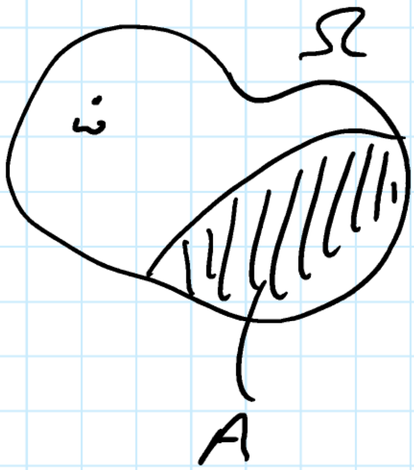
$(\Omega, \mathcal{F}, \mathbb{P})$
$\;\;\Psi \;\;\; \Psi$
$\;\;\omega \;\;\; A$



$X : \Omega \to \mathbb{R}$

$\mathbb{P}(X \in B)$

$= \mathbb{P}(\{\omega \in \Omega : X(\omega) \in B\})$, $B \subseteq \mathbb{R}$.

e.g. 1) score of 1st die
score of 2nd die } of two dice

2) time when we reach the 17^{th} slide

integer part in minutes

fractional part

$F_X(x) = \mathbb{P}(X \le x)$

discrete r.v. : $p_X(x) = \mathbb{P}(X = x)$, $x \in \mathbb{R}$

$$\sum_x p_X(x) = 1$$

cont. r.v. : $f(x) = F_X'(x)$, $\int_{-\infty}^{\infty} f(x)\,dx = 1$

$$\mathbb{P}(X \le x) = \int_{-\infty}^{x} f(u)\,du$$

$$\mathbb{E}(g(X)) = \begin{cases} \sum_x g(x)\, p_X(x) \\ \int_{-\infty}^{\infty} g(x) f(x)\,dx \end{cases}$$

$\mathrm{Var}(X)$, $\mathrm{Cov}(X, Y)$

$(A_i)_{i \in I}$ indep. if $\mathbb{P}\left(\bigcap_{i \in J} A_i\right) \overset{(*)}{=} \prod_{i \in J} \mathbb{P}(A_i)$

for $J \subseteq I$ finite

$(X_i)_{i \in I}$ indep. if $(*)$ for all $A_i = \{X_i \in B_i\}$

(suffices to take $\{X_i \le x_i\}$

integer part of time of slide 17 is 12

fractional part is 46 seconds

We have observed random variables

by carrying out an experiment

Let X and Y be two random variables (r.v.).
What might it mean to say that X and Y are close?

(1) We observe X and Y, and
(on this occasion, or on any occasion)
$|X-Y| < \varepsilon$, i.e. $|X(\omega)-Y(\omega)| < \varepsilon$.

(2) Something about the joint distribution of X and Y, e.g. $\mathbb{P}(|X-Y|<\varepsilon) > 1-\varepsilon$
or $\mathbb{E}(|X-Y|) < \varepsilon$

(3) Something comparing the dist$^{\underline{n}}$ of X and the dist$^{\underline{n}}$ of Y, e.g.
$|F_X(x) - F_Y(x)| < \varepsilon$ for all x

Let $X_1, X_2, \dots$ and X be r.v.s.

Here are three notions of convergence:

(1) $\{X_n \xrightarrow[n\to\infty]{} X\}$ <u>is an event</u>

$$\| $$

$$\{\omega \in \Omega : X_n(\omega) \xrightarrow[n\to\infty]{} X(\omega)\}$$

$$\| $$

$$\bigcap_{\cancel{\varepsilon>0}\ m\geq 1} \bigcup_{n_0 \geq 1} \bigcap_{n \geq n_0} \{|X_n - X| < \cancel{\varepsilon}\ \tfrac{1}{m}\}$$

$$X_n \xrightarrow[n\to\infty]{a.s.} X \ :\Longleftrightarrow\ \mathbb{P}(X_n \xrightarrow[n\to\infty]{} X) = 1$$

"X_n converges to X <u>almost</u> surely"

(2) $X_n \xrightarrow[n\to\infty]{P} X \ :\Longleftrightarrow\ \forall_{\varepsilon>0}\ \mathbb{P}(|X_n - X| < \varepsilon) \xrightarrow[n\to\infty]{} 1$

"X_n converges to X in probability".

(3) $X_n \xrightarrow[n\to\infty]{d} X \;:\Longleftrightarrow\; F_n(x) \xrightarrow[n\to\infty]{} F(x)$ for all $x \in \mathbb{R}$ for which F is continuous at x

where F_n and F are the cumulative distribution function (cdf) of X_n and X.

<u>Thm</u>: $X_n \xrightarrow[n\to\infty]{a.s.} X \Rightarrow X_n \xrightarrow[n\to\infty]{p} X \Rightarrow X_n \xrightarrow[n\to\infty]{d} X$

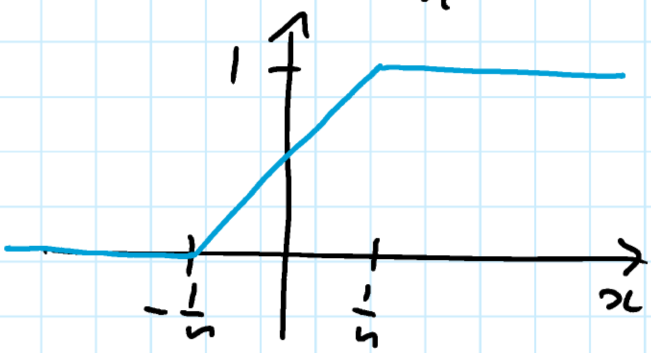
The reverse implications do not hold, in general.

$X_n \xrightarrow[n\to\infty]{d} X$ "X_n converges to X in distribution"

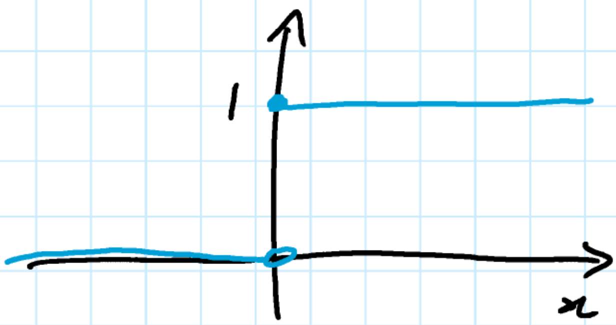
Why did we restrict to x s.th. F cont. at x?

<u>Example 1:</u> Let $X_n \sim \mathrm{Unif}([-\frac{1}{n}, \frac{1}{n}])$

Show that $X_n \xrightarrow{d} 0$.



$F_n(x) = \mathbb{P}(X_n \leq x)$



$$F(x) = \mathbb{P}(0 \leq x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases}$$

$F_n(x) \longrightarrow 0 = F(x)$ if $x < 0$

$F_n(x) \longrightarrow 1 = F(x)$ if $x > 0$

So $F_n(x) \to F(x)$ except at $x = 0$, where F is not continuous.

Note $F_n(0) = \frac{1}{2} \not\to F(0) = 1$

This is OK since F not cont. at 0.

Remark: Notice (3) only depends on the dist$^{\underline{ns}}$ of X_n and X, not directly on the r.v.s. The r.v.s do not even need to be defined on the same probability space (unlike (1) and (2)).

(3) really defines a notion of convergence of distributions rather than r.v.s.

"weak convergence"

Notation: Sometimes we write a dist^{n} on the RHS, e.g.

$$X_n \xrightarrow{d} N(0, \sigma^2)$$

$$X_n \xrightarrow{d} U([0,1])$$

Example 2: Let Y_n be geometric with parameter $\frac{\lambda}{n}$, i.e.

$$\mathbb{P}(Y_n = k) = \left(1 - \frac{\lambda}{n}\right)^{k-1} \frac{\lambda}{n}, \quad k \geq 1$$

$$\mathbb{P}(Y_n > k) = \left(1 - \frac{\lambda}{n}\right)^{k}, \quad k \geq 0.$$

Show $\frac{Y_n}{n} \xrightarrow[n \to \infty]{d} Exp(\lambda)$

Let $Y \sim Exp(\lambda)$ "Y has dist^{n} $Exp(\lambda)$"

$$\mathbb{P}\left(\frac{Y_n}{n} > x\right) \underset{x \geq 0}{=\!=} \mathbb{P}(Y_n > nx)$$

$$= \mathbb{P}(Y_n > \lfloor nx \rfloor)$$

$$= \left(1 - \frac{\lambda}{n}\right)^{\lfloor nx \rfloor}$$

$$\sim \left(\left(1 - \frac{\lambda}{n}\right)^n\right)^x$$

$$\xrightarrow[n \to \infty]{} e^{-\lambda x} = \mathbb{P}(Y > x)$$

So also $\mathbb{P}\left(\frac{Y_n}{n} \leq x\right) \xrightarrow[n \to \infty]{} \mathbb{P}(Y \leq x)$

for all $x \geq 0$.

For $x < 0$, both sides are 0.

0 included,

as required!

Prop: $X_n \xrightarrow[n\to\infty]{P} X \overset{(1)}{\Rightarrow} X_n \xrightarrow[n\to\infty]{d} X$

$\underset{(2)}{\not\Leftarrow}$

Proof: (1) Suppose $X_n \xrightarrow[n\to\infty]{P} X$.

Let F_n and F the cdfs of X_n and X.

Fix x s.th. F is continuous at x. Let $\varepsilon > 0$.

If $X_n \le x$, then $X \le x+\varepsilon$ or $|X_n - X| > \varepsilon$

So $\underbrace{\mathbb{P}(X_n \le x)}_{=F_n(x)} \le \mathbb{P}(\{X \le x+\varepsilon\} \cup \{|X_n - X| > \varepsilon\})$

$\le \mathbb{P}(X \le x+\varepsilon) + \underbrace{\mathbb{P}(|X_n - X| > \varepsilon)}$

$\xrightarrow[n\to\infty]{} F(x+\varepsilon) \qquad \xrightarrow[n\to\infty]{} 0$

So $F_n(x) \le F(x+\varepsilon) + \varepsilon$ for n suff large.

