Lec 1

Prerequisits : combinatorics, calculus ...
Homework : digest, develop gut feeling for probability

Topic specifcs

- Combinatorics problems trickly to translate into math

How to follow class

- Go lecture
- Review notes, before next lec & before homework
- Some memorisation
- Stuck → try, little hints

Probability Space

{all possible outcome}

Probability distribution aka measure aka weight aka mass

$$(\Omega , F , P)$$

← Some mathematical description of experiment with random outcome

Collection of subsets of Ω, "events", often $F = \mathcal{P}(\Omega)$, but not always depending on various reasons
- relevance, maybe only some $F \subset \mathcal{P}(\Omega)$ is relevant
- some not admissible for technical reason

$$P : F \to [0,1]$$
$$A \mapsto P[A]$$

with properties
1. $P[\Omega] = 1$
2. A, B disjoint $\Rightarrow$ $\left.\begin{array}{l} A \cap B = \emptyset \\ \text{and } P[A \cup B] = P[A] + P[B] \end{array}\right\}$ addictivity

⋮ (out of time)

Ex Binary experiment Rain or No Rain

$\Omega = \{0, 1\}$ R N

1 0

Ex A die 1 or 2 or 3 or 4 or 5 or 6

$\Omega = \{1, \dots, 6\}$

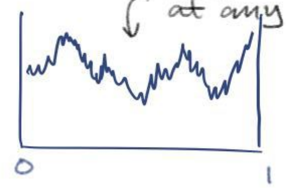
Event $A = \{2, 4, 6\} \subseteq \Omega$

If any of 2, 4, 6 occured, we say "A occured"

Ex Stock price

$\Omega = \{ f(t) \mid f : [0, 1] \to \mathbb{R}^+ \}$

Model = not differentiable at any point



Lec 2

Count.

	requirements
Ω	
F	$\emptyset, \Omega \in F$ ...
P	σ-addictivity

← like adding area, volume, mass

└ $P[A \cup B] = P[A] + P[B]$ if $A \cap B = \emptyset$

→ Mass analogy of probability ← Both can have uneven distribution; both addictive

We require:

↙ Numberable by natural number

P is Countable addictivity (aka σ-addictivity)

if $A_1, A_2, \ldots$ disjoint, then $P[\bigcup_{k=1}^{\infty} A_k] = \sum_{k=1}^{\infty} P(A_k) = \lim_{n\to\infty} \sum_{k=1}^{n} P(A_k)$

P is probability measure viz. $P[\Omega] = 1$

F is approprite subset of $\mathcal{P}(\Omega)$ ... requires $\emptyset, \Omega \in F$ but also:

- closed w.r.t countably many set theory operations (on elems of F)

↑ Call it "σ-algebra" or "σ-field"

complement, union, intersection

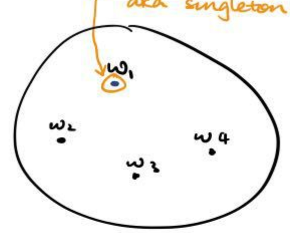
$$(\Omega, F, P)$$

↑ possible outcome ↑ σ-field, the events ↑ probality measure

Hmm ... why not just model with just Ω and P?

Discrete models

$\{\omega_1\}$ Elemental events aka singleton



Assume / define:
- Ω countable (finite or countably infinite)
- $F = \mathcal{P}(\Omega)$
- $P[\{\omega_k\}] = p_k$ $\quad (\sum p_k = 1)$

Then ... P is completely determined by all the p_k

Take any event A, $P(A) = P\left[\bigcup_{k:\omega_k\in A} \{\omega_k\}\right] = \sum_{k:\omega_k \in A} P[\{\omega_k\}]$

Ex. flip coin n times

$\Omega = \{ \underbrace{(\omega_1, \dots, \omega_n)}_{\text{binary seq of len } n} \mid \omega_i \in \{0,1\} \} = \{0,1\}^n$

$\mathcal{F} = \mathcal{P}(\Omega)$

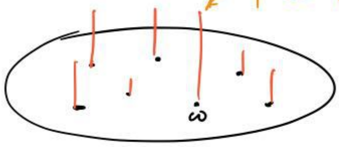
P depends on the coin and how to throw

fair coint, independent throws → normal then $P[\{\omega\}] = 2^{-n}$, uniform dist.

Lec 3

Discrete model (cont.)

prob for indiv. outcome

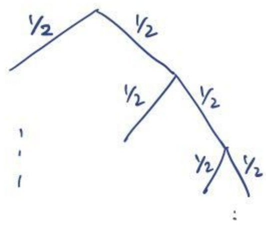


where $\sum | = 1$

$| = P[\{\omega_k\}] = p_k$ ← Now we have to specify this

Ex. Fair indep. coin flip N times — $p_k = \frac{1}{2^N}$

- by symmetry: each outcome equally likely ⇒ uniform dist.
- by $\underbrace{\frac{1}{2}\cdot\frac{1}{2}\cdot\dots\cdot\frac{1}{2}}_{N \text{ times}}$



Def For $A, B \in F$, A and B are indep. $\Leftrightarrow P(A \cap B) = P(A)P(B)$

Modified model — non-discrete

Set $N = \infty$, $\Omega = \{(x_1, x_2, \dots) \mid x_i \in \{0,1\}\}$ ← Not countable

$P[\{\omega_i\}] = 0 = \lim_{N\to\infty} \prod_{i=1}^{N} \frac{1}{2}$

Hmm

But we also want $\underbrace{\sum P[\{\omega_i\}]}_{} = 1$

Ω not countable, this doesn't make sense

Aside proving Ω uncountable. Suppose it's countable so $\Omega = \{\omega_1, \dots\}$
Let y s.t. $y \neq \omega_k$ for all $\omega_k \in \Omega$ (just flip kth bit of ω_k)

So we can't define P just using singletons $P[\{\omega_i\}]$.
→ Define it in terms of subset of Ω. eg. finite prefixes

$P[\underline{\{\omega = (x_1, \dots) \mid x_1 = 1\}}] = \frac{1}{2}$

Alt not. $(1, *, *, \dots)$

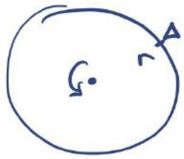
$P[(1, 0, *, 1, *, \dots)] = \frac{1}{8}$

these etc. implies unique P

But $F = \mathcal{P}(\Omega)$ also breaks here

Instead, $F = \sigma(\{(x_1, x_2, \dots, x_n, \dots) \mid n \geq 1\})$

Ex. Continuous roulette



$r \in [0,1)$

$\Omega = [0,1)$

$F = \dots ?$

$P[\{r\}] = 0$

$\mathbb{Q}$ countable

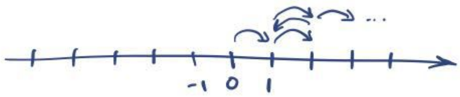
$P[\mathbb{Q}] = 0 = P[\bigcup \{q_k\}] = \sum^{\infty} 0 = 0$

$P[(a,b]] = b - a$

Lec 4

Cont (random walk (RW))

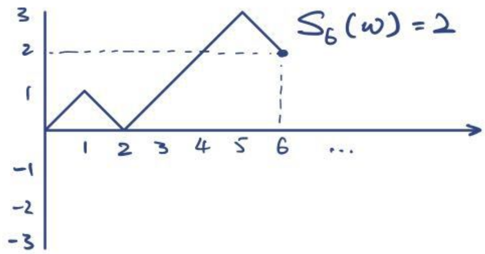
Ex. Consider particle moving +1 or −1 on number line determined by independent fair coin toss



$\Omega = \{ \omega = (\omega_1, \omega_2, \dots) \mid \omega_i = \{-1, 1\} \}$

$= \{-1, 1\}^{\infty}$

$= \{ w = (w_0, w_1, \dots) \mid w_i \in \mathbb{Z},$
$w_0 = 0, |w_i - w_{i+1}| = 1 \}$



Notation, random variable

$X_0(\omega) := 0$ – not random yet

$X_k(\omega) := \omega_k$ for $k \geq 1$

└ random variable for direction taken at step k.

looks like a function!

Random variable

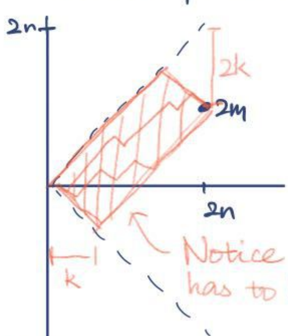
<u>Def</u> deterministic function $X : \Omega \to (\mathbb{R} \mid \mathbb{R}^d \mid \dots)$

Notice... $X(\omega)$

this ↑ is random ?!

So
$$\begin{cases} S_0(\omega) = 0 \\ S_n(\omega) = \sum_{k=1}^{n} X_k(\omega) \end{cases}$$

Ex. What's probability $P[S_n = m]$ f.s. $-n \leq m \leq n$?



Notice the path has to be in here

→ $P[S_2 = 1] = 0$

there's parity going on

What about $P[S_{2n} = 2m]$?

Suppose we step down k times in rectangle, $S_{2n} = 2n - 2k$

We want $S_{2n} = 2n - 2k = 2m$

$\Rightarrow k = n - m$

So we want ω s.t. $(\omega_1, \dots, \omega_{2n})$ has k step downs.

So $\binom{2n}{n-m}$ out of 2^{2n} possible prefixes

But ... $\{-1, 1\}^N$ ... $N \neq 2n$

Well we just want $A \subseteq \Omega$ s.t. A has the prefixes we want.

Say $2n = 4$, $2m = 0$ e.g. $A_{j_1, j_2} = \{(1, -1, 1, -1, *, *, \dots)\}$ for $1 \leq j_1 < j_2 \leq 2k$

$P[A_{j_1, j_2}] = \frac{1}{2^{2n}}$. ← $j_1 = 2$, $j_2 = 4$ in example

$$P[\underbrace{S_{2n} = 2m}_{=\ \bigcup_{1 \leq j_1 < \cdots < j_k \leq 2n} A_{j_1 \cdots j_k}}] = \sum_{j_1 \dots j_k} P[A_{j_1 \dots j_k}] = \sum_{j_1 \dots j_k} \frac{1}{2^{2n}}$$

disjoint union notation

$$= \frac{1}{2^{2n}} \sum_{j_1 \cdots j_k} 1$$

$$= \left(\frac{1}{2^{2n}}\right) \binom{2n}{k}$$

Independent

<u>Thm</u> $\mathbb{R}$ random vars X and Y independent

$\Leftrightarrow \forall A, B \subseteq \mathbb{R}$, the event $\underbrace{\{X \in A\}}_{\{\omega \mid X(\omega) \in A\}}$, $\{Y \in B\}$ independent

Lec 5

Independent

Def $\forall k, l, x, y$, $P[X_k = x, X_l = y] = P[X_k = x]\,P[X_l = y]$.
More generally, $X_{k_1}, \dots, X_{k_l}$ independent iff
$$P[\underbrace{X_{k_1}}_{k_1\text{-th flip}} \in A_1, \dots, X_{k_l} \in A_l] = \prod P[X_{k_j} = A_j] \quad \forall k_1 \dots k_l,\ A_1 \dots A_l$$

Recall RW

$$S_n(\omega) = \sum_{k=1}^{n} X_k(\omega)$$

Write $S_\bullet(\omega)$ to not specify k, so it's a random path.

Usefulness of looking at infinite system

Given large enough system and sufficient independence between components, something depending on many of these systems may become deterministic

Ex. Air bumping almost randomly → Statistical mechanics
most molecules don't interact, high independence
Pressure, temperature ... stable
Almost deterministic

Ex. Flip fair coin enough of time → 50% head 50% tail
$\frac{1}{n}\sum_k^n S_k(\omega) \to 0$

... What about tolerance $T = \left\{ \omega \,\middle|\, \left| \frac{1}{n}\sum_k^n S_k(\omega) - 0 \right| \leq \delta \right\}$ fixed tolerance > 0

$\lim_{n\to\infty} P[T] = 1$ ← Asymptotically approaching 1

Thm This is the weak law of large number (wLLN)

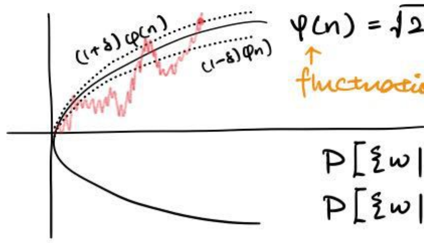
Thm strong LLN (sLLN)

Consider $N = \infty$ instead...

$$P\left[\left\{ \omega \,\middle|\, \frac{1}{n} S_n(\omega) \xrightarrow[n\to\infty]{} 0 \right\}\right] = 1$$ ← so it becomes deterministic

so fluctuation in $S_n(\omega)$ is lower than n

Consider



$\varphi(n) = \sqrt{2n \log\log n}$ $(n \geq 2)$

fluctuation of random walk

$P[\{\omega \mid S_n(\omega) \geq (1+\delta)\varphi(n) \text{ as of the time}\}] = 0$

$P[\{\omega \mid S_n(\omega) \geq (1-\delta)\varphi(n) \text{ as of the time}\}] = 1$

Consequences of σ-addictivity

Suppose $A, B \in F$ in a (Ω, F, P) system, then:

1. $B \subseteq A \Rightarrow P[A \setminus B] = P[A] - P[B]$

   「$B \uplus (A \setminus B) = A$」

2. $B \subseteq A \Rightarrow P[B] \leq P[A]$, so P is monotonically increasing
3. $P[A'] = P[\Omega \setminus A] = 1 - P[A]$
4. $P[A \cup B] = P[A] + P[B] - P[A \cap B]$ $= P[A \uplus (B \setminus A)]$
6. P is monotoneous continuous. Let $A_1 \subseteq A_2 \subseteq \cdots \subseteq \Omega$

$$P\left[\bigcup_{k \geq 1} A_k\right] = \lim_k P[A_k]$$

Lec 6 Disjoint Finite

Uniform dist for disjoint finite

Ω is disjoint finite
P is uniform dist so $P[\{\omega\}] = \frac{1}{|\Omega|}$

So $P[A] = \sum_{\omega\in A} P[\{\omega\}] = \frac{1}{|\Omega|} \sum_{\omega \in A} 1 = \frac{|A|}{|\Omega|}$

Permutation

Ex. arrange 52 cards. Let $n=52$

Mathematically ... we can model by $\pi : \underbrace{\{1,2,\dots,n\}}_{\text{orig position}} \to \underbrace{\{1,2,\dots,n\}}_{\text{new position}}$,
each π being a permutation.

Notice π is bijective.

Then set of all perms is $S_n = \{\pi : \{1,2,\dots,n\} \to \{1,2,\dots,n\} \mid \pi \text{ bijective}\}$

$|S_n| = n!$

Power set size

$|\mathcal{P}(\{1,2,\dots,n\})| = 2^n$

$\mathcal{P}([n]) \xleftrightarrow{\text{biject}} \{0,1\}^n$
$|\{0,1\}^n| = 2^n$.

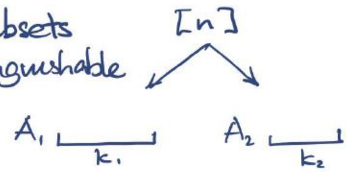
Choosing size k subset

$|\{A \subseteq [n] \mid |A| = k\}|$

$n(n-1)\cdots(n-(k-1)) = \frac{n!}{(n-k)!}$? Nope... we picked in order

$\binom{n}{k} = \frac{n!}{(n-k)!\,k!}$ overcount compensation

Ex. choosing two subsets
s.t. A_1, A_2 are distinguishable



Ambiguous... what if $k_1 = k_2$ and A_1, A_2 not distinguishable.

$$\binom{n}{k_2} \cdot \binom{n-k_1}{k_2} = \frac{n!}{\cancel{(n-k_1)!}\, k_1!} \, \frac{\cancel{(n-k_1)!}}{(n-k_1-k_2)!\, k_2!}$$

$$= \frac{n!}{(n-k_1-k_2)!\, k_2!\, k_1!}$$

$$= \binom{n}{k_1,\ k_2}$$ ← Notation to choosing multiple subsets

But ...

Partitioning

Ex. partition [7] into 4 non-empty, non-numerated parts
so we need 4 disjoint subsets that union to the set

Case on possible partition sizes

$1, 2, 2, 2 \Rightarrow \binom{7}{2,\ 2,\ 2}$

$4, 1, 1, 1 \Rightarrow \binom{7}{4}$

$3, 2, 1, 1 \Rightarrow \binom{7}{3,\ 2}$

$\boxed{\text{Lec 7}}$

Partition continued

Q: how many ways to split n things into $\overbrace{k \text{ non-empty partitions}}^{\text{Not enumerated}}$

we want k non-empty subsets with non-1 size that are disjoint but union to everything

... uh oh ... counting this is hopeless. brute force. No closed form solution.

reduction: figure out all possible size distribution, count each.

$1,2,2,2 \Rightarrow \binom{7}{2,2,2}$ or $\binom{7}{2\ 2\ 2\ 1}\frac{1}{3!}$

$4,1,1,1 \Rightarrow \binom{7}{4}$ or $\binom{7}{4\ \underbrace{1\ 1\ 1}_{3! \text{ ways to overcount}}}\frac{1}{3!}$

$3,2,1,1 \Rightarrow \binom{7}{3,2}$ or $\binom{7}{3,2,1,1}\frac{1}{2!}$

Now look at random partitions of $n=7$ $k=4$.

Ω = all the possible partitions $|\Omega| = 350 = \Sigma$
P = uniform dist

Q: What's $P[\{1,2,3\}$ in some partition$]$

$3,2,1,1 \Rightarrow \binom{4}{2}$

$4,1,1,1 \Rightarrow \binom{4}{1}$

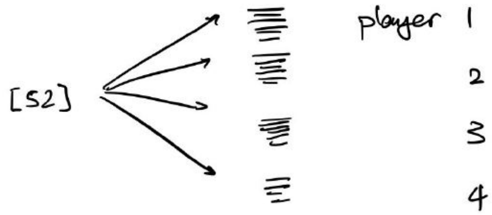
Card games

Dack of 52 cards $\{1, \dots, 52\}$

Suppose $\underbrace{\text{4 players}}_{\text{whether to enumerate depends on context?}}$ each 13 cards

Ω_{shuffle} = all perms of the 52 cards

P_{shuffle} = uniform $\qquad P[\{\omega\}] = \frac{1}{52!}$



Let K = all hearts.

$$P[\underbrace{H_2(\omega)} = K] = P[\underbrace{\{\omega \mid H_2(\omega) = K\}}_{\text{call it } A}]$$

↑ hand of 2^{nd} player

$H_2 : \Omega \to \left\{ \begin{array}{l} \text{all subsets} \\ \text{of size 13} \end{array} \right\}$

$$= \frac{|A|}{|\Omega|}$$

Counting A :

$\Leftarrow$ $52-13$, $52-13-1$

$\Leftarrow$ 13 choice, 12, ⋮

$$\binom{39}{13 \;\; 13 \;\; 13} \frac{1}{3!} \quad ?$$

$39!\,13!$

$$= \frac{39!\,13!}{51!}$$

$\boxed{\text{Lec 8}}$

Card shuffling (cont.)

Ω = { permutations of deck of 52 cards }
P = uniform dist.

Consider the dist. of As.

