this costs $\$B$ (towing costs, etc). What is our LR cost per time?

Lecture 22 (xix)

$$\mathrm{E}\tau = \int_0^T t h(t) dt + T \underbrace{\int_T^\infty h(t) dt}_{P(\text{car lasts} > T)}$$

$$\mathrm{E}r = A + B \qquad \underbrace{\int_0^T h(t) dt}_{P(\text{car breaks before time } T)}$$

(Additional $\$B$ if car breaks down, $\$A$ spent for new car either way.)

Lecture 22 (xx)

$\therefore$ LR cost per unit time is

$$\lim_{t\to\infty} \frac{R_t}{t} = \frac{\mathrm{E}r}{\mathrm{E}\tau}$$

$$= \frac{A + B\int_0^T h(t)dt}{\int_0^T th(t)dt + T\int_T^\infty h(t)dt}$$

Eg: If $A = 10$, $B = 3$ (in thousands of \$), $\tau \sim \mathrm{Uniform}(0, 10)$ years,

Lecture 22 (xxi)

$$\lim_{t\to\infty} \frac{R_t}{t} = \frac{10 + .3T}{T - .05T^2}$$

Optimal T (by calculus) is

$$T_* = \frac{1 + \sqrt{1.6}}{0.03} \approx 8.83 \text{ years}$$

Lecture 23

Recall:

A **Renewal Process** $(N_t, t \geq 0)$ has IID inter-arrival times $\tau_1, \tau_2, \ldots$ with $\mu = \mathrm{E}\tau$.

Main theorem:

$$\frac{N_t}{t} \to \frac{1}{\mu} \text{ as } t \to \infty$$

(On average, wait μ for next point, so $\frac{1}{\mu}$ points per unit time.)

Lecture 23 (ii)

Renewal-reward process

IID seq. of "rewards" $r_1, r_2, \ldots$ indep. of $\tau_1, \tau_2, \ldots$. Reward r_n is received at time $T_n = \sum_{i=1}^{n} \tau_i$.

$$R_t = \sum_{i=1}^{N_t} r_i$$
$$= \text{total reward by time } t$$

Main theorem

$$\frac{R_t}{t} \to \frac{Er}{\mu}$$

Lecture 23 (iii)

[D] § 3.2: Applications of Renewal Processes to queueing theory

1st example: GI/G/1 queue

GI = General input
G = General service times
1 = One server.

Lecture 23 (iv)

$\tau_1, \tau_2, \ldots$ inter-arrival times with some CDF $F(x) = P(\tau \leq x)$ and $\mathrm{E}\tau = \frac{1}{\lambda}$.

By previous theorem,

$$\frac{N_t}{t} \to \lambda \text{ as } t \to \infty$$

where $N_t = \#$ arrivals by time t.

Lecture 23 (v)

Next, suppose service times $s_1, s_2, \dots$ are IID with CDF $G(x) = P(S \le x)$ and $\mathrm{E}S = \frac{1}{\mu}$.

$\therefore$ customers arrive at rate λ
and customers are served at rate μ

Our first result is very intuitive: Namely, if $\lambda < \mu$ then server can handle the customers well:

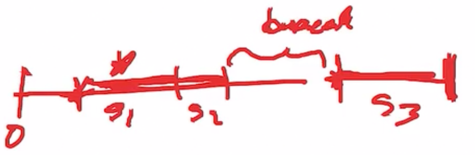
Lecture 23 (vi)

Theorem. If $\lambda < \mu$ then the long run proportion of time the server is busy is $\leq \frac{\lambda}{\mu} < 1$.

Proof. $T_n = \sum_{i=1}^{n} \tau_i$ = arrival time of n^{th} customer

By SLLN, $\frac{T_n}{n} \to \frac{1}{\lambda}$.

Similarly, let $S_n = \sum_{i=1}^{n} s_i$ = total time in service after n customers served.



Lecture 23 (vii)

By SLLN $\frac{S_n}{n} \to \frac{1}{\mu}$. Also, at time T_n, server has been busy $\leq S_n$ time.

& $\frac{S_n}{T_n} \to \frac{\lambda}{\mu}$.

Skipping some details (see p. 131). We'll see this another way later on ...

Lecture 23 (viii)

Cost equations (**any** queue)

Let X_s = # customers in system at time s (being served or waiting in queue).

Then LR average # customers in system is

$$L = \lim_{t \to \infty} \frac{1}{t} \int_0^t X_s ds$$

Lecture 23 (ix)

LR average amount of time a customer spends in system:

$$W = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} W_m$$

where W_m = amount of time spent in system by mth customer.

Lecture 23 (x)

Finally, LR average rate at which customers arrive:

$$\lambda_a = \lim_{t\to\infty} \frac{N_a(t)}{t} = \lambda^{[113]}$$

where $N_a(t)$ = # customers by time t.

Little's formula

$$L = \lambda_a W$$

[113] if GI/G/1 queue

Lecture 24

[D] § 3.2: Applications of Renewal Processes to queueing theory

1st example: GI/G/1 queue

GI = General input
G = General service times
1 = One server

Lecture 24 (ii)

Customers arrive at rate λ ($\mathrm{E}\tau = \frac{1}{\lambda}$)
Customers served at rate μ ($\mathrm{E}s = \frac{1}{\mu}$)

Theorem. If $\lambda < \mu$ then the long run proportion of time the server is busy is $\leq \frac{\lambda}{\mu} < 1$.

Lecture 24 (iii)

Cost Equations (any queue)

Let X_s = # customers in system at time s (being served or waiting in queue).

Then LR average # customers in system is

$$L = \lim_{t\to\infty} \frac{1}{t} \int_0^t X_s ds$$

Lecture 24 (iv)

LR average amount of time a customer spends in system:

$$W = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} W_m$$

Where W_m = amount of time spent in system by mth customer.

Lecture 24 (v)

Finally, LR average rate at which customers arrive:

$$\lambda_a = \lim_{t \to \infty} \frac{N_a(t)}{t} = \lambda^{[114]}$$

Where $N_{a(t)}$ = # customers by time t.

Little's Formula[115]

$$L = \lambda_a W$$

[114] if GI/G/1 Queue

[115] holds for any queue

Lecture 24 (vi)

Proof. Suppose customers pay \$1/min while in system.

Then in LR, system earns \$L/min.

Seen another way, each customer on average spends \$W. Therefore, since customers arrive at a rate λ_a, system earns $\$W\lambda_a/$ min.

Lecture 24 (vii)

An application:

Let W_Q = average time spent in queue (waiting to be served).

Note $W_Q = \underbrace{W}_{\text{ave. time in system}} - \underbrace{ES}_{\text{ave. time being served}}$

Lecture 24 (viii)

Let L_Q = LR average queue length (not counting customer being served).

If instead, only pay \$1 when in queue, then

$$L_Q = \lambda_a W_Q$$

The length of queue is 0 if no customers, and otherwise 1 less than # customers in system.

Lecture 24 (ix)

$$\begin{aligned}\therefore L_Q &= (L-1)(1-\pi_0) + L\pi_0 \\ &= L - 1 + \pi_0\end{aligned}$$

where π_0 = LR prob. of no customers altogether.

Altogether,

Lecture 24 (x)

$$W_Q = W - \mathrm{E}S$$

$$L_Q = \lambda_a W_Q \quad \& \quad L = \lambda_a W$$

$$L_Q = L - 1 + \pi_0$$

$$\begin{aligned} \Rightarrow \pi_0 &= L_Q - (L - 1) \\ &= 1 + L_Q - L \\ &= 1 + \lambda_a \big(W_Q - W \big) \\ &= 1 - \lambda_a \ \mathrm{E}S \end{aligned}$$

Lecture 24 (xi)

$\therefore$ if GI/G/1 Queue, $\pi_0 = 1 - \frac{\lambda}{\mu}$[116]

$\therefore 1 - \pi_0 =$ LR prop. busy $= \frac{\lambda}{\mu}$

This shows result of previous theorem (LR prop. busy $\leq \frac{\lambda}{\mu}$) is sharp.

[116] $\lambda_a = \lambda, \mathrm{E}S = \frac{1}{\mu}, \lambda_a\ \mathrm{E}S = \frac{\lambda}{\mu}$

Lecture 24 (xii)

Important special case of a GI/G/1 queue is

M/G/1 where we assume customers arrive according to a $\mathrm{PP}(\lambda)$. (M $\equiv$ Markovian).

X_n = # customers in queue when n^{th} customer starts being served.

Lecture 24 (xiii)

Thus MC can be constructed as follows:

Prob. exactly k customers arrive during any given service time is:

$$a_k = \int_0^\infty \underbrace{e^{-\lambda t}\frac{(\lambda t)^k}{k!}}_{[117]} \underbrace{dG(t)}_{[118]}$$

[117]Prob. k more customers arrive during this service time, $P(\mathrm{Poi}(\lambda t) = k)$

[118]G = CDF of service time, g = PDF, $dG(t) = g(t)dt$

Lecture 24 (xiv)

Let $\xi_1, \xi_2, \ldots$ be IID RVs.

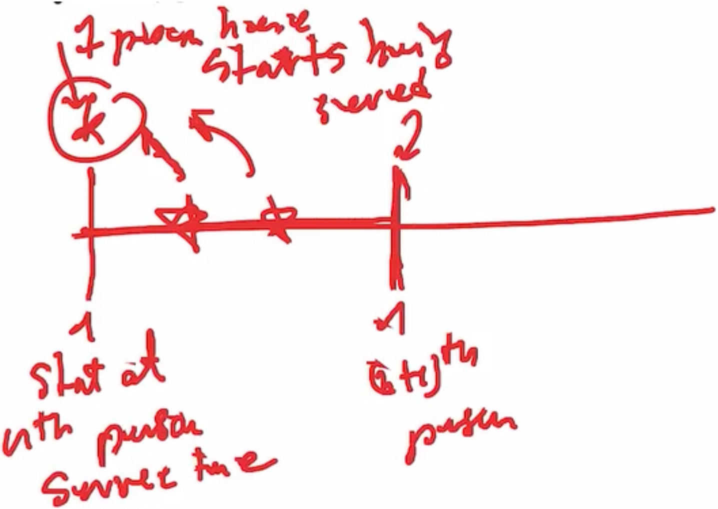
$$P(\xi = k) = a_k$$

ξ_n = # customers arriving during n^{th} service time.

If $X_n = \xi_n = 0$[119], then $X_{n+1} = 0$.
Otherwise: $X_{n+1} = X_n + \xi_n \underbrace{-1}_{n^{\text{th}} \text{ person has done being served}}$

[119] $X_n = 0$, there is no people in the queue when the n^{th} people starts to be served

Lecture 24 (xv)



Lecture 24 (xvi)

Therefore,

$$P = \begin{pmatrix} a_0 + a_1 & a_2 & a_3 & a_4 & \dots \\ a_0 & a_1 & a_2 & a_3 & \dots \\ 0 & a_0 & a_1 & a_2 & \dots \\ 0 & 0 & a_0 & a_1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

$$P_{00} = a_0 + a_1$$

$$P_{01} = a_2$$

$$P_{10} = a_0$$

Lecture 24 (xvii)

Theorem. In M/G/1 Queue,

X_n = # in queue when n^{th} customer starts service

$\lambda < \mu : (X_n)$ Pos Rec & $\mathrm{E}_0 T_0$[120] $= \frac{\mu}{\mu - \lambda}$[121]

$\lambda = \mu : (X_n)$ Null Rec

$\lambda > \mu : (X_n)$ Trans

[120] starting with 0 people in the queue, the expected length of time we have to wait to clear out the queue

[121] $\pi_0 = 1 - \frac{\lambda}{\mu}$

Lecture 24 (xviii)

Proof Consider the customers that arrive during n^{th} service time the "children" of this customer. Then we can compare (X_n) with a BP with offspring distribution

$$P(\xi = k) = a_k$$

Lecture 24 (xix)

$$
\begin{aligned}
\mathrm{E}\xi &= \sum_k k a_k \\
&= \sum_k k \int_0^\infty \underbrace{e^{-\lambda t}\frac{(\lambda t)^k}{k!}}_{P(\mathrm{Poi}(\lambda t)=k)} dG(t) \\
&= \int_0^\infty \underbrace{\lambda t}_{[122]} dG(t) \\
&= \lambda \, \mathrm{E}S \\
&= \frac{\lambda}{\mu}
\end{aligned}
$$

[122] $\sum_k kP(\mathrm{Poi}(\lambda t) = k) = \mathrm{E}(\mathrm{Poi}(\lambda t)) = \lambda t$

Lecture 24 (xx)

∴ Trans, Pos Rec, Null Rec follow by BP[123] theory.