Similarly $F_n(x) \ge F(x-\varepsilon) - \varepsilon$ for n suff. large

Since $\varepsilon > 0$ is arbitrary, and F is cont. at x, we get $F_n(x) \longrightarrow F(x)$, as $n \to \infty$.

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

(2) Take Y, Z two r.v.s. with the same dist$^{\underline{n}}$ but s.th. they are not equal w.p. 1.

Let $X_n = Y$, $n \geq 1$, and $X = Z$.

Thus $X_n \xrightarrow[n\to\infty]{d} X$, but $X_n \overset{p}{\not\longrightarrow} X$.

<u>Lemma</u>: Let $(A_n, n \geq 1)$ be an increasing sequence of events, i.e. $A_1 \subseteq A_2 \subseteq \cdots$

Then $\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}\left(\bigcup_{n\geq 1} A_n\right)$.

<u>Proof</u>: $\mathbb{P}\left(\bigcup_{n\geq 1} A_n\right) = \mathbb{P}\Big(\underbrace{A_1 \cup \bigcup_{i\geq 1} (A_{i+1} \setminus A_i)}_{\text{disjoint union}}\Big)$

countable additivity

$$= \mathbb{P}(A_1) + \sum_{i\geq 1} \mathbb{P}(A_{i+1} \setminus A_i)$$

$$= \mathbb{P}(A_1) + \lim_{n\to\infty} \sum_{i=1}^{n-1} \mathbb{P}(A_{i+1} \setminus A_i)$$

$$= \lim_{n\to\infty} \left(\mathbb{P}(A_1) + \sum_{i=1}^{n-1} \mathbb{P}(A_{i+1} \setminus A_i)\right)$$

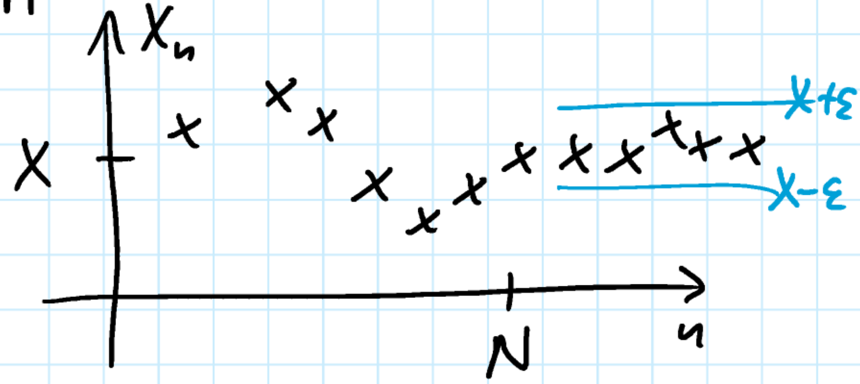
$$= \lim_{n\to\infty} \mathbb{P}(A_n)$$

$\square$

Prop: $X_n \xrightarrow[n\to\infty]{a.s.} X \overset{(3)}{\Longrightarrow} X_n \xrightarrow[n\to\infty]{P} X$

$\underset{(4)}{\not\Longleftarrow}$

Proof: (3) Suppose $X_n \xrightarrow[n\to\infty]{a.s.} X$. Let $\varepsilon > 0$.



for $N \in \mathbb{N}$ define the event A_N by

$$A_N = \{ |X_n - X| < \varepsilon \text{ for all } n \geq N \}$$

If $\{X_n \to X\}$ occurs then A_N occurs for some N

i.e. $\bigcup_{N\geq 1} A_N$ occurs. Since $\mathbb{P}(X_n \to X) = 1$,

also $\mathbb{P}\left(\bigcup_{N\geq 1} A_N\right) = 1$.

Since $(A_N, N \geq 1)$ is increasing.

By the lemma, $\mathbb{P}(A_N) \to \mathbb{P}\left(\bigcup_{n\geq 1} A_n\right) = 1$.

Since $\mathbb{P}(|X_N - X| < \varepsilon) \geq \mathbb{P}(A_N)$

So $\mathbb{P}(|X_N - X| < \varepsilon) \xrightarrow[N\to\infty]{} 1$ also.

Since $\varepsilon > 0$ was arbitrary, $X_n \xrightarrow[n\to\infty]{P} X$.

(4) Example when $X_n \xrightarrow[n\to\infty]{P} X$, but not

$X_n \xrightarrow[n\to\infty]{a.s.} X$

Consider $X_1, X_2, \ldots$ independent with

$$\mathbb{P}(X_n = 1) = \frac{1}{n}, \quad \mathbb{P}(X_n = 0) = \frac{n-1}{n}.$$

Then $X_n \xrightarrow[n\to\infty]{P} 0$ (check!)

Since X_n only takes values 0 and 1,

$$\mathbb{P}(X_n \to 0) = \mathbb{P}(X_n = 0 \text{ for all } n \text{ large enough})$$
$$= \mathbb{P}\Big(\bigcup_{N \geq 1} B_N\Big)$$

where $B_N = \{X_n = 0 \text{ for all } n > N\}$

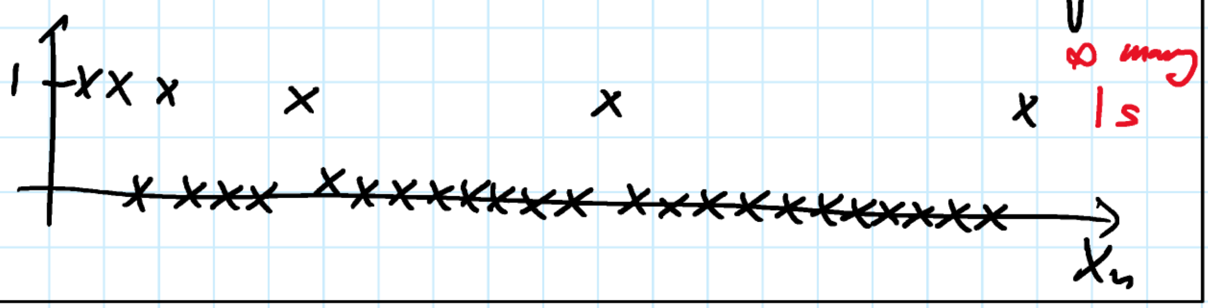
What is $\mathbb{P}(B_N)$? Fix K. Then

$$\mathbb{P}(B_N) \leq \mathbb{P}(X_n = 0 \text{ for } n = N+1, \dots, N+K)$$
$$= \frac{N}{\cancel{N+1}} \frac{\cancel{N+1}}{\cancel{N+2}} \cdots \frac{\cancel{N+K-1}}{N+K} = \frac{N}{N+K}$$

Since K is arbitrary, $\mathbb{P}(B_N) = 0$.

Since $(B_N, N \geq 1)$ is increasing, the lemma yields $\mathbb{P}\Big(\bigcup_{N \geq 1} B_N\Big) = \lim_{N \to \infty} \mathbb{P}(B_N) = 0$.

So $\mathbb{P}(X_n \to 0) = 0$, so a.s. conv. fails.



Thm: (WLLN). Let $X_1, X_2, \ldots$ be a sequence of iid r.v.s with mean μ.
Let $S_n = X_1 + \cdots + X_n$. Then

$$\frac{S_n}{n} \xrightarrow[n\to\infty]{P} \mu$$

i.e. $\forall_{\varepsilon>0}\ \mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| < \varepsilon\right) \xrightarrow[n\to\infty]{} 1.$

Prop: (Markov inequality). If $\mathbb{P}(X \geq 0) = 1$, then for all $c > 0$, $\mathbb{P}(X \geq c) \leq \frac{\mathbb{E}(X)}{c}$.

Cor: (Chebyshev inequality). If Y has finite variance, $\mathbb{P}(|Y - \mathbb{E}(Y)| \geq \varepsilon) \leq \frac{\mathrm{Var}(Y)}{\varepsilon^2}$

Thm: (SLLN). In the setting of the WLLN,

$$\frac{S_n}{n} \xrightarrow[n\to\infty]{a.s.} \mu, \text{ i.e. } \mathbb{P}\left(\frac{S_n}{n} \xrightarrow[n\to\infty]{} \mu\right) = 1.$$

Remark: Interpretation

- WLLN: for n large enough, $\frac{S_n}{n}$ is likely to be close to μ.
- SLLN: for n large enough, $\frac{S_n}{n}$ is close to μ.

We will prove SLLN under the additional assumption $\mathbb{E}(X_1^4) < \infty$. Recall that the Chebyshev proof of WLLN also req^s an additional assumption $\mathbb{E}(X_1^2) < \infty$. But WLLN and SLLN both hold without these additional assumptions (→ B8.1 Lecture 16)

<u>Sketch of proof:</u> Assume $\mathbb{E}(X_1^4) < \infty$.

$$\mathbb{E}\left(\sum_{n=1}^{\infty}\left(\frac{S_n}{n}-\mu\right)^4\right)$$

$$= \sum_{n=1}^{\infty} \mathbb{E}\left(\left(\frac{S_n}{n}-\mu\right)^4\right)$$

$$= \sum_{n=1}^{\infty} \frac{1}{n^4}\, \mathbb{E}\left(\boxed{\left(\underbrace{S_n - n\mu}_{=\sum_{i=1}^{n}\underbrace{(X_i-\mu)}_{W_i}}\right)^4}\right)$$

$$\sum_{i=1}^{n} W_i \sum_{j=1}^{n} W_j \sum_{k=1}^{n} W_k \sum_{l=1}^{n} W_l$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} \mathbb{E}(W_i W_j W_k W_l)$$

$$= n\,\mathbb{E}(W_1^4) + \mathbb{E}(W_1^3)\overbrace{\mathbb{E}(W_2)}^{=0} + \dots$$
$$+ 6\binom{n}{2}\mathbb{E}(W_1^2 W_2^2)$$

$< \infty$ since $\displaystyle\sum_{n=1}^{\infty} \frac{\text{const}}{n^2} < \infty$

Hence $\mathbb{E}\left(\sum_{n=1}^{\infty}\left(\frac{S_n}{n} - \mu\right)^4\right) < \infty$.