Most likely:
→ 1-1-1-1 ?
→ 2-1-1-0 ? ← The actual typical ... higher entropy

] notation: unordered players

Consider A = { 2-1-1-0 distribution of As }

__Sysmetically counting__

Concrete toy ex.	Players	1	2	3	4	Symbol
	Cardes	♡♤	◇		♧	γ

⋮

Notice $B_\gamma \subseteq A \subseteq \Omega$

$$P[B_\gamma] = \frac{|B_\gamma|}{52!} = \frac{\binom{13}{2}\cdot 2\cdot\binom{13}{1}\cdot\binom{13}{1}\cdot 48!}{52!}$$

Consider Player 1: $\binom{13}{2}\cdot 2$ ways to insert other cards
P 2 $\binom{13}{1}$ - - - -
3
P 4 $\binom{13}{1}$ - - - -

Let Γ be all possible symbols like γ
Then $A = \bigcup_{\gamma\in\Gamma} B_\gamma$

$$P[A] = \sum_{\gamma\in\Gamma} P[B_\gamma] = |\Gamma|\, P[B_\gamma].$$

To count $|\Gamma|$ ...
1. Choose who gets 2 A's and who gets 0
   - Choose who gets 2
   - Choose who gets 0

   $\binom{4}{1}\binom{3}{1} = 12$
2. Decide where A's go
   - choose 2 for one player
   - permute other two / choose 1 for another

   — $\binom{4}{2}\binom{2}{1} = 12$

So $|\Gamma| = 12^2$

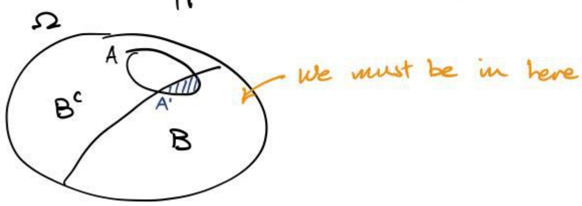
$$P[A] = 12^2 \cdot P[B_\gamma] = \frac{13^3\ 12^3\ 48!}{52!} \approx 0.57$$

Lec 9 Conditional probability

Conditional prob

Given a priori model (Ω, F, P) ... we know nothing

... now suppose we saw $B \subseteq \Omega$ occured ... build a posterior (Ω', F', P')



$\Omega' = B$

$F' = \{ A \cap B \mid A \in F \}$ Let $A' = A \cap B$

$$P'(A') = \frac{P(A')}{P(\Omega')} = \frac{P(A \cap B)}{P(B)}$$ Note this requires $P(B) > 0$

But that's complicated ... try:

New posterior model, only update F

(Ω, F, Q)

<u>Def</u> $Q(A) = \dfrac{P(A \cap B)}{P(B)} = P(A \mid B)$

<u>Claim</u> Q is prob. measure

$Q(A \mid B)$ ✓

$$Q(\Omega) = \frac{P(\Omega \cap B)}{P(B)} = 1 \quad ✓$$

$$Q\left(\bigcup_k A_k \right) = \frac{1}{P(B)} P\left(B \cap \bigcup_k A_k \right)$$

$$= \frac{1}{P(B)} P\left(\bigcup_k A_k \cap B \right)$$

$$= \frac{1}{P(B)} \sum_k P\left(A_k \cap B \right)$$

$$= \sum_k \frac{P(A_k \cap B)}{P(B)}$$ ✓ σ - addictivity

<u>Nota</u> $P(\cdot \mid B)$ is conditional prob measure of P on B

<u>Conseq</u> $P(B^c \mid B) = 0$

$$P(A \cup C \mid B) = P(A \mid B) + P(C \mid B) - P(A \cap C \mid B)$$

$$P(A^c \mid B) = 1 - P(A \mid B)$$

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \Leftrightarrow P(A \cap B) = P(A \mid B)\, P(B)$$

$$\Leftrightarrow P(B \cap A) = P(B \mid A) = P(A)$$

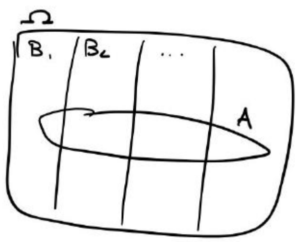
$$\Rightarrow P(B \mid A) = P(A \mid B) \frac{P(B)}{P(A)}$$

Ex. weather.

A = 20% rain predicted yesterday
B = cloudy

$$P(A \mid B) \overset{\text{probably}}{>} P(A)$$

More partitioning



$$\Omega = \bigcup_{k}^{N} B_k \quad , \ N \leq \infty$$

$$A = \bigcup_{k}^{n} A \cap B_k$$

$$P(A) = \sum P(A \cap B_k)$$

$$= \sum P(A \mid B_k)\ P(B_k)$$

weighted avg of conditional probs.

So having exhaustive scenarios B_k s and $P(A \mid B_k)$ can help find $P(A)$

This looks like ... $\underbrace{\sum_k p_k \cdot a_k}_{\text{weighted average}} = 1$

Now what if we want $P(B_j \mid A)$ if we know $P(A \mid B_j)$.

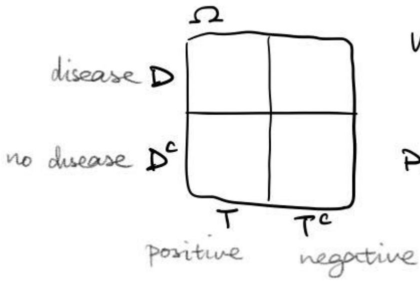
natural question: which partition are we in if we observed A?

$$P(B_j \mid A) = \frac{P(A \mid B_j)\ P(B_j)}{P(A)} = \frac{P(A \mid B_j)\ P(B_j)}{\sum P(A \mid B_k)\ P(B_k)}$$

Ex. Getting tested positive on some medical test

→ Look up reliability of the test

Pretend we don't know test result yet. We want to find $P(D|T)$



We could find $P(T|D) = 99\%$

$P(T^c|D^c) = 97\%$

$$P(D|T) = P(D|T)\frac{P(T)}{P(D)}$$

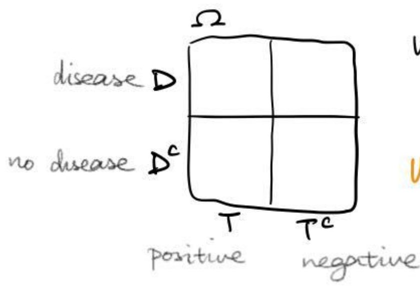
Lec 10 Conditional Prob. & Random Variable

Recall disease test

Ex. Getting tested positive on some medical test

→ Look up reliability of the test

Pretend we don't know test result yet. We want to find $P(D|T)$



We could find $P(T|D) = 99\%$

$P(T^c|D^c) = 97\%$

We want $P(D|T) = P(T|D)\dfrac{P(T)}{P(D)}$

Imperative way: give test to ppl with D and see how many gets positive

Realiability $P(T|D) = 99\%$

Specificity $P(T^c|D^c) = 97\%$

$P(D) = 0.1\%$

from googling

frequency of disease

$$P(D|T) = \frac{P[D\cap T]}{P[T]} = \frac{P(T|D)\,P(D)}{P(T)}$$

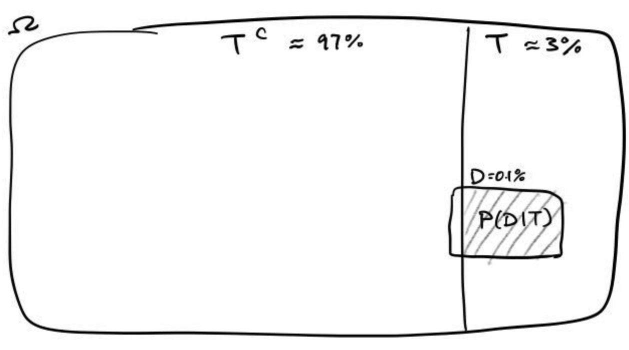
frequency of positive across all ppl tested

$$= \frac{P(T|D)\,P(D)}{P(T|D)P(D) + P(T|D^c)\,P(D^c)}$$

$$\approx \frac{1}{31}$$

That's low. Usually they send you to another test.
The two tests are ideally independent but usually not really.

Intuition



Consider 2 tests. T_1, T_2

$P(T_2 \mid T_1) > P(T_2)$

Usually more likely to have disease so 2nd test more likely to be positive

Conditional independence!

Assume T_1, T_2 independent w.r.t. $P(\cdot \mid D)$

Random variable & their dist (denote μ_X)

Random variable : function $X : \Omega \to S$ for some set S

$\Omega \xrightarrow{X} S$ eg. $\Omega \xrightarrow{X} \mathbb{R}$

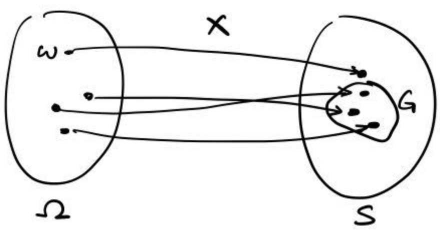
Let $\mathcal{G}$ be σ-fields on S

Notat. $P[X \in G] = P[\{\omega \mid X(\omega) \in G\}]$
$= \underbrace{P[X^{-1}(G)]}$

Prob of all ω that maps into G by X

$= \mu_X(G)$

This is itself a prob measure on $(S, \mathcal{G})$



$(S, \mathcal{G}, \mu_X)$ acts like another random system

Special case : X is discrete

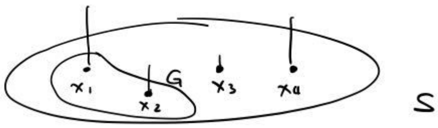
X discrete $\Leftrightarrow$ $\{X(\omega) \mid \omega \in \Omega\}$ is countable
$= \{x_1, \cdots, x_n\}$ are possible values

Naturally we consider $P[X_k = x_k] =: p_k$

Notice $X = x_k$ disjoint for ks and every ω goes to some x_k.

So $\bigcup_{k\geq 1} \{X = x_i\} = \Omega \Rightarrow \sum_k p_k = 1$

Discrete point measure



then $\mu_X(G) = \sum_{\substack{k, \\ x_k \in G}} P[X = x_k] = \sum_{\substack{k, \\ x_k \in G}} p_k$

Binomial Dist.

$B(n, p)$. Consider n coin flips with head prob p.

$X_1 \ \dots \ X_n$ $\qquad X_k \begin{cases} 1 & p \\ 0 & 1-p \end{cases}$

Sum $S_n(\omega) := \sum_{i=1}^{n} X_i(\omega)$ $\leftarrow$ # of 1's in n flips

Want $P[S_n = k]$ Well... $k \in S_n = \{0, 1, \dots, n\}$
each outcomes have different prob

$$= p^k (1-p)^{n-k} \binom{n}{k}$$

<u>Notat</u> $S_n \sim B(n, p)$
$\uparrow$ Distributed

$\boxed{\text{Lec 11}}$

RV cont.

(Ω, F, P) $\quad \Omega \xrightarrow{X} S$ $\quad \mathcal{G} \leftarrow$ subsets of S, a σ-field

$\mu_X(G) = P[X \in G] = P[\{\omega \mid X(\omega) \in G\}] = P \circ X^{-1}(G)$

$S \supseteq G \in \mathcal{G}$ $\quad \Rightarrow \mu_X$ is a prob measure on $(S, \mathcal{G})$

P transferred over to μ, $(S, \mathcal{G})$ retains structure.

so (Ω, F, P)
$\downarrow$
$(S, \mathcal{G}, \mu_X)$

⚠ Note this requires $\forall G \in \mathcal{G}$, $X^{-1}(G) \in F$. This is usually assumed.

<u>Def</u> μ_X is the distribution of X w.r.t. P and a prob. measure on $(S, \mathcal{G})$

<u>Def</u> X is discrete $\Leftrightarrow \{X(\omega) \mid \omega \in \Omega\} = \{x_1, \dots, x_n\} \subseteq S$ is countable

Then it's sufficient to look at prob of singletons

$\mu(\{x_k\}) = P[X = x_k] = p_k$

$$\Rightarrow P[X \in G] = P[\bigcup_{\substack{k \\ x_k \in G}} \{X = x_k\}] = \sum_{x \in G} P[\{X = x\}] = \sum_{\substack{k \\ x_k \in G}} p_k$$

Example dist.

Distribute 13 of 52 cards to 1 player P_1, consider the hand

$\Omega = \{$ all perms of 52 cards $\}$
$F = \{ A \subseteq \{1, \dots, 52\} \mid |A| = 13 \}$

$X : \Omega \to S$ by taking first 13 cards, putting it inside a set, and giving it to P_1.

Want μ_X. Let A be some subset of $\{1, \dots, 52\}$, $|A| = 13$

$$\mu_X(\{A\}) = \underset{\substack{\text{III} \\ P[\underbrace{\{\omega \mid X(\omega) = A\}}_{\text{Count this}}]}}{P[X = A]} = \frac{X = A}{|\Omega|} = \frac{13!\ 39!}{52!} = \frac{1}{\binom{52}{13}}$$

But any such A will yield this result. So μ_X is uniform.

Ex. random vars

① binomial $B(n, p)$

$X_1 \dots X_n$

$\downarrow$ ← independent, identical dist.

iid $\sim B(p) = \left\langle \begin{array}{ll} 1 & p \\ 0 & 1-p \end{array} \right.$

$\Vert\; B(1, p)$

$S_n = \sum_k^n X_k \qquad P[S_n = k] = \binom{n}{k} p^k (1-p)^{n-k}$

②

$X_1 X_2 \dots \qquad$ iid $\sim B(p)$

$T(\omega) := \min \{ k \geq 1 \mid X_k(\omega) = 1 \}$ ← Index of the first 1
Waiting time for first success

$0\ 0\ 0\ 1\ 0\ 1\ 1\ 0 \dots$

$T(\omega) = 4 = \min \{4, 6, 7, \dots\}$

Notice $T(\omega) \in \mathbb{N}^+$

$$\begin{aligned} \mu_T(\{k\}) = p_k &= P[T = k] \\ &= P[\{X_1 = 0\} \cap \{X_2 = 0\} \cap \dots \cap \{X_{k-1} = 0\} \cap \{X_k = 1\}] \\ &= P[\{X_1 = 0\}] P[\{X_2 = 0\}] \dots P[\{X_{k-1} = 0\}] P[\{X_k = 1\}] \\ &= (1-p)^{k-1} p \end{aligned}$$

Sanity check all p_k sum up to 1 : $\sum_{k \in \mathbb{N}^+} (1-p)^{k-1} p$

③ Negative Binomial

$X_1 \dots X_n \quad$ iid $\left\langle \begin{array}{ll} 1 & p \\ 0 & 1-p \end{array} \right.$

Fix $1 \leq n$.

$\underbrace{0\ 0\ 1\ 0\ 0\ 1\ 0}_{\text{want 2 succ before } k}\ \overset{3^{rd}}{1}$

$T(\omega) :=$ time until n^{th} success

Want $P[T_n = k]$ for some $k \geq n$.

$P[T_n = k] = p^n (1-p)^{k-n} \binom{k-1}{n-1}$

④ Poison dist. $X = \{0, 1, 2, \dots\}$

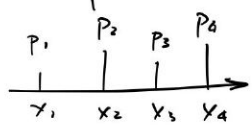
Poison (λ) $P[X = k] = e^{-\lambda} \frac{\lambda^k}{k!}$

Lec 12 Expected Value

Given some discrete RV $X \in \mathbb{R}$.

Idea: want to replace X with a single, deterministic number.
simplification, reduction

Attempt 1 — weighted avg



$$W(X) = p_1 x_1 + p_2 x_2 + p_3 x_3 + p_4 x_4 = \sum_k p_k x_k$$

weighted avg

Attempt 2 — prediction with least square error

Prediction $= b \Rightarrow$ Error $= |X(\omega) - b|$

Want $\min_{b \in \mathbb{R}} |X(\omega) - b|$

Try $\min_{b \in \mathbb{R}} W(|X(\omega) - b|^2)$

Then the minimiser b_0 is optimal pred.

<u>Fact</u> b_0 is unique

Attempt 3 — Statistics

Take many samples $X_1, X_2, \dots$ iid with $X_k \sim X$

Take average $\frac{1}{n} \sum_{k=1}^{n} X_k(\omega)$

By law of large num...

$$P\left[\left\{ \omega \mid \frac{1}{n} \sum_{k=1}^{n} X_k(\omega) \to c \right\} \right] = 1$$

converges to some constant with probability 1

Expected val

Turns out attempts $1 \equiv 2 \equiv 3$. Define $\mathbb{E}[X] = W(X) = b_0 = c$

Properties of expected val

Consider $\mathbb{E}[\cdot]$ to be func on RVs
$\mathbb{E}[\cdot] : \{RVs\} \to [-\infty, \infty]$

Note not every RV has exp. val. eg. when we need $-\infty + \infty$

1. Exp. val. is extension of prob. measure

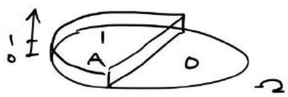
Let $A \in \mathcal{F}$, $1_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A \end{cases}$ ← indicator RV

$$\mathbb{E}[1_A(\omega)] = 1 \cdot P[1_A = 1] + 0 \cdot P[1_A = 0]$$
$$= P[A]$$

So $(\Omega, \mathcal{F}, P)$ automatically generates $\mathbb{E}$

carry over :
- σ-additivity
- monotone cont.

"integral w.r.t. P"



2. $\mathbb{E}[\cdot]$ is linear

$$\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$$
$$\mathbb{E}[cX] = c\,\mathbb{E}[X]$$

<u>Proof</u> $\mathbb{E}[X+Y] = \sum_{z \in Im(X+Y)} z \cdot P[X+Y=z]$

$$= \sum_{\substack{x+y=z \\ x \in Im X \\ y \in Im Y}} z \cdot P[X=x, Y=y]$$

$$= \sum_{\substack{x \in Im X \\ y \in Im Y}} (x+y) P[X=x, Y=y]$$

$$= \sum_x \sum_y (x+y) P[X=x, Y=y]$$

$$= \sum_x x \sum_y P[X=x, Y=y]$$

$\vdots$

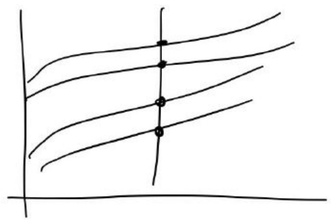
$$= \mathbb{E}X + \mathbb{E}Y$$

3. $\forall \omega, \quad X(\omega) \geq Y(\omega) \Rightarrow \mathbb{E}X \geq \mathbb{E}Y$

Proof $\quad \mathbb{E}[\underbrace{X-Y}_{\geq 0}] = \mathbb{E}X - \mathbb{E}Y$

4. $\mathbb{E}[\cdot]$ monotone cont.

<u>Thm</u> $\quad (0 \leq X_k(\omega) \nearrow \ \forall \omega) \Rightarrow \mathbb{E}[\lim \nearrow X_n(\omega)] = \lim \mathbb{E}(X_n)$



\# $\mathbb{E}$ of binom dist.