Only remains to show
$\mathrm{E}_0 T_0 = \frac{\mu}{\mu - \lambda}$ in Pos Rec case.

Note BP transitions in discrete

[123]branching process

Lecture 24 (xxi)

steps, whereas (X_n) transitions in varying continuous steps $\tau_1, \tau_2, \ldots$ of the underlying $\mathrm{PP}(\lambda)$.

This did not matter for determining Trans, Pos Rec, Null Rec. But for $\mathrm{E}_0 T_0$ we need to look at time T_0,

Lecture 24 (xxii)

which is different in (X_n) than in BP.

Recall that for GI/G/1 we showed $\pi_0 = 1 - \frac{\lambda}{\mu} = \frac{\mu - \lambda}{\mu}$ if customers arrive rate λ.

That is case here for $\mathrm{PP}(\lambda)$.

$$\therefore \mathrm{E}_0 T_0 = \frac{1}{\pi_0} = \frac{\mu}{\mu - \lambda}$$

Lecture 24 (xxiii)

More to say about M/G/1 queue. We will skip this. See p.134-136 if interested.

Lecture 25

[D] §3-3 – Age & Residual Life.

(N_t) a RP with inter-arrival times IID $\tau_1, \tau_2, \ldots$

In a RP we have to deal with inter-arrival times that **do not** have lack of memory.

Lecture 25 (ii)

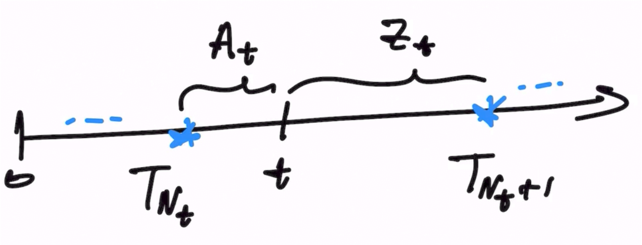
N_t = # points by time t

Recall each time a point arrives is a **renewal time**:

$$(N_t, t \geq 0) \stackrel{d}{=} \left(N_{T_n+t} - N_{T_n}, t \geq 0\right)$$

[For $\mathrm{PP}(\lambda)$ this is true at **any** point s, not only arrival times $s = T_n$]

Lecture 25 (iii)



$$A_t = t - T_{N_t}$$

$$= \text{Age of lightbulb in use at time } t$$

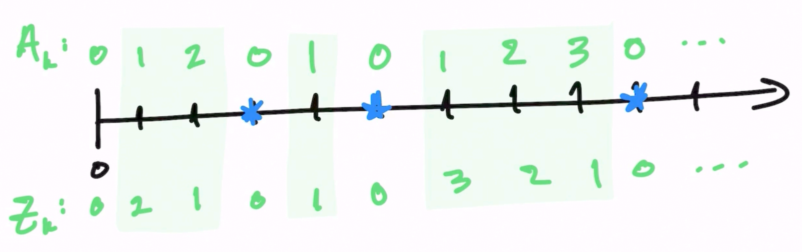
$$Z_t = T_{N_t+1} - t$$

$$= \text{Residual lifetime of lightbulb in use at time } t$$

Lecture 25 (iv)

Discrete case

Assume inter-arrival times $\tau_1, \tau_2, \ldots$ IID only taking values in $\{1, 2, 3, \ldots\}$.



$\therefore$ Enough to study one of (A_k) or (Z_k).

Lecture 25 (v)

We choose (Z_k).

Note: $Z_k = i > 0 \Rightarrow Z_{k+1} = i - 1$

If $Z_k = 0$ this is a renewal time.

(Z_k) is a MC:

$$\begin{cases} p_{0,j} = f_{j+1}{}^{[124]} & j \geq 0 \\ p_{i,i-1} = 1 & i > 0 \\ p_{i,j} = 0 & \text{o/w} \end{cases}$$

[124]PMF of inter-arrival times

Lecture 25 (vi)

We can find the SD of this MC using the "cycle trick". (We discussed this in §1 – see Theorem 1.24 3rd ed [D]).

Starting from 0, one visits any $i > 0$ at most once before returning and this happens $\Leftrightarrow$ first jump from 0 is to some $j \geq i$.

$\therefore \mu_i = P(T > i)$ is a stationary measure.

Lecture 25 (vii)

Note $\sum_i \underbrace{P(T > i)}_{\mu_i} =^{[125]} E(\tau)$.

$\therefore \pi_i = \frac{P(\tau > i)}{E(\tau)}$ is SD.

[Assuming $E(\tau) < \infty$].

So by MC theory

[125] tail formula for expectation

Lecture 25 (viii)

Theorem (Assuming (Z_n) is **IRR** & **APER**) we have

$$\lim_{n\to\infty} P({Z_n}^{[126]} = i) = \frac{P(\tau \geq i)}{E(\tau)}$$

This gives the limiting distribution of the (Z_n) in terms of distribution of inter-arrival times τ.

[126] Also for A_n

Lecture 25 (ix)

Continuous Case

Done in detail in [D]. We'll just touch on it briefly.

In discrete case, we showed $\pi_i = \frac{P(\tau > i)}{E(\tau)}$ is limiting distribution for (Z_k) [& also (A_k)].

Lecture 25 (x)

Similarly, in the continuous case, one can show that

$$g(x) = \frac{P(\tau > x)}{E(\tau)}$$

is limiting distribution for $(Z_s, s \geq 0)$ & $(A_s, s \geq 0)$.

Using this

Lecture 25 (xi)

Inspection Paradox

$$\int_0^\infty z g(z) dz = \int_0^\infty z \frac{P(\tau > z)}{E(\tau)} dz$$

$$= \frac{1}{E(\tau)} \int_0^\infty z P(\tau > z) dz$$

$$=^{[127]} \frac{1}{E(\tau)} \frac{E(\tau^2)}{2}$$

[127]Recall: $E(\tau) = \int_0^\infty P(\tau > z) dz$. Similarly, $E\left(\tau^k\right) = \int_0^\infty k z^{k-1} P(\tau > z) dz$

Lecture 25 (xii)

$\therefore$ For large t,

$$L_t{}^{[128]} = A_t + Z_t$$

$$\therefore E(L_t) \approx \cancel{2} \cdot \frac{1}{E(\tau)} \frac{E(\tau^2)}{\cancel{2}}$$

$$= \frac{E(\tau^2)}{E(\tau)} > E(\tau)$$

(since $E(\tau^2) - (E(\tau))^2 = \mathrm{Var}(\tau) > 0$[129])

[128]Lifetime of bulb in use at time t

[129]If $\mathrm{Var}(\tau) = 0$, then all $\tau_i \equiv c$. This case isn't interesting.

Lecture 25 (xiii)

This is an (**apparent**) paradox, since each lightbulb has mean $E(\tau)$.

However, lightbulbs with longer lifetimes are more likely to be the ones in use when we happen to make an inspection.

Continuous Time Markov Chains

Lecture 26

Recall that discrete time MC's (X_n) have "Markov Property":

$$P(X_{n+m} = j \mid X_n = i, X_{n-1} = i_{n-1}, ..., X_0 = i_0)$$
$$= P(X_{n+m} = j \mid X_n = i)$$

for all n, m & $j, i, i_{n-1}, ..., i_0 \in S$.

I.e. given present, past & future are independent. Only need to know present to determine probability of future behavior.

Lecture 26 (ii)

In §4, we extend this to processes $(X_t, t \geq 0)$ evolving continuously in time. We still assume, however, that the state space (possible values of X_t) is discrete.

We extend the MP as follows:

Lecture 26 (iii)

Def

$(X_t, t \geq 0)$ is a **continuious** time MC on $\Omega \subset Z$ if

$$P\Big(X_{t+s} = j \mid X_s = i, X_{s_n} = i_n, \ldots, X_{s_0} = i_0\Big)$$
$$= P\big(X_{t+s} = j \mid X_s = i\big)$$

For all $0 \leq s_0 < s_1 < \ldots < s$ & $0 \leq t$ & $j, i, i_n, i_{n-1}, \ldots, i_0 \in \Omega$.

Lecture 26 (iv)

In fact, we have already seen some examples of such processes, e.g. the Poisson Process.

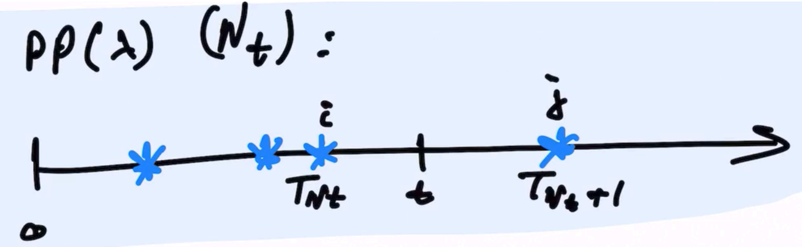
Lecture 26 (v)

E.g. (N_t) a PP(λ) & (Y_n) a discrete time MC with TR probs r_{ij}. Let $X_t = Y_{N_t}$. The (X_t) is a continuous time MC.

(For instance, if $r_{i,i+1} = 1$ & $r_{ij} = 0$ o/w, we see (N_t) itself is a continuous time process.)

Lecture 26 (vi)

Proof



If $X_t = i$, then by LOM & MP of (Y_n), the next transition will occur after an $\mathrm{Exp}(\lambda)$ amount of time to some j w.p. r_{ij}, independently of anything at times $s < t$.

Lecture 26 (vii)

For discrete time MC everything we wanted to know came from the "1-step transition probabilities" $p_{ij} = P(X_{n+1} = j | X_n = i)$.

In continuous time, there is no first step, **but** we still have $p_t(i, j) = P(X_t = j \mid X_0 = i)$[130].

[130] currently at state i, prob. of being at state j after time t

Lecture 26 (viii)

E.g. in previous example:

$$p_t(i,j) \overset{\text{LoTP}}{=} \sum_{n=0}^{\infty} \underbrace{e^{-\lambda t}\frac{(\lambda t)^n}{n!}}_{P(n\text{ *'s in }(0,t])} \underbrace{(r^n)_{ij}}_{\substack{n\text{-step pr.}\\ P(Y_n=j \mid Y_0=i)}}$$

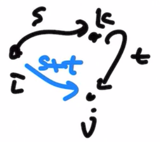
Lecture 26 (ix)

And we still have the Chapman-Kolmogorov Equations:

$$\sum_k p_s(i,k)p_t(k,j)$$
$$= p_{s+t}(i,j)$$

Proof

Again by MP + LoTP, at time $s \in [0, t+s]$, the MC has to be a some k.



Lecture 26 (x)

Although there is no 1^{st} step, it is intuitive that in this continuous context we should instead take a limit $t \to 0$ to distill all relevant info about the transition probabilities of a continuous MC.

Lecture 26 (xi)

More specifically, the derivative at $t = 0$ gives us the "jump rates".

$$q_{ij} = \lim_{t \to 0} \frac{p_t(i,j)}{t} \quad (i \neq j)$$

This is the **rate** at which (X_t) jumps from i to j.

Lecture 26 (xii)

E.g. in the previous examples:

$$
\begin{aligned}
&P(\geq 1 \text{ *'s in } (0,t]) \\
&= \underbrace{P(= 1 \text{ * in } (0,t])}_{e^{-\lambda t}\lambda t} + o(t)^{[131]} \\
&\therefore q_{ij} = \lim_{t\to 0} \frac{p_t(i,j)}{t} = \lambda r_{ij}{}^{[132]}
\end{aligned}
$$

(since $e^{-\lambda t} \to 1$ as $t \to 0$)

---

[131] Some function going $\to 0$ as $t \to 0$

[132] we will do this in more detail in §4.2

Lecture 26 (xiii)

This makes sense:

*‘s appear in PP(λ) at rate λ & $i \to j$ happens w.p. r_{ij}.
Therefore, in (X_t), transitions $i \to j$ occur at rate λr_{ij}.

E.g. in PP(λ) itself, we just have $q_{i,i+1} = \lambda$.

Lecture 26 (xiv)

Other simple (but important) examples:

E.g. M/M/s queue s servers
Arrivals are IID $\mathrm{Exp}(\lambda)$
Service times IID $\mathrm{Exp}(\mu)$

$$q_{i,i+1} = \lambda$$
$$q_{i,i-1} = \min\{i, s\}\mu$$

Lecture 27

Last class, we began discussing continuous time MC's.

For these processes, there is no "1st step" after any given time. But we still have transition probabilities

$$p_t(i,j) = P\big(X_{s+t} = j \mid X_s = i\big)$$

Lecture 27 (ii)

Recall that for discrete time MC's the 1-step probabilities p_{ij} in P held all info we needed to study it.