In particular, $\mathbb{P}\left(\sum_{n=1}^{\infty}\left(\frac{S_n}{n} - \mu\right)^4 < \infty\right) = 1$

$$\Rightarrow \mathbb{P}\left(\frac{S_n}{n} \xrightarrow[n\to\infty]{} \mu\right) = 1$$

$\square$

Remark: $Z_n : \Omega \longrightarrow [0,\infty)$, $n \geq 1$

$\Rightarrow T = \sum_{n=1}^{\infty} Z_n : \Omega \longrightarrow [0,\infty]$

$\mathbb{P}(T=\infty) > 0 \Rightarrow \mathbb{E}(T) = \infty$

$\not\Leftarrow$

e.g. $\mathbb{P}(T=k) = \frac{1}{k(k+1)}$, $k \geq 1$

has $\mathbb{E}(T) = \infty$.

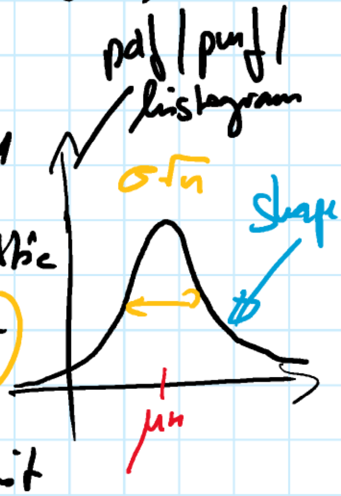
Thm: (CLT). Let X_i, $i \geq 1$, be iid with $\mathbb{E}(X_i) = \mu$ and $Var(X_i) \in (0, \infty)$. Then

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} N(0,1).$$

Check: $\mathbb{E}\left(\frac{S_n - n\mu}{\sigma\sqrt{n}}\right) = 0$

$$Var\left(\frac{S_n - n\mu}{\sigma\sqrt{n}}\right) = \frac{1}{\sigma^2 n} Var(S_n - n\mu) = \frac{1}{\sigma^2 n} Var(S_n) = 1.$$

Summary:
- S_n concentrates around μn (1st order behaviour deterministic
- fluctuations around μn are of order $\sqrt{n}$
- fluctuations are random and the limit law is "universal" normal



Remark: • There are lots of ways to rephrase the CLT, e.g.

$$\mathbb{P}\left(a \le \frac{S_n - n\mu}{\sigma\sqrt{n}} \le b\right) \longrightarrow \int_a^b \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx = \Phi(b) - \Phi(a)$$

cdf of $N(0,1)$

$$= \int_{\sigma a}^{\sigma b} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{z^2}{2\sigma^2}\right) dz$$

Hence $\dfrac{S_n - n\mu}{\sqrt{n}} \xrightarrow[n\to\infty]{d} N(0, \sigma^2)$

• In CLT, $\xrightarrow[n\to\infty]{d}$ cannot be strengthened to $\xrightarrow[n\to\infty]{P}$. Consider $\dfrac{S_n - n\mu}{\sigma\sqrt{n}} - \dfrac{S_{2n} - 2n\mu}{\sigma\sqrt{2n}}$.

Example: $n = 10{,}000$ similar car insurance policies. X_i = claim amount on i^{th} policy with mean £240 and standard dev. £800.

Aim: $\mathbb{P}$(total reserve > total claim amount) ≥ 0.99
from premiums

Solutions:

Set $S_n = \sum_{i=1}^{n} X_i$, $\mu = 240$, $\sigma^2 = 800$

$\Phi^{-1}(0.99) \approx 2.326$

Total claim amount $= S_n$

CLT $\Rightarrow \mathbb{P}\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq \Phi^{-1}(0.99)\right) \longrightarrow 0.99$

$$\Rightarrow \mathbb{P}\Big(S_n \leq \underbrace{2.326\sigma\sqrt{n} + n\mu}_{\text{appropriate reserve}}\Big) \approx 0.99$$

£ 2,586,080

So, per policy we need $240 + \underline{\underline{18.61}}$

Example: Binomial distribution

(1) Normal approximation: Let A_i, $i \geq 1$, be independent events, each occurring w.p. p.

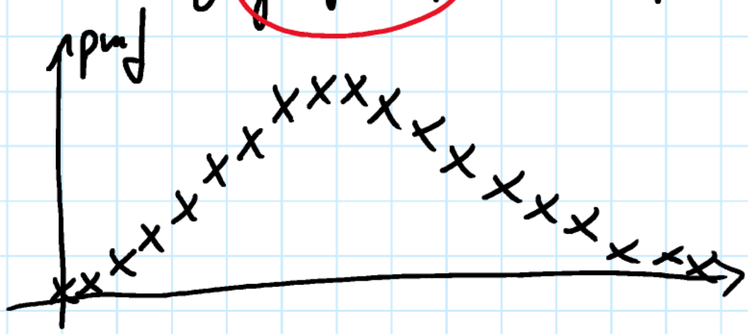
Then $S_n = \#\{i \in \{1, \ldots, n\} : A_i \text{ occurs}\}$

$$= \sum_{i=1}^{n} \mathbb{1}_{A_i} \sim Bin(n, p)$$

with $\mu = \mathbb{E}(\mathbb{1}_{A_i}) = p$, $\sigma^2 = Var(\mathbb{1}_{A_i}) = p(1-p)$

By CLT, $$\frac{S_n - np}{\sqrt{np(1-p)}} \xrightarrow[n \to \infty]{d} N(0,1)$$

loosely, for fixed p $Bin(n,p) \approx N(np, np(1-p))$ for large n



(11) Poisson approximation: Instead of fixed p, we fix the mean $np = \lambda$. Then

$$S_n \xrightarrow[n\to\infty]{d} \text{Poisson}(\lambda)$$

To show this, it's enough (check!) to show

$$\mathbb{P}(S_n = k) \xrightarrow[n\to\infty]{} \frac{\lambda^k}{k!} e^{-\lambda}, \quad k \geq 0.$$

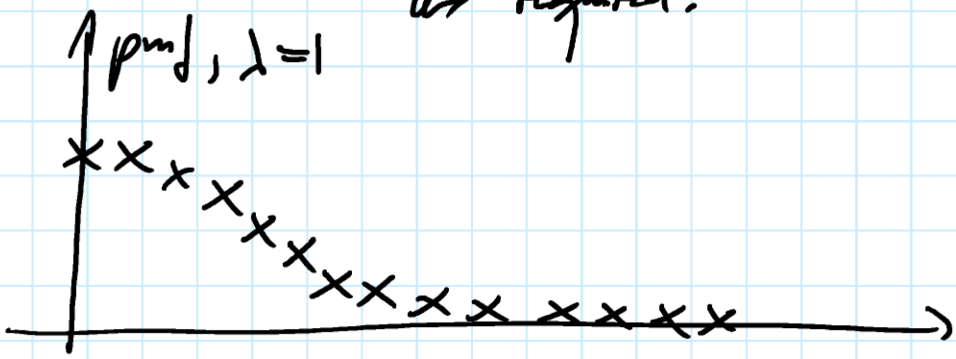
So $\mathbb{P}(S_n = k) = \binom{n}{k}\left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k}$

$$= \frac{n(n-1)\cdots(n-k+1)}{k!}\frac{\lambda^k}{n^k}\underbrace{\left(1-\frac{\lambda}{n}\right)^n}_{\to e^{-\lambda}}\underbrace{\left(1-\frac{\lambda}{n}\right)^{-k}}_{\to 1}$$

$\to 1$

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

as required.



Review: For X $\mathbb{N}_0$-valued,

$$G_X(z) = \mathbb{E}(z^X) = \sum_{k=0}^{\infty} \mathbb{P}(X=k) z^k$$

for $z \in [0,1]$ or $z \in [-1,1]$

or $z \in \mathbb{R}$ s.th. series converges

Uniqueness Theorem: For $\mathbb{N}_0$-valued r.v.s X and Y,

$G_X(z) = G_Y(z)\ \forall_{z\in[0,1]} \iff X$ and Y have the same dist^n

Convergence Theorem: For $\mathbb{N}_0$-valued r.v.s. X_n, $n \geq 1$ and X,

$$G_{X_n}(z) \longrightarrow G_X(z)\ \forall_{z\in[0,1]} \iff X_n \xrightarrow[n\to\infty]{d} X$$

Also: $X_n \xrightarrow[n\to\infty]{d} X \iff p_{X_n}(k) \longrightarrow p_X(k)\ \forall_{k\in\mathbb{N}_0}$

For $\mathbb{R}$-valued X

$$M_X(t) = \mathbb{E}\left(e^{tX}\right) \in (0,\infty] \quad , \; t \in \mathbb{R}$$

(or we restrict to $t \in \mathbb{R}$ s.th. expectation finite)

Uniqueness Theorem: For $\mathbb{R}$-valued r.v.s X and Y,

$M_X(t) = M_Y(t)$ $\boxed{\forall_{t\in[-t_0,t_0]} \text{ for some } t_0>0}$ $\Rightarrow$ X and Y have the same dist$^{\underline{n}}$

Convergence Theorem: For $\mathbb{R}$-valued X_n, $n \geq 1$, and X,

$M_{X_n}(t) \xrightarrow[n\to\infty]{} M_X(t)$ $\boxed{\forall_{t\in[-t_0,t_0]} \text{ for some } t_0>0}$ $\Rightarrow$ $X_n \xrightarrow[n\to\infty]{d} X$

Some remarks about proofs: Uniqueness Theorem

cf. "Integral Transforms" using results from "Integration"

$\rightarrow$ also inversion formula

<u>Example</u>: mgf of Exponential $(\lambda) \sim X$

$$M_X(t) = \mathbb{E}(e^{tX}) = \int_0^\infty e^{tx} \lambda e^{-\lambda x} dx$$

$$= \frac{\lambda}{\lambda - t} \int_0^\infty \underbrace{(\lambda-t) e^{-(\lambda-t)x}}_{\substack{\text{pdf of } Exp(\boxed{\lambda - t}) \\ \text{integrates to } 1}} dx$$

if $\lambda - t > 0$

$$= \frac{\lambda}{\lambda - t} \quad \text{for } t < \lambda$$

(for $t \geq \lambda$, $M_X(t) = \infty$)

For Unig Thm, we can take $t_0 = \frac{\lambda}{2} > 0$.