$S \sim B(n,p) \qquad \mathbb{E}S = \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k}$

Try $\quad S \sim \hat{S} = \sum_{k=1}^{n} \hat{X}_k \qquad \mathbb{E}\hat{S} = \sum_{k=1}^{n} \mathbb{E}\tilde{X}_k$

$= np$

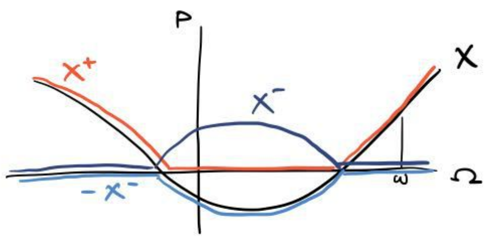
Lec 13

Recall ...

$\mathbb{E}[X]$ — linear, monotone

If $X \geq 0$, $\underbrace{\mathbb{E}[X]}_{\sum_k p_k x_k} \in [0, \infty]$ ← this is ok

If X is not always positive, we can say $X = X^+ - X^-$

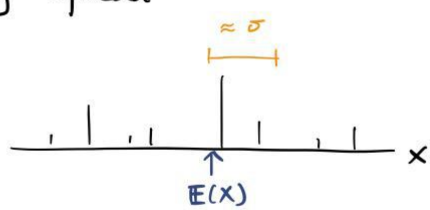


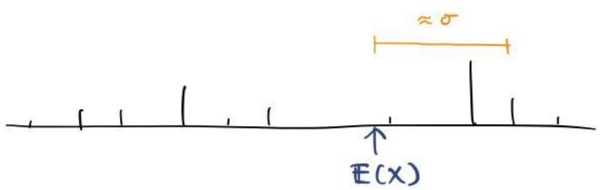
Then $\mathbb{E}[X] = \mathbb{E}[X^+] - \mathbb{E}[X^-]$

$\mathbb{E}(X^+) \in [0, \infty]$
$\mathbb{E}(X^-) \in [0, \infty]$

Note if $\mathbb{E}(X^+) = \mathbb{E}(X^-) = \infty$, $\mathbb{E}(X)$ not well defined

Describing spread





$\mathrm{var}(X) = \sigma^2(X) = \sigma^2 := \mathbb{E}\Big[\underbrace{(X(\omega) - \mathbb{E}(X))^2}_{\text{square error}}\Big]$ ← Variance

$\sigma := \sqrt{\mathrm{var}(X)}$ ← Standard deviation

<u>Variance properties</u>

① $\mathrm{var}(\alpha X) = \alpha^2 \,\mathrm{var}(X)$

② $\mathbb{E}[(X(\omega) - \mathbb{E}(X))^2] = \mathbb{E}[X^2 + (\mathbb{E}(X))^2 - 2(\mathbb{E}(X))X]$

$$= \mathbb{E}[X^2] + \mathbb{E}[\underbrace{(\mathbb{E}(X))^2}] - 2(\mathbb{E}(X))\cdot\mathbb{E}(X)$$

This is constant
RV, viz. $P[Y=c] = 1$

$$= \mathbb{E}[X^2] - (\mathbb{E}(X))^2$$

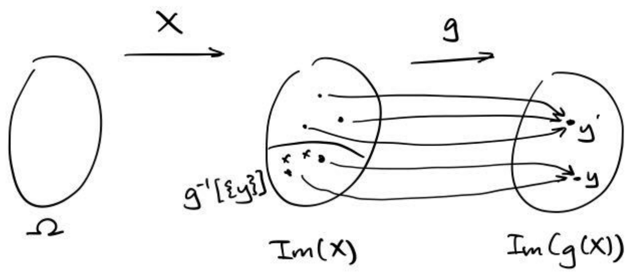
2nd moment minus expected squared

<u>Transformation formula</u> $g : \mathbb{R} \to \mathbb{R}$

$$\mathbb{E}[g(X)] = \sum_{x \in Im(X)} g(x) \cdot P[X = x] \quad \leftarrow \text{works}$$

$$= \sum_{y \in Im(g(X))} y \cdot P[g(x) = y] \quad \leftarrow \text{by definition}$$

Showing they are equal



So $var(X) = \mathbb{E}(\underbrace{X^2}_{\text{Consider } g(x) = x^2}) - (\mathbb{E}X)^2$

$$\sum_{x \in Im(X)} g(x) \cdot P[X = x] = \sum_{y \in Im(g(X))} \sum_{x \in g^{-1}[\{y\}]} g(x) P[X = x]$$

$$= \sum_{y \in Im(g(X))} \sum_{x \in g^{-1}[\{y\}]} y P[X = x]$$

$$= \sum_{y \in Im(g(X))} y \sum_{x \in g^{-1}[\{y\}]} P[X = x]$$

$$= \sum_{y \in Im(g(X))} y \, P[g(X) = y]$$

③ Assume $\mathbb{E}X = \mathbb{E}Y = 0$

$$\begin{aligned}
var(X+Y) &= \mathbb{E}[(X+Y)^2] - \cancelto{0}{(\mathbb{E}[X+Y])^2} \\
&= \mathbb{E}[X^2] + \mathbb{E}[Y^2] + 2\,\mathbb{E}[XY] \\
&= var(X) + var(Y) + 2\mathbb{E}[XY]
\end{aligned}$$

Observe $var(X+c) = var(X)$

$$\mathbb{E}(X^2 + c^2 + 2Xc) - (\mathbb{E}(X+c))^2$$

$$\cdots = var(X)$$

Then

$$\begin{aligned}
& var(\tilde{X} + \tilde{Y}) \\
&= var(X + Y + \mathbb{E}\tilde{X} + \mathbb{E}\tilde{Y}) \\
&= var(X+Y) \\
&= var(X) + var(Y) + 2\mathbb{E}(XY) \\
&= var(X) + var(Y) + 2\underbrace{\mathbb{E}\left[(\tilde{X} - \mathbb{E}\tilde{X})(\tilde{Y} - \mathbb{E}\tilde{Y})\right]}_{\text{Covariance}\ !!\ Cov(\tilde{X},\tilde{Y})} \\
&= var(\tilde{X}) + var(\tilde{Y}) + 2Cov(\tilde{X},\tilde{Y})
\end{aligned}$$

| define $X = \tilde{X} - \mathbb{E}\tilde{X}$
$Y = \tilde{Y} - \mathbb{E}\tilde{Y}$

Lec 14

Recall variance

$$\mathrm{var}(X) = \mathbb{E}[(X - \mathbb{E}X)^2] = \mathbb{E}(X^2) - (\mathbb{E}X)^2$$

Observation: $\mathrm{var}(X)$ doesn't depend on $\mathbb{E}X$.

$$\mathrm{var}(X) = \mathrm{var}(\tilde{X}) = \mathbb{E}[\tilde{X}^2]$$

Covariance

$$\mathrm{cov}(X | Y) = \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)]$$

Observe $\mathrm{cov}(\cdot | \cdot)$ as function is symmetric and bilinear

linear for each arg

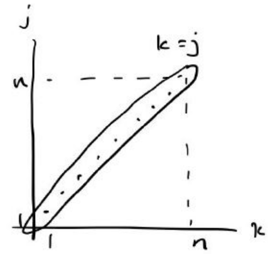
Observe $\mathrm{var}(X) = \mathrm{cov}(X | X)$

Ex. $S = \sum_{k=1}^{n} X_k$

$$\mathrm{var}(S) = \mathrm{cov}\left(\sum_{k=1}^{n} X_k \,\Big|\, \sum_{j=1}^{n} X_j\right)$$

$$= \sum_k \mathrm{cov}\left(X_k \,\Big|\, \sum_{j=1}^{n} X_j\right)$$

$$= \sum_k \sum_j \mathrm{cov}(X_k | X_j)$$



$$= \sum_k \mathrm{cov}(X_k | X_k) + \sum_{k \neq j} \mathrm{cov}(X_k | X_j)$$

$$= \sum_k \mathrm{var}(X_k) + 2 \sum_{k<j} \mathrm{cov}(X_k | X_j)$$

Variance of sums of indep variables

<u>Def</u> X, Y independent $\Leftrightarrow \forall A, B \subseteq \mathbb{R}$, $\{X \in A\}$, $\{Y \in B\}$ indep.

$\Rightarrow P[X \in A \mid Y \in B] = P[X \in A]$

$\Leftrightarrow$ for discrete X, Y, $\forall k, \ell$, $P[X = x_k, Y = y_\ell] = P[X = x_k] P[Y = x_\ell]$

Expected value of product

$$\mathbb{E}[XY] = \sum_k \sum_l x_k y_l \, P[X = x_k, Y = y_l]$$

Notice
$cov(X|Y) = \mathbb{E}(XY) - \mathbb{E}X \cdot \mathbb{E}Y$

Consider

Indicator func

$$\mathbb{E}[XY] = \mathbb{E}\Big[\Big(\underbrace{\sum_k x_k \cdot 1_{\{X = x_k\}}(\omega)}_{= X(\omega)}\Big)\Big(\sum_l y_l \cdot 1_{\{Y = y_l\}}(\omega)\Big)\Big]$$

only if both indicators are 1, the inner is $x_k y_l$. Else it's 0

$$= \mathbb{E}\Big[\sum_k \sum_l x_k \cdot y_l \cdot \underbrace{1_{\{Y = y_l\}}(\omega) \cdot 1_{\{X = x_k\}}(\omega)}_{\text{indicator for } X \cap Y}\Big]$$

$$= \mathbb{E}\Big[\sum_k \sum_l x_k \cdot y_l \cdot 1_{\{Y = y_l,\, X = x_k\}}(\omega)\Big]$$

$$= \sum_k \sum_l x_k \cdot y_l \; \mathbb{E}\Big[1_{\{Y = y_l,\, X = x_k\}}(\omega)\Big]$$

$$= \sum_k \sum_l x_k \cdot y_l \; P[Y = y_l, X = x_k]$$

Special case: consider independent X, Y.

$$\mathbb{E}[XY] = \sum_k \sum_l x_k \cdot y_l \; P[Y = y_l] \, P[X = x_k]$$

$$= \sum_k x_k \, P[X = x_k] \cdot \sum_l y_l \, P[Y = y_l]$$

$$= \mathbb{E}(X) \cdot \mathbb{E}(Y)$$

Back to covariance

If X, Y, X_k independent $\Rightarrow$ Note $\Leftarrow$ is not true

$$cov(X|Y) = \mathbb{E}(XY) - \mathbb{E}X \cdot \mathbb{E}Y$$

$$= \mathbb{E}X \cdot \mathbb{E}Y - \mathbb{E}X \cdot \mathbb{E}Y$$

$$= 0 \qquad \text{uncorrelated.}$$

So $var(\sum X_k) = \sum_k var(X_k) + 0$

Variance of distributions

① $S \sim B(n, p)$ $\qquad\qquad \mathbb{E}S = np$

$$S \sim S' := \sum_{k=1}^{n} \underset{\substack{\uparrow \\ \text{iid}}}{X_k}$$

$$\begin{aligned}
var(S) &= var(S') = var\left(\sum_{k}^{n} X_k\right) \\
&= \sum_{k=1}^{n} var(X^2) \\
&= \sum_{k=1}^{n} \left(\mathbb{E}(X^2) - (\mathbb{E}X)^2 \right) \\
&= \sum_{k=1}^{n} (p - p^2) \\
&= n(p - p^2) \\
&= np(1-p)
\end{aligned}$$

$\boxed{\text{Lec 15}}$

Taylor expansion

$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$

Poisson distribution

$X \sim Poi(\lambda)$

$\mathbb{E}[X] = \sum_{k\geq 0} k e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda}\lambda \sum_{k\geq 1} \frac{\lambda^{k-1}}{k(k-1)!} = e^{-\lambda}\lambda \sum_{j\geq 0} \frac{\lambda^j}{j!} \overset{*}{=} e^{-\lambda}\lambda e^{\lambda} = \lambda$

$var(X) \overset{?}{=}$

$\mathbb{E}[X(X-1)] = \sum_{k\geq 0} k(k+1) e^{-\lambda} \frac{\lambda^k}{k!} \cdots = \lambda^2 e^{-\lambda} e^{\lambda} = \lambda^2$

$\mathbb{E}[X^2] = \mathbb{E}[X(X-1) + X] = \lambda^2 + \lambda$

$var(X) = \lambda$

Geometric distribution

$geom(p) \sim X \qquad P[X=k] = (1-p)^{k-1} p \qquad \text{for } k \geq 1$

$\mathbb{E}[X] = \sum_{k\geq 1} k p (1-p)^{k-1} = p \sum k x^{k-1} \Big|_{x=1-p}$

$= p \sum (x^k)' \Big|_{x=1-p}$

$= p \left(\sum_{k\geq 1} x^k \right)' \Big|_{x=1-p}$ ← power series $0 \leq 1-p < 1$ it should converge

$= p \left(\sum_{k\geq 0} x^k \right)' \Big|_{x=1-p}$

$= p \left(\frac{1}{1-x} \right)' \Big|_{1-p}$

$= p \frac{1}{(1-x)^2} \Big|_{1-p}$

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

$= p \frac{1}{p^2}$

$= \frac{1}{p}$

Alt. If $X \in \mathbb{N}$ $\Rightarrow$ $\mathbb{E}[X] = \sum_{k \geq 0} P[X > k] = \sum_{j \geq 1} P[X \geq j]$

$$= \sum_{k \geq 0} (1-p)^k$$

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

$$= \frac{1}{p}$$

Conditional Expectation

apriori P, observation A, $P_A = P(\cdot \mid A)$

$\mathbb{E}$ $\qquad$ $\mathbb{E}_A = \mathbb{E}(\cdot \mid A)$

$\mathbb{E}[X] = \sum_k x_k P[X = x_k]$ (for discrete X)

$\mathbb{E}_A[X] = \sum_k x_k P_A[X = x_k] = \sum_k x_k P[X = x_k \mid A]$

$\mathbb{E}_A[X] = \mathbb{E}[X \mid A]$

$$\begin{aligned} var_A(X) &= \mathbb{E}_A[(X - \mathbb{E}_A(X)^2)] \\ &= \mathbb{E}[(X - \mathbb{E}(X \mid A)^2) \mid A] \\ &= \mathbb{E}_A[X^2] - (\mathbb{E}_A X)^2 \end{aligned}$$

Partition Thm Expectation Ver

$P[A] = \sum_k P[A \cap B_k] = \sum_k P[A \mid B] P[B_k]$

$\mathbb{E}[X] = \sum_k \mathbb{E}[X \cdot 1_{B_k}] = \sum_k \mathbb{E}[X \mid B_k] P[B_k]$

Recall discrete := countable codomain

Joint Distribution

$\Omega \xrightarrow{X_k} S = \mathbb{R}$ for discrete RVs X_k, $k \in 1..n$

Let $\vec{X}(\omega) = (X_1, \dots, X_n) \in \mathbb{R}^n$ which is still a discrete RV.

Joint distribution

$\mu_{\vec{X}}(A) = P[\vec{X} \in A] = \sum_{(x_1, \dots, x_n) \in A} P[X_1 = x_1, \dots, X_n = x_n]$

Lec 16

MT1 next Monday – no cheat sheet

Joint distribution (cont.)

Let X, Y be RVs $\Omega \to S$. Define $\vec{X}(\omega) = (X(\omega), Y(\omega))$

$\Omega \to S \times S$

$S^2 \to \mathbb{R}$

$$\mu_{\vec{X}}(A) = P[\vec{X} \in A] = \sum_{(x,y) \in A} P[\vec{X} = (x,y)]$$

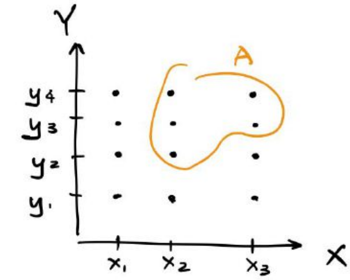
'' (notation)

$$\mu_{X,Y}(A) \quad = \sum_{(x,y)\in A} P[X = x, Y = y]$$

$$= \sum_{(x,y)\in A} \mu_{X,Y}(\{(x,y)\})$$

(notation)

$$= \sum_{(x,y)\in A} \mu_{X,Y}(x,y)$$



So $\mu_{X,Y}$ is completely determined by all $P[X = x, Y = y]$, $x \in Im(X)$, $y \in Im(Y)$

"joint" "marginals"

Say $X, Y \sim \mu_{X,Y}$. Can one recover μ_X, μ_Y?

$$\mu_X(x) = P[X = x] = \sum_{y \in Im(Y)} P[X = x, Y = y] = \sum_{y\in Im(Y)} \mu_{X,Y}(x,y)$$

Say we know μ_X, μ_Y, is $\mu_{X,Y}$ recoverable?

Y \ X	0 ($\frac{1}{2}$)	1 ($\frac{1}{2}$)
0 ($\frac{1}{2}$)	$\frac{1}{4}$	$\frac{1}{4}$
1 ($\frac{1}{2}$)	$\frac{1}{4}$	$\frac{1}{4}$

← Independent coin flips

Y \ X	0 ($\frac{1}{2}$)	1 ($\frac{1}{2}$)
0 ($\frac{1}{2}$)	$\frac{1}{2}$	0
1 ($\frac{1}{2}$)	0	$\frac{1}{2}$

← Not independent coin flips

So not recoverable in general, but recoverable iff independent.