There is no 1st step in continuous time, so instead we take $t \to 0$ to distill info needed to study a continuous time MC:

Lecture 27 (iii)

More specifically, the derivative at $t = 0$ gives us the “jump rate.”

$$q_{ij} = \lim_{t \to 0} \frac{p_t(i, j)}{t} \quad (i \neq j)$$

This is the **rate** at which (X_t) jumps from i to j.

Lecture 27 (iv)

Example:

E.g. Poisson Process (N_t) with rate λ. Here:

$$q_{i,i+1} = \lambda \quad \text{for } i \geq 0.$$

Because new points appear in time at rate λ.

Lecture 27 (v)

E.g. M/M/s queue s servers

Arrivals are IID Exp(λ).

Service times IID Exp(μ).

X_t = # customers in system (in queue or being served) at time t.

$$\begin{cases} q_{i,i+1} = \lambda \\ q_{i,i-1} = \min\{i, s\}^{[133]}\ \mu \end{cases}$$

[133] At most s can be served at once

Lecture 27 (vi)

E.g. Branching Process

A population where particles die at rate μ & give birth at rate λ. (Yule process if $\mu = 0$.)

$$\begin{cases} q_{i,i+1} = \lambda_i \\ q_{i,i-1} = \mu_i \end{cases}$$

Lecture 27 (vii)

Often it is simple/natural to write down jump rates for a given MC, as in previous examples.

But how to actually construct the MC, given the q_{ij}?

Lecture 27 (viii)

Suppose we know:

$$q_{ij} = \text{rate } i \to j$$

Then put:

$$\lambda_i = \sum_{j \neq i} q_{ij} = \text{rate leave } i$$

$$r_{ij} = \begin{cases} \frac{q_{ij}}{\lambda_i} & \lambda_i > 0 \\ 0 & \lambda_i = 0 \end{cases}$$

Lecture 27 (ix)

If $\lambda_i = 0$ then MC never leaves if it hits i. I.e. state i is absorbing.

To construct MC (X_t), we proceed as follows:

Case 1 If ever MC reaches an absorbing state $(\lambda_i = 0)$ then MC stays forever, & so construction is done.

Lecture 27 (x)

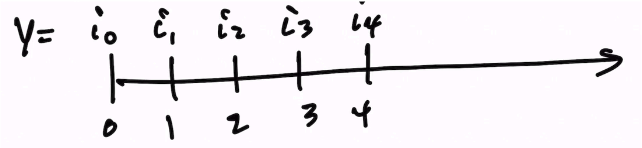
Otherwise:

Case 2 If MC most recently jumped to some i with $\lambda_i > 0$, then select an $\mathrm{Exp}(\lambda_i)$ RV. After this amount of time, jump to some other state j with prob. $r_{ij} = \frac{q_{ij}}{\lambda_i}$.

Lecture 27 (xi)

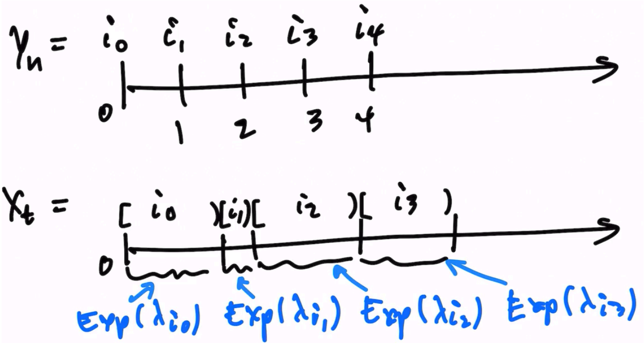
Another way of thinking of this is:

(1) Start with a discrete MC (Y_n) with transition probabilities r_{ij}.



Lecture 27 (xii)

Then stretch/pull the unit intervals according to independent exponentials.



Lecture 27 (xiii)

(Y_n) is sometimes called the "skeleton" of (X_t).

[D] p.151 explains how you can simulate this on a computer:

1. Get IID $T_0, T_1, \ldots$ $\text{Exp}(1)$.
2. Start at i_0 at time 0. Jump to j w.p. $r_{i_0,j}$ after time $t_1 = \frac{T_0}{\lambda_{i_0}}$.

Lecture 27 (xiv)

⋮

After k jumps in state i_k
After time $T_k = \sum_{i=1}^{k} t_i$
After another time $t_{k+1} = \frac{\tau_{k+1}}{\lambda_{i_k}}$ amount of time, jump to j w.p. $r_{i_k,j}$.

⋮

& so on until an absorbing state is reached (if ever).

Lecture 27 (xv)

Why is this the right construction?

Good question – but beyond scope of Stat 150.

Briefly: it can be shown that the definition of the Poisson Process given before:

Lecture 27 (xvi)

$$\begin{cases} 1. \quad N_0 = 0 \\ 2. \quad \text{All } N_{t+s} - N_s \sim \text{Poisson}(\lambda t) \\ 3. \quad \text{Indep. Incr.} \end{cases}$$

is equivalent to:

$$\begin{cases} 1. \quad N_0 = 0 \\ 2. \quad P\left(N_{t+h} - N_t = 1 \mid N_t = i\right) = \lambda h + o(h) \;\text{ as }\; h \to 0 \\ 3. \quad P\left(N_{t+h} - N_t = 0 \mid N_t = i\right) = 1 - \lambda h + o(h) \;\text{ as }\; h \to 0 \end{cases}$$

Lecture 27 (xvii)

I.e. $N_0 = 0$ & $q_{i,i+1} = \lambda$ defines the Poisson Process.

Now, recall that the Poisson process is constructed using IID $\mathrm{Exp}(\lambda)$ inter-arrival times.

Lecture 27 (xviii)

$\therefore$ It is natural to expect that the above construction will give a process (X_t) with

$$P(X_{t+h} = j \mid X_t = i) = q_{ij}h + o(h)(i \neq j) \quad \text{as } h \to 0$$

$$P(X_{t+h} = i \mid X_t = i) = 1 - \lambda_i h + o(h) \quad \text{as } h \to 0$$

$\left(\lambda_i = \sum_j q_{ij}\right)$

Lecture 28

Recall from last class, we discussed how to construct continuous time MC's.

$$p_{t(i,j)} = P\big(X_{s+t} = j \mid X_s = i\big)$$

$$q_{ij} = \lim_{t \to 0} \frac{p_{t(i,j)}}{t} \quad i \neq j$$

$$= \text{jump rate } i \to j$$

$$\lambda_i = \sum_{j \neq i} q_{ij} = \text{jump rate out of } i$$

Lecture 28 (ii)

It turns out (we will not prove this in Stat 150) the way to do this is as follows:

If currently in state i, hold MC in i for an $\mathrm{Exp}(\lambda_i)$ amount of time, and then jump to some j w.p. $r_{ij} = \frac{q_{ij}}{\lambda_i}$.

– Unless $\lambda_i = 0$, then MC stays in i forever (absorbing state).

Lecture 28 (iii)

To see why this is the correct construction, it may be helpful to consider the following simple example:

Lecture 28 (iv)

Birds visit a feeder according to independent Poisson processes:

$$(N_t^R)$$

Robins rate λ_R

$$(N_t^B)$$

Blackbirds rate λ_B

$$(X_t) = (N_t^R, N_t^B)$$

1st coord at rate λ_R 2nd coord at rate λ_B

Lecture 28 (v)

Times between birds are iid $\mathrm{Exp}(\lambda_R + \lambda_B)$. And

$$P(\mathrm{Robin}) = \frac{\lambda_R}{\lambda_R + \lambda_B}$$

$$P(\mathrm{Blackbird}) = \frac{\lambda_B}{\lambda_R + \lambda_B}$$

So hold (X_t) for $\mathrm{Exp}(\lambda_R + \lambda_B)$ time & then increase 1st or 2nd co-ord w.p. $\frac{\lambda_R}{\lambda_R+\lambda_B}$ and $\frac{\lambda_B}{\lambda_R+\lambda_B}$.

Lecture 28 (vi)

[D] §4.2 — Transition Probabilities

For discrete MC's we could find n-step p_{ij}^n by multiplying P n times: $p_{ij}^n = (P^n)_{ij}$.

This was proved by CK-eqns.

For continuous MC, how to find $p_{t(i,j)}$ probabilities from q_{ij}?

Lecture 28 (vii)

For continuous MC, how to find $p_{t(i,j)}$ probabilities from q_{ij}?

Recall jump rates

$$q_{ij} = \lim_{h \to 0} \frac{p_{h(i,j)}}{h}$$

play the role of p_{ij} in the discrete case.

Once again, the answer is to use the CK-eqns, **but** more complicated:

Lecture 28 (viii)

By CK-eqns:

$$p_{t+h}(i,j) - p_{t(i,j)}$$

$$= \left[\sum_k p_{h(i,k)} p_{t(k,j)}\right] - p_{t(i,j)}$$

$$= \left[\sum_{k \neq i} p_{h(i,k)} p_{t(k,j)}\right] + \Big(p_{h(i,i)} - 1\Big) p_{t(i,j)}$$

Now divide by h, then take $h \to 0$:

Lecture 28 (ix)

$$\frac{p_{t+h}(i,j) - p_{t(i,j)}}{h} \to p'_{t(i,j)}$$

as $h \to 0$

$$= \left[\sum_{k \neq i} \frac{p_{h(i,k)}}{h} p_{t(k,j)}\right] - \left(\frac{p_{h(i,i)} - 1}{h}\right) p_{t(i,j)}$$

As $h \to 0$, $\frac{p_{h(i,k)}}{h}$ becomes q_{ik}.

As $h \to 0$, what is $\frac{p_{h(i,i)}-1}{h}$?

Lecture 28 (x)

$$p_{h(i,i)} = 1 - \sum_{j \neq i} p_{h(i,j)}$$

$$\therefore \frac{p_{h(i,i)} - 1}{h} = -\sum_{j \neq i} \frac{p_{h(i,j)}}{h}$$

As $h \to 0$, the sum on the right becomes $\lambda_i = \sum_{j \neq i} q_{ij}$.

Hence, altogether,

$$p'_{t(i,j)} = \sum_{k \neq i} q_{ik} p_{t(k,j)} - \lambda_i p_{t(i,j)}$$

Lecture 28 (xi)

If we define matrix Q with entries

$$Q_{ij} = \begin{cases} q_{ij} & \text{if } i \neq j \\ -\lambda_i & \text{if } i = j \end{cases}$$

Then previous equations can be written as:

Kolmogorov's Backward Equation:

$$P_t' = QP_t$$

Lecture 28 (xii)

$$p_{h(i,i)} = 1 - \sum_{j \neq i} p_{h(i,j)}$$

$$\therefore \frac{p_{h(i,i)} - 1}{h} = - \sum_{j \neq i} \frac{p_{h(i,j)}}{h} \to -\lambda_i$$

as $h \to 0$

Hence, altogether,

DE for $P_{t(i,j)}$

$$p'_{t(i,j)} = \sum_{k \neq i} q_{ik} p_{t(k,j)} - \lambda_i p_{t(i,j)}$$

Lecture 28 (xiii)

If we define matrix Q with entries

$$Q_{ij} = \begin{cases} q_{ij} & \text{if } i \neq j \\ -\lambda_i & \text{if } i = j \end{cases}$$

Then previous equations can be written as:

Kolmogorov's Backward Equation:

$$P_t' = QP_t$$

Lecture 28 (xiv)

By the same argument, **but**

$$p_{t+h}(i,j) - p_{t(i,j)} = \left[\sum_k p_{t(i,k)} p_{h(k,j)}\right] - p_{t(i,j)}$$

instead of

$$p_{t+h}(i,j) - p_{t(i,j)} = \left[\sum_k p_{h(i,k)} p_{t(k,j)}\right] - p_{t(i,j)}$$

Lecture 28 (xv)

Kolmogorov's Forward Equation:

$$P_t' = P_t Q$$

Note: By Kolmogorov Forward & Backward (K-eqns)

$$QP_t = P_t Q$$

Lecture 28 (xvi)

This is remarkable, since $AB \neq BA$ in general for matrices A, B.

The reason why $P_t Q = Q P_t$ is that these matrices are made up of powers of Q:

Lecture 28 (xvii)

Recall $y' = ay \Rightarrow y = e^{ax} = \sum_{k=0}^{\infty} \frac{(ax)^k}{k!}$

Here we have $P_t' = QP_t$. These are matrices, but still natural to expect

$$P_t = e^{Qt} := \sum_{k=0}^{\infty} \frac{(Qt)^k}{k!}$$

Lecture 28 (xviii)

Indeed,

$$\frac{d}{dt}\sum_{k=0}^{\infty}\frac{(Qt)^k}{k!} = \sum_{k=1}^{\infty}\frac{Q^k t^{k-1}}{(k-1)!}$$

$$= Q\sum_{k=0}^{\infty}\frac{(Qt)^k}{k!}$$

$$= \left(\sum_{k=0}^{\infty}\frac{(Qt)^k}{k!}\right)Q$$

Lecture 28 (xix)

The rest of §4.2 is devoted to examples of MC's where K-eqns can be solved to find $P_{t(i,j)}$ explicitly.