<u>Thm</u>: (a) $Y = aX + b \Rightarrow M_Y(t) = e^{bt} M_X(at)$

(b) $X_1, \dots, X_n$ indep $\Rightarrow M_{X_1+\dots+X_n}(t) = M_{X_1}(t) \cdots M_{X_n}(t)$

Proof of (b): $\text{LHS} = \mathbb{E}\left(e^{t(X_1+\dots+X_n)}\right)$

$$= \mathbb{E}\left(e^{tX_1} \cdots e^{tX_n}\right)$$

$$\underset{\text{indep}}{=} \mathbb{E}\left(e^{tX_1}\right) \cdots \mathbb{E}\left(e^{tX_n}\right) = \text{RHS} \quad \square$$

<u>Lemma</u>: $\boxed{\exists_{t_0>0} \forall_{t \in [-t_0, t_0]} \; M_X(t) < \infty}$

$$\Longleftrightarrow \boxed{\exists_{t_0>0} \; \mathbb{E}\left(e^{t_0 |X|}\right) < \infty.} \quad (*)$$

Furthermore, in this case, $\mathbb{E}(X^k)$ exists $\forall_{k \geq 1}$.

<u>Proof</u>: $0 \leq e^{tX} \leq e^{|tX|} \leq e^{t_0|X|} \leq e^{t_0 X} + e^{-t_0 X}$.

Now take expectations.

Similarly, $0 \leq |X|^k \leq \frac{k!}{t_0^k} e^{t_0|X|}$.

For odd k, $X^k = \underbrace{|X|^k \mathbb{1}_{\{X>0\}} - |X|^k \mathbb{1}_{\{X<0\}}}_{\text{both have finite expectation}}$ □

Thm: Taylor expansion. Suppose (*). Then

(a) $M_X(t) = \sum_{k=0}^{\infty} \mathbb{E}(X^k) \frac{t^k}{k!}$ for $|t| \le t_0$

(b) $M_X^{(k)}(0) = \mathbb{E}(X^k)$

Informal proof: $M_X(t) = \mathbb{E}(e^{tX}) = \mathbb{E}\Big(\sum_{k=0}^{\infty} \underbrace{X^k}_{\le |X|^k} \underbrace{\frac{t^k}{k!}}_{\le t_0^k}\Big)$ □

Example: mgf of $N(\mu, \sigma^2)$.

(1) First let $Z \sim N(0,1)$

$$M_Z(t) = \mathbb{E}(e^{tZ}) = \int_{-\infty}^{\infty} e^{tz} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(z^2 - 2tz\right)\right) dz$$

$$= \int_{-\infty}^{\infty} \exp\left(\frac{t^2}{2}\right) \underbrace{\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(z-t)^2\right)}_{\text{this is pdf of } N(t,1)} dz$$

$$= e^{\frac{t^2}{2}}, \quad t \in \mathbb{R}$$

(ii) Then for $X \sim N(\mu, \sigma^2)$, put $X = \sigma Z + \mu$

so by the theorem proved earlier,

$$M_X(t) = e^{\mu t} M_Z(\sigma t) = \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right)$$

Note that $M_X'(0) = \mu$

and $M_X''(0) = \cdots = \sigma^2 + \mu^2$

(iii) For $\left.\begin{array}{l} X \sim N(\mu_1, \sigma_1^2) \\ Y \sim N(\mu_2, \sigma_2^2) \end{array}\right\}$ indep.

then $X + Y \sim N(\mu_1 + \mu_2,\ \sigma_1^2 + \sigma_2^2)$

Here we can show this using mgfs:

$$\boxed{\begin{aligned} M_{X+Y}(t) &= M_X(t) M_Y(t) \quad \text{by part (b) of the theorem.} \\ &= \exp\left(\mu_1 t + \tfrac{1}{2}\sigma_1^2 t^2 + \mu_2 t + \tfrac{1}{2}\sigma_2^2 t^2\right) \\ &= \exp\left((\mu_1+\mu_2)t + \tfrac{1}{2}(\sigma_1^2+\sigma_2^2)t^2\right) \end{aligned}}$$

This is the mgf of $N(\mu_1+\mu_2,\ \sigma_1^2+\sigma_2^2)$

This calculation holds for all $t \in \mathbb{R}$.

By the Uniqueness Thm, $X+Y \sim N(\mu_1+\mu_2,\ \sigma_1^2+\sigma_2^2)$

<u>Recall</u> $f(h) = o(g(h))$ as $h \to 0$

means $\dfrac{f(h)}{g(h)} \longrightarrow 0$ as $h \to 0$

similarly $f(n) = o(g(n))$ as $n \to \infty$

Also $f(h) = f(0) + h f'(0) + o(h)$
as $h \to 0$

means $\dfrac{f(h) - f(0) - h f'(0)}{h} \longrightarrow 0$

Now let $X, X_1, X_2, \ldots$ be iid with
$\boxed{\mathbb{E}(e^{t_0|X|}) < \infty \text{ for some } t_0 > 0}$.
Let $S_n = X_1 + \cdots + X_n$. Let $\mu = \mathbb{E}(X)$, $\sigma^2 = \mathrm{Var}(X)$

Let $M_X(t) = \mathbb{E}(e^{tX})$.

Using the Taylor expansion of M_X

$$M_X(h) = M_X(0) + h M_X'(0) + o(h)$$
$$= 1 + h\mu + o(h) \quad \text{as } h \to 0$$

Then

$$M_{\frac{S_n}{n}}(t) = \mathbb{E}(e^{\frac{tS_n}{n}})$$
$$= \mathbb{E}(e^{\frac{tX_1}{n}} \cdots e^{\frac{tX_n}{n}})$$
$$= \left(M_X\left(\tfrac{t}{n}\right)\right)^n \quad \text{by indep.}$$
$$= \left(1 + \tfrac{t}{n}\mu + o\left(\tfrac{t}{n}\right)\right)^n \quad \text{as } n \to \infty$$
$$\xrightarrow[n\to\infty]{} e^{t\mu} = \mathbb{E}(e^{tY}), \ \mathbb{P}(Y=\mu) = 1.$$

By the Convergence Thm, $\frac{S_n}{n} \xrightarrow[n\to\infty]{d} \mu$, since $\mathbb{E}(e^{t_0 X}) < \infty$.

Check that this implies $\frac{S_n}{n} \xrightarrow[n\to\infty]{P} \mu$

since μ is deterministic.

For the CLT, let $Y_i = X_i - \mu$, mean 0
variance σ^2

Then

$$M_Y(h) = M_Y(0) + h M_Y'(0) + \frac{h^2}{2} M_Y''(0) + o(h^2)$$

$$= 1 + h \underbrace{\mathbb{E}(Y)}_{=0} + \frac{h^2}{2} \underbrace{\mathbb{E}(Y^2)}_{=\sigma^2 \text{ since } \mathbb{E}(Y)=0} + o(h^2) \text{ as } h \to 0$$

$$= 1 + \frac{h^2}{2}\sigma^2 + o(h^2)$$

Let $\tilde{M}_n$ be the mgf of $\frac{S_n - n\mu}{\sigma\sqrt{n}}$

$$\tilde{M}_n(t) = \mathbb{E}\left(\exp\left(\frac{(S_n - \mu n)t}{\sigma\sqrt{n}}\right)\right)$$

$$= \mathbb{E}\left(\exp\left(\frac{Y_1 t}{\sigma\sqrt{n}}\right) \cdots \exp\left(\frac{Y_n t}{\sigma\sqrt{n}}\right)\right)$$

$$= \left(M_Y\left(\frac{t}{\sigma\sqrt{n}}\right)\right)^n \quad \text{by indep.}$$

$$= \left(1 + \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right)^n \quad \text{as } n \to \infty$$

$$\xrightarrow[n\to\infty]{} e^{t^2/2} \quad \text{mgf of } N(0,1)$$

Since we assumed $\boxed{\mathbb{E}(e^{t_0|X|}) < \infty}$, the Convergence Theorem applies and

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow[n\to\infty]{d} N(0,1)$$

$\square$

<u>Example:</u> X_i iid $\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = -1) = \frac{1}{2}$

$S_n = X_1 + \cdots + X_n$, $n \geq 1$.

We know that $S_n \approx O(\sqrt{n})$

So $\mathbb{P}(|S_n| > na)$ should decay as $n \to \infty$

Chebyshev's ineq

$$\mathbb{P}(|S_n| > na) \leq \frac{\mathrm{Var}(S_n)}{(na)^2}$$

$$= \frac{1}{na^2} \xrightarrow[n\to\infty]{} 0$$

Mgf $\mathbb{E}(e^{tX_i}) = \dfrac{e^t + e^{-t}}{2} = \underbrace{\cosh(t)}_{= \sum_{k\geq 0}^{\infty} \frac{t^{2k}}{(2k)!}}$

$$\leq \exp\left(\frac{t^2}{2}\right) = \sum_{k\geq 0}^{\infty} \frac{t^{2k}}{2^k k!}$$

Then for $t>0$

$$\mathbb{P}(S_n > na) = \mathbb{P}(\exp(tS_n) > \exp(tna))$$

$$\underset{\text{Markov}}{\leq} \frac{\mathbb{E}(\exp(tS_n))}{\exp(tna)} \underset{\text{indep.}}{=} \left(\frac{\mathbb{E}(\exp(tX_1))}{\exp(ta)}\right)^n$$

$$\leq \left(\exp\left(\frac{t^2}{2} - ta\right)\right)^n$$

This is true for all $t>0$. Choose t to minimise $\exp\left(\frac{t^2}{2} - ta\right)$. This gives $t=a$.

$$\mathbb{P}(S_n > na) \leq \exp\left(-n\frac{a^2}{2}\right)$$

By symmetry,

$$\mathbb{P}(|S_n| > na) \leq 2\exp\left(-n\frac{a^2}{2}\right)$$

This is exponential decay as $n \longrightarrow \infty$!