<u>Thm</u> X, Y indep $\Leftrightarrow$ $\mu_{X,Y}(x,y) = \mu_X(x)\mu_Y(y)$

With conditional

Y \ X	-1	1
-1	1/6	1/8
0	1/6	1/4
1	1/6	1/8

$$\mu_{Y|X=-1}(y) = \text{uniform}$$
$$= P[Y=y \mid X=1]$$
$$= \frac{P[Y=y,\ X=1]}{P[X=1]}$$
$$= \frac{1/6}{1/2}$$

Notation $\mu_{X,Y}(x,y)$

$$\mu_{Y|X}(a,b) := P[Y=b \mid X=a]$$
$$\mu_{X|Y}(a,b) := P[X=a \mid Y=b]$$

Ex. Given $\mu_{X,Y}$, find μ_{X+Y}

$$\mu_{X+Y}(z) = P[X+Y=z]$$

$z \in \mathrm{Im}(X+Y) = \{x+y \mid x \in \mathrm{Im}\,X,\ y \in \mathrm{Im}\,Y\}$

$$= \sum_{x \in \mathrm{Im}X} P[X+Y=z,\ X=x]$$
$$= \sum_{x \in \mathrm{Im}X} P[x+Y=z,\ X=x]$$
$$= \sum_{x \in \mathrm{Im}X} P[Y=z-x,\ X=x]$$
$$= \sum_{x \in \mathrm{Im}X} \mu_{X,Y}(x,\ z-x)$$

If X,Y indep

$$= \sum_{x \in \mathrm{Im}X} \mu_X(x)\,\mu_Y(z-x)$$

← "convolution of μ_X, μ_Y"

$\boxed{\text{Lec 17}}$

Joint dist. (cont)

$$\Omega \xrightarrow{X,Y} S$$
$$\Omega \xrightarrow{\vec{X} = (X,Y)} S \times S$$

Discrete case: sufficient to just look at singletons $\{(x,y)\} \subseteq S \times S$

Transformation formula

$$\mathbb{E}[g(X,Y)] = \mathbb{E}[g(\vec{X})] = \sum_{(x,y)\in \mathrm{Im}(\vec{X})} g(x,y)\, P[X=x, Y=y]$$

$$= \sum_{\vec{x}\in \mathrm{Im}(\vec{X})} g(\vec{x})\, P[\vec{X} = \vec{x}]$$

Conditional

$$\mu_{Y|X}(x, \cdot) = P[Y = \cdot \mid X = x]$$

$$\mathbb{E}[g(Y) \mid X = x] = \sum_{y \in \mathrm{Im}(Y)} g(y) \cdot P[Y = y \mid X = x]$$

$$\mathbb{E}[g(X,Y) \mid X = x] = \sum_{y \in \mathrm{Im}(Y)} g(x,y) \cdot P[Y = y \mid X = x]$$

Independence

By def, $P[\bigcap_{k=1}^{n} \{\omega \mid X_k(\omega) \in A_k\}] = \prod_k P[X_k \in A_k]$

$\Updownarrow$

$\forall x_1 \in X_1, \ldots, x_n \in X_n$,
$P[X_1 = x_1, \ldots, X_n = x_n] = \prod_k P[X_k = x_k]$

joint dist equals product of marginal dist

Ex. $Z_1, Z_2, \dots$ iid biased coin flips] independent
$N \sim \text{Poi}(\lambda)$

$$X(\omega) = \sum_{k=1..N} Z_k(\omega) \quad \leftarrow \text{ \# of heads in first } N \text{ flips}$$

$$Y(\omega) = N - X \quad \leftarrow \text{ \# of tails in first } N \text{ flips}$$

$\lambda = 10$, $p = \frac{1}{2}$ $\Rightarrow$ $\mathbb{E}X = 5$, $\mathbb{E}Y = 5$

In general $\quad \mathbb{E}X = \lambda p$

$\rightarrow$ $\mathbb{E}[Y \mid X = 100] \overset{?}{=} 100$ for fair coin Way off !
$= 5!$

Because X, Y independent so $\mathbb{E}[Y \mid X=100] = \mathbb{E}[Y] = 5$

Doing the computation

$$\begin{aligned}
\mu_X(k) = P[X = k] &= \sum_{n \geq k} P[X = k, N = n] \\
&= \sum_{n \geq k} P[X = k \mid N = n] P[N = n] \\
&= \sum_{n \geq k} P[X = k \mid N = n] \, e^{-\lambda} \frac{\lambda^n}{n} \\
&= \sum_{n \geq k} \binom{n}{k} p^k q^{n-k} \, e^{-\lambda} \frac{\lambda^n}{n} \qquad (q = 1-p)
\end{aligned}$$

$\vdots$

$\Downarrow$

$X \sim \text{Poi}(\lambda p)$
$Y \sim \text{Poi}(\lambda q)$

$$\begin{aligned}
P[X = k, Y = j] &= \sum_{n \geq k+j} P[X = k, Y = j \mid N = n] P[N = n] \\
&= \sum_{n \geq k+j} P[X = k, Y = j \mid N = k+j] P[N = k+j] \\
&= P[X = k \mid N = k+j] P[N = k+j] \\
&= P[X = k \mid N = k+j] P[N = k+j] \\
&= \binom{k+j}{k} p^k q^j \quad e^{-\lambda} \frac{\lambda^{k+j}}{k+j} \\
&\vdots \\
&= \mu_X(k) \cdot \mu_Y(j)
\end{aligned}$$

$N \mid X = k \quad \sim \quad k + \text{Poi}(*)$

Lec 18 Continuous Prob

* What still applies

$X : \Omega \to \mathbb{R}$ for $(\Omega, F, \mathbb{R})$

$\mu_X(B) = P[X \in B]$ $B \subseteq \mathbb{R}$

more specifically $B \in \mathcal{B} = \sigma(\text{intervals})$

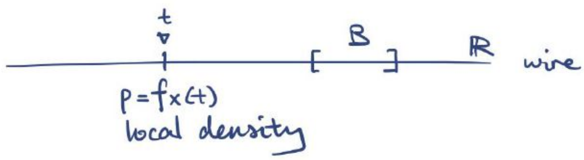
practically $\mathcal{B} \cong \mathcal{P}(\mathbb{R})$

we shall assume this for now

<u>Def</u> X is absolutely continuous $\Leftrightarrow \exists f_X(t) \geq 0$, $f_X : \mathbb{R} \to [0, \infty)$,

$\forall B \in \mathcal{B}$, $P[X \in B] = \int_B f_X(t)\, dt$

<u>Analogy</u>



mass of $B = \int_B f_X(t)\, dt$

<u>Def</u> such f_X is the <u>prob density func</u> of X

$$P[X \in B] \overset{\text{discrete}}{=} \sum_{\substack{x \in \mathrm{Im}X \\ x \in B}} \underbrace{P[X = x]}_{p(x)} = \sum_{x \in \mathrm{Im}X} 1_B(x) \cdot p(x)$$

$$\overset{\text{continuous}}{=} \quad = \int_{\mathbb{R}} 1_B(x)\, f_X(x)\, dx = \overset{\text{notation}}{\int_{\mathbb{R}} 1_B(x)\, \mu_X(dx)}$$

$$\mathbb{E}[X] = \int_{\mathbb{R}} x \cdot f_X(x)\, dx$$

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}} g(t) \cdot f_X(t)\, dt$$

Properties of f_X: Typically f_X needs to be stepwise cont.

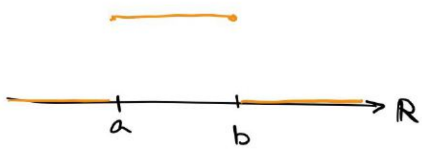
1. $f_X(t) \geq 0$
2. $\int_{\mathbb{R}} f_X(t)\, dt = P[X \in \mathbb{R}] = 1$ (assuming X is real ... sometimes $X \in [-\infty, \infty]$ then this breaks)
3. $\int_B f_X(t)\, dt$ has to be well defined for all B

$$\operatorname{var} X = \mathbb{E}[(X-\mathbb{E}X)^2] = \mathbb{E}(X^2) - (\mathbb{E}X)^2$$

$$= \int_{\mathbb{R}} x^2 f_X(x)\, dx - (\mathbb{E}X)^2$$

Distributions

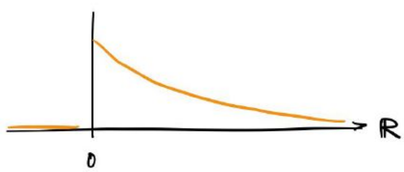
$$X \sim \text{Uniform}([a,b]) \iff f_X(x) = \frac{1}{b-a} 1_{\{x \in [a,b]\}}$$



$$\mathbb{E}X = \frac{b+a}{2} \quad \left(= \int_a^b t \cdot \frac{1}{b-a}\, dt = \frac{1}{b-a} \frac{b^2-a^2}{2} = \frac{b+a}{2} \right)$$

analogue to geometric

$$X \sim \exp(\lambda) \iff f_X(x) = \lambda e^{-\lambda t} \quad 1_{\{x \in [0,\infty)\}}$$



$$\mathbb{E}X = \frac{1}{\lambda} \quad \left(= \int_0^\infty t\, \lambda e^{-\lambda t}\, dt = \cdots \quad \text{ouch by part} \right)$$

$$= \left[t \cdot -e^{-\lambda t} \right]_0^\infty - \int 1 \cdot -e^{-\lambda t}\, dt$$

$$= 0 - -\frac{1}{\lambda}$$

$$\mathbb{E}(X^2) = \int_0^\infty t^2 \lambda e^{-\lambda t}\, dt = \cdots = \text{"whatever that is"}$$

standard normal

$$X = N(0,1) \Rightarrow f_X(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$$

$$\int_{\mathbb{R}} e^{-t^2/2}\, dt =: C$$

$$C^2 = \left(\int_{\mathbb{R}} e^{-x^2/2}\, dx \right) \left(\int_{\mathbb{R}} e^{-y^2/2}\, dy \right)$$

radius of circle!

$$= \int_{\mathbb{R}} \int_{\mathbb{R}} e^{-(x^2+y^2)/2} \, dx \, dy$$

$$= \int_0^{2\pi} \int_0^{\infty} r e^{-r^2/2} \, dr \, d\theta$$

$$= \int_0^{2\pi} 1 \, d\theta$$

$$= 2\pi$$

Lec 19

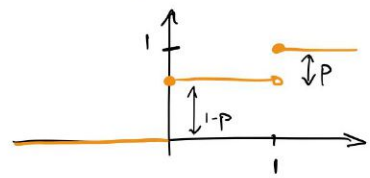
Cumulative distribution function

Let X be RV $\in \mathbb{R}$

$F_X(t) = P[X \leq t]$

↑ CDF

$X \sim B(p)$



Properties of $F_X : \mathbb{R} \to [0,1]$

1. Monotone increasing

2. $\lim_{t\to\infty} F_X(t) = \lim_{t\to\infty} P[X \leq t] = 1$

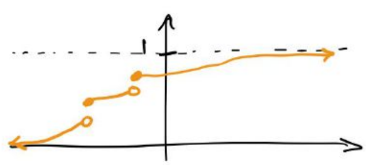
$$= \lim_{t\to\infty} P[\{X \leq t\}]$$

$$= \lim_{t\to\infty} P[\bigcup_t \{X \leq t\}] \quad \text{monotone continuity!}$$

$$= P[\Omega]$$

$$= 1$$

Generally:



3. $\lim_{t\to -\infty} F_X(t) = 0$

4. F_X is right continuous

Thm μ_X and F_X uniquely determine each other

If X is absolutely cont. with density f_X,

$$\frac{d}{dt} F_X(t) = \frac{d}{dt} P[X \leq t] = \frac{d}{dt} \int_{-\infty}^{t} f_X(x)\, dx \underset{\uparrow}{=} f_X(t)$$

fundamental thm of calculus

$\boxed{\text{Lec 20}}$ Continuous Joint Dist

Joint Dist

$\Omega \xrightarrow{X} \mathbb{R}$ $\quad X$ abs. cont. $\Leftrightarrow \exists f_X(x)\,,\ P[X\in B] = \int_B f_X(x)\,dx$

$\quad \llcorner\ f_X \geq 0$

$\quad \llcorner\ \int_{\mathbb{R}} f_X(x)\,dx = 1$

Consider:

$\Omega \xrightarrow{\vec{X}} \mathbb{R}^n$ $\quad \vec{X}$ abs. cont. $\Leftrightarrow \exists f_{\vec{X}} : \vec{X} \to \mathbb{R}\,,\ P[\vec{X}\in B] = \int_B f_{\vec{X}}(\vec{x})\,d\vec{x}$

$\vec{X} = (X_1, \dots, X_n)$ $\quad f_{\vec{X}}$

$\quad \llcorner\ f_{\vec{X}} \geq 0$

$\quad \llcorner\ \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f_{\vec{X}}(\vec{x})\,dx_1 \cdots dx_n$

$\quad = \int_{\mathbb{R}^n} f_{\vec{X}}(\vec{x})\,d_x \ \ d\vec{x}$

$$\begin{array}{ccccc}
\Omega & \xrightarrow{X} & \mathbb{R} & \xrightarrow{\varphi} & \mathbb{R} \\
\vdots & & \vdots & & \\
P & & \mu_X = P\circ X^{-1} & & \\
\vdots & & \vdots & & \\
\mathbb{E}[\cdot] = \int_{\Omega} \cdot\, dP & & \int_{\mathbb{R}} \cdot\ \mu_X(dx) & &
\end{array}$$

$$\mathbb{E}[\varphi(X)] = \int_{\Omega} \varphi\circ X(\omega)\cdot P(d\omega)$$

$$= \int_{\mathbb{R}} \varphi(x)\ \mu_X(dx)$$

If $X \sim f_X$ (X abs. cont.) $\Rightarrow \mathbb{E}[\varphi(X)] = \int_{\mathbb{R}} \varphi(x)\, f_X(x)\,dx$

Analoguous to ...

$$\begin{array}{ccccc}
\Omega & \xrightarrow{\vec{X}} & \mathbb{R}^n & \xrightarrow{\varphi} & \mathbb{R} \\
\vdots & & \vdots & & \\
P & & \mu_{\vec{X}} = P\circ \vec{X}^{-1} & &
\end{array}$$

full generality

$$\mathbb{E}[\varphi(\vec{X})] = \int_{\mathbb{R}^n} \varphi(\vec{x})\ \mu_{\vec{X}}(d\vec{x})$$

If $\vec{X} \sim f_{\vec{X}}$ ($\vec{X}$ abs. cont.) : $\quad \mathbb{E}[\varphi(\vec{X})] = \int_{\mathbb{R}^n} \varphi(\vec{x})\, f_{\vec{X}}(\vec{x})\,d\vec{x}$

$$= \int_{\mathbb{R}} dx_1 \cdots \int_{\mathbb{R}} dx_n\ \varphi(x_1,\dots,x_n)\ f_{\vec{X}}(x_1,\dots,x_n)$$

Ex. $\varphi = 1_B$, $B \subseteq \mathbb{R}^n$

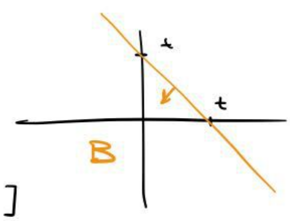
$$E[\varphi(\vec{X})] = P[\vec{X} \in B] = \mu_{\vec{X}}(B) \quad \text{(indicator way)}$$
$$= \int_{\mathbb{R}^n} 1_B(\vec{x})\, f_{\vec{X}}(\vec{x})\, d\vec{x} \quad \text{(transformation formula)}$$
$$= \int_B f_{\vec{X}}(\vec{x})\, d\vec{x}$$

Ex. $X, Y \sim f_{X,Y}$ given

$S := X + Y$

Q: if S abs. cont. ?

Examine cumm. dist $F_S(t) = P[S \leq t]$
then $f_S = \frac{d}{dt} F_S(t)$



want $\frac{d}{dt} P[X + Y \leq t]$

$$= \frac{d}{dt} P[(X,Y) \in B = \{(x,y) \in \mathbb{R} \mid x+y \leq t\}]$$
$$= \frac{d}{dt} \int_{\mathbb{R}} \underbrace{\int_{-\infty}^{t-x} f_{X,Y}(x,y)\, dy}_{\text{sth depending on } x \text{ and } t}\, dx$$
$$= \frac{d}{dt} \int_{\mathbb{R}} g_t(x)\, dx$$
$$= \int_{\mathbb{R}} dx\, \frac{d}{dt} \int_{-\infty}^{t-x} f_{X,Y}(x,y)\, dy$$
$$= \int_{\mathbb{R}} dx\, \underline{\frac{d}{dt} G(t-x)} \qquad G(s) := \int_{-\infty}^{s} f_{X,Y}(x,y)\, dy$$
$$= \int_{\mathbb{R}} dx\, G'(t-x)\, \cancel{\frac{d}{dt}(t-x)}^{\,1} \qquad G'(s) = f_{X,Y}(x,s)$$
$$= \int_{\mathbb{R}} f_{X,Y}(x, t-x)\, dx$$
$$= f_{X+Y}(t)$$

Fact X, Y indep. & abs. cont. w f_X, f_Y $\Leftrightarrow$ $f_{X,Y}(x,y) = f_X(x)\, f_Y(y)$

Also $\Rightarrow$ $f_{X+Y}(t) = \int f_X(x)\, f_Y(t-x)\, dx = \overbrace{(f_X * f_Y)}^{\text{"convolution"}}(t)$

Ex. (X,Y) abs. cont. w $f_{X,Y}$

Q: can we get X w f_X ?

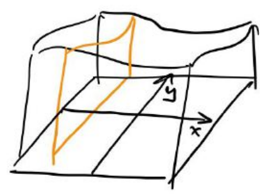
$$\frac{d}{dt} F_X(t) = \frac{d}{dt} P[X \le t] = \frac{d}{dt} P[(X,Y) \in B = \{(x,y) \in \mathbb{R} \mid x \le t\}]$$

$$= \frac{d}{dt} \int_{\mathbb{R}} dy \int_{-\infty}^{t} dx \quad f_{X,Y}(x,y)$$

$$= \int_{\mathbb{R}} dy \quad \frac{d}{dt} \int_{-\infty}^{t} dx \quad f_{X,Y}(x,y)$$

$$= \int_{\mathbb{R}} dy \quad f_{X,Y}(t,y)$$ ← integrate over line

$$= f_X(t)$$



$\boxed{\text{Lec 21}}$

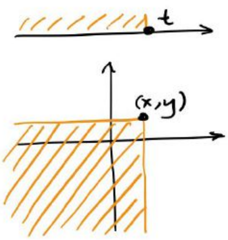
Recall joint cont. dist.

$$\mu_{X,Y}(B) = \iint_B f_{X,Y}(x,y)\,dx\,dy$$

$$\mathbb{E}[g(x,y)] = \iint_B g(x,y)\, f_{X,Y}(x,y)\,dx\,dy$$

Finding f_X

$$f_X(t) = \frac{d}{dt}F_X(t) = \frac{d}{dt}\mathbb{P}[X \le t]$$



$$f_{X,Y}(x,y) = \frac{\partial}{\partial x}\frac{\partial}{\partial y}\mathbb{P}[X \le x,\ Y \le y]$$

$$= \frac{\partial}{\partial x}\frac{\partial}{\partial y}\int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(u,v)\,du\,dv$$

Conditional Dist

Under $\mathbb{P}[\,\cdot \mid Y = y]$, there's cond. density $f_{X|Y=y}(\cdot)$

$= f_{X|Y}(\cdot \mid y)$

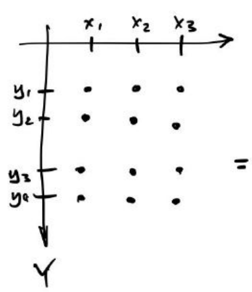
$$\mathbb{P}[X \in B \mid Y = y] = \int_B f_{X|Y}(x,y)\,dx$$

$$\mathbb{E}[g(Y) \mid X = t] = \int_{\mathbb{R}} g(y)\, f_{Y|X}(t,y)\,dy$$

$$\mathbb{E}[g(X,Y) \mid Y = s] = \mathbb{E}[g(X,s) \mid Y = s]$$

$$= \int_{\mathbb{R}} g(x,s)\, f_{X|Y}(x,s)\,dx$$

In discrete



$$\mu_{X|Y}(x,y) = \frac{\mu_{X,Y}(x,y)}{\mu_Y(Y)}$$

Cont

<u>Def</u>

$$f_{X|Y}(x,y) = \begin{cases} \dfrac{f_{X,Y}(x,y)}{f_Y(y)} & \text{if } f_Y(y) > 0 \\ \left.\begin{cases}\text{Not defined} \leftarrow \text{formal} \\ 0 \leftarrow \text{practical} \\ \text{Whatever you like} \leftarrow \text{doesn't matter}\end{cases}\right\} & \text{if } f_Y(y) = 0 \end{cases}$$

<u>Fact</u> if $Y \sim a.c.$ $\quad \forall a, \ P[Y=a] = \int_a^a \text{whatever} = 0$

So... $\mathbb{P}[X \in B \mid Y=t] = \dfrac{\mathbb{P}[X\in B, Y=t] \leftarrow 0}{\mathbb{P}[Y=t] \leftarrow 0}$:C

\# Checking other things

$\mathbb{P}[X \in B] \stackrel{?}{=} \int f_Y(y)\, \mathbb{P}[X\in B \mid Y=y]\, dy$ | assume $f_Y > 0$

$$\text{RHS} = \int_{\mathbb{R}} f_Y(y) \int_B f_{X|Y}(x,y)\, dx\, dy$$

$$= \int_{\mathbb{R}} \cancel{f_Y(y)} \int_B \frac{f_{X,Y}(x,y)}{\cancel{f_Y(y)}}\, dx\, dy$$

$$= \int_{\mathbb{R}} \int_B f_{X,Y}(x,y)\, dx\, dy$$

$$= \int_{\mathbb{R}} \int_{\mathbb{R}} 1_B(x)\, f_{X,Y}(x,y)\, dx\, dy$$

$$= \mathbb{E}[1_B(X)]$$

$$= \mathbb{P}[X\in B]$$

So conditioning still works!