We'll do **some** of these.

Lecture 28 (xx)

Eg Poisson Process.

We know

$$p_{t(i,j)} = P(\mathrm{Poi}(\lambda t) = j - i)$$

$$= e^{-\lambda t} \frac{(\lambda t)^{j-i}}{(j-i)!}$$

By BKE we should have:

$$p'_{t(i,j)} = \lambda p_{t(i,j+1)} - \lambda p_{t(i,j)}$$

Lecture 28 (xxi)

This is easy to check!

$$p_{t(i,j)} = e^{-\lambda t} \frac{(\lambda t)^{j-i}}{(j-i)!}$$

$$p'_{t(i,j)} = \lambda e^{-\lambda t} \frac{(\lambda t)^{j-i-1}}{(j-i-1)!} - \lambda e^{-\lambda t} \frac{(\lambda t)^{j-i}}{(j-i)!}$$

$$= \lambda p_{t(i,j+1)} - \lambda p_{t(i,j)}$$

$$\checkmark$$

Lecture 29

(1) To construct (X_t):

The inductive procedure works as follows: If last jump was to i, then hold the MC at i for an $\mathrm{Exp}(\lambda_i)$ time then jump to j w.p. $\frac{q_{ij}}{\lambda_i}$ [Assuming $\lambda_i > 0$, o/w if $\lambda_i = 0$ the MC stays at i forever, i.e. i is absorbing].

Lecture 29 (ii)

(2) We can use the jump rates q_{ij} to find $p_{t(i,j)}$:

K-BE $P_t' = QP_t$ K-FE $P_t' = P_tQ$

where $(Q)_{ij} = \begin{cases} -\lambda_i & \text{if } i=j \\ q_{ij} & \text{if } i\neq j \end{cases}$

$(P_t)_{ij} = p_{t(i,j)}$ & $(P_t')_{ij} = p'_{t(i,j)}$

Lecture 29 (iii)

Eg 2-state MC.

Jump $1 \to 2$ rate λ $2 \to 1$ rate μ

$q_{12} = \lambda$ $q_{21} = \mu$

$Q = \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix}$

To find $p_{t(i,j)}$ we solve

$$P_t' = QP_t$$

Lecture 29 (iv)

I.e.

$$\begin{pmatrix} p'_{t(11)} & p'_{t(12)} \\ p'_{t(21)} & p_{t(22)} \end{pmatrix} = \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix} \begin{pmatrix} p_{t(11)} & p_{t(12)} \\ p_{t(21)} & p_{t(22)} \end{pmatrix}$$

$$\Rightarrow p'_{t(11)} = -\lambda\Big(p_{t(11)} - p_{t(21)}\Big)$$

$$p'_{t(21)} = \mu\Big(p_{t(11)} - p_{t(21)}\Big)$$

$$\Big(p_{t(11)} - p_{t(21)}\Big)' = -(\lambda + \mu)\Big(p_{t(11)} - p_{t(21)}\Big)$$

Lecture 29 (v)

$$\Rightarrow p_{t(11)} - p_{t(21)} = e^{-(\lambda+\mu)t}$$

(Using I.C.s $p_0(11) = 1$ & $p_0(21) = 0$)

Plugging back into above:

$$p'_{t(11)} = -\lambda e^{-(\lambda+\mu)t}$$

$$\Rightarrow p_{t(11)} - p_0(11) = -\lambda \int_0^t e^{-(\lambda+\mu)s} ds$$

$$= \frac{\lambda}{\lambda+\mu}\left(e^{-(\lambda+\mu)t} - 1\right)$$

Lecture 29 (vi)

$$\Rightarrow p_{t(11)} = \frac{\lambda}{\lambda+\mu}e^{-(\lambda+\mu)t} - \frac{\lambda}{\lambda+\mu} + 1$$

$$= \frac{\lambda}{\lambda+\mu}e^{-(\lambda+\mu)t} + \frac{\mu}{\lambda+\mu}$$

Similarly,

$$p_{t(21)} = -\frac{\lambda}{\lambda+\mu}e^{-(\lambda+\mu)t} + \frac{\mu}{\lambda+\mu}$$

Lecture 29 (vii)

Note that both

$$\lim_{t\to\infty} p_{t(11)} = \frac{\mu}{\lambda+\mu}$$

$$\lim_{t\to\infty} p_{t(21)} = \frac{\mu}{\lambda+\mu}$$

LR behavior of continuous time mc's will be our next topic — after slides.

Lecture 29 (viii)

Eg Branching Processes

First we consider the special case of no death: Yule Process

$q_{i,i+1} = \beta i$

Here particles split into two at rate β. Particles live forever, so population will continue to grow exponentially.

$Y_t =$ # particles at time t.

Lecture 29 (ix)

Theorem. For Yule Process (Y_t):

$$p_{t(1,j)} = e^{-\beta t}\left(1 - e^{-\beta t}\right)^{j-1} \quad (j \geq 1)$$

$$= P\left(\mathrm{Geo}\left(e^{-\beta t}\right) = j\right)$$

&

$$p_{t(i,j)} = \binom{j-1}{i-1}\left(e^{-\beta t}\right)^{i}\left(1 - e^{-\beta t}\right)^{j-i}$$

Lecture 29 (x)

Note that by $p_{t(1,j)} = e^{-\beta t}\left(1 - e^{-\beta t}\right)^{j-1}$,

$$P\left(e^{-\beta t} Y_t > x\right) = P\left(Y_t > x e^{\beta t}\right)$$

$$= \left(1 - e^{-\beta t}\right)^{x e^{\beta t}}$$

$$\to e^{-x} \quad \text{as} \quad t \to \infty$$

$$\therefore e^{-\beta t} Y_t \xrightarrow{d} \mathrm{Exp}(1) \quad \text{as} \quad t \to \infty$$

So starting with 1 particle, the pup. grows exponentially, $Y_t \approx \mathrm{Exp}(1) e^{\beta t}$.

Lecture 29 (xi)

We won't prove the theorem in detail, but note by K-FE:

$$\text{only} \quad q_{j-1,j} \neq 0$$

$$p'_{t(1,j)} = \sum_{k \neq j} p_{t(1,k)} q_{kj} - p_{t(1,j)} \lambda_j$$

$$= p_{t(1,j-1)} \beta(j-1) - p_{t(1,j)} \beta j$$

You can check that this is satisfied by $p_{t(1,j)} = e^{-\beta t}\left(1 - e^{-\beta t}\right)^{j-1}$.

Lecture 29 (xii)

Once $p_{t(1,j)}$ is proved, it is easy to get $p_{t(i,j)}$.

By independence, we can study the "family trees" of the initial i particles separately:

$$p_{t(i,j)} = \sum_{n_1+\ldots+n_i=j, n_k \geq 1} \mathrm{prod}_{k=1}^{i} p_{t(1,n_k)}$$

Lecture 29 (xiii)

For any such $n_1, \ldots, n_i$

$$\text{prod}_{k=1}^{i} p_{t(1,n_k)} = \text{prod}_{k=1}^{i} e^{-\beta t}\left(1 - e^{-\beta t}\right)^{n_k - 1}$$

$$= \left(e^{-\beta t}\right)^{i}\left(1 - e^{-\beta t}\right)^{j-i}$$

There are $\binom{j-1}{i-1}$ such $n_1, \ldots, n_i$.

v **...** v 1 2 ... j

Choosing $i - 1$ of these $j - 1$ v's determines $n_1, \ldots, n_i \geq 1$ & $\sum n_i = j$.

Lecture 29 (xiv)

$$\therefore p_{t(i,j)} = \binom{j-1}{i-1} \left(e^{-\beta t}\right)^{i} \left(1 - e^{-\beta t}\right)^{j-i}$$

See also p159 of Durrett for proof of $p_{t(1,j)} = e^{-\beta t}\left(1 - e^{-\beta t}\right)^{j-1}$ using LoM property of exponentials.

Lecture 29 (xv)

General case BP with birth & death (called a B&D chain)

$$q_{i,i+1} = \lambda i$$

$$q_{i,i-1} = \mu i$$

Think: particles split into two at rate λ, however die at rate μ.

Lecture 29 (xvi)

This case is much more complicated.

Theorem. $Z_t = \#$ particles in general BP, birth rate λ, death rate μ. Suppose $\lambda > \mu$. Then Z_t has "generalized geometric distribution".

$$p_0 = \alpha, \quad p_n = (1-\alpha)(1-\rho)\rho^{n-1} \quad (n \geq 1)$$

Lecture 29 (xvii)

Where

$$\alpha = \frac{\mu e^{(\lambda-\mu)t} - \mu}{\lambda e^{(\lambda-\mu)t} - \mu}$$

$$\rho = \frac{\lambda e^{(\lambda-\mu)t} - \lambda}{\mu e^{(\lambda-\mu)t} - \mu}$$

See Durrett (p 159-162) if interested.

Lecture 30

[D] §4.3 – Limiting behavior of continuous time MC's.

Good News: Much of the theory for discrete time MC's carries over. Also due to Exp holding times, we do not need to worry about periodicity issues.

Lecture 30 (ii)

In continuous time, we extend the notion of irreducibility as follows:

Def: (Y_t) is irr if for any i, j there are $k_0 = i, k_1, ..., k_n = j$ with all $q_{k_{m-1},k_m} > 0$. In other words it is possible to get from i to j in a finite # of jumps.

Lecture 30 (iii)

This implies $P_{t(i,j)} > 0$ for some $t > 0$. (See p162 in Durrett for proof.)

Stationary distributions of continuous time MC's

In discrete time, π is SD if $\pi = \pi P$. Recall that this implies $\pi = \pi P^n$ for all n.

Lecture 30 (iv)

I.e. if MC is in SD, then stays in SD forever.

$$P(X_0 = i) = \pi_i \Rightarrow P(X_n = i) = \pi_i$$

any $n \geq 1$.

In continuous time there is no first step. So we extend the notion of SD as follows:

Lecture 30 (v)

Def: π is a SD if for all $t > 0$ $\pi = \pi \underbrace{P_t}$, where $\left(\underbrace{P_t}\right)_{ij} =$

$P_{t(i,j)}$.

However, this is difficult to check since it involves **all** t.

Fortunately,

Lecture 30 (vi)

Theorem.

$$\pi \quad \mathrm{SD} \Leftrightarrow \pi = \pi \underbrace{Q}_{}$$

where

$$\left(\underbrace{Q}_{}\right)_{ij} = \left\{-\lambda_i \quad \text{if} \quad i = j; q_{ij} \quad \text{if} \quad i \neq j; \right\}$$

(This is another instance of the q_{ij} playing role of p_{ij} in continuous time.)

Lecture 30 (vii)

"Proof."

Think of π_i as proportion of sand at i in equilibrium.

$$\pi \underbrace{Q} = 0 \Leftrightarrow \underbrace{\sum_{k \neq j} \pi_k q_{kj}}_{\text{Rate at which sand arrives at j}} = \underbrace{\pi_j \lambda_j}_{\text{Rate at which sand leaves j}}$$

Equilibrium if Rate in = Rate out.

Lecture 30 (viii)

Proof. See p163.

It can be shown, as for discrete time MC's:

Theorem. If (X_t) is irr & has SD π, then

$$\lim_{t\to\infty} P_{t(i,j)} = \pi_j$$

Lecture 30 (ix)

An important special case where easier to find π than $0 = \pi Q$ is

Detailed balance: If $\pi_i q_{ij} = \pi_j q_{ji} \quad \forall i \neq j$ Then $0 = \pi \underbrace{Q}$.

$\Rightarrow \pi$ SD

Lecture 30 (x)

Thinking of sand again, DB implies flow of sand between any two states is balanced.

$$\underbrace{\pi_i q_{ij}}_{\text{Rate of sand moved } i\to j} = \underbrace{\pi_j q_{ji}}_{\text{Rate of sand moved } j\to i}$$

Lecture 30 (xi)

Not all MC's have DB, but many important ones do, e.g. B&D chain

$$q_{i,i+1} = \lambda_i$$

$$q_{i,i-1} = \mu_i$$

Jump Rates: A process on $\{0, ..., N\}$ with jumps only to adjacent states. The rate of jumping $i \to i+1$ is λ_i and $i \to i-1$ is μ_i.

$$\#(\text{jumps} \quad i \to i+1) = \#(\text{jumps} \quad i+1 \to i) + -1$$

For all times t.