Comment: Tails of random variables

When does a mgf exist on some $[-t_0, t_0]$?

We need $\mathbb{E}(e^{t_0|X|}) < \infty$ for some $t_0 > 0$

Equivalently, $\mathbb{P}(|X| > x) < e^{-bx}$

for some $b > 0$

"exponential tails"

Classification of distributions acc. to tails

mgf on $\mathbb{R}$	Bernoulli, binomial, uniform	bounded support	tail is eventually 0
	Normal, Poisson	super-exponential	e^{-cx^2}
mgf on interval	Exponential, geometric	exponential tail	e^{-cx}

no mgf

Pareto dist^{n} polynomial tail

$$\mathbb{P}(X > x) = x^{-\alpha - 1}$$

Cauchy dist^{n}

pdf $f_X(x) = \frac{1}{1+x^2}\frac{1}{\pi}$

Dist^{ns} with polynomial tails only have moments $\mathbb{E}(X^k)$ for some $k < \alpha$.

For $\mathbb{R}$-valued X

$$\phi_X(t) = \mathbb{E}(e^{itX}), \quad \phi_X : \mathbb{R} \to \mathbb{C}$$
$$= \mathbb{E}(\cos(tX)) + i\,\mathbb{E}(\sin(tX))$$

So $|\phi_X(t)| \leq 1$ for all $t \in \mathbb{R}$.

<u>Facts</u>:
- $\phi_{X+Y}(t) = \phi_X(t)\,\phi_Y(t)$ for X, Y indep.
- $\phi_X(t) = 1 + it\,\mathbb{E}(X) + i^2 \frac{t^2}{2}\mathbb{E}(X^2) + o(t^2)$ as $t \to \infty$ if $\mathbb{E}(X^2)$ exists
- Uniqueness and Convergence Theorems replacing $\forall_{t \in [t_0, t_0]} \ldots$ by $\forall_{t \in \mathbb{R}}$
- Proofs of WLLN and CLT can be adapted.

Note that no assumption of $\mathbb{E}(e^{t_0|X|})<\infty$ was needed here!

When mgf is finite on $(-t_0, t_0)$, $t_0>0$, the theory of analytic continuation of complex-analytic functions gives that

$$\phi_X(t) = M_X(it)$$

<u>Examples</u>: (a) $Z \sim N(0,1)$, mgf $e^{\frac{t^2}{2}}$

c.f. $e^{(it)^2/2} = e^{-t^2/2}$

(b) $\mathrm{Exp}(\lambda)$ mgf $\frac{\lambda}{\lambda - t}$, c.f. $\frac{\lambda}{\lambda - it}$

(c) Consider the Cauchy distr

$$f(x) = \frac{1}{\pi(1+x^2)}$$, then $\mathbb{E}(|X|)=\infty$

There is no mgf!

$$\phi_X(t) = \mathbb{E}(e^{itX}) = \int_{-\infty}^{\infty} \frac{e^{itx}}{\pi(1+x^2)}\, dx$$

$$= \cdots = e^{-|t|}$$

by contour integration

Note $\phi_X'(0)$ does not exist

(since $\mathbb{E}(|X|) = \infty$)

Consider $X_1, X_2, \ldots$ iid Cauchy

$S_n = X_1 + \cdots + X_n$

$$\phi_{\frac{S_n}{n}}(t) = \left(\phi\left(\tfrac{t}{n}\right)\right)^n = \left(e^{-\left|\frac{t}{n}\right|}\right)^n = e^{-|t|}$$

By the Uniq Thm, $\boxed{\frac{S_n}{n} \sim \text{Cauchy.}}$

Hence, no WLLN, no CLT

X and Y jointly continuous if

$$\mathbb{P}(X \le x, Y \le y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u,v)\, du\, dv$$

for some f, called the joint pdf of X and Y, denoted by $f_{X,Y}$. Then

$$\mathbb{P}((X,Y) \in A) = \iint_A f(x,y)\, dx\, dy$$

for (nice) $A \subseteq \mathbb{R}^2$. Marginal pdfs

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\, dy, \quad f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\, dx.$$

X, Y independent if $f_{X,Y}$ can be chosen as

$$f_{X,Y}(x,y) = f_X(x) f_Y(y), \quad (x,y) \in \mathbb{R}^2.$$

Example: Polar coordinates $(x,y) \longmapsto (r,\theta)$

bijection from $D = \mathbb{R}^2 \setminus \{(0,0)\}$ to $R = (0,\infty) \times [0,2\pi)$

Theorem: $D, R \subseteq \mathbb{R}^2$, $T : D \to R$ invertible with diff$^{\text{ble}}$ inverse, $(x,y) \mapsto (u,v)$ with $(x,y) \in D$, $(u,v) \in R$

$$J(u,v) = \det \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{pmatrix}$$

If (X,Y) jointly continuous, then so are $(U,V) = T(X,Y)$, with joint pdf

$$f_{U,V}(u,v) = \begin{cases} f_{X,Y}(x(u,v), y(u,v)) \, |J(u,v)| & (u,v) \in R \\ 0 & \text{o/w} \end{cases}$$

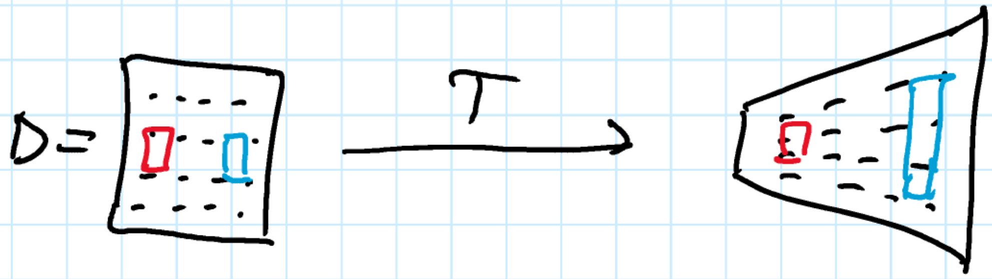
<u>Proof:</u> Let $B \subseteq R$, $A := T^{-1}(B) \subseteq D$

$$\mathbb{P}((U,V) \in B) = \mathbb{P}((X,Y) \in A)$$

$$= \iint_A f_{X,Y}(x,y)\, dx\, dy$$

$$= \iint_B \underbrace{f_{X,Y}(x(u,v), y(u,v))\, |J(u,v)|}\, du\, dv$$

Since this holds for all $B \subseteq R$, in particular for $B = (-\infty, a] \times (-\infty, b]$, we identify the integrand as $f_{U,V}(u,v)$ $\square$



<u>Example 1</u>: Let X, Y be iid $\mathrm{Exp}(\lambda)$

Let $U = \dfrac{X}{X+Y}$, $V = X+Y$.

What is the $\boxed{\text{joint dist}^{\text{n}} \text{ of } (U,V)}$?

<u>Solution</u>:

$$f_{X,Y}(x,y) = \lambda e^{-\lambda x} \lambda e^{-\lambda y}$$
$$= \lambda^2 e^{-\lambda(x+y)}, \quad (x,y) \in (0,\infty)^2$$

The map $(u,v) = T(x,y) = \left(\dfrac{x}{x+y}, x+y\right)$

takes $D = (0,\infty)^2$ to $(0,1) \times (0,\infty) = R$

Inverse, $x = uv$, $y = v(1-u)$

Jacobian

$$J(u,v) = \det\begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{pmatrix} = \det\begin{pmatrix} v & u \\ -v & 1-u \end{pmatrix} = v$$

By the Transformation Formula for pdfs,

$$f_{U,V}(u,v) = f_{X,Y}\big(x(u,v), y(u,v)\big)\,|J(u,v)|$$

$$= \lambda^2 e^{-\lambda(x(u,v)+y(u,v))}\,|v|$$

$$= \lambda^2 v e^{-\lambda v} \quad \text{for } (u,v)\in(0,1)\times(0,\infty)$$

This is a product. So U, V are indep.

with $f_U(u) = 1 \qquad u\in(0,1)$

and $f_V(v) = \lambda^2 v e^{-\lambda v} \qquad v\in(0,\infty)$.

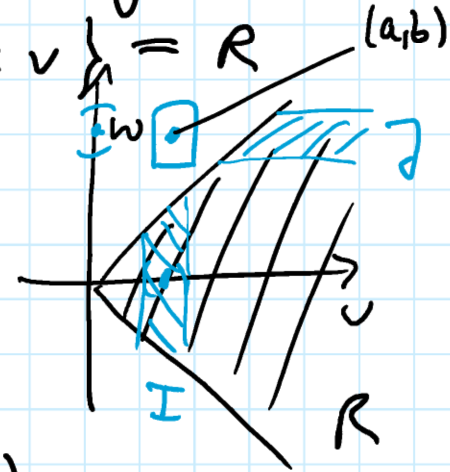
So $\boxed{U \sim \mathrm{Unif}(0,1)}$, $\boxed{V \sim \mathrm{Gamma}(2,\lambda)}$ $\Big\}$ $\boxed{\text{independently}}$

Example 2: Let X, Y indep. $\mathrm{Exp}(\lambda)$.

Now let $V = X+Y$, $W = X-Y$

taking $D = (0,\infty)^2$ bijectively to

$\{(v,w) \in \mathbb{R}^2 : |w| < v\} = R$



Inverse $x = \dfrac{v+w}{2}$

$y = \dfrac{v-w}{2}$

Jacobian

$$J(v,w) = \det \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{pmatrix} = -\frac{1}{2}$$

By the Transformation Formula for pdfs

$$f_{V,W}(v,w) = \begin{cases} f_{X,Y}\left(\frac{v+w}{2}, \frac{v-w}{2}\right) |J(v,w)| & \text{for } |w| < v \\ 0 & \text{o/w} \end{cases}$$

$$= \begin{cases} \frac{1}{2}\lambda^2 e^{-\lambda v}, & |w| < v \\ 0 & o/w \end{cases}$$

$$= \frac{1}{2}\lambda^2 e^{-\lambda v} \mathbb{1}_{\{|w|<v\}}$$

Is this a product? No! Because of the restriction $|w| < v$.