<u>Thm</u> Let $X \geq 0$ $\quad \mathbb{E}[X] \stackrel{\text{total generality}}{=} \int_0^\infty \mathbb{P}[X>t]\, dt \quad \left(= \int_0^\infty \mathbb{P}[X\geq t]\, dt \right)$

$$\text{RHS} = \int_0^\infty \mathbb{E}[\,1_{\{X>t\}}(\omega)\,]\, dt$$

$$= \int_0^\infty \int_\Omega \mathbb{P}[d\omega]\, 1_{\{X>t\}}(\omega)\, dt$$

$$= \int_\Omega \int_0^\infty \mathbb{P}[d\omega]\, 1_{\{X>t\}}(\omega)\, dt \longleftarrow \text{integrating non neg stuff}$$

$$= \int_\Omega \mathbb{P}[d\omega] \int_0^\infty 1_{\{t\in(-\infty, X(\omega))\}}(t)\, dt$$

$$= \int_\Omega \mathbb{P}[d\omega]\, X(\omega)$$

$$= \mathbb{E}[X]$$

Lec 22

\# Some shortcut for transformation

$(X_1, X_2) \sim f_{X_1,X_2}$

$$\vec{\Phi} : \underset{G}{\mathbb{R}^2} \longrightarrow \underset{D}{\mathbb{R}^2}, \quad G \mapsto D \quad (G \subseteq \mathbb{R}^2,\ D \subseteq \mathbb{R}^2)$$

$$\vec{\Phi}(X_1(\omega), X_2(\omega)) = (U_1(\omega), U_2(\omega))$$

Assume:

1. Φ bijective
2. Φ differentiable $\Leftrightarrow$ locally linear, approximate with plane, representable by matrix — in fact Jacobian matrix

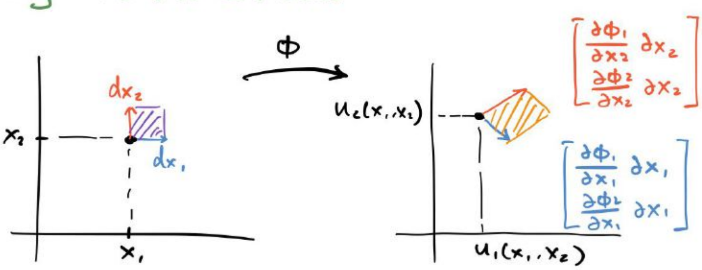
$$\phi = \begin{bmatrix} \phi_1(x_1, x_2) \\ \phi_2(x_1, x_2) \end{bmatrix}$$

$$\text{jacobian} = \begin{bmatrix} \frac{\partial \phi_1}{\partial x_1} & \frac{\partial \phi_2}{\partial x_1} \\ \frac{\partial \phi_1}{\partial x_2} & \frac{\partial \phi_2}{\partial x_2} \end{bmatrix} \Bigg|_{(x_1, x_2)} = \text{local derivative}$$

Want f_{U_1,U_2}

→ If we want $F_{X,Y}$ ... double integral & differentiate :C

→ Try matrix calculus



$$\mathbb{P}[(X_1, X_2) \in \square] = \mathbb{P}[(U_1, U_2) \in \lozenge]$$

area so small, density doesn't change

$$f_{X_1,X_2}(x_1, x_2) \cdot \text{area}(\square) = f_{U_1,U_2}(u_1, u_2) \cdot \underbrace{\text{area}(\lozenge)}_{\text{use determinant}}$$

$$\Rightarrow f_{U_1,U_2}(u_1, u_2) = f_{X_1,X_2}(x_1(u_1,u_2), x_2(u_1,u_2)) \cdot \frac{1}{|\det D\phi(x_1(u_1,u_2), x_2(u_1,u_2))|}$$

$$= f_{X_1,X_2}(x_1(u_1,u_2), x_2(u_1,u_2)) \cdot |\det D\phi^{-1}(u_1(x_1,x_2), u_2(x_1,x_2))|$$

$$= f_{X_1,X_2}(\Phi^{-1}(u_1,u_2)) \cdot |\det D\Phi^{-1}(u_1,u_2)|$$

Ex. (X, Y) iid $N(0,1)$ $\xrightarrow{\text{polar}}$ $R = \sqrt{X^2+Y^2} \quad \in [0,\infty)$
with $f_{X,Y}$ $\quad\quad\quad\quad\quad \Theta = \tan^{-1}(Y/X) \quad \in [0,2\pi)$

$$\Phi : \mathbb{R}^2 \longleftrightarrow [0,\infty) \times [0, 2\pi)$$

$$\Phi(x,y) = \begin{bmatrix} \sqrt{x^2+y^2} \\ \tan^{-1}(y/x) \end{bmatrix} \qquad \Phi^{-1}(\theta, r) = \begin{bmatrix} r\cos\theta \\ r\sin\theta \end{bmatrix}$$

$$\det D\Phi^{-1} = \det \begin{bmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{bmatrix} = r(\cos^2\theta + \sin^2\theta) = r$$

$$f_{R,\Theta}(r,\theta) = f_{X,Y}(x(\theta,r), y(\theta,r)) \cdot |\det D\Phi^{-1}|$$

$$= \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)} \cdot r$$

$$= \frac{1}{2\pi} e^{-\frac{1}{2}(r^2\cos^2\theta + r^2\sin^2\theta)} \, r$$

$$= \frac{r}{2\pi} e^{-\frac{1}{2}r^2}$$ ← independent of θ. rotationally invariant

Clean up:

$$f_{R,\Theta}(r,\theta) = 1_{[0,\infty)}(r) \, 1_{[0,2\pi)}(\theta) \, \frac{r}{2\pi} e^{-\frac{1}{2}r^2}$$

Turns out here R and Θ independent.

Proof:

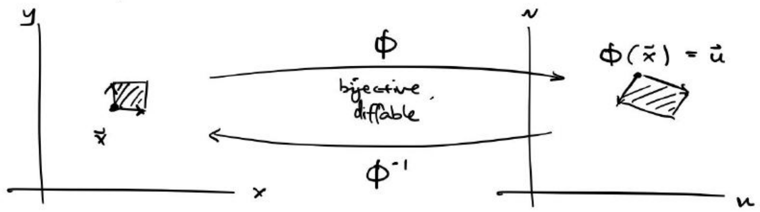
$$f_{R,\Theta}(r,\theta) = 1_{[0,\infty)}(r) \, 1_{[0,2\pi)}(\theta) \, \frac{r}{2\pi} e^{-\frac{1}{2}r^2}$$

$$= \left[1_{[0,2\pi)}(\theta) \, \frac{1}{2\pi}\right] \cdot \left[1_{[0,\infty)}(r) \, r \, e^{-\frac{1}{2}r^2}\right]$$

$$\overset{\text{check}}{=} f_\Theta(\theta) \, f_R(r)$$

Lec 23

Recall transformation trick



X, Y　　　　　　　　U, V

Realise $\phi(\vec{x} + d\vec{x}) - \phi(\vec{x}) \cong [D\phi]_{\vec{x}} \cdot d\vec{x}$

$\frac{\partial(u,v)}{\partial(x,y)}$, local linearisation of ϕ

$$= \begin{bmatrix} \frac{\partial \phi_1}{\partial x} & \frac{\partial \phi_1}{\partial y} \\ \frac{\partial \phi_2}{\partial x} & \frac{\partial \phi_2}{\partial y} \end{bmatrix}$$

Also $[D(\phi^{-1})]_{\vec{u}} = [(D\phi)^{-1}]_{\vec{x}}$ where $\vec{x} = \phi^{-1}(\vec{u})$

inverse func thm

Shortcut $f_{U,V}(\vec{u}) = \dfrac{1}{|[\det D\phi]_{\vec{x}}|} \cdot f_{X,Y}(\vec{x})$

$= |[\det D(\phi^{-1})]_{\vec{u}}| \cdot f_{X,Y}(\vec{x})$

Multivar normal dist

Single var:
$X \sim \mathcal{N}(0,1)$
$\sigma X + b =: Y \sim \mathcal{N}(b, \sigma^2)$

Multi
$\vec{X} = (X_1, \dots, X_n)$　X_i iid $\sim \mathcal{N}(0,1)$

$\vec{Y} = A \cdot \vec{X}$ ← linearly transformed

so $\phi: \mathbb{R}^n \to \mathbb{R}^n$, $\phi(\vec{x}) = A \cdot \vec{x}$
also $[D\phi]_{\vec{x}} = A \quad \forall \vec{x}$ since ϕ already linear

$$f_{\vec{Y}}(\vec{y}) = \frac{1}{|\det A|} f_{\vec{X}}(\vec{x})$$

$$= \frac{1}{|\det A|} \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}(x_1^2 + \cdots + x_n^2)} \quad \text{where} \quad x_i = \#i\ \phi^{-1}(\vec{y})$$

$$= \frac{1}{|\det A|} \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}(\vec{x}^T \cdot \vec{x})}$$

$$= \frac{1}{|\det A|} \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}((A^{-1} \cdot \vec{y})^T \cdot (A^{-1} \cdot \vec{y}))}$$

$$= \frac{1}{|\det A|} \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2} \vec{y}^T (A^{-1})^T A^{-1} \vec{y}}$$

$$= \frac{1}{|\det A|} \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2} \vec{y}^T (A^T)^{-1} A^{-1} \vec{y}}$$

$$= \frac{1}{|\det A|} \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2} \vec{y}^T (\underbrace{A A^T}) ^{-1} \vec{y}}$$

$\mathbb{E}\vec{Y} = \vec{0} \Leftarrow$ each X_i centered, linear combo of them centered

$$\mathrm{cov}(y_k | y_j) = \mathrm{cov}\Big(\sum_l A_{kl} X_l \,\Big|\, \sum_i A_{ji} X_i\Big)$$

$$= \sum_{l,i} A_{kl} A_{ji} \underbrace{\mathrm{cov}(X_l | X_i)} = \begin{cases} 1 & l = i \\ 0 & \text{else} \end{cases}$$

$$= \sum_l A_{kl} A_{jl} \; 1$$

$$= \sum_l A_{kl} A^T_{lj} \; 1$$

$$= \underbrace{(AA^T)}_{\text{interesting}}{}_{k,j} \qquad =: C_{k,j} \qquad \text{"covar matrix"}$$

$$C = AA^T$$

$$\det C = \det A \; \det A^T$$

$$= (\det A)^2 \qquad > 0$$

$$\sqrt{\det C} = |\det A|$$

$$f_{\vec{Y}}(\vec{y}) = \frac{1}{\sqrt{\det C}\,(\sqrt{2\pi})^n}\; e^{-\frac{1}{2}\vec{y}^T C^{-1}\vec{y}}$$

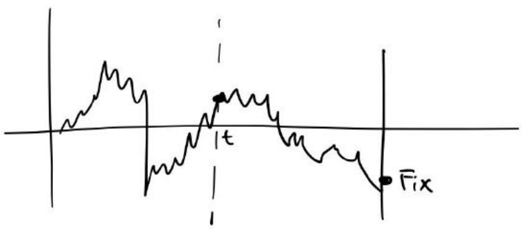
multivariate normal dist

symmetric positive definite matrix

More general one can do $\vec{Y} := A\vec{X} + \vec{b}$

multivar normal RV vector

Browneean motion, conditioned on destination



what's dist of height at time t ?

Lec 24

\# Ex.

symmetric positive definite

(X, Y) joint normal $C = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$, $\rho = \text{cov}(X, Y) = \text{corr}(X, Y)$

Def Correlation $\text{corr}(X, Y) = \dfrac{\text{cov}(X, Y)}{\sqrt{\text{var}X}\sqrt{\text{var}Y}}$ ← normalise, so that corr just captures correlation.

Note cov $\in (-\infty, \infty)$
corr $\in [-1, 1]$

in extreme case, $\text{corr}(X, X) = \dfrac{\text{cov}(X, X)}{\text{var}X} = 1$

$\Rightarrow \rho$ must be in $(-1, 1)$ (safe to assume $\rho \neq -1, \rho \neq 1$?)

Question: what's conditional dist of Y given $X = x$ viz. $f_{Y|X}$

... very messy ... try method ① let $Z, X \sim N(0, 1)$ iid

$Y := \alpha X + \beta Z$, adjust α, β st. $C_{X,Y} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$

$$\begin{aligned} \text{cov}(X, Y) &= \text{cov}(X, \alpha X + \beta Z) \\ &= \alpha \,\text{cov}(X, X) + \beta (X, Z) \\ &= \alpha = \rho \end{aligned}$$

$\text{var}\, Y = \text{var}(\alpha X + \beta Y)$

$1 \stackrel{?}{=} \text{var}(Y) = \text{var}(\rho X + \beta Z) = \rho^2 + \beta^2 \Rightarrow \beta = \sqrt{1 - \rho^2}$

then $Y = \rho X + \sqrt{1 - \rho^2}\, Z$

so
$(X, Y) = N(\vec{0}, C)$

$Y|_{X=x} = \rho x + \sqrt{1-\rho^2}\, \underbrace{Z}_{N(0,1)}$ $\Rightarrow f_{Y|X}(x, y) = \dfrac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} e^{-\frac{1}{2}\left(\frac{y - \rho x}{\sqrt{1-\rho^2}}\right)^2}$

$\sim N(\rho x, 1 - \rho^2)$

method ②

$$C^{-1} = \frac{1}{1-\rho^2}\begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix}$$

$$f_{X,Y} = \frac{1}{\sqrt{1-\rho^2}\;2\pi} e^{-\frac{1}{2}\frac{1}{1-\rho^2}\vec{x}^T \begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix}\vec{x}}$$

$$= \quad '' \quad e^{-\frac{1}{2}\frac{1}{1-\rho^2}(x^2+y^2-2\rho xy)}$$

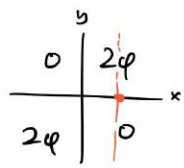
$$f_X(x) = \int_{\mathbb{R}} dy \; \frac{1}{\sqrt{1-\rho^2}\;2\pi} e^{-\frac{1}{2}\frac{1}{1-\rho^2}\left((y-\rho x)^2 + x^2(1-\rho^2)\right)}$$

"it's totally trivial"

⚠ warning

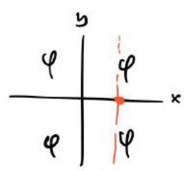
$X_1, \ldots, X_n$ normal $\not\Rightarrow$ they are joint normal

Ex. $\varphi(x,y) = \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)}$

$\psi :=$ 

. Obviously $\int_{\mathbb{R}^2} \psi = 1$

not joint normal

$\psi' :=$ 

but $\int_{|} \psi = \int_{|} \psi'$

joint normal

Lec 25

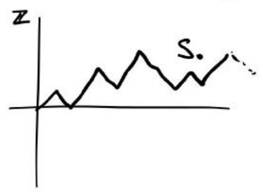
Stochastic process

Def Stoc. process on (Ω, F, P) is a bunch of random outcomes from same space, like time evolution

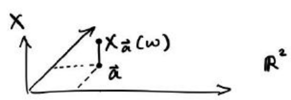
$$(X_\alpha(\omega))_{\alpha \geq I} \qquad \text{typically } I = \mathbb{N},\ \mathbb{R}^+$$

Ex. 1. random walk $(S_k(\omega))_{k\geq 0}$ $\quad S_n(\omega) = \sum_{k\leq n} X_k(\omega)$ $\quad X_k \sim$ Bernoulli (0.5)

$S_\cdot(\omega)$ ← single path, 1 realisation of the process



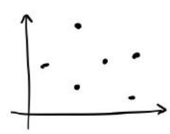
2. $I = \mathbb{R}^2$ $\quad (X_{\vec{a}}(\omega))_{\vec{a}\in\mathbb{R}^2}$ ← random landscape



If all $X_{\vec{a}}$ independent, we get noise

If we want object-like surface, something more clever.

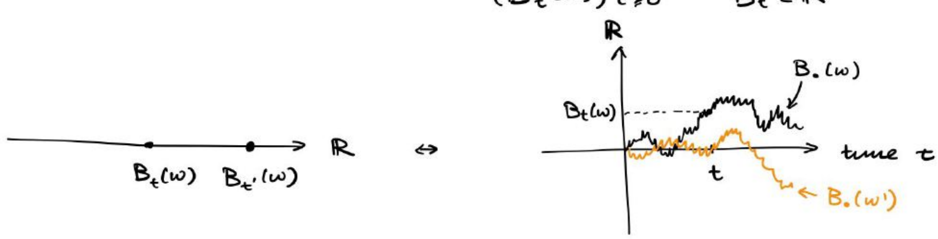
- Point process — make most $X_{\vec{a}}$ zero, and get sparse dots



← like flower at rand places

Brownian motion (BM)

$$(B_t(\omega))_{t\geq 0} \qquad B_t \in \mathbb{R}$$



Def B is a BM iff 1. $B_0 = 0$

2. $\forall n$, $t_0 < t_1 < t_2, \dots, t_n$

"first incr"

Increments of the process → $\underbrace{(B_{t_1} - B_{t_0})}_{D_1}, \underbrace{(B_{t_2} - B_{t_1})}_{D_2}, \dots, \underbrace{(B_{t_n} - B_{t_{n-1}})}_{D_n}$

all independent

and

variance = time difference

$\forall k$, $D_k \sim N(0, t_k - t_{k-1})$

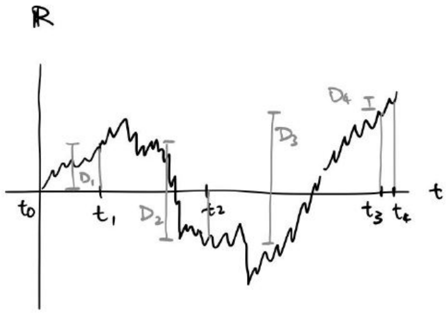
3. $\forall \omega$, $B_\cdot(\omega): t \mapsto B_t(\omega)$ is continuous

Notice $B_{t_0} = 0$ so $D_1 = B_{t_1} \sim N(0, t_1)$

then $B_{t_1+t_2} = N(0, t_1+t_2)$

$B_{t_1} + (B_{t_2} - B_{t_1})$

so across time we retain normal dist



viz. future std scales with $\sqrt{\text{time}}$

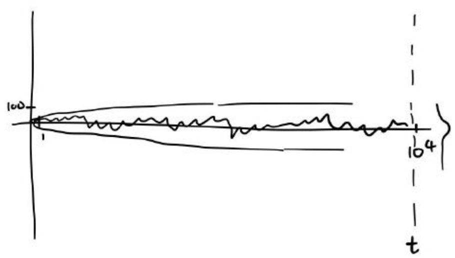
variance scales with time

One construction:

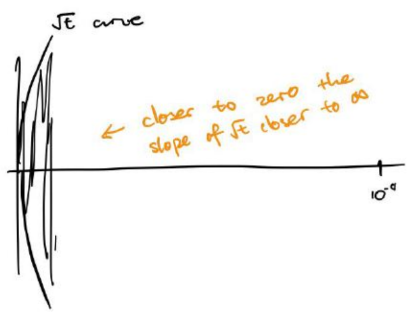
$X_k(\omega) \sin(kx)$ ← take infinite fourier series with random coefficient

└ $N(0,1)$

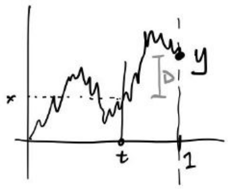
Large scale



at t $\sim N(0,10)$



Conditioned BM?