Lecture 30 (xii)

Using DB can show for B&D on $\{0, 1, \ldots, N\}$,

$$\pi_i = \pi_0 \frac{\lambda_{i-1}\lambda_{i-2}\ldots\lambda_0}{\mu_i\mu_{i-1}\ldots\mu_1}$$

[Can find π_0 using $\sum \pi_i = 1$.]

Lecture 30 (xiii)

Proof that DB $\Rightarrow$ SD:

Suppose $\pi_i q_{ij} = \pi_j q_{ji} \quad \forall i \neq j.$

Then $\sum_{i \neq j} \pi_i q_{ij} = \pi_j \sum_{i \neq j} q_{ji} = \pi_j \lambda_j$

$$\Rightarrow \left(\pi \underbrace{Q}\right)_j = \sum_{i \neq j} \pi_i q_{ij} - \pi_j \lambda_j = 0$$

$$\Rightarrow \pi \underbrace{Q} = 0 \Rightarrow \pi \quad \text{SD}.$$

$\square$

Lecture 30 (xiv)

There are many worked examples in §4.2 for you to look at where you can find SD by either (1) $\pi Q = 0$ or (2) DB.

Lecture 30 (xv)

Recall:

E.g. 2-State MC.

$$q_{12} = \lambda$$

$$q_{21} = \mu$$

$$Q = \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix}$$

$$(\pi_1, \pi_2)Q = (0,0) \Rightarrow \pi = \left(\frac{\mu}{\lambda + \mu}, \frac{\lambda}{\lambda + \mu} \right)$$

Recall: We found $P_{t(1,1)}$ & $P_{t(2,1)} \to \frac{\mu}{\lambda+\mu}$ as $t \to \infty$.

Lecture 30 (xvi)

E.g. M/M/∞ queue.

$$q_{i,i+1} = \lambda$$

$$q_{i,i-1} = i\mu$$

DB $\Rightarrow \pi_i = \pi_0 \frac{\left(\frac{\lambda}{\mu}\right)^i}{i!}$

Normalizing:

$$\pi_i = e^{-\frac{\lambda}{\mu}} \frac{\left(\frac{\lambda}{\mu}\right)^i}{i!} = P\left(\mathrm{Poi}\left(\frac{\lambda}{\mu}\right) = i\right).$$

Lecture 30 (xvii)

E.g. LA weather chain.

1 = Sunny 2 = Smoggy 3 = Rainy

- Sunny for $\mathrm{Exp}\left(\frac{1}{3}\right)$ amount of day then smoggy.
- Smoggy for $\mathrm{Exp}\left(\frac{1}{4}\right)$ days until rain.
- Rain for $\mathrm{Exp}(1)$ days then sun.

Lecture 30 (xviii)

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

$$\pi Q = 0 \Rightarrow \pi = \frac{1}{8}(3, 4, 1)$$

$\therefore$ LR prop. of day that are sunny is $\frac{3}{8}$.

Lecture 30 (xix)

Alternatively, consider the discrete skeleton chain

$$P = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \text{ SD} = \frac{1}{3}(1, 1, 1)$$

When in state 1, 2, 3 held on average 3, 4, 1 days. So LR prop. of days in 1, 2, 3 should be $\frac{\left(\frac{3}{3}, \frac{4}{3}, \frac{1}{3}\right)}{\frac{3+4+1}{3}} = \pi$.

Lecture 31

[D] §4.4 — Exit distributions & exit times

We can extend the results of §1.9–1.10 to the continuous setting.

The key is to consider the discrete skeleton MC with transition probs.

$$r_{ij} = \frac{q_{ij}}{\lambda_i}$$

.

Lecture 31 (ii)

Notation

$$V_a = \text{Time of 1st visit to a}$$

$$= \min\{t \geq 0 : X_t = a\}$$

$$T_a = \text{Time of 1st return to a}$$

$$= \min\{t \geq 0 : X_t = a \;\;\&\;\; X_s \neq a \;\;\text{for some}\;\; s < t\}$$

Lecture 31 (iii)

Exit distributions:

$$h(i) = P_{i(V_A < V_B)}$$

$$= P(\text{Hit A before B} \mid X_0 = i)$$

Note $V_A < V_B$ in $X_t \Leftrightarrow V_A < V_B$ in discrete skeleton. So by FSA

$$h(i) = \sum_{j \neq i} h(j) r_{ij} \quad i \text{ notin } A \cup B.$$

Lecture 31 (iv)

$$h(i) = \sum_{j \neq i} h(j) r_{ij} \quad i \text{ notin } A \cup B.$$

Multiplying both sides by λ_i, we find

$$\sum_j Q_{ij} h(j) = 0. \quad i \text{ notin } A \cup B$$

where $Q_{ij} = \begin{cases} -\lambda_i \\ \quad\quad i=j \\ q_{ij} \\ \quad\quad i \neq j \end{cases}$

Lecture 31 (v)

Eg Branching process

$$q_{i,i+1} = \lambda i$$

$$q_{i,i-1} = \mu i$$

State space =

$$\{0, 1, 2, ...\}$$

0 is absorbing. Suppose $X_0 = 1$.

If $\mu \geq \lambda$ extinction (i.e. hit 0) occurs w.p. 1.

For $\lambda < \mu$, $\rho = P(\text{extinction}) =$?

Lecture 31 (vi)

$$\rho = \overbrace{\frac{\mu}{\lambda+\mu}}^{\text{Initial particle dies before giving branching} \to 0} + \overbrace{\frac{\lambda}{\lambda+\mu}\rho^2}^{\text{Initial particle branches} \to \text{These 2 family trees evolve independently. Both eventually die out w.p. } \rho.}$$

Solving:

$$\rho = \frac{\mu}{\lambda}$$

.

Lecture 31 (vii)

Exit times:

$$g(i) = E_i V_A$$

$$= E(\text{time of 1st visit to A} \mid X_0 = i)$$

If $i \in A$, then $g(i) = 0$, since in A at time 0. Otherwise, mean time held in i is $\frac{1}{\lambda_i}$ before 1st jump. So by FSA

Lecture 31 (viii)

$$g(i) = \frac{1}{\lambda_i} + \sum_{j \neq i} r_{ij} g(j). \quad i \text{ notin } A$$

Again, multiplying by λ_i, we find

$$\sum_j Q_{ij} g(j) = -1 \quad i \text{ notin } A$$

Lecture 31 (ix)

[D] §4.5 - Markovian queues.

Eg M/M/1 Queue.

$$\rightarrow$$

Customer inter-arrivals IID $\text{Exp}(\lambda)$

$$\rightarrow$$

Service times IID $\text{Exp}(\mu)$

$$\rightarrow$$

Only 1 server.

$$q_{i,i+1} = \lambda$$

Lecture 31 (x)

$$q_{i,i-1} = \mu$$

Simple B&D MC.

Lecture 31 (xi)

By detailed balance:

$$\pi_n = \left(\frac{\lambda}{\mu}\right)^n \pi_0$$

$$\sum_{n=1}^{\infty} \left(\frac{\lambda}{\mu}\right)^n = \frac{1}{1 - \frac{\lambda}{\mu}}$$

$$\implies \pi_n = \left(1 - \frac{\lambda}{\mu}\right)\left(\frac{\lambda}{\mu}\right)^n, \quad n \geq 0$$

(Recall in §3, we found $\pi_0 = 1 - \frac{\lambda}{\mu}$.)

Lecture 31 (xii)

By Little's formula ($L = \lambda W$),

Lecture 31 (xiii)

Eg M/M/1 queue with a waiting room

Same as above, but system has capacity N. If N people in system (1 being served, N-1 in queue) then if a new customer arrives they cannot enter system, so they leave never to return.

Lecture 31 (xiv)

Theorem. Suppose (X_t) satisfies DB with SD π. Let $\left(X_t^A\right)$ be the same MC, but not allowed to leave A. I.e.

$$q_{ij}^A = \begin{cases} q_{ij} & \\ & i \\ j \in A & \\ 0 & \\ & \text{o/w.} \end{cases}$$

Then

$$\pi_i^A = \frac{\pi_i}{\sum_{j \in A} \pi_j}$$

Lecture 31 (xv)

It turns out, for this MC

$$\pi_n = \frac{\left(1 - \frac{\lambda}{\mu}\right)\left(\frac{\lambda}{\mu}\right)^n}{1 - \left(\frac{\lambda}{\mu}\right)^{N+1}}.$$

This follows by previous example + the following useful observation:

Lecture 31 (xvi)

Proof.

$$(X_t) \text{ DB} \Longrightarrow \pi_i q_{ij} = \pi_j q_{ji}$$

$$\therefore$$

Same holds for π^A:

If i or j notin A then $0 = 0$.

o/w

$$\pi_i^A q_{ij} = \pi_j^A q_{ij}$$

.

Lecture 31 (xvii)

$$\square$$

Lecture 31 (xviii)

Theorem. Suppose (X_t) satisfies DB with SD π. Let $\left(X_t^A\right)$ be the same MC, but not allowed to leave A. I.e.

$$q_{ij}^A = \begin{cases} q_{ij} & \\ & i \\ j \in A & \\ 0 & \\ & \text{o/w.} \end{cases}$$

Then

$$\pi_i^A = \frac{\pi_i}{\sum_{j \in A} \pi_j}$$

Lecture 31 (xix)

Proof.

$$(X_t) \text{ DB} \Longrightarrow \underline{\pi_i q_{ij} = \pi_j q_{ji}}$$

$$\therefore$$

Same holds for π^A:

If i or j notin A then $0 = 0$.

o/w

$$\pi_i^A q_{ij} = \pi_j^A q_{ij}$$

.

Lecture 31 (xx)

□

Lecture 31 (xxi)

Eg Barber cuts hair rate 3 Customers arrive rate 2. Only 2 waiting chairs. $N = 3$ in previous example, so

$$\pi = \frac{1}{65}(27, 18, 12, 8).$$

$$L = \text{LR ave. \# people} = \frac{1 * 18 + 2 * 12 + 3 * 8}{65}$$

in system

$$= \frac{66}{65}.$$

Lecture 31 (xxii)

It turns out, for this MC

$$\pi_n = \frac{\left(1 - \frac{\lambda}{\mu}\right)\left(\frac{\lambda}{\mu}\right)^n}{1 - \left(\frac{\lambda}{\mu}\right)^{N+1}}.$$

This follows by previous example + the following useful observation:

Lecture 31 (xxiii)

Eg Barber cuts hair rate $3 = \mu$ Customers arrive rate $2 = \lambda$
Only 2 waiting chairs. $N = 3$ in previous example, so

$$\pi = \frac{1}{65}(27, 18, 12, 8).$$

$$L = \text{LR ave. \# people} = \frac{1 \cdot 18 + 2 \cdot 12 + 3 \cdot 8}{65}$$

in system

$$= \frac{66}{65}.$$

Lecture 31 (xxiv)

Little's formula :

$$L = \lambda_a W$$

where λ_a = LR ave. rate customer join system.

Here $\lambda_a = 2(1 - \pi_3) = \frac{114}{65}$

$$\therefore W = \text{LR ave time customers}$$

$$\text{spend at barber shop}$$

$$= \frac{L}{\lambda_a} = \frac{66}{114} \approx .58 \quad \text{hours.}$$

Lecture 31 (xxv)

Eg M/M/s multiple servers.

For instance: Bank with s tellers.

$$q_{i,i+1} = \lambda$$

$$q_{i,i-1} = \mu \min\{i, s\}.$$

If $\lambda < s\mu$ there is a SD π. Use DB, but a little messy, see p181.

Lecture 32

Martingales capture the notion of a **fair game**:

Suppose a series of outcomes $X_0, X_1, \ldots$ are bet on. Let $M_n = \$$ won after n bets.

(M_n) is a martingale (MG) with respect to (X_n) if

Lecture 32 (ii)

1. All $E(|M_n|) < \infty$ (usually not an issue.)
2. M_n can be determined by $X_n, ..., X_0, M_0$ (i.e. \$ won after n bets depends on initial fortune M_0 & first n bets only.)
3. $E\left(M_{n+1} - M_n \mid X_n = x_n, ..., X_0 = x_0, M_0 = m_0\right) = 0$

for any $x_n, ..., x_0, m_0$.

(i.e. neither gambler nor the house has an advantage)

Lecture 32 (iii)

Often in our examples we will have (X_n) a MC & $M_n = f(X_n, n)$, i.e. some (deterministic) function of X_n and n.

Lecture 32 (iv)

Martingales are extremely useful tools in probability.

For example, we can use them to get much easier proofs about exit distributions / exit times.

eg: Gambler's Ruin: Instead of solving tricky recurrence relations (HW#1), MG's provide a slick proof.

Lecture 32 (v)

To study MG's we'll need to use info about **conditional expectation** from Stat 134.