Certainly, V, W are not independent.

$\mathbb{P}(|W| < V) = 1.$

$\mathbb{P}((V,W) \in I \times J) = 0 \neq \mathbb{P}(V \in I)\,\mathbb{P}(W \in J)$

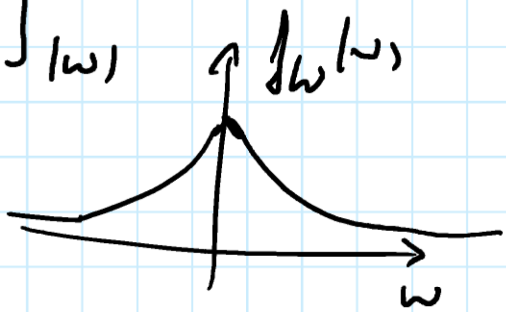
or $f_{V,W}(a,b) = 0 \neq f_V(a) f_W(b).$

From Example 1, $V \sim \text{Gamma}(2, \lambda).$

$$f_W(w) = \int_{|w|}^{\infty} \frac{1}{2}\lambda^2 e^{-\lambda v}\, dv$$

$$= \left[-\frac{1}{2}\lambda^2 e^{-\lambda w} \right]_{|w|}^{\infty}$$

$$= \frac{1}{2}\lambda e^{-\lambda |w|}$$



Dist^n of W is symmetric around 0

Dist^n of $|W|$ has pdf $\lambda e^{-\lambda w}$, $w > 0$

so $|W| \sim Exp(\lambda)$.

The marginal dist^{ns} do not fully describe the joint dist^n. In Example 1 we had <u>independence</u> as the further piece of information. Here, we only have the joint pdf.

Example 3: Sum of two cont. r.v.s

X and Y have joint pdf $f_{X,Y}$.

What is the distr^{n} of $X+Y$?

We can change variables to $U=X+Y$, $V=X$

Jacobian: 1 (check)

By the Transformation Formula for pdfs

$$f_{U,V}(u,v) = f_{X,Y}(v, u-v)$$

Marginal distr^{n} of $U=X+Y$:

$$f_{X+Y}(u) = \int_{-\infty}^{\infty} f_{X,Y}(v, u-v)\, dv$$

Special case: If X, Y indep, "convolution formula":

$$f_{X+Y}(u) = \int_{-\infty}^{\infty} f_X(v) f_Y(u-v)\, dv.$$

Theorem: $D, R \subseteq \mathbb{R}^n$, $T: D \to R$ invertible with diff^{ble} inverse $\left\{ \begin{array}{ll} z = (z_1, \dots, z_n) & (w_1, \dots, w_n) = w \\ \in D & \in R \end{array} \right.$

$$J(w) = \det\left(DT^{-1}(w) \right) = \det\left(\frac{\partial z_i}{\partial w_j} \right)_{1 \le i,j \le n}$$

If $(Z_1, \dots, Z_n)$ jointly continuous, then so are $(W_1, \dots, W_n) = T(Z_1, \dots, Z_n)$, with joint pdf

$$f_W(w) = \begin{cases} f_Z\left(T^{-1}(w)\right) \left| J(w) \right| & w \in R \\ 0 & \text{o/w} \end{cases}$$

Example: Multivariate normal distribution

Let $Z = (Z_1, \dots, Z_n)^T$ be a vector of iid $Z_i \sim N(0,1)$, $i = 1, \dots, n$.

Joint pdf $f_Z(z) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z_i^2}{2}\right)$

$$= \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2} z^T z\right)$$

Define $W_1, \dots, W_n$ by

$$\begin{pmatrix} W_1 \\ \vdots \\ W_n \end{pmatrix} = A \begin{pmatrix} Z_1 \\ \vdots \\ Z_n \end{pmatrix} + \underbrace{\begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix}}_{\mu}$$

where A is an $n \times n$ matrix.

Assume A invertible. Then the Transformation formula, with Jacobian $J(w) = \frac{1}{\det A}$, yields

$$f_W(w) = \frac{1}{(2\pi)^{n/2} |\det A|} \exp\left(-\frac{1}{2} (w-\mu)^T (AA^T)^{-1} (w-\mu)\right)$$

$\Sigma = AA^T$ is the "covariance matrix" of W.

$$\operatorname{Cov}(W_i, W_j) = (AA^T)_{ij}$$

$(W_1, \dots, W_n)$ are multivariate normal with mean vector μ and covariance matrix Σ.

For $n=2$, $(X, Y) = (W_1, W_2)$,

$$f_{X,Y}(x,y) = \frac{1}{2\pi \sigma_X \sigma_Y \sqrt{1-\rho^2}}$$

$$\times \exp\left(-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_X)^2}{\sigma_X^2} - \frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} - \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right)\right)$$

where $\mu_X = \mathbb{E}(X) \qquad \sigma_X^2 = \operatorname{Var}(X)$

$\mu_Y = \mathbb{E}(Y) \qquad \sigma_Y^2 = \operatorname{Var}(Y)$

$\rho = \dfrac{\operatorname{Cov}(X,Y)}{\sigma_X \sigma_Y} \in (-1,1)$ correlation coefficient, $\Sigma = \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}$

4.4.1 Conditioning on events of positive probab.

Motivation: A, B events, $\mathbb{P}(A) > 0$

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

Random variable X:

$$\underbrace{\mathbb{P}(X \leq x \mid A)}_{\text{conditional cdf } F_{X|A}(x)} = \frac{\mathbb{P}(\{X \leq x\} \cap A)}{\mathbb{P}(A)}$$

Discrete case: conditional mass f:

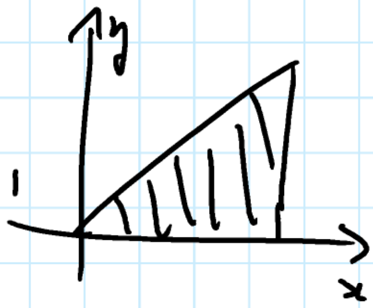
$$p_{X|A}(x) = \mathbb{P}(X = x \mid A)$$

Cont. case: conditional pdf

$f_{X|A}(x)$ s.th.

$$\mathbb{P}(X \leq x \mid A) = \int_{-\infty}^{x} f_{X|A}(u)\, du$$

Example: X, Y uniform distⁿ on the set $\{0 < y < x < 1\}$

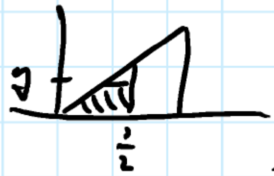


Density $f_{X,Y}(x,y) = \begin{cases} 2, & 0<y<x<1 \\ 0 & \text{o/w} \end{cases}$

What is conditional distⁿ of Y given $X \le \frac{1}{2}$?

For $0 \le y \le \frac{1}{2}$,

$$\mathbb{P}\left(Y \le y \mid X \le \tfrac{1}{2}\right) = \frac{\mathbb{P}\left(Y \le y,\ X \le \frac{1}{2}\right)}{\mathbb{P}\left(X \le \frac{1}{2}\right)}$$

= area() / area()

$$= 1 - 4\left(\tfrac{1}{2} - y\right)^2$$

Conditional density of Y given $X \le \frac{1}{2}$

$$f_{Y|X\le\frac{1}{2}}(y) = \begin{cases} 4-8y & y \in (0,\frac{1}{2}) \\ 0 & o/w \end{cases}$$

Conditional expectation

$$\mathbb{E}(Y \mid X \le \tfrac{1}{2}) = \int_0^{\frac{1}{2}} y\,(4-8y)\,dy = \frac{1}{6}$$

<u>Common situation:</u> two r.v. X and Y.

We observe $X = x$. What does this tell us about Y?

If X is discrete, then $\mathbb{P}(X=x) > 0$, use the approach above

If X is continuous, then $\mathbb{P}(X=x) = 0$:

$$\mathbb{P}(Y \le y \mid X = x) = \frac{\mathbb{P}(Y \le y, X = x)}{\mathbb{P}(X=x)} = \frac{0}{0} = ??$$

Let (X,Y) be jointly continuous.

Idea: To condition Y on $\{X=x\}$ despite $\mathbb{P}(X=x)=0$, look at the dist^{n} of Y conditional on $\{x \le X \le x+\varepsilon\}$ and let $\varepsilon \downarrow 0$.

$$\mathbb{P}(Y \le y \mid x \le X \le x+\varepsilon) = \frac{\int_{v=-\infty}^{y} \int_{u=x}^{x+\varepsilon} f_{X,Y}(u,v)\,du\,dv}{\int_{u=x}^{x+\varepsilon} f_X(u)\,du}$$

$$\sim \frac{\varepsilon \int_{v=-\infty}^{y} f_{X,Y}(x,v)\,dv}{\varepsilon f_X(x)}$$

assuming $f_{X,Y}$ is sufficiently smooth

$$= \int_{v=-\infty}^{y} \frac{f_{X,Y}(x,v)}{f_X(x)}\,dv =: F_{Y|X=x}(y)$$

the conditional cdf of Y given $X=x$

Now we define the conditional pdf of Y given $X=x$ as the integrand

$$\boxed{f_{Y|X=x}(y) = \frac{f_{X,Y}(x,y)}{f_X(x)}}$$

This makes sense whenever $f_X(x) > 0$.

Then $f_{Y|X=x}(y)$ is indeed a density function, i.e. $f_{Y|X=x}(y) \geq 0$ with

$$\int_{-\infty}^{\infty} f_{Y|X=x}(y)\, dy = \frac{\int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy}{f_X(x)} = 1$$

because $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy$.

Observe the resemblance with discrete $\boxed{\mathbb{P}(Y=y \mid X=x) = \frac{\mathbb{P}(Y=y, X=x)}{\mathbb{P}(X=x)}}$

Note $f_{X,Y}(x,y) = f_X(x)\, f_{Y|X=x}(y)$.