Force $B_\cdot$ to arrive at y at $t=1$ $\quad \mathbb{P}[\cdot \mid B_1 = y]$

What's $B_t \sim ?$ under $B_1 = y$?

$$f_{B_t|B_1}(x,y) = \frac{f_{B_t,B_1}(x,y)}{f_{B_1}(y)}$$

(Bayes trick) $$= \frac{f_{B_1|B_t}(x,y)\cdot f_{B_t}(x)}{f_{B_1}(y)}$$

But $B_1 = B_t + D$

$= x + D \sim N(x, 1-t)$

Notation $\varphi_t(x) := \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}}$

variance

$$= \frac{\varphi_{1-t}(y-x)\,\varphi_t(x)}{\varphi_1(y)}$$

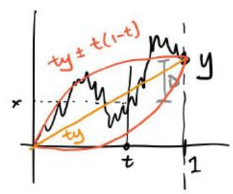
$= \cdots$

$$= \frac{1}{\sqrt{2\pi\, t(1-t)}} e^{-\frac{1}{2}\frac{1}{t(1-t)}(x-ty)^2}$$

$\sim N(ty,\ t(1-t))$

So $\mathbb{E}[B_t \mid B_1 = y] = ty$

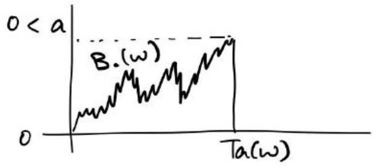
$\mathrm{var}[B_t \mid B_1 = y] = t(1-t)$



"brownian bridge"

Lec 26

Functional of BM



$T_a \in (0, \infty]$

↑ time until hitting a

Thm $\mathbb{P}[T_a < \infty] = 1$

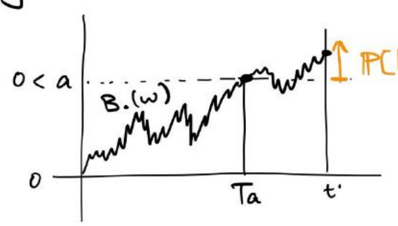
$\mathbb{E}[T_a] \overset{?!}{=} \infty$ ← if ∞ ever show up in weighted avg ... boom
prob of large Ta doesn't decay fast enough

⇢ maybe median more reasonable here

Question: $T_a \sim ?$

$\mathbb{P}[T_a \leq t] =$ hopeless

Try:



$\mathbb{P}[\text{landing here}] = \frac{1}{2}$

↖ strong markov property

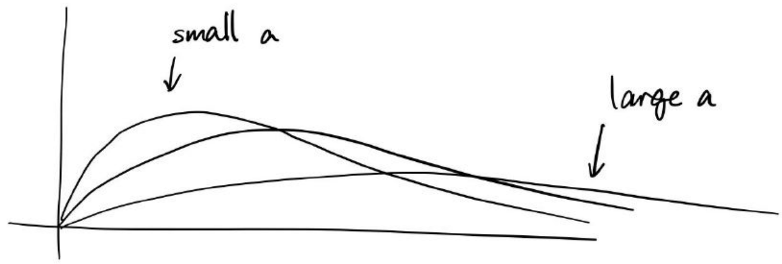
$$\mathbb{P}[B_t > a] = \mathbb{P}[B_t > a, \ T_a \leq t]$$
$$= \mathbb{P}[B_t > a \mid T_a \leq t] \ \mathbb{P}[T_a \leq t]$$
$$= \frac{1}{2} \mathbb{P}[T_a \leq t]$$

$$\frac{d}{dt} \mathbb{P}[T_a \leq t] = \frac{d}{dt} 2 \mathbb{P}[B_t > a]$$
$$= \frac{d}{dt} 2 \int_a^\infty \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}} dx$$
$$= \frac{d}{dt} 2 \int_{\frac{a}{\sqrt{t}}}^\infty \frac{1}{\sqrt{t}} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \sqrt{t} \, du$$

$u = \frac{x}{\sqrt{t}}$ $\quad du = \frac{1}{\sqrt{t}}$

$\frac{a}{\sqrt{t}} \leq u \leq \infty$

$$= \frac{d}{dt} 2 \left(1 - \Phi\left(\frac{a}{\sqrt{t}}\right)\right)$$
$$= 2\left(-\Phi'\left(\frac{a}{\sqrt{t}}\right)\right)\left(\frac{-a}{t^{3/2}}\right)\frac{1}{2}$$
$$f_{T_a}(t) = 1_{(0,\infty]}(t) \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}a^2\frac{1}{t}} \frac{a}{t^{3/2}}$$



Markov property: if past doesn't influence future
viz. future depend only on present

Central Limit Thm (CLT)

Thm Given $(X_k)_{k\geq1}$ i.i.d. with finite 1st and 2nd moment
i.e. $m = \mathbb{E}X_k$ and $\sigma^2 = var(X_k)$ finite

$S_n(\omega) = \sum^n X_k(\omega)$

$\mathbb{E}\, S_n = n\,\mathbb{E}X_k = nm$

$var\, S_n = n \cdot var\, X_k = n\sigma^2$

Interested in $\tilde{S}_n := \frac{S_n - \mathbb{E}S_n}{\sqrt{var\, S_n}}$ ← squeeze them back $m=0$, $\sigma^2=1$

$\tilde{S}_n \sim \mu_n$

$\mathbb{E}\,\tilde{S}_n = 0$

μ_k can be many things as long finite m and σ^2

$var\,\tilde{S}_n = 1$

Then $\lim_{n\to\infty} \mu_n \sim N(0,1)$

In other words:

$$\mathbb{P}[\tilde{S}_n \in [a,b]] \xrightarrow[n\to\infty]{} \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\,dx = \phi(b) - \phi(a)$$

Applications

- Flip biased coin Ber(0.6), $n=100$

$\tilde{S}_n \approx N(0,1)$

$$\mathbb{P}[S_n > 65] = \mathbb{P}\Big[\underbrace{\frac{S_n - \mathbb{E}S_n}{\sqrt{var\,\tilde{S}_n}}}_{N(0,1)} > \underbrace{\frac{65 - \overbrace{\mathbb{E}S_n}^{60}}{\sqrt{var\,\tilde{S}_n}}}_{\approx 1.02}\Big]$$

$$\approx 1 - \phi(1.02)$$

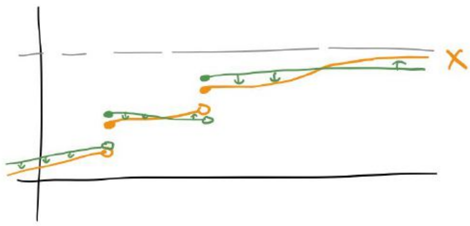
$\boxed{\text{Lec 27}}$ Convergence in distribution

General Framework

Given X , X_n

distribution converge

<u>Def</u> $X_n \xrightarrow[(n\to\infty)]{d} X \iff \forall t$, if $F(t)$ not jump, $F_{X_n}(t) \longrightarrow F_X(t)$



Ex. X : RV $\quad X_n(\omega) = X(\omega) + \frac{1}{n}$

$F_{X_n}(t) \xrightarrow{d} F_X(t)$

$= \mathbb{P}[X_n \le t]$

$= \mathbb{P}[X + \frac{1}{n} \le t]$

$= \mathbb{P}[X \le t - \frac{1}{n}]$

$= F_X(t - \frac{1}{n}) \qquad \lim_{n\to\infty} F_X(t-\frac{1}{n}) = F_X(t^-) \ne F_X(t)$ if $F_X(t)$ jumps

<u>Thm</u> these equivalent

1. $X_n \xrightarrow[n\to\infty]{d} X$

1'. if $X_n \sim \mu_n$, $X \sim \mu$, $\quad \mu_n \xrightarrow[n\to\infty]{d} \mu$

⚠ cannot say $\mathbb{E}X_n$ converge because $\varphi(x) = x$ not bounded

2. $\forall \varphi$, φ cont. and bounded, $\mathbb{E}[\varphi(X_n)] \xrightarrow[n\to\infty]{} \mathbb{E}[\varphi(X)]$

└ finite range bounded top & bottom

2'. $\forall t \in \mathbb{R}$, $\mathbb{E}[e^{itX_n}] \xrightarrow[n\to\infty]{} \underbrace{\mathbb{E}[e^{itX}]}$ $\qquad \varphi(x) = e^{itx} = \cos tx + i \sin tx$

$= \mathbb{E}[\cos tX] + i\,\mathbb{E}[\sin tX]$

Fourier transform of X

$\Phi_X(t) = \mathbb{E}[e^{itX}] \in \mathbb{C}$

__CLT__ X_n iid m, σ finite

$$\tilde{S}_n \xrightarrow{d} N(0,1) \iff \forall t,\ \mathbb{P}[\tilde{S}_n \le t] \xrightarrow[n\to\infty]{} \int_{-\infty}^{t} \phi_1(x)\,dx = \Phi_1(t)$$

$$\mathbb{P}[a \le \tilde{S}_n \le b] = \mathbb{P}[\tilde{S}_n \le b] - \mathbb{P}[\tilde{S}_n \le a] \longrightarrow \Phi_1(b) - \Phi(a)$$

__Fact__ If $X_n, X \in \mathbb{Z}$

$$X_n \xrightarrow[n\to\infty]{d} X \iff \forall k \in \mathbb{Z}\ \ \mathbb{P}[X_n = k] \xrightarrow[n\to\infty]{} \mathbb{P}[X = k]$$

for discrete case, just check jump points

__Law of small numbers__ aka __law of rare events__

Fix n, $X_1, \dots, X_n$ iid $\sim \mathrm{Ber}(\frac{\lambda}{n})$ $\lambda > 0$

$$S_n = \sum_{i=1}^{n} X_i \sim \mathrm{Binom}(n, \tfrac{\lambda}{n}) \quad \text{so } \mathbb{E} S_n = \lambda$$

Question $S_n \xrightarrow[n\to\infty]{} ?$

$$\mathbb{P}[S_n = k] = \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

$$= \frac{n(n-1)\cdots(n-(k-1))}{k!} \frac{\lambda^k}{n^k} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k}$$

$$= \frac{\lambda^k}{k!} \frac{n}{n} \frac{(n-1)}{n} \cdots \frac{(n-(k-1))}{n} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k}$$

$$\xrightarrow[n\to\infty]{} \frac{\lambda^k}{k!} \frac{1}{1} \frac{1}{1} \cdots \frac{1}{1} e^{-\lambda}\ (1)$$

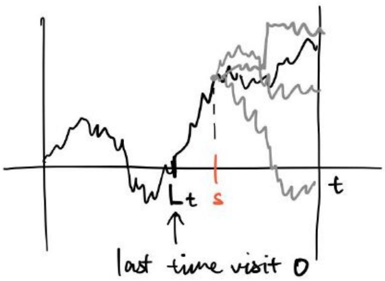
$$= \frac{\lambda^k}{k!} e^{-\lambda}$$

So $B(n, \underbrace{\tfrac{\lambda}{n}}_{\text{rare event}}) \xrightarrow[n\to\infty]{} \mathrm{Poi}(\lambda)$

"poission approx. of binomial"

⚠ In general $\left(1 + \frac{z}{n}\right)^n \xrightarrow[n\to\infty]{} e^z$

HW Hint



$$\mathbb{P}[\ L_t \le s\ ]$$
$$= \mathbb{P}[\ \text{not hit } 0 \text{ btwn } s \text{ and } t\ ]$$
$$=: \mathbb{P}[\ N_{s,t}\ ]$$
$$= \int_{\mathbb{R}} \mathbb{P}[\ N_{s,t} \mid B_s = x\ ]\ f_{B_s}(x)\ dx$$
$$= \int_{\mathbb{R}} \mathbb{P}[\text{not BM hit } -x \text{ within } (t-s) \text{ time}]\ \ f_{B_s}(x)\ dx$$
$$= \int_{\mathbb{R}} \mathbb{P}[\ T_{-x} > t-s\ ]\ f_{B_s}(x)\ dx$$
$$= \int_{\mathbb{R}} \mathbb{P}[\ T_{x} > t-s\ ]\ f_{B_s}(x)\ dx$$

$\boxed{\text{Lec 28}}$

Recall convergence

$X_n \xrightarrow[n\to\infty]{d} X \iff F_{X_n}(t) \xrightarrow[n\to\infty]{} F_X(t)$ for t without jump in F_X

$\iff F_{X_n}(t) \xrightarrow[n\to\infty]{} F_\mu(t) = \mu(-\infty, t]$ $\qquad \mu(B) = \mathbb{P} \circ X^{-1}(B)$

$F_X(t) = \mathbb{P}[X \le t] = \mu(-\infty, t]$

$\mu_n \longrightarrow \mu \iff F_{\mu_n}(t) \longrightarrow F_\mu(t) \quad \forall t$ with $\underbrace{\mu(\{t\}) = 0}_{\text{no jump}}$

Ex. $X_1, \dots, X_n$ iid with $m, \sigma^2 < \infty$

$\tilde{S}_n \xrightarrow{d} N(0,1) \iff \forall t, \ \mathbb{P}[\tilde{S}_n \le t] = \int_{-\infty}^{t} \varphi_1(x)\,dx = \Phi(t)$

CLT with error bound

cubic variance

Given $X_1, \dots, X_n$ iid with finite m, σ^2, $\rho = \mathbb{E}[|X - \mathbb{E}X|^3]$

$\Rightarrow \left| \mathbb{P}[\tilde{S}_n \le t] - \Phi(t) \right| \le \varepsilon_n := \underbrace{\frac{1}{\sqrt{n}}}_{\text{very slow decay}} \frac{\rho}{\sigma^3} \cdot 3$ smaller ρ, smaller error

$\iff \mathbb{P}[\tilde{S}_n \le t] = \Phi(t) \pm \varepsilon_n$

Ex. $X_{1..n}$ iid $\sim \text{Ber}(\tfrac{1}{2}) \qquad m = \frac{1}{2} \quad \sigma^2 = \frac{1}{4} \quad \rho = \frac{1}{8}$

$n := 100$

$\mathbb{P}[S_n > 55] = \mathbb{P}\left[\tilde{S}_n > \frac{50 - 50}{\sqrt{n \cdot \frac{1}{4}}}\right] \pm \varepsilon_n$

$= \mathbb{P}[\tilde{S}_n > 1] \pm \varepsilon_n$

$= 1 - \Phi(1) \pm \frac{1}{\sqrt{100}} \cdot \frac{\frac{1}{8}}{\frac{1}{8}} \cdot 3$

$\approx 1 - 0.84 \pm \frac{3}{10}$

$= 0.16 \pm \underbrace{0.3}_{\text{uh oh}}$

$\in [\cancel{-0.26}\ 0.0, 0.46]$

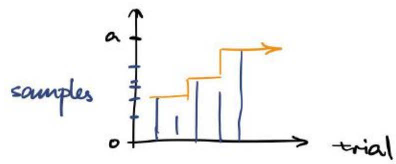
If take $n = 10000$ get 0.16 ± 0.03

$\qquad 1000000 \qquad 0.16 \pm 0.003$

Review prob example

$X_{1..n}$ iid $\sim$ Uniform $[0, a]$

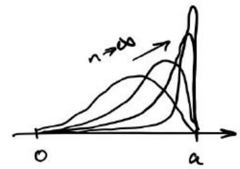
$Z_n = \max(X_1, \dots, X_n)$



$Z_n \xrightarrow[n\to\infty]{d} a$ ie. $(a - Z_n) \xrightarrow[n\to\infty]{} 0$

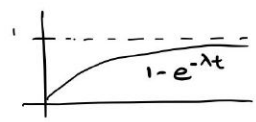
Consider $U_n = n(a - Z_n)$

<u>Claim</u> $U_n \xrightarrow[n\to\infty]{} \exp(\lambda)$



<u>Proof</u> WTS $F_{U_n}(t) \xrightarrow[n\to\infty]{} F_{\exp(\lambda)}(t)$

$$= 1 - e^{-\lambda t}$$



$$\begin{aligned}
F_{U_n}(t) &= \mathbb{P}[\, n(a - Z_n) \le t \,] \\
&= \mathbb{P}\left[\, Z_n \ge -\frac{t}{n} + a \,\right] \\
&= 1 - \mathbb{P}\left[\, Z_n \le a - \frac{t}{n} \right] \\
&= 1 - \mathbb{P}\left[\bigcap_{k=1}^{n} \{ X_k \le a - \frac{t}{n} \} \right] \\
&= 1 - \left(\mathbb{P}\left[X_1 \le a - \frac{t}{n} \right] \right)^n \\
&= 1 - \left(\frac{a - \frac{t}{n}}{a - 0} \right)^n \\
&= 1 - \left(1 - \frac{t}{na} \right)^n \\
&= 1 - e^{-\frac{1}{a} t}
\end{aligned}$$

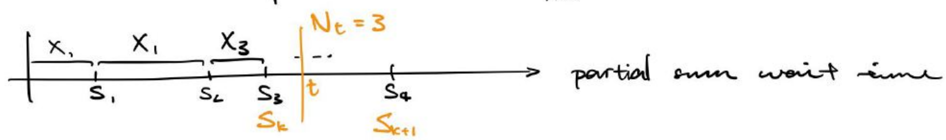
$$\lim_{n\to\infty} \left(1 + \frac{z}{n}\right)^n = e^z$$

So $\lambda = \frac{1}{a}$

$U_n \sim \exp\left(\frac{1}{a}\right)$

Another

$X_{1..n}$ iid $\sim \exp(\lambda)$ $\quad S_n = \sum_{1..n} X_i$



$N_t(\omega) = $ # of points $\le t$ $\qquad N_t \in \mathbb{N}$

<u>Claim</u> $N_t \sim \mathrm{Poi}(\lambda)$

$$\mathbb{P}[N_t = k] = \mathbb{P}[S_k \leqslant t, S_{k+1} > t]$$

$$= \int_{\mathbb{R}^+} \mathbb{P}[S_k \leqslant t, S_{k+1} > t \mid S_k = x] f_{S_k}(x)\,dx$$

$$= \int_{\mathbb{R}^+} \mathbb{P}[x \leqslant t, x + X_{k+1} > t \mid S_k = x] f_{S_k}(x)\,dx$$

$$\boxed{\text{Lec 29}}$$

RV value convergence

X_n, X

<u>Def</u> ⓪ $X_n \xrightarrow[n\to\infty]{} X$ "surely" $\Leftrightarrow \forall \omega,\ X_n(\omega) \xrightarrow[n\to\infty]{} X(\omega)$