Please review your Stat 134 notes and/or read §5.1 Soj p201-204.

Lecture 32 (vi)

Shorthand notation to make life easier.

For a vector $v = (x_n, ..., x_0, m_0)$, Let $A_v = \{X_n = x_n, ..., X_0 = x_0, M_0 = m_0\}$.

Property (3) in definition of MG says $E(M_{n+1} - M_n \mid A_v) = 0$ for all v.

Lecture 32 (vii)

Eg SRW. X_i IID. $\mu = EX$.

$$S_n = S_0 + \sum_{i=1}^{n} X_i$$

.

$$M_n = S_n - n\mu$$

MG wrt X_n

$$E(M_{n+1} - M_n \mid A_v) = E(X_{n+1} - \mu \mid A_v)$$

$$= E(X_{n+1}) - \mu = 0$$

Lecture 32 (viii)

Eg X_i IID, $EX = 0$, $\sigma^2 = \text{Var } X$.

$$M_n = S_n^2 - n\sigma^2$$

MG wrt X_n

$$M_{n+1} - M_n = \left(S_n + X_{n+1}\right)^2 - S_n^2 - \sigma^2$$

$$= 2S_n X_{n+1} + X_{n+1}^2 - \sigma^2$$

$$\therefore E\left[M_{n+1} - M_n \mid A_v\right]$$

$$= 2S_n \underbrace{E\left(X_{n+1}\right)}_{=0} + \underbrace{E\left(X_{n+1}^2\right)}_{=\text{sigma\^{}2}} - \sigma^2$$

Lecture 32 (ix)

$$= 0$$

✓

Lecture 32 (x)

Eg Stock evolving in time:

X_i IID, $EX = 1$.

$$M_n = \underbrace{X_n X_{n-1} \dots X_1}_{\text{fluctuations in value}} \underbrace{M_0}_{\text{initial value}}$$

where M_n is the value at time n.

$$E(M_{n+1} - M_n \mid A_v) = E[M_n(X_{n+1} - 1) \mid A_v]$$

$$= M_n E(X_{n+1} - 1) = 0$$

✓

Lecture 32 (xi)

Eg Exponential MG.

Y_i IID, $\varphi(\theta) = Ee^{\theta Y} < \infty$.

$$S_n = S_0 + \sum_{i=1}^{n} Y_i$$

.

Then $M_n = \frac{e^{\theta S_n}}{\varphi(\theta)^n}$ is a MG wrt (Y_n).

Proof: Use last example with $X_i = \frac{e^{\theta Y_i}}{\varphi(\theta)}$. $\square$

Lecture 32 (xii)

Eg Gambler's Ruin.

Y_i IID, $P(X=1)=p$, $P(X=-1)=1-p$.

Suppose $p \neq \frac{1}{2}$. Then

$$M_n = \left(\frac{q}{p}\right)^{S_n}$$

is a MG, where $q = 1-p$ & $S_n = S_0 + \sum_i^n X_i$.

Proof. Select θ such that $e^\theta = \frac{q}{p}$. Then $\varphi(\theta) = p \cdot \frac{q}{p} + q \cdot \frac{p}{q} = 1$. So by previous example $e^{\theta S_n} = \left(\frac{q}{p}\right)^{S_n}$ is a mg. $\square$

Lecture 33

Recall: (M_n) a mg wrt (X_n) if

$$E(M_{n+1} - M_n \mid A_n) = 0$$

where A_n is defined by $X_n = x_n, ..., X_0 = x_0, M_0 = m_0$.

Eg Exponential mg.

$$Y_i$$

: IID, $\varphi(\theta) = Ee^{\theta Y} < \infty$.

$$S_n = S_0 + \sum_{i=1}^{n} Y_i$$

Lecture 33 (ii)

. Then $M_n = \frac{e^{\theta S_n}}{(\varphi(\theta))^n}$ is a mg.

Lecture 33 (iii)

Eg Gambler's ruin.

$$Y_i$$

: IID $P(X=-1)=p$, $P(X=-1)=1-p$.

Suppose $p \neq \frac{1}{2}$. Then

$$M_n = \left(\frac{q}{p}\right)^{S_n}$$

is a mg, where $q = 1-p$ & $S_n = S_0 + \sum_i^n X_i$.

Proof. Select θ such that $e^\theta = \frac{q}{p}$. Then $\varphi(\theta) = p \cdot \frac{q}{p} + q \cdot \frac{p}{q} = 1$. So by previous example $e^{\theta S_n} = \left(\frac{q}{p}\right)^{S_n}$ is a mg.

Lecture 33 (iv)

In many applications, the game / process is **not** fair.

Super-mg if

$$E(M_{n+1} - M_n \mid A_n) \le 0$$

$\Rightarrow$ House has advantage

Sub-mg if

$$E(M_{n+1} - M_n \mid A_n) \ge 0$$

$\Rightarrow$ Gambler has advantage.

Lecture 33 (v)

Theorem. (X_n) a MC, $(P)_{ij} = p_{ij}$. Suppose

$$f(i,n) \geq \sum_j p_{ij} f(j,n+1):$$

Then $M_n = f(X_n, n)$ super mg.

Note: M_n super mg $\Leftrightarrow -M_n$ sub mg. So similar result holds for sub mgs but with "$\leq$" instead of "$\geq$".

Lecture 33 (vi)

In many applications, the game / process is **not** fair.

Super-mg if

$$E(M_{n+1} - M_n \mid A_n) \leq 0$$

$\Rightarrow$ House has advantage

Sub-mg if

$$E(M_{n+1} - M_n \mid A_n) \geq 0$$

$\Rightarrow$ Gambler has advantage.

Lecture 33 (vii)

Proof.

$$E\big[f(X_{n+1}, n+1) \mid A_n\big]$$

$\to$ FSA

$$= \sum_y P_{X_n,y} f(y, n+1)$$

$$\leq f(X_n, n)$$

$$\therefore E\big[M_{n+1} - M_n \mid A_n\big] \leq 0.$$

So (M_n) a super mg.

Lecture 33 (viii)

[D] §5.3 — Stopping times & gambling strategies.

In this section, we'll prove the most famous result about mg's "You cannot beat an unfolable game"

(Theorem 5.9 [D]) — Bad news for gamblers

Lecture 33 (ix)

First observation: If game is unfavorable (a super mg) then fortune will tend to $\downarrow$ in time.

Theorem. (M_n) a supermg. Then

$$EM_n \le EM_m$$

for any $n \ge m$.

$\Rightarrow$ (M_n) a mg. Then all $EM_n = EM_0$.

Proof. Try on your own. or p 207. ■

Lecture 33 (x)

To lead up to our main result, we first consider the following betting strategy (popular in 18th century France) where **martingales** got their name.

Lecture 33 (xi)

Series of iid coin flips. Bet on H. If win double your wager. Otherwise lose your wager. After every loss, double your wager, so that when you eventually win, you recover your losses + win a profit equal to size of original bet.

Eg

$$
\begin{array}{ccccc}
X & X & X & \checkmark & \\
1 & 2 & 4 & 8 & \to = 15
\end{array}
$$

spent

Problem: You might need a lot of $$$ to continue along while you wait for a win.

Lecture 33 (xii)

Eg Buy lottery tickets. Every time you lose, buy twice as many next time. Would anyone want to do this?

At the stopping time $T = 1^{\text{st}}$ win we are sure to get rich — but the process is still unfavorable (a super mg) $E(M_{n+1} \mid A_n) \leq M_n$ for all (deterministic) times n. Can show directly (see [D]) — also special case of:

Lecture 33 (xiii)

Theorem. (M_n) a super mg wrt (X_n). H_n "predictable" (depends only on info available **before** time n: $X_{n-1}, \ldots, X_0, M_0$) & bounded $0 \le H_n \le C_n$. Then:

$$\underbrace{W_n}_{\text{th after n bets}} = \underbrace{W_0}_{\text{initial wealth}} + \sum_{m=1}^{n} \underbrace{H_m}_{\text{our bet on value of } M_m - M_{m-1}} (M_m - M_{m}$$

is a super mg.

Lecture 34

Finally, we'll prove the most famous result in MG theory:

"You can't beat an unfavorable game."

Lecture 34 (ii)

Theorem Let (M_n) be a super-martingale with respect to (X_n). Let H_n be "predictable" (depends only on info available before time n: $X_{n-1}, ..., X_0, M_0$) & bounded, $0 \leq H_n \leq C_n$.

Then:

$$W_n = W_0 + \sum_{m=1}^{n} H_m(M_m - M_{m-1})$$

where

- W_n is the wealth after n bets
- W_0 is the initial wealth

Lecture 34 (iii)

- H_m is our bet on the value of $M_m - M_{m-1}$

is a **super-martingale**.

Lecture 34 (iv)

i.e. No way to bet on an unfavorable game to make it favorable if you:

Cannot know the future ($\boldsymbol{H_n}$ is predictable) Have a bounded amount of money to wager: $0 \leq H_n \leq C_n$

Lecture 34 (v)

To see that the previous “doubling strategy” is a super-mg, let $M_n = \sum_{i=1}^{n} X_i$ be the wealth of a gambler that places \$1 on all bets.

The wealth of a gambler that follows a doubling strategy selects $H_n = 2H_{n-1}$ if in the last step, other gamblers...

Lecture 34 (vi)

Proof.

$$W_{n+1} - W_n = H_{n+1}(M_{n+1} - M_n)$$

So,

$$\begin{aligned}
&E[W_{n+1} - W_n \mid A_n] \\
&= E[H_{n+1}(M_{n+1} - M_n) \mid A_n] \\
&= H_{n+1}E[M_{n+1} - M_n \mid A_n] \qquad \text{\$H_(n+1)\$ is constant on \$A_n\$} \\
&\leq C_{n+1} \cdot 0 \qquad \text{since \$(M_n)\$ is a super-mg} \\
&= 0
\end{aligned}$$

Lecture 34 (vii)

Theorem Let $(\ast\, M_n \,\ast)$ be a **super-martingale** wrt (X_n). Let $\ast\, H_n \,\ast$ be "predictable" (depends only on info available before time n: $X_{n-1}, \ldots, X_0, M_0$) & bounded, $0 \le \ast\, H_n \,\ast \le C_n$. (Constant depending on n).

Then: $\ast\, W_n \,\ast = \ast\, W_0 \,\ast + \sum_{m=1}^{n} \ast\, H_m(M_m - M_{m-1}) \,\ast$ where

- W_n is the wealth after n bets
- W_0 is the initial wealth
- H_m is our bet on the value of $M_m - M_{m-1}$

is a **super-martingale**.

Lecture 34 (viii)

Proof.

$$W_{n+1} - W_n = H_{n+1}(M_{n+1} - M_n)$$

So

$$
\begin{aligned}
&E[\texttt{underline}[W_{n+1} - W_n] \mid * A_n *] \\
&= E[H_{n+1}(M_{n+1} - M_n) \mid A_n] \\
&= H_{n+1} E[M_{n+1} - M_n \mid A_n] \qquad \text{\$H\_(n+1)\$ is constant on \$A\_n\$} \\
&\leq C_{n+1} \cdot 0 \qquad \text{since \$(M\_n)\$ is a super-mg} \\
&= 0
\end{aligned}
$$

Lecture 34 (ix)

This theorem has a very important consequence.

Def: T is a stopping time wrt (X_n) if whether $\{T = n\}$ has occurred can be determined from the info by time n: $n, X_n, ..., X_0, M_0$.

We'll see many applications of the following theorem in §5.4.

Lecture 34 (x)

Theorem Let (M_n) be a super-martingale wrt (X_n). Let T be a stopping time (ST).

Then the stopped process is a super-martingale wrt (X_n).

In particular, $E[M_{n\wedge T}] \leq E[M_0]$.

If (M_n) is a martingale, $E[M_{n\wedge T}] = E[M_0]$.

Lecture 34 (xi)

Proof.

Follows from the previous theorem by taking $H_n = 1$ up to time T, and $H_n = 0$ after the stopping time T. Then,

$$W_n = M_{n \wedge T}$$

is still a super-mg.

Lecture 34 (xii)

Warning

As we saw for the doubling strategy, the same is not necessarily true for M_T.

i.e. It is possible for M_n to be a martingale and T a stopping time, but $EM_T \neq EM_0$.

This is a very common mistake!

Lecture 34 (xiii)

"Bad mg" where $EM_T \neq EM_0$

Eg: X_i IID $= +-1$ with probability $\frac{1}{2}$.

$$S_n = S_0 + \sum_{i=1}^{n} X_i$$

Suppose $S_0 = 1$. $V_0 =$ time of 1st visit to 0. $T = V_0$ is a ST. Since (S_n) is recurrent, $P_1(T < \infty) = 1$. $\therefore S_T = 0 \Rightarrow ES_T = 0 \neq 1 = ES_0$.