<u>Interpretation</u>: The following are equivalent:

(1) generate (X,Y) acc. to $f_{X,Y}$

(2) first generate X acc. to f_X,
then having observed $X=x$,
generate Y acc. to $f_{Y|X=x}$.

<u>Example</u>: In the setting of the previous example, what is the dist$^{\underline{n}}$ of Y given $X=x$?

for $x \in (0,1)$

$$f_{Y|X=x}(y) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

$$= \begin{cases} \frac{2}{f_X(x)} & 0<y<x \\ 0 & \text{o/w} \end{cases}$$

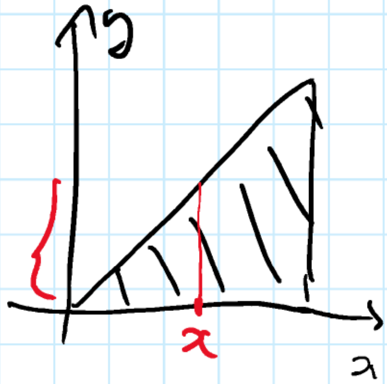
No need to calculate $f_X(x)$, since the conditional density is constant on $(0,x)$.

$\Rightarrow$ Given $X=x$, Y is Uniform $([0,x])$:

$$f_{Y|X=x}(y) = \begin{cases} \frac{1}{x} & 0<y<x \\ 0 & o/w. \end{cases}$$

Conditional mean

$$E(Y|X=x) = \frac{x}{2}$$



Let X, Y be jointly normal with means μ_1 and μ_2, variances σ_1^2 and σ_2^2, correlation $\rho = \frac{\mathrm{Cov}(X,Y)}{\sigma_1 \sigma_2}$.

What is the cond. dist$^{\underline{n}}$ of Y given $X=x$

Approach 1: look at $\frac{f_{X,Y}(x,y)}{f_X(x)}$ directly

...

Approach 2: Let Z_1, Z_2 be iid $N(0,1)$.

Define $X = \alpha Z_1 + \mu_1$

$Y = \beta Z_1 + \gamma Z_2 + \mu_2$

What should α, β, γ be?

We want
$$\Leftarrow \begin{cases} \sigma_1^2 = Var(X) = \alpha^2 \\ \sigma_2^2 = Var(Y) = \beta^2 + \gamma^2 \\ \rho = \dfrac{Cov(X,Y)}{\sigma_1 \sigma_2} = \dfrac{\alpha\beta}{\sigma_1\sigma_2} \end{cases}$$

$\alpha = \sigma_1$

$\beta = \rho\sigma_2$

$\gamma = \sqrt{1-\rho^2}\,\sigma_2$

$$\Rightarrow \begin{cases} X = \sigma_1 Z_1 + \mu_1 \\ Y = \rho\sigma_2 Z_1 + \sqrt{1-\rho^2}\,\sigma_2 Z_2 + \mu_2 \end{cases}$$

$$= \underbrace{\rho\frac{\sigma_2}{\sigma_1}(X-\mu_1)}_{\text{function of } X} + \mu_2 + \underbrace{\sqrt{1-\rho^2}\,\sigma_2 Z_2}_{\text{indep. of } X}$$

Cond. on $X=x$, Y is normal with mean $\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1)+\mu_2$ and variance $(1-\rho^2)\sigma_2^2$

Stochastic process: $X_t : \Omega \to I$ r.v. for each $t \in J$

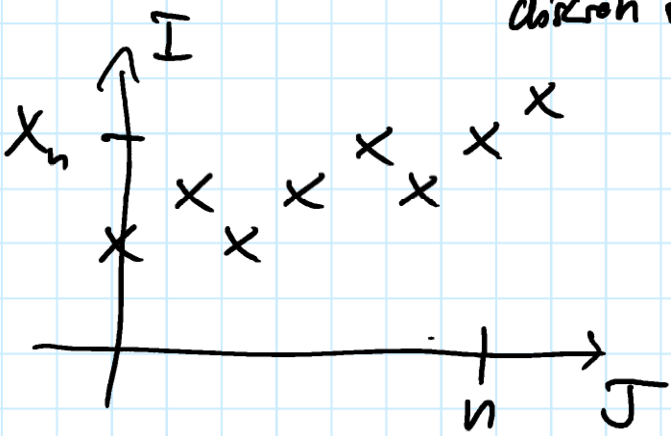
state space, e.g. $\{1,\dots,N\}$, $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Z}^2$, vertices on a graph

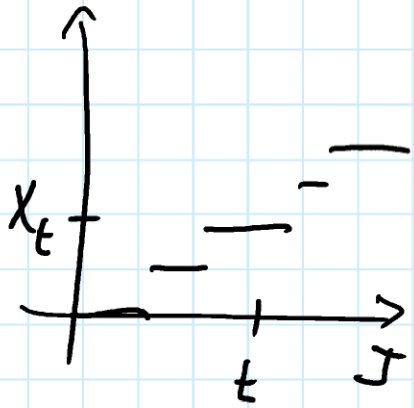
for us: countable I

finite or countably infinite

time set: either $J = \mathbb{N}$ or $J = [0,\infty)$

discrete time   continuous time





<u>Markov chains</u>: I countable

A (probability) distribution on I is

$\lambda = (\lambda_i, i \in I)$ with $\begin{cases} \lambda_i \geq 0 & \forall i \in I \\ \sum_{i \in I} \lambda_i = 1 \end{cases}$

We will often think of λ as a row vector.

Y has dist$^{\text{n}}$ λ if $\mathbb{P}(Y=i) = \lambda_i \quad \forall i \in I$.

<u>Def$^{\text{n}}$</u>: $X = (X_0, X_1, X_2, \dots) = (X_n, n \geq 0)$

Markov chain if for every $n \geq 0$ and every

$i_0, \dots, i_{n+1} \in I$,

$$\boxed{\mathbb{P}(X_{n+1} = i_{n+1} \mid X_n = i_n, \cancel{\dots, X_0 = i_0}) = \mathbb{P}(X_{n+1} = i_{n+1} \mid X_n = i_n).}$$

This MC is called time-homogeneous if

in addition, $\mathbb{P}(X_{n+1}=j \mid X_n=i)$ depends on i and j, but not on n.

Thus we write $p_{ij} = \mathbb{P}(X_{n+1}=j \mid X_n=i)$ and refer to p_{ij} as "transition probability"

$p_{ij} = p_{i,j}$

We'll only work with time-homogeneous MCs.

To specify the joint dist$^{\underline{n}}$ of $(X_n, n\geq 0)$, we have specified

(a) the initial dist$^{\underline{n}}$ of X_0: $\lambda_i = \mathbb{P}(X_0=i)$, $i\in I$

(b) the transition matrix $P = (p_{ij})_{i,j\in I}$

P is square (maybe infinite)

rows and columns indexed by I

P is a stochastic matrix:

- all entries ≥ 0
- every row sums to 1

i.e. every row is a distⁿ

The i^{th} row is the conditional distⁿ of X_{n+1} given $X_n = i$

<u>Theorem</u>: $\mathbb{P}_\lambda(X_0 = i_0, X_1 = i_1, \ldots, X_{n-1} = i_{n-1}, X_n = i_n)$

$$= \lambda_{i_0} p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{n-1} i_n}$$

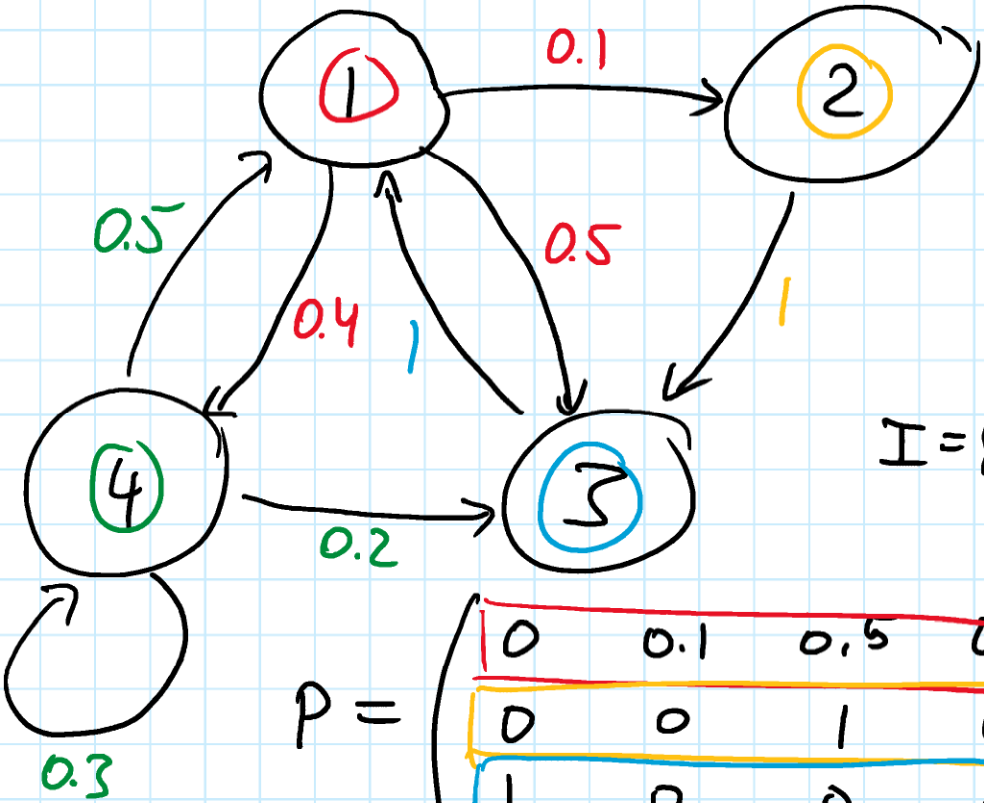
<u>Notation</u>: subscript $\mathbb{P}_\lambda$ to indicate initial distⁿ; also $\mathbb{P}_i$ if $\lambda_i = 1$, $\lambda_j = 0$ $\forall j \neq i$.