↑ makes sense, but not often used

relax

③ $X_n \xrightarrow[n\to\infty]{} X$ "$\mathbb{P}$-almost surely" $\Leftrightarrow \mathbb{P}[\{\omega \mid X_n(\omega) \xrightarrow[n\to\infty]{} X(\omega)\}] = 1$

↑ allow for non-empty non-convergence

$\Leftrightarrow \mathbb{P}[\{\omega \mid |X_n(\omega) - X(\omega)| \xrightarrow[n\to\infty]{} 1\}] = 1$

② $X_n \xrightarrow[n\to\infty]{} X$ "in L^p" $\Leftrightarrow \mathbb{E}[|X_n - X|^p] \xrightarrow[n\to\infty]{} 0$ for fixed $p \geq 1$

e.g. $\mathbb{E}[|X_n - X|] \xrightarrow[n\to\infty]{} 0$ with $p = 1$

① $X_n \xrightarrow[n\to\infty]{} X$ "in probability" $\Leftrightarrow \forall \delta > 0,\ \mathbb{P}[|X_n - X| > \delta] \xrightarrow[n\to\infty]{} 0$

<u>Fact</u>

$$\begin{tikzcd}
③ \arrow[r, shift left, Rightarrow] \arrow[rr, bend right, Rightarrow, "/" marking] & ① \arrow[l, shift left, Rightarrow, "/" marking] \arrow[r, shift right, Rightarrow, "/" marking] & ② \arrow[l, shift right, Rightarrow] \arrow[ll, bend right, Rightarrow, "/" marking]
\end{tikzcd}$$

Ex. $X_n = X + \frac{1}{n}$ $\qquad X_n \xrightarrow{?} X$

① ✓ ① ✓ ⓪ ✓ ③ ✓

<u>Thm</u> weak LLN (wLLN)

$(X_n)_{n \geq k}$ $\quad \mathbb{E}[X_k] \overset{\forall k}{=} m \quad \text{var } X_k \overset{\forall k}{=} \sigma^2 < \infty, \quad \text{cov}(X_i, X_j) = 0$

$\Rightarrow \frac{1}{n}\sum_k^n X_k \xrightarrow[n\to\infty]{} m$ in L^2 (and thus also in prob)

$$\mathbb{E}\left[\left(\frac{1}{n}\sum_k^n X_k - m\right)^2\right] = \mathbb{E}\left[\left(\frac{\sum X_k - nm}{n}\right)^2\right]$$

$$= \mathbb{E}\left[\frac{1}{n^2}\left(\sum(X_k - m)\right)^2\right]$$

$$= \frac{1}{n^2}\mathbb{E}\left[\left(\sum \tilde{X}_k\right)^2\right]$$

$$= \frac{1}{n^2}\text{var}\left[\sum \tilde{X}_k\right]$$

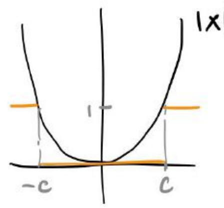
$$= \frac{1}{n^2} \sum \text{var}[X_k] \quad \leftarrow \text{all covariances } 0$$

$$= \frac{1}{n^2} n\sigma^2$$

$$= \frac{\sigma^2}{n}$$

$\square$ L^2 convergence ✓

<u>Chebyshev's Inequality</u>



for $p > 0$, $c > 0$,

$$\mathbb{P}[|X| \geq c] \leq \frac{1}{c^p} \mathbb{E}[|X|^p]$$

Intuition: consider $\varphi(x) = c^p \cdot \mathbb{1}_{--}(x)$

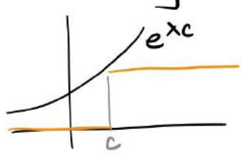
$$\forall x, \quad \varphi(x) \leq |x|^p$$

$$\forall \omega, \quad \varphi(X) \leq |X|^p$$

$$c^p \, \mathbb{P}[|X| \geq c] = \mathbb{E}[\varphi(X)] \leq \mathbb{E}[|X|^p]$$

$$\boxed{\mathbb{P}[|X| \geq c] \leq \frac{\mathbb{E}[|X|^p]}{c^p}}$$

Alternatively ...



$$\mathbb{P}[X \geq c] \leq \frac{1}{e^{\lambda c}} \mathbb{E}[e^{\lambda X}]$$

③ $\Rightarrow$ ① $\quad \mathbb{P}[|X_n - X| > \delta] \leq \frac{1}{\delta^p} \mathbb{E}[|X_n - X|^p] \xrightarrow[n \to \infty]{} 0$

Lec 30

Chebyshev Application

Recall $\mathbb{P}[|X| \geq c] \leq \frac{1}{c^p} \mathbb{E}[|X|^p]$ for $p > 0$

Let X be RV, $\mathbb{E}[|X|] = 0$ ($\Rightarrow$ feels like $X(\omega) \overset{\forall \omega}{=} 0$ or... $\mathbb{P}[X=0] = 1$)

→ Maybe $\forall \omega$, $X(\omega) = 0$? False!

Counterexample: $\Omega = [0,1]$, $\mathbb{P}$ = uniform (length of $B \subseteq [0,1]$)
$X(\omega)$ = if $\omega = 0.5$ then 1 else 0

$$\begin{aligned} \mathbb{E}|X| = \mathbb{E}X &= 1 \cdot \mathbb{P}[X=1] + 0 \cdot \mathbb{P}[X=0] \\ &= \mathbb{P}[\{0.5\}] \\ &= 0 \end{aligned}$$

But $\exists \omega$, $X(\omega) \neq 0$

→ Instead $\mathbb{P}[X=0] = 1 \iff \mathbb{P}[X>0] = 0$

$$\{|X| > 0\} = \bigcup_k \{|X| > \tfrac{1}{k}\}$$

($\Leftarrow$) Trivial

($\Rightarrow$) Let $\omega \in \{|X| > 0\}$,
Pick large enough k s.t. $\frac{1}{k} < |X(\omega)|$
Then $\omega \in$ RHS

Observe $\bigcup_k \{|X| > \frac{1}{k}\}$ is monotone $\{|X| > \frac{1}{k_1}\} \subseteq \{|X| > \frac{1}{k_2}\}$ for $k_2 \geq k_1$

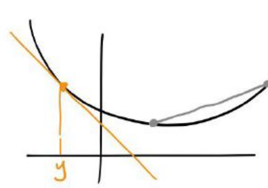
$$\text{Then } \mathbb{P}\left[\bigcup_k \{|X| > \tfrac{1}{k}\}\right] = \lim_{k \to \infty} \underbrace{\mathbb{P}[|X| > \tfrac{1}{k}]}_{\leq \frac{1}{\frac{1}{k}} \mathbb{E}|X| \;=\; 0} = 0$$

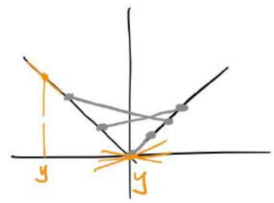
Jensen's Inequality

Let $\varphi(x)$ be convex func, X be RV with finite $\mathbb{E}X$.

Thm $\mathbb{E}[\varphi(X)] \geq \varphi(\mathbb{E}X)$

Def convex func looks like





take any 2 points, connect them, never goes below curve

always cont, diffable at most points

alt: $\forall y$, make line and push up a support line l_y,
- $l_y(x) \le \varphi(x) \quad \forall x$
- $l_y(y) = \varphi(y)$

convex if $\forall y$ we can make such line

Proof (for Thm)

$$\varphi(x) \overset{\forall x}{\geq} l_y(x) \Rightarrow \varphi(X) \overset{\forall w}{\geq} l_y(X)$$

choose $y = \mathbb{E}X$

$$\Rightarrow \mathbb{E}[\varphi(X)] \geq \mathbb{E}[\underbrace{l_y(X)}_{aX+b}] = a\,\mathbb{E}X + b = l_y(\mathbb{E}X)$$
$$= l_y(y)$$
$$= \varphi(y)$$
$$= \varphi(\mathbb{E}X)$$

Moments

with $p \geq 1$, $\mathbb{E}[|X|^p]$ is p^{th} moment

$\mathbb{E}[|X|^p]^{\frac{1}{p}} = \|X\|_p$ is p^{th} norm

$$\mathcal{L}^p := \{X \mid \underbrace{\mathbb{E}[|X|^p] < \infty}_{\|X\|_p < \infty}\}$$

Claim if $1 \leq q < p$ then $\mathcal{L}^q \supseteq \mathcal{L}^p$

i.e. $\mathcal{L}^1 \supseteq \mathcal{L}^{1.5} \supseteq \mathcal{L}^2 \ldots$
so finite higher th moment
$\Rightarrow$ finite lower th moment

Choose $\varphi(x) = |x|^{p/q}$ ← convex, $Y = |X|^q$

Then $\mathbb{E}[(|X|^q)^{p/q}] \geq (\mathbb{E}[|X|^q])^{p/q}$

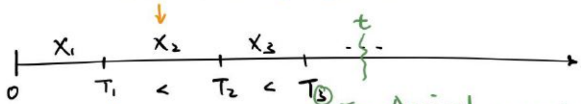
$$\mathbb{E}[|X|^p]^{\frac{1}{p}} \geq \mathbb{E}[|X|^q]^{\frac{1}{q}}$$
$$\|X\|_p \geq \|X\|_q$$

Lec 32 Poission Process

Time intervals

$(X_k)_{k\geq 1}$ iid $\sim \exp(\lambda)$

Inter arrival times



Arrival count before t

Arrival times $(T_k)_{k\geq 1}$ ← not independent!

① $T_n := \sum_{k=1}^{n} X_k$ shown $f_{T_n}(x) = 1_{[0,\infty)}(x)\, \lambda e^{-\lambda x} \frac{(\lambda x)^{n-1}}{(n-1)!}$

② $N_t := \max\left(\max\{n \geq 1 \mid T_n \leq t\},\ 0\right)$

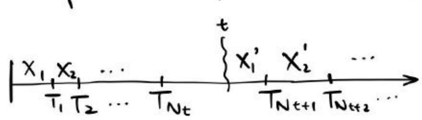
shown $\mathbb{P}[N_t = k] = e^{-\lambda t} \frac{(\lambda t)^k}{k!}$

$= \mathbb{P}[T_k \leq t,\ T_{k+1} > t]$

then condition on T_k

Markov property +

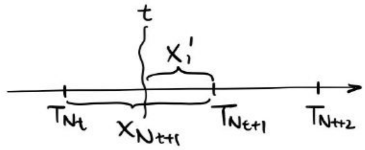
<u>Thm</u> $\forall t$, the process after t is still a poi(λ) process and is independent from what happened before t.



$\forall n$, w.r.t. $\mathbb{P}[\,\cdot \mid N_t = n]$, $(X'_k)_{k\geq 1}$ iid $\sim \exp(\lambda)$

viz. $(X'_k)_{k\geq 1}$ indep of N_t.

<u>Proof</u> Observe $X_2', X_3', \dots$ iid $\sim \exp(\lambda)$



how is $X_1' \sim \exp(\lambda)$?

this actually is not $\sim \exp(\lambda)$, since N_{t+1} depends on other X

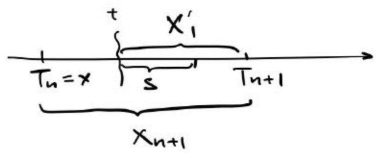
in fact larger than $\exp(\lambda)$

heuristic: more chance to put t in big gap given those gaps

$$\mathbb{P}[X_1' > s \mid N_t = n] = \mathbb{P}[X_1' > s,\ T_n \le t,\ T_{n+1} > s]\ \frac{1}{\mathbb{P}[N_t = n]} \overset{\text{WTS}}{=} e^{-\lambda s}$$

$$= \int_0^\infty dx\ f_{T_n}(x)\ \mathbb{P}[X_1' > s,\ \overset{x}{\cancel{T_n}} \le t,\ T_{n+1} > t \mid T_n = x]$$

$$= \int_0^t dx\ f_{T_n}(x)\ \mathbb{P}[\underbrace{X_1' > s}_{T_{n+1} > t+s},\ T_{n+1} > t \mid T_n = x]$$



$$= \int_0^t dx\ f_{T_n}(x)\ \mathbb{P}[\underbrace{T_{n+1} > t+s}_{X_{n+1} > s + (t-x)} \mid T_n = x]$$

← depends on $\sigma(X_1, \dots, X_n)$

← indep from $X_1, \dots, X_n$

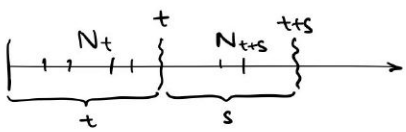
$$= \int_0^t dx\ f_{T_n}(x)\ \mathbb{P}[X_{n+1} > s + t - x]$$

$$= \int_0^t f_{T_n}(x)\ e^{-\lambda(s+t-x)}\ dx$$

$$\mathbb{P}[X_1' > s \mid N_t = n] = \frac{1}{\mathbb{P}[N_t = n]} \int_0^t f_{T_n}(x)\ e^{-\lambda(s+t-x)}\ dx$$

Thm (simplified version)

$(\underbrace{N_t}_{poi(\lambda t)}, (\underbrace{N_{t+s} - N_t}_{poi(\lambda s)}))$ indep



Def Poission process / counting process completely characterised by:

$N_{t_1},\ \underbrace{N_{t_2} - N_{t_1}}_{Poi(\lambda(t_2 - t_1))},\ \underbrace{N_{t_3} - N_{t_2}}_{Poi(\lambda(t_3 - t_2))}, \dots$ indep

Next step WTS from this def we recover exponential interarrival time

Lec 33 Point process (locally finite)

Modelling

Assume finite — finite # of points in finite interval

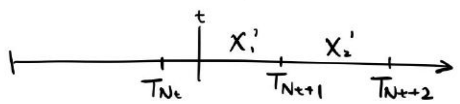
$(N_t) :=$ # points in $(0, t)$
$(T_k) :=$ location of k^{th} point
$X_k :=$ dist btwn points

$$N_{t_1}, \quad N_{t_2} - N_{t_1}, \quad \underbrace{N_{t_3} - N_{t_2}}_{\text{all } \sim \text{Poi}(\lambda(t_k - t_{k-1}))}, \dots \qquad \text{iid} \qquad ①$$

$$\Updownarrow$$

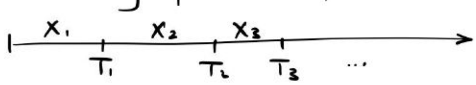
$$(X_k)_{k \geq 1} \quad \text{iid} \sim \exp(\lambda) \qquad ②$$

Recall



w.r.t. $\mathbb{P}[\cdot \mid N_t = n]$, $(X'_k)_{k \geq 1} \sim$ iid $\exp(\lambda)$

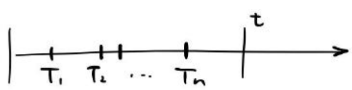
Proof ① counting process def $\Rightarrow$ ② $\exp(\lambda)$ def



$$X_1 \overset{?}{\sim} \exp(\lambda)$$

$$\begin{aligned} \mathbb{P}[X_1 \leq t] &= 1 - \mathbb{P}[X_1 < t] \\ &= 1 - \mathbb{P}[N_t = 0] \\ &= 1 - e^{-\lambda t} \frac{(\lambda t)^0}{0!} \\ &= 1 - e^{-\lambda t} \quad \leftarrow \text{CDF of } \exp(\lambda) \end{aligned}$$

Fact



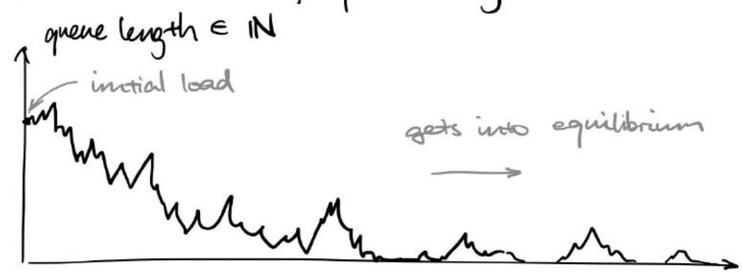
$\mathbb{P}[\cdot \mid N_t = n]$ ← under this, what's dist of $(T_1, \dots, T_n)$?

Make n points $U_{1..n}$ in Unif$[0, t]$ iid, then enumerate them in order $V_1, \dots, V_n$

Queues

random arrival time $\longrightarrow$ single processor (random process time) $\longrightarrow$

Interested in dist of queue length.



<u>Recall</u> $f(x)$ in $o(x)$ as $x \to 0 \iff \frac{f(x)}{x} \xrightarrow[x\to 0]{} 0$

1. $x^2 = f(x)$ is $o(x)$ as $(x \to 0)$
   so $x^2 = o(x)$ as $x \to 0$

2. $f(x) = \alpha x$ is not $o(x)$

3. $0 \le f(x) \le g(x)$ and $g(x)$ is $o(x)$ $\Rightarrow$ $f(x)$ is $o(x)$

Checking differentiability

write $f(t+x) = f(t) + ax + r(x)$
$r(x) := f(t+x) - f(t) - ax$

f diffable at t with deri a $\iff$ $r(x)$ is $o(x)$

Proof $$\frac{r(x)}{x} = \frac{f(t+x) - f(t) - ax}{x} = \frac{f(t+x) - f(t)}{x} - a$$

$$\frac{r(x)}{x} \to 0 \iff \frac{f(t+x) - f(t)}{x} - a \to 0$$

Lec 34

Small-o

Recall f diffable at t with derivative a iff

$$f(t+x) = f(t) + ax + o(x)$$

└ going to 0 faster than x as $x \to 0$

Ex $f(x) = x^{1+\varepsilon} \Rightarrow f(x)$ is $o(x)$

$f(x) = \alpha x$, $\alpha \neq 0 \Rightarrow f(x)$ is not $o(x)$

f, g both $o(x) \Rightarrow f(x) + g(x)$ in $o(x)$

$0 \le f \le g$ and g is $o(x) \Rightarrow f$ is $o(x)$

$o(x) \xrightarrow[x \to 0]{} 0$ (otherwise $\frac{o(x)}{x} \not\to 0$

$e^{-\lambda x} = 1 - \lambda x + o(x)$ (at $t=0$)

$$\lambda x e^{-\lambda x} = \lambda x - \underbrace{\lambda^2 x^2}_{o(x)} + \underbrace{\lambda x \, o(x)}_{o(x)}$$

$$= \lambda x + o(x)$$

Poission Process Characterisation

Let $(N_t)_t$ be point process with indep increments and $\forall t$,

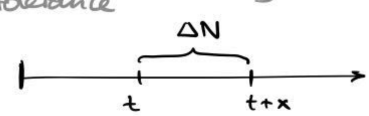
1. $\mathbb{P}[N_{t+x} - N_t = 1] = \mathbb{P}[\Delta N = 1] = \lambda x + o(x)$ as $x \to 0$

   ↖ approx. constant intensity, with small tolerance

2. $\mathbb{P}[\Delta N \geq 2] = o(x)$ ↩ not many points at the same time

   ↙ 1 and 2 implies this

3. $\mathbb{P}[\Delta N = 0] = 1 - \lambda x + o(x)$ ↖ most of the time no point.