Lecture 34 (xiv)

The main problem is that S_n can take very large values before finally visiting 0. In fact,

$$E\left(\max_{0\leq m\leq T} S_m\right) = \sum_{i=1}^{\infty} P_1(V_m < V_0)$$

$$= \sum_{m=1}^{\infty} \frac{1}{m} = \infty$$

Lecture 34 (xv)

On the other hand, if M_n cannot get too large before T, then $EM_T = EM_0$:

Optional Stopping Theorem (OST)

(M_n) a martingale

T a stopping time

$P(T < \infty) = 1$, and for some $k < \infty$, $|M_{T \wedge n}| \leq k$, then

$$EM_T = EM_0$$

.

(There are other versions of the OST.)

Lecture 34 (xvi)

Proof.

Since T is a ST, $EM_{T\wedge n} = EM_0$.

$$EM_{T\wedge n} = E\big(M_T \mathbb{1}_{T\leq n}\big) + E\big(M_n \mathbb{1}_{T>n}\big)$$

As $n \to \infty$, $E\big(M_n \mathbb{1}_{T>n}\big) \leq kP(T > n) \to 0$.

& $E\big(M_T \mathbb{1}_{T\leq n}\big) \to EM_T$.

$\therefore$ Taking $n \to \infty$, we find

$$EM_T = EM_0$$

.

Lecture 35

Final chapter on MGs

[D] § 5.4 — Applications of MGs

We apply previous results about MGs to get quicker solutions to questions about

- Exit distributions
- Exit times
- Extinction probabilities.

Lecture 35 (ii)

eg Gambler's ruin, $p = \frac{1}{2}$.

$$S_n = S_0 + \sum_{i=1}^{n} X_i$$

is a MG.

$$T = \min\{n \geq 0 : S_n = a \quad \text{or} \quad b\}.$$

$$a \leq S_{n \wedge T} \leq b$$

is bounded. Saw REC so $P_{x(T<\infty)} = 1$. $\therefore$ By OST,

$$x = E_x S_T = aP_{x(S_T=a)} + bP_{x(S_T=b)} \Longrightarrow$$

Lecture 35 (iii)

$$x = aP_{x(S_T=a)} + b\left(1 - P_{x(S_T=a)}\right)$$

$$\implies P_{x(S_T=a)} = \frac{b-x}{b-a}$$

.

When $a = 0$, $b = N$,

$$P_{x(\text{Jackpot})} = \frac{x}{N}$$

.

So much easier than by recurrence relations!

Lecture 35 (iv)

eg Asymmetric case, $p \neq \frac{1}{2}$.

$$P(X_i = +-1) = \begin{cases} p \\ q = 1 - p \end{cases}$$

$$M_n = \left(\frac{q}{p}\right)^{S_n}$$

is a MG.

$$M_{n \wedge T}$$

is bounded, and since SRW transient $P_{x(T<\infty)} = 1$.

$$\Rightarrow$$

Lecture 35 (v)

by OST,

$$\left(\frac{q}{p}\right)^x = \left(\frac{q}{p}\right)^a P_{x(S_T=a)} + \left(\frac{q}{p}\right)^b P_{x(S_T=b)}$$

Lecture 35 (vi)

$$\implies \underline{P_{x(S_T=a)}} = \frac{\left(\frac{q}{p}\right)^b - \left(\frac{q}{p}\right)^x}{\left(\frac{q}{p}\right)^b - \left(\frac{q}{p}\right)^a}$$

For $a = 0, b = N$,

$$P_{x(\text{Jackpot})} = \frac{1 - \left(\frac{q}{p}\right)^x}{1 - \left(\frac{q}{p}\right)^N}$$

.

Recall!

Lecture 35 (vii)

eg B&D MC. Birth & Death

In order to find a MG, recall:

$$(X_n)$$

MC with p_{ij}'s. If

$$f(x) = \sum_y p_{xy} f(y) = E_x f(X_1)$$

then $f(X_n)$ is a MG.

Lecture 35 (viii)

We need to find f so that

$$f(i) = q_i f(i-1) + p_i f(i+1) + (1 - q_i - p_i) f(i)$$

$$\implies (q_i + p_i) f(i) = q_i f(i-1) + p_i f(i+1)$$

$$\implies f(i+1) - f(i) = \frac{q_i}{p_i}(f(i) - f(i-1))$$

$$\implies f(i) = \sum_{j=1}^{i} \prod_{k=1}^{j-1} \frac{q_k}{p_k}$$

Lecture 35 (ix)

$$f(X_n)$$

is a MG.

Then, using OST,

$$P_{x(V_a<V_b)} = \frac{f(x) - f(a)}{f(b) - f(a)}.$$

Lecture 35 (x)

Similarly, MGs useful for calculating mean exit times.

eg Gambler's ruin, $p = \frac{1}{2}$.

$$T =$$

time game over.

By translation, we can assume

$$a < 0$$

$$b > 0$$

We'll show $E_0 T = -ab$.

Lecture 35 (xi)

This implies if $a < b \in Z$ then $E_x T = (x - a)(b - x)$.

$$M_n = S_n^2 - n$$

is a MG. However,

$$M_{n \wedge T} = S_{n \wedge T}^2 - (n \wedge T)$$

is not bounded. So can not

Lecture 35 (xii)

apply OST. But, we always have $EM_{n\wedge T} = EM_0 = 0$. We'll use this (just have to work a bit harder).

$$0 = E_0\big(S^2_{T\wedge n} - (T \wedge n)\big)$$

$$P_0(S_T = a, T \leq n) + b^2 P_0(S_T = b, T \leq n) + E_0\big(S_n^2 1_{T>n}\big) - E_0(T$$

Lecture 35 (xiii)

$$S_n^2$$

is bounded & $P(T < \infty) = 1$, so $E_0\big(S_n^2 1_{T>n}\big) \to 0$ as $n \to \infty$. Also $E_0(T \wedge n) \to E_0(T)$ as $n \to \infty$. $\therefore$ Taking $n \to \infty$,

$$E_0 T = a^2 \underbrace{P_0(S_T = a)}_{=\frac{b}{b-a}} + b^2 \underbrace{P_0(S_T = b)}_{=-\frac{a}{b-a}}$$

$$= \frac{a^2 b - b^2 a}{b-a} = -ab$$

Lecture 35 (xiv)

Wald's Equation

$$S_n = S_0 + \sum_1^n X_i$$

$$S_0 \in R$$

, not random.

$$T$$

a ST,

$$ET < \infty$$

.

Lecture 35 (xv)

Then,

$$E(S_T - S_0) = \mu ET$$

Note: This doesn't just follow by LOE, because T & (X_i) **not** indep.

Lecture 35 (xvi)

Informal proof:

$$M_n = S_n - n\mu$$

is a MG OST:

$$EM_T = EM_0$$

$$\implies S_0 = ES_T - \mu ET$$

$$\implies E(S_T - S_0) = \mu ET$$

However M_n not bounded, so OST does not apply.

Lecture 35 (xvii)

But this can be made rigorous by using

$$EM_{T \wedge n} = EM_0$$

Always holds instead, as we did in previous example.

See p 215 for details.

Lecture 35 (xviii)

Applying this to Gambler's Ruin, $p \neq \frac{1}{2}$.

Suppose $p < \frac{1}{2}$. $V_0 =$ Time of 1st visit to 0. Recall $M_n =$

$S_n - \overbrace{(p-q)}^{\mu} n$ is a MG.

$$\therefore x = E_{x\left(S_{V_0 \wedge n} - (p-q)(n \wedge V_0)\right)}$$

$$\geq (q-p) E_{x(n \wedge V_0)}$$

Lecture 35 (xix)

Taking $n \to \infty$,

$$E_x V_0 \leq \frac{x}{q-p} < \infty$$

.

$$\therefore$$

By Wald's Equation

$$(\to E(S_T - S_0) = \mu E T)$$

$$E_{x\left(S_{V_0} - x\right)} = (p-q) E_x V_0$$

Lecture 35 (xx)

$$\implies E_x V_0 = \frac{x}{q-p}$$

.

Lecture 36

Extinction probabilities using M.G.'s

Def: X_i iid $\mathbb{Z}$-valued, $\mu = EX \geq 0$, $P(X \geq -1) = 1$, & $P(X = -1) > 0$. Then $S_n = S_0 + \sum_{i=1}^{n} X_i$ is called a **left-continuous walk**.

Can make large jumps to $\rightarrow$. But can ever jump 1 to $\leftarrow$.

Lecture 36 (ii)

Theorem.

$$P_{x(V_0<\infty)} = e^{\alpha x_0}$$

where $\alpha < 0$ solves $\varphi(\alpha) = 1$ with $\varphi(\theta) = Ee^{\theta X}$.

Proof. First note that α is well-defined: $\varphi(0) = 1$ and $\varphi'(\theta) = E\big(Xe^{\theta X}\big)$, so $\varphi'(0) = EX \geq 0$.

Lecture 36 (iii)

Also for $\theta < 0$,

$$\varphi(\theta) \geq e^{-\theta} P(X = -1) \to \infty$$

as $\theta \to -\infty$

This choice of α makes $M_n = e^{\alpha S_n}$ a M.G.

Lecture 36 (iv)

$$E(M_{n+1} \mid \mathcal{A}_n)$$

$$= E(e^{\alpha X_{n+1}} M_n \mid \mathcal{A}_n)$$

$$= M_n E e^{\alpha X_{n+1}}$$

since $Ee^{\alpha X_{n+1}} = 1$.

M_n is not bounded, so again OST does not directly apply. We must use $EM_{\min(n,V_0)} = EM_0$ and then take $n \to \infty$.

Lecture 36 (v)

(But ignoring this, and using OST anyway,

$$e^{\alpha x} = E_x M_0 = E_x M_{V_0}$$

$$= P_{x(V_0 < \infty)}.$$

we do get the right answer.)

To make this rigorous,

Lecture 36 (vi)

$$e^{\alpha x} = E_x e^{\alpha S_{\min(V_0,n)}}$$

$$= P_{x(V_0 \leq n)} + E_{x\left(e^{\alpha S_n} 1_{V_0 > n}\right)}$$

Take $n \to \infty$: $E_{x\left(e^{\alpha S_n} 1_{V_0 > n}\right)} \to 0$ since by SLLN $S_n \approx n\mu$ & $\alpha < 0$ so $e^{\alpha S_n} \to 0$. Hence,

$$e^{\alpha x} = P_{x(V_0 < \infty)}$$

. $\blacksquare$

Lecture 36 (vii)

This result can be used to obtain a “slick” proof of the branching process theorem — using connection with left-cts walk & BP (Think: Test 1, #2).

See [D] p. 217 for details.

Lecture 36 (viii)

[PK] §2-5 — Martingale Maximal Inequality.

If (M_n) a MG, then all $EM_n = EM_0$. So by Markov's Inequality,

$$P(M_n \geq \lambda) \leq \frac{EM_0}{\lambda}.$$

Lecture 36 (ix)

However a much more powerful inequality holds:

MG maximal inequality

$$P\left(\max_{0\leq n\leq m} M_n \geq \lambda\right) \leq \frac{EM_0}{\lambda},$$

And so,

$$P\left(\max_{n\geq 0} M_n \geq \lambda\right) \leq \frac{EM_0}{\lambda}.$$

Lecture 36 (x)

Eg A gambler bets on a series of fair games (coin tosses, for example), always betting proportion $\alpha \in (0, 1)$ of their current fortune.

$$X_n =$$

fortune after n bets.

$$X_{n+1} = \begin{cases} (1+\alpha)X_n & \text{w.p.} \quad \frac{1}{2} \\ (1-\alpha)X_n & \text{w.p.} \quad \frac{1}{2} \end{cases}$$

Lecture 36 (xi)

Q Find an upper bound for prob. that they ever at some point double their money, that holds for an α.