<u>Claim</u> $N_t \sim \mathrm{Poi}(\lambda t)$ and (N_t) is $\mathrm{Poi}(\lambda)$ process

Let $k \geq 0$. $\mathbb{P}[N_t = k] =: p_k(t)$ ... can we find $p'_k(t)$?

$$p_k(t+x) = \mathbb{P}[N_{t+x} = k, \overbrace{\Delta N}^{N_{t+x} - N_t} = 0] + \mathbb{P}[N_{t+x} = k, \Delta N = 1] + \mathbb{P}[N_{t+x} = k, \Delta N \geq 2]$$

$$= \mathbb{P}[N_t = k, \Delta N = 0] + \mathbb{P}[N_{t+x} = k, \Delta N = 1] + \mathbb{P}[N_{t+x} = k, \Delta N \geq 2]$$

$$= \underbrace{\mathbb{P}[\Delta N = 0 \mid N_t = k]}_{} \mathbb{P}[N_t = k] \quad \boxed{\begin{array}{l} + \mathbb{P}[N_{t+x} = k,\ \Delta N = 1] \\ + \underbrace{\mathbb{P}[N_{t+x} = k,\ \Delta N \geq 2]}_{0 \leq \quad \leq \mathbb{P}[\Delta N \geq 2]} \end{array}}$$

* if $k = 0$ these don't exist

$o(x)$

$$= \overbrace{(1 - \lambda x + o(x))}\, p_k(t) + \mathbb{P}[N_t = k-1,\ \Delta N = 1] + o(x)$$

$$= (1 - \lambda x + o(x))\, p_k(t) + \mathbb{P}[\Delta N = 1 \mid N_t = k-1]\,\mathbb{P}[N_t = k-1] + o(x)$$

$$= (1 - \lambda x + o(x))\, p_k(t) + (\lambda x + o(x))\, p_{k-1}(t) + o(x)$$

$$\frac{p_k(t+x) - p_k(t)}{x} = \frac{(\cancel{1} - \lambda x + o(x))\, p_k(t) + (\lambda x + o(x))\, p_{k-1}(t) + o(x) \cancel{- p_k(t)}}{x}$$

$$= -\lambda p_k(t) + \lambda p_{k-1}(t) + \frac{o(x)}{x}$$

$$\xrightarrow[x \to 0]{} -\lambda p_k(t) + \lambda p_{k-1}(t)$$

$$= p_k'(t)$$

$$\boxed{p_k'(t) = -\lambda p_k(t) + \lambda p_{k-1}(t)} \quad \text{for } k \geq 1$$

$$p_k'(t) = -\lambda p_k(t) \quad \text{for } k = 0$$

$$\begin{cases} \dfrac{d}{dt} \vec{p}(t) = \begin{bmatrix} p_0'(t) \\ p_1'(t) \\ p_2'(t) \\ \vdots \end{bmatrix} = \begin{bmatrix} -\lambda & 0 & 0 & 0 & \cdots \\ \lambda & -\lambda & 0 & 0 & \cdots \\ 0 & \lambda & -\lambda & 0 & \cdots \\ 0 & 0 & \lambda & -\lambda & \cdots \\ & & \vdots & & \end{bmatrix} \begin{bmatrix} p_0(t) \\ p_1(t) \\ p_2(t) \\ \vdots \end{bmatrix} \\ \\ \vec{p}_0(0) = \mathbb{P}[N_0 = 0] = 1 \\ \vec{p}_0(k) = \mathbb{P}[N_0 = 0] = 0 \qquad \text{for } k \geq 1 \end{cases}$$

⋮ solve

$$\Rightarrow \forall k \geq 0, \quad p_k(t) = e^{-\lambda t} \frac{(\lambda t)^k}{k!}$$

Lec 35

M/M/1 Queues

exponential dequeue

M/M/1

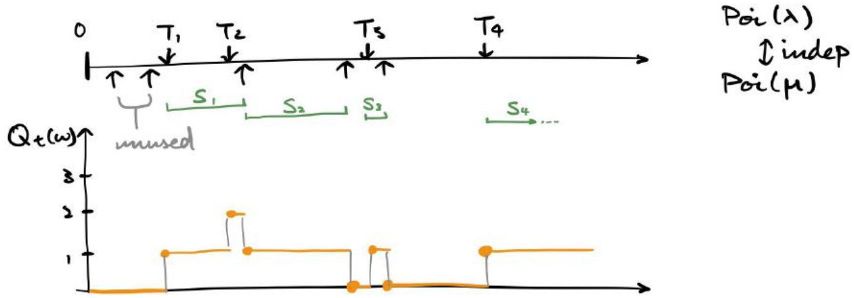
exponential enqueue   1 processor to dequeue

$Q_t(\omega)$ — queue size at time t

$T_1, T_2, \dots$ ~ $Poi(\lambda)$ process enqueue time

$(S_k)_{k \geq 1}$ iid ~ $\exp(\mu)$ service time

Assume $\mu > \lambda$



Indeed $S_k \sim \exp(\lambda)$ because we can condition on T_k

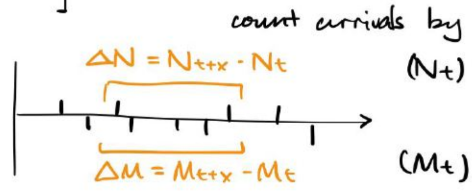
$$\begin{aligned}
\mathbb{P}[S_1 > u] &= \int_0^\infty \mathbb{P}[S_1 > u \mid T_1 = t]\, f_{T_1}(t)\, dt \\
&= \int_0^\infty \mathbb{P}[S_1 > u]\, f_{T_1}(t)\, dt \\
&= \int_0^\infty e^{-\mu u}\, f_{T_1}(t)\, dt \\
&= e^{-\mu u}
\end{aligned}$$

Distribution of $Q_t \in \mathbb{N}$

Rewriting trick:

$\mathbb{P}[Q_t = k] = p_k(t)$ $\quad\quad$ $\mu_{Q_t} = \vec{p}(t) = \begin{bmatrix} p_1(t) \\ p_2(t) \\ \vdots \end{bmatrix}$

Consider $k \geq 1$



$$\begin{aligned} p_k(t+x) &= \mathbb{P}[Q_{t+x} = k] \\ &= \mathbb{P}[Q_t = k, \Delta N = 0, \Delta M = 0] \\ &+ \mathbb{P}[Q_t = k-1, \Delta N = 1, \Delta M = 0] \\ &+ \mathbb{P}[Q_t = k+1, \Delta N = 0, \Delta M = 1] \\ &+ \mathbb{P}[Q_{t+x} = k, \Delta N = 1, \Delta M = 1] \\ &+ \underbrace{\mathbb{P}[Q_{t+x} = k, \Delta N \geq 2, \Delta M \geq 2]} \end{aligned}$$

$\leq \mathbb{P}[\Delta N \geq 2] + \mathbb{P}[\Delta M \geq 2]$ which is small

$\leq o(x)$

$= \mathbb{P}[\Delta N = 0, \Delta M = 0 \mid Q_t = k]\,\mathbb{P}[Q_t = k]$

$+ \vdots$ similar

$= \mathbb{P}[\Delta N = 0]\,\mathbb{P}[\Delta M = 0]\,\mathbb{P}[Q_t = k]$

$+ \vdots$ similar

$$\begin{aligned} &= (1 - \lambda x + o(x))(1 - \mu x + o(x))\, p_k(t) \\ &+ (\lambda x + o(x))(1 - \mu x + o(x))\, p_{k-1}(t) \\ &+ (1 - \lambda x + o(x))(\mu x + o(x))\, p_{k+1}(t) \\ &+ (\lambda x + o(x))(\mu x + o(x))\, p_k(t) \end{aligned}$$

$$\begin{aligned} &= p_k(t) - (\lambda + \mu) x\, p_k(t) + o(x) \\ &+ \lambda x\, p_{k-1}(t) + o(x) \\ &+ \mu x\, p_{k+1}(t) + o(x) \\ &+ o(x) \end{aligned}$$

$$
\begin{aligned}
p_k(t) = {} & (\lambda+\mu) x\, p_k(t) + o(x) \\
& + \lambda x\, p_{k-1}(t) + o(x) \\
& + x\, p_{k+1}(t) + o(x) \\
& + o(x)
\end{aligned}
$$

$$\frac{1}{x}\Big(p_k(t+x) - p_k(t)\Big) = \frac{1}{x}\Big(-(\lambda+\mu)x\; p_k(t) + \lambda x\, p_{k-1}(t) + \mu x\, p_{k+1}(t) + o(x)\Big)$$

as $x \to 0$,

$$\Big(p_k(t+x) - p_k(t)\Big) \to \Big(-(\lambda+\mu)\;\; p_k(t) + \lambda\, p_{k-1}(t) + \mu\; p_{k+1}(t)\Big)$$

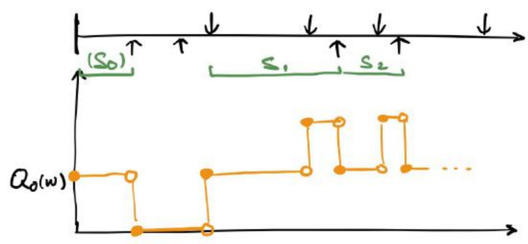
$$\begin{bmatrix} \vdots \\ \frac{d}{dt}\vec{p}(t) \\ \vdots \end{bmatrix} = \begin{bmatrix} -\lambda & \mu & 0 & & \cdots \\ \lambda & -(\lambda+\mu) & \mu & 0 & \\ 0 & \lambda & -(\lambda+\mu) & \mu & 0 \\ \vdots & 0 & \lambda & -(\lambda+\mu) & \mu \\ & \vdots & & & \ddots \end{bmatrix} \begin{bmatrix} \vdots \\ \vec{p}(t) \\ \vdots \end{bmatrix}$$

<u>Fact</u> Q_t always converge to equilibrium distribution

Solve with $\frac{d}{dt}\vec{p}(t) = 0$ for $\vec{p}(t)$

Lec 36 Queuing Process

M/M/1 reminder



Poi (λ)

Poi (μ)
potential departures

$\mu > \lambda$

$Q_0 \sim \pi^{(0)}$

$\uparrow$
arbitrary dist in $\mathbb{N}$

random initial queue size

Last time :

$$\begin{bmatrix} p_0'(t) \\ p_1'(t) \\ p_2'(t) \\ \vdots \end{bmatrix} = \begin{bmatrix} -\lambda & \mu & 0 & & \cdots \\ \lambda & -(\lambda+\mu) & \mu & 0 & \\ 0 & \lambda & -(\lambda+\mu) & \mu & 0 \\ \vdots & 0 & \lambda & -(\lambda+\mu) & \mu \\ \vdots & \vdots & & & \ddots \end{bmatrix} \begin{bmatrix} p_0(t) \\ p_1(t) \\ p_2(t) \\ \vdots \end{bmatrix}$$

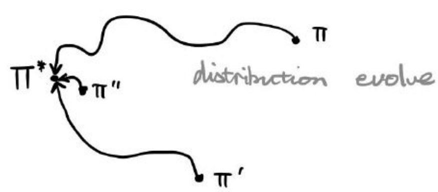
$$\frac{d}{dt}\vec{p}(t) = A \quad \vec{p}(t)$$

We know $\vec{p}(0) = \pi = \begin{bmatrix} \pi_0 \\ \vdots \end{bmatrix} = \begin{bmatrix} \mathbb{P}[Q_0 = 0] \\ \vdots \end{bmatrix}$

Solving this

<u>Thm</u> $\exists!$ solution $\vec{p}(t) = [e^{tA}]\, \vec{p}(0)$

$\exists!$ π^*, $\vec{p}(t) \xrightarrow[t\to\infty]{} \pi^*$



Consider $\vec{p}(0) = \pi^*$. Then $\forall t$, $\vec{p}(t) = \pi^*$

$$\frac{d}{dt}\vec{p}(t) = \vec{0} = A\,\vec{p}(t) = A\pi^* = 0$$

solve for π^*,

$$\begin{cases} 0 = -\lambda \pi_0^* + \mu \pi_1^* & ① \\ 0 = \lambda \pi_0^* - (\lambda+\mu) \pi_1^* + \mu \pi_2^* & ② \\ \vdots \end{cases}$$

① $\Rightarrow$ $\pi_1^* = \left(\frac{\lambda}{\mu}\right) \pi_0^*$

② $\Rightarrow$ $\pi_2^* = \left(\frac{\lambda}{\mu}\right)^2 \pi_0^*$

ⓚ $\Rightarrow$ $\pi_k^* = \left(\frac{\lambda}{\mu}\right)^k \pi_0^*$

$$\begin{aligned} 1 &= \sum \pi_k^* \\ &= \pi_0^* \sum_{k \geq 0} \left(\frac{\lambda}{\mu}\right)^k \\ &= \pi_0^* \frac{1}{1 - \frac{\lambda}{\mu}} \end{aligned}$$

So $\pi_0^* = 1 - \frac{\lambda}{\mu}$

$\pi_k^* = \left(1 - \frac{\lambda}{\mu}\right)\left(\frac{\lambda}{\mu}\right)^k$

Let $T \sim \text{geom}\left(1 - \frac{\lambda}{\mu}\right)$

$\mathbb{P}[T = k+1] = \left(\frac{\lambda}{\mu}\right)^{k+1-1}\left(1 - \frac{\lambda}{\mu}\right) = \mathbb{P}[Q = k]$

Let $Q = T - 1$

So $\pi^* \sim Q$

$\mathbb{E} Q = \frac{\mu}{\mu - \lambda}$

$\text{var } Q = \text{var } T = \frac{\frac{\lambda}{\mu}}{\left(1 - \frac{\lambda}{\mu}\right)^2}$

To simulate stable queue generate $Q_0 \sim \pi^*$

Markov Property

<u>Def</u> Process (X_t) has Markov if (equivalent definitions)

1. Condition on state at time t $(X_t = \cdot)$, then past and future are independent

2. The future depend on only the present among {past, present} events

Ex $\mathbb{P}[X_4 > y \mid \underbrace{N_t = 2}_{\text{present}}, \underbrace{T_1 \le x}_{\text{past}}] = \mathbb{P}[X_4 > y \mid N_t = 2]$

Fact Poi process, queuing process, BM all have this property

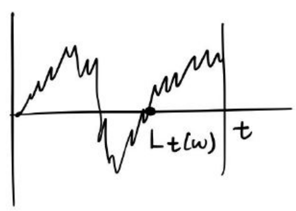
Def Strong Markov property

If the Markov property is still true for random time.

Markov: indep w.r.t. fixed t
Strong Markov: indep w.r.t. random time e.g. T_k

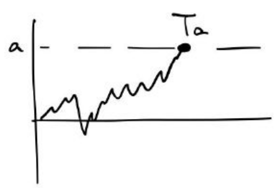
Thm this holds if T_k doesn't depend on own future
"stopping time"

Ex.



$L_t(w)$ – last visit to 0 depends on own future, since there cannot be further visit to 0 before t.
not a stopping time!

Ex.



T_a – hitting time to a. Once hit the future doesn't matter
stopping time

(see HW)

U_k – start of service of k^{th} customer in queue
stopping time

Lec 37 Kolmogorov's 0-1 Law

"In the realm of abstract nonsense"

How does one justify the existence of randomness?

How does random, chaotic, independent agents give rise to determinism

- → Statistical mechanics
- → Fluid
- → Economics
- → Population dynamics

Full randomness to complete deterministic

Defininctions

Consider abstract RVs $(X_k)_{k\geq 1}$ describing state of some system

$$X_k : \Omega \to (S, \mathcal{B})$$

state ↗ ↖ space

$\sigma(X_k) = \{\{\omega \mid X_k(\omega) \in B\} \mid B \in \mathcal{B}\}$

= "events depending on X_k only"
If X_k know, we know if they are in $\sigma(X_k)$ or not

Think: k — time
S — $\mathbb{R}$
$\mathcal{B}$ — intervals (σ-field)

$\sigma(X_1, X_2) := \sigma(\underset{\substack{\parallel \\ (X_1, X_2)}}{\vec{X}}) = \{\{\omega \mid \vec{X}(\omega) \in B\} \mid B \in \mathcal{B}^2\}$

Note $\sigma(X_1) \cup \sigma(X_2) \subseteq \sigma(X_1, X_2)$
$\sigma(\sigma(X_1), \sigma(X_2)) = \sigma(X_1, X_2)$

Consider $\underbrace{X_1, \ldots, X_n}, X_{n+1}, \ldots$

$\mathcal{F}_n := \sigma(X_1, \ldots, X_n)$
= "sigma field up to n"
= "observable events when knowing $X_1, \ldots, X_n$"

$\mathcal{F}^n := \sigma(X_{n+1}, \ldots)$
= "after time observable events"

$\mathcal{F}_\infty := \sigma(X_1, \dots)$
= "all observable events"

$\mathcal{F}^* := \bigcap_n \mathcal{F}^n$ "asymptotic σ-field"
"tail field"

Ex. $X_1, \dots \in \mathbb{R}$
$A = \{\omega \mid \exists$ infinitely k s.t. $X_k(\omega) \geq 0\}$

claim $\forall n,\ A \in \mathcal{F}^n \Rightarrow A \in \mathcal{F}^*$
whether $\exists$ infinitely many only depends on future.
it doesn't depend on any finite collection of X_k s.

Ex. $A = \{\omega \mid X_k(\omega) \to a\}$
convergence only depends on tail

Observe $\mathcal{F}^* \subseteq \mathcal{F}^n \subseteq \mathcal{F}_\infty$

Kolmogorov's

Consider indep. RVs $(X_k)_{k \geq 1}$ indep

<u>Thm</u> $A \in \mathcal{F}^* \Rightarrow P(A) \in \{0,1\}$

<u>Proof</u> Consider
$\sigma(X_1), \dots, \sigma(X_n), \underbrace{\sigma(X_{n+1}), \dots}_{\mathcal{F}^n}$

independent from $\mathcal{F}^n$

So $\sigma(X_1), \dots, \sigma(X_n)$ and $\mathcal{F}^*$ still indep since $\mathcal{F}^* \subseteq \mathcal{F}^n$

Then $\sigma(X_1), \dots, \sigma(X_{n+1}) \xleftrightarrow{\text{indep}} \mathcal{F}^*$
$\sigma(X_1), \dots, \sigma(X_{n+2}) \xleftrightarrow{\text{indep}} \mathcal{F}^*$
$\vdots$

Then $\mathcal{F}^* \xleftrightarrow{\text{indep}} \underbrace{\sigma(X_1, \dots, X_n)}_{= \mathcal{F}_\infty} \quad \forall n$

But $\mathcal{F}_\infty \xleftrightarrow{\text{indep}} \mathcal{F}^*$. Yet $\mathcal{F}_\infty \ni A \in \mathcal{F}^*$, so A indep with itself

$\mathbb{P}[A \cap A] = \mathbb{P}[A]\mathbb{P}[A]$ Anything in $\mathcal{F}^*$ is indep with itself !
$\mathbb{P}[A] = \mathbb{P}[A]^2$
$\mathbb{P}[A] \in \{0,1\}$