That is,

$$P(\text{ever} \quad X_n \geq 2X_0) \leq ?$$

Lecture 36 (xii)

$$\therefore E(X_{n+1} \mid \mathcal{A}_n)$$

$$= \frac{1}{2}[(1+\alpha)X_n + (1-\alpha)X_n]$$

$$= X_n$$

$$\therefore$$

by MG max ineq.

$$P(\text{ever double})$$

$$= P\left(\max_n X_n \geq 2X_0\right) \leq \frac{1}{2}.$$

Lecture 36 (xiii)

Proof of max ineq.

$$EX_m = \sum_{n=0}^{m} E\Big(X_m 1_{X_0,\ldots,X_{n-1}<\lambda, X_n\geq\lambda}\Big)$$

$$+E\Big(X_m 1_{X_0,\ldots,X_m<\lambda}\Big)$$

(since $1 = \sum_{n=0}^{m} 1_{X_0,\ldots,X_{n-1}<\lambda, X_n\geq\lambda} + 1_{X_0,\ldots,X_m<\lambda}$)

$$\geq \sum_{n=0}^{m} E\Big(X_m 1_{X_0,\ldots,X_{n-1}<\lambda, X_n\geq\lambda}\Big)$$

Lecture 36 (xiv)

$$EX_m \geq \sum_{n=0}^{m} E\Big(X_m 1_{X_0,\ldots,X_{n-1}<\lambda, X_n \geq \lambda}\Big)$$

$$= \sum_{n=0}^{m} E\Big(X_n 1_{X_0,\ldots,X_{n-1}<\lambda, X_n \geq \lambda}\Big)$$

$$E\Big(X_m 1_{X_0,\ldots,X_{n-1}<\lambda, X_n \geq \lambda}\Big)$$

$$= E\Big[E(X_m \mid X_0, \ldots, X_n) 1_{X_0,\ldots,X_{n-1}<\lambda, X_n \geq \lambda}\Big]$$

$$= E\Big(X_n 1_{X_0,\ldots,X_{n-1}<\lambda, X_n \geq \lambda}\Big)$$

Lecture 36 (xv)

$$\geq \lambda \sum_{n=0}^{m} P(X_0, ..., X_{n-1} < \lambda, X_n \geq \lambda)$$

$$= \lambda P\left(\max_{0 \leq n \leq m} X_m \geq \lambda\right) \blacksquare$$

Lecture 37

New Topic: Brownian Motion

§8 [PK]

- A random, continuous path in time & space.
- Here in $\mathbb{R}^2$

Lecture 37 (ii)

Background

- Roman Philosopher Lucretius 60 BC observed Brownian motion, looking at dust particles dance in the light:

  Observe what happens when sunbeams are admitted into a building and shed light on its shadowy places. You will see a multitude of tiny particles mingling in a multitude of ways... their dancing is an actual indication of underlying movements of matter that are hidden from our sight... It originates with the atoms which move of themselves [i.e., spontaneously]. Then those small compound bodies that are least removed from the impetus of the atoms are set in

Lecture 37 (iii)

motion by the impact of their invisible blows and in turn cannon against slightly larger bodies. So the movement mounts up from the atoms and gradually emerges to the level of our senses so that those bodies are in motion that we see in sunbeams, moved by blows that remain invisible.

Lecture 37 (iv)

- Botanist Robert Brown 1827 observed Brownian motion through a microscope, tracking the trajectory of a pollen grain in a container of water.
- Einstein 1905 began a more mathematical study of Brownian motion. Explained that this rough/random path is caused by continual bombardment in many directions by water molecules upon the grain of pollen.

Lecture 37 (v)

- Einstein's work gave good evidence for the existence of atoms. $\rightarrow$ 1908: Perrin explores this further & wins Nobel Prize (1926)
- The mathematical study then began to flourish, beginning with e.g. Wiener, Lévy.

Lecture 37 (vi)

- Brownian motion remains an area of active research. It is also an important building block used to construct more complicated processes in use today.
- Has many applications, perhaps most notably, mathematical finance.

Lecture 37 (vii)

Plan

- We'll cover **some** of §8 in [PK].
- See also "Brownian Motion" by Peter Mörters & Yuval Peres if you want to see proofs and many more facts about Brownian Motion (418 page book!)

Lecture 37 (viii)

Brownian motion has many interesting/beautiful properties. In particular, it is a **fractal**:

Lévy ↓ Paley, Wiener & Zygmund ↓

Continuous, but **nowhere** differentiable.

"It is a path, but very rough"

(Proofs in Mörters-Peres, beyond Stat 150 level.)

Lecture 37 (ix)

- If you zoom into **anywhere** along a polynomial, it looks like a straight line (the tangent line).

Lecture 37 (x)

- On the other hand, Brownian motion is "self-similar" (fractal). No matter how close you zoom in, it looks the similar to when you started!

↓

Lecture 38

Einstein:

Let $(B_t, t \geq 0)$ be position of particle evolving in time.

Let $p_{t(x,y)}$ be PDF, the density with which particle moves from x to y in time t.

He argued, by physics, that

$$\frac{\text{del } p}{\text{del } t} = \frac{\sigma^2}{2} \frac{\text{del}^2 p}{\text{del } x^2}$$

for some "diffusion coefficient" σ^2.

Lecture 38 (ii)

This implies (see [PK] Exercise 8.1.3) that

$$p_{t(x,y)} = \frac{1}{\sqrt{2\pi t\sigma}} e^{-\frac{(y-x)^2}{2t\sigma^2}},$$

a normal density.

- If $\sigma^2 = 1$, $p_{t(x,y)} = \varphi(y - x)$,

and we call $(B_t, t \geq 0)$ **standard** Brownian motion.

Lecture 38 (iii)

Mathematical definition of Brownian motion:

Definition

$(B_t, t \geq 0)$ is BM with diffusion coeff. σ^2 started from x if

- a) All increments normal: $B_{s+t} - B_s \sim \mathrm{Normal}(0, \sigma^2 t)$
- b) Indep. increments
- c) $B_0 = x$ & $t \to B_t$ is a continuous function.

Lecture 38 (iv)

Unless otherwise stated, we'll always assume $(B_t, t \geq 0)$ is standard, $\sigma^2 = 1$.

Lecture 38 (v)

Brownian Motion & Random Walk (Donsker's Invariance Principle)

- Brownian motion is the "scaling limit" of random walk.

Let $X_1, X_2, \dots$ be iid. $EX = 0$, Var $X = 1$.

$$S_m = \sum_{i}^{n} X_i$$

: SRW.

Lecture 38 (vi)

Let B_n be the linear interpolation of S_n, scaled by $\frac{1}{\sqrt{n}}$.

$$B_{n(t)} = \frac{S_{nt}}{\sqrt{n}}$$

Lecture 38 (vii)

Theorem (Donsker)

$$\left(B_{n(t)} = \frac{S_{nt}}{\sqrt{n}}, 0 \leq t \leq 1\right)$$

$$\text{longrightarrow.b}(n \to \infty)(B_t, 0 \leq t \leq 1)$$

Lecture 38 (viii)

This is a very substantial improvement upon the CLT:

Lecture 38 (ix)

That is,

Lecture 38 (x)

Some results about BM can be anticipated due to connection with SRW.

Eg X_i iid $P(X = +-1) = \frac{1}{2}$. $a < 0 < b$.

$$P_0(\text{hit } a \ \text{ before } \ b) = \frac{b}{a+b}$$

.

Therefore:

$$P_0(B_t \ \text{ hits } \ a \ \text{ before } \ b) = \lim_{n\to\infty} \frac{b\sqrt{n}}{a\sqrt{n} + b\sqrt{n}} = \frac{b}{a+b}$$

Lecture 38 (xi)

.

Lecture 38 (xii)

§8.2: Maximum of BM & The Reflection Principle.

- Many useful properties follow just from continuity of BM + symmetry of Normal PDF.

Lecture 39

§8.2: Maximum of BM & the Reflection Principle.

- Many useful properties follow just from continuity of BM + symmetry of Normal PDF.

Lecture 39 (ii)

Theorem (Reflection Principle)

For any $x > 0$:

$$P\left(\max_{0 \leq s \leq t} B_s > x\right) = 2P(B_t > x)$$

Lecture 39 (iii)

$$\therefore P(B_t > x)$$

is only $\frac{1}{2}$ the probability that $P\left(\max_{0 \le s \le t} B_s > x\right)$.

$$\square$$

Lecture 39 (iv)

The RP has many important consequences.

Eg **Zeros of Brownian motion**

$$P(B_u = 0, \text{some} \;\; t \le u \le t + s) = \frac{2}{\pi} \arctan \sqrt{\frac{s}{t}}$$

Lecture 39 (v)

Proof.

$$T_x =$$

hitting time of x.

By RP:

$$\begin{aligned} P(T_x \leq t) &= P\left(\max_{0\leq u\leq t} B_u \geq x\right) \\ &= 2P(B_t \geq x) \\ &= \frac{2}{\sqrt{2\pi t}} \int_x^\infty e^{-\frac{y^2}{2t}} dy \end{aligned}$$

Lecture 39 (vi)

$$= \sqrt{\frac{2}{\pi}} \int_{\frac{x}{\sqrt{t}}}^{\infty} e^{-\frac{u^2}{2}} du \quad \left(u = \frac{y}{\sqrt{t}} \right)$$

$$\therefore f_{T_x}(t) = \frac{x t^{-\frac{3}{2}}}{\sqrt{2\pi}} e^{-\frac{x^2}{2t}}$$

Now, let $H_{t(z,x)} = P_{z(T_x \in t)}$.

Clearly $H_{t(0,x)} = H_{t(x,0)}$ (Symm. of Normal)

Lecture 39 (vii)

$$\therefore H_{t(0,x)}$$

$$= H_{t(x,0)}$$

$$= \int_0^t \frac{x\xi^{-\frac{3}{2}}}{\sqrt{2\pi}} e^{-\frac{x^2}{2\xi}} d\xi$$

$$P(B_u = 0, \text{some} \ \ t \leq u \leq t + s)$$

$$= 2\int_0^\infty H_{s(x,0)} \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}} dx$$

by sym.

Lecture 39 (viii)

$$= 2\int_0^\infty \left[\int_0^s \frac{x\xi^{-\frac{3}{2}}}{\sqrt{2\pi}} e^{-\frac{x^2}{2\xi}} d\xi\right] \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}} dx$$

$$= \frac{1}{\pi\sqrt{t}} \int_0^s \left(\int_0^\infty x e^{-\frac{x^2}{2\xi}-\frac{x^2}{2t}} dx\right) \xi^{-\frac{3}{2}} d\xi$$

Calc. Change of variables p. 409

Trig: Exercise 8.2.2.

$$= \frac{2}{\pi} \arccos\sqrt{\frac{t}{t+s}} = \frac{2}{\pi} \arctan\sqrt{\frac{s}{t}}$$

Lecture 39 (ix)

$$\square$$

Lecture 39 (x)

It can be shown (more advanced classes in prob. or see Mörters-Peres) that the set

$$Z = \{t : B_t = 0\}$$

is

- **Infinite** by RP
- **Uncountable**
- **No isolated points**
- **Measure zero**
- **Fractal (Hausdorff) dimension 1/2.**

Lecture 39 (xi)

Other useful extensions discussed in §8.3-8.4.

Reflected BM: $|\boldsymbol{B_t}|$ Absorbed: $B_0 = x$, stopped at 0.

$$A_t = \begin{cases} B_t \text{ if } & t \leq T_x \\ 0 \quad \text{if } & t > T_x \end{cases}$$

- With drift μ: $X_t = \mu t + \sigma B_t$

Incr. Normal($\mu s, \sigma^2 s$).

Lecture 39 (xii)

- Brownian Bridge:

$(B^0, 0 \leq t \leq 1)$ is BM conditioned on $\{B_0 = B_1 = 0\}$.

A slick way of obtaining this is:

$$B_t^0 = B_t - tB_1$$

Lecture 39 (xiii)

Application (used in e.g. Non-parametric Stats) $X_1, X_2, \ldots$ IID

Empirical CDF

$$F_{N(t)} = \frac{1}{N} \sum_{i}^{N} \xi_{i(t)}, \xi_i = 1_{\{X_i \leq t\}}$$

Use this to estimate (unknown) CDF $F(t) = P(X \leq t)$.

Lecture 39 (xiv)

Eg if X_i IID Uniform(0,1)

$$E\xi_{i(t)} = P(X \leq t) = t$$

$$X_{N(t)} = \frac{\sum_1^N \left(\xi_{i(t)} - t\right)}{\sqrt{N}}$$

$$= \sqrt{N}\left(F_{N(t)} - t\right)$$

$$\to B_t^0$$

$$\therefore F_{N(t)} = t + \frac{B_t^0}{\sqrt{N}}$$

Lecture 39 (xv)

So error is a Brownian bridge, scaled by $\sqrt{N}$.

Lecture 39 (xvi)

It is an important, but somewhat tricky, matter to show that Brownian motion exists – i.e. it is possible to construct a process $(B_t, t \geq 0)$ satisfying conditions 1-3 above.

Lecture 39 (xvii)

Paul Lévy's Construction

Dyadic Rationals

$$D_n = \left\{ \frac{k}{2^n} : 0 \leq k \leq 2^n \right\}.$$

Are dense in $[0, 1]$.

Select IID $Z_t \sim \mathrm{Normal}(0, 1)$, one for each $t \in D_n$.

Lecture 39 (xviii)

Fig. 1.2. The first three steps in the construction of Brownian motion

See Mörters-Peres p 9-12.