<u>Proof</u>: $\mathbb{P}(X_0 = i_0, \ldots, X_n = i_n)$

$$= \mathbb{P}(X_0 = i_0)\mathbb{P}(X_1 = i_1 | X_0 = i_0) \cdots \mathbb{P}(X_n = i_n | \cancel{X_0 = i_0, \ldots,} X_{n-1} = i_{n-1})$$

$$= \lambda_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i_n} \quad \text{since } (X_n, n \geq 0) \text{ MC.} \ \square$$

Example 1: Frogs on lily pads



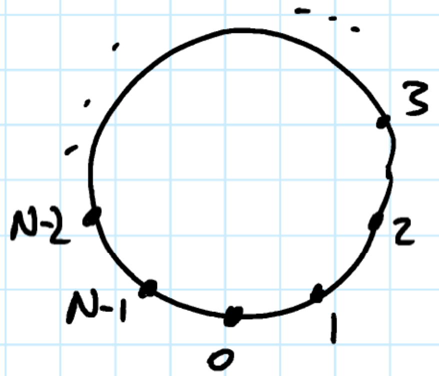
$I = \{1, 2, 3, 4\}$

$$P = \begin{pmatrix} 0 & 0.1 & 0.5 & 0.4 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0.5 & 0 & 0.2 & 0.3 \end{pmatrix}$$

$\lambda = (0.5, 0.5, 0, 0)$ initial dist^{2}

$$\mathbb{P}(X_0 = 1, X_1 = 2) = \lambda_1 p_{12} = \frac{1}{2}\frac{1}{10} = \frac{1}{20}$$

Example 2: Random walk on a cycle



$I = \{0, \ldots, N-1\}$

at each step

go anti-clockwise w.p. p

clockwise w.p. $1-p$

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

better description of $P = (p_{ij})$

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

Example 3: Simple symmetric random on $\mathbb{Z}^d$

At each step, move to a randomly chosen neighbour: (a) $p_{ij} = \begin{cases} \frac{1}{2d} & \text{if } |i-j| = 1 \\ 0 & \text{otherwise} \end{cases}$

or (b) $p_{ij} = \begin{cases} \frac{1}{2^d} & \text{if } |i_k - j_k| = 1 \text{ for all } 1 \le k \le d \\ 0 & \text{otherwise} \end{cases}$

where $i = (i_1, \dots, i_d)$, $j = (j_1, \dots, j_d)$

<u>Markov property</u>: $\underbrace{(X_0,\dots,X_n)}_{\text{past}} \overset{\text{present}}{\underset{X_n=i}{\perp\!\!\!\perp}} \underbrace{(X_n, X_{n+1},\dots)}_{\text{future}}$

"is conditionally indep. given $X_n=i$, of"

i.e. $\mathbb{P}\big((X_n,X_{n+1},\dots)\in B \mid X_n=i,\ \cancel{(X_0,\dots,X_n)\in A}\big)$

$$= \mathbb{P}_i\big((X_0,X_1,\dots)\in B\big)$$

This generalises the def^{n} of the MC, where

$$B = I\times\{i_{n+1}\}\times I^{\mathbb{N}}, \quad A=\{(i_0,\dots,i_n)\}.$$

<u>Sketch proof</u> of Markov property

In fact, it suffices to show

$$\mathbb{P}(X_{n+1}\in A_{n+1},\dots,X_{n+m}\in A_{n+m} \mid X_0\in A_0,\dots,X_{n-1}\in A_{n-1}, X_n=i)$$

$$= \mathbb{P}_i\,(X_1\in A_{n+1},\dots,X_m\in A_{n+m})$$

For this, if $A_k=\{i_k\}$, apply the Theorem on p. 70.

For general A_k, sum over i_k for $k \geq n+1$

For general A_k, $k \leq n$, prove and apply

If $\mathbb{P}(E \mid F_1) = \mathbb{P}(E \mid F_2) = \mathbb{P}(E \mid G)$

for $F_1, F_2 \subseteq G$

then $\mathbb{P}(E \mid F_1 \cup F_2) = \mathbb{P}(E \mid G)$. $\square$

<u>n-step probabilities</u>: Write $p_{ij}^{(n)} = \mathbb{P}(X_{m+n} = j \mid X_m = i)$

<u>Theorem:</u> Chapman-Kolmogorov equation

(i) $p_{ik}^{(n+m)} = \sum_{j \in I} p_{ij}^{(n)} p_{jk}^{(m)}$

(ii) $p_{ij}^{(n)} = (P^n)_{ij}$ where $P = (p_{ij}) = (p_{ij}^{(1)})$

is the transition matrix

Proof: (i) $\mathbb{P}(X_{n+m}=k \mid X_0=i)$

$$= \sum_{j\in I} \mathbb{P}(X_{n+m}=k, X_n=j \mid X_0=i)$$

$$= \sum_{j\in I} \mathbb{P}(X_n=j \mid X_0=i)\, \mathbb{P}(X_{n+m}=k \mid \cancel{X_0=i}, X_n=j)$$

by the Markov property

$$= \sum_{j\in I} p_{ij}^{(n)}\, p_{jk}^{(m)}$$

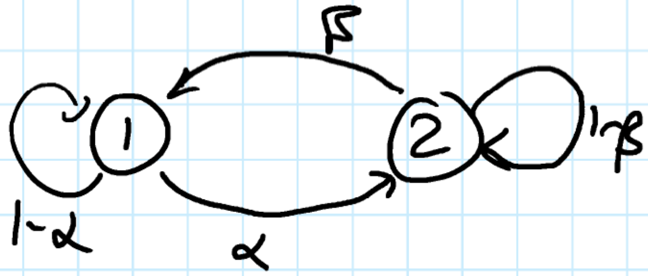
(ii) By induction

$$p_{ik}^{(2)} = \sum_j p_{ij}\, p_{jk} = (P^2)_{ik}$$

$$p_{ik}^{(n+1)} = \sum_j p_{ij}^{(n)}\, p_{jk} = \sum_j (P^n)_{ij}\, p_{jk} = (P^{n+1})_{ik}$$

$\square$

$$P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$$



What is $p_{11}^{(n)}$?

P has eigenvalues 1 and $1-\alpha-\beta$ (check!)

So $P = U \begin{pmatrix} 1 & 0 \\ 0 & 1-\alpha-\beta \end{pmatrix} U^{-1}$

$$\Rightarrow P^n = \boxed{U} \begin{pmatrix} 1 & 0 \\ 0 & (1-\alpha-\beta)^n \end{pmatrix} \boxed{U}^{-1}$$

Get $p_{11}^{(n)} = (P^n)_{11} = A + B(1-\alpha-\beta)^n$ for some A, B

We know $p_{11}^{(0)} = 1$
and $p_{11}^{(1)} = 1-\alpha$

Solve for A, B: $p_{11}^{(n)} = \frac{\beta}{\alpha+\beta} + \frac{\alpha}{\alpha+\beta}(1-\alpha-\beta)^n, \; n \geq 0.$

Example, when the Markov property fails?

Let X_i iid $\mathbb{P}(X_i = 1) = p$, $\mathbb{P}(X_i = -1) = 1-p$

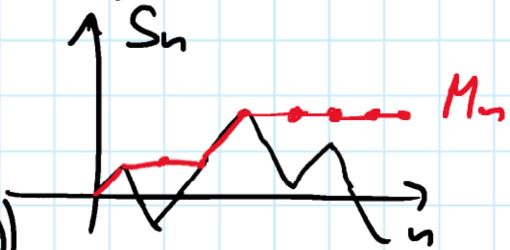
Let $S_0 = 0$, $S_n = \sum_{i=1}^{n} X_i$ simple RW on $\mathbb{Z}$

(1) (X_i) is a MC

(2) (S_n) is a MC

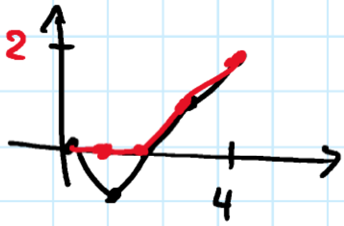
We'll get back to this.

(3) $M_n = \max\{S_m,\ 0 \le m \le n\}$



$a_1 = \mathbb{P}(M_5 = 3 \mid (M_0, \dots, \boxed{M_4}) = (0,0,0,1,\boxed{2}))$

$= \mathbb{P}(X_5 = 1) = p$

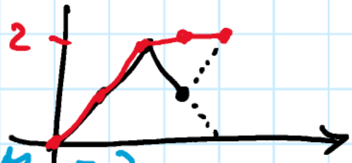


$a_2 = \mathbb{P}(M_5 = 3 \mid (M_0, \dots, \boxed{M_4}) = (0,1,2,2,\boxed{2}))$

either $\boxed{S_4 = 0}$ or $S_4 = 2$

$< \mathbb{P}(X_5 = 1) = p$

not enough for $M_5 = 3$



Hence a_1 and a_2 cannot both equal

$$\mathbb{P}(M_5 = 3 \mid M_4 = 2)$$

and so the $\boxed{\text{Markov property (cf p.68)}}$ fails

The path of $(M_n, n \geq 0)$ to $M_4 = 2$ is relevant for the next step, so the future and past are not conditionally independent given the present.

Proposition: Suppose that for each n we can write

$$\boxed{Y_{n+1} = f(Y_n, X_{n+1})}$$

where X_{n+1} is independent of $(Y_0, \dots, Y_n)$

Then (Y_n) is a MC

Proof: $\mathbb{P}(Y_{n+1} = i_{n+1} \mid Y_n = i_n, \dots, Y_0 = i_0)$

$$= \mathbb{P}(\underbrace{f(i_n, X_{n+1}) = i_{n+1}}_{} \mid \underbrace{Y_n = i_n, \dots, Y_0 = i_0}_{})$$

indep.

$$= \mathbb{P}\big(f(i_n, X_{n+1}) = i_{n+1}\big)$$
$$= \mathbb{P}\big(f(i_n, X_{n+1}) = i_{n+1} \mid Y_n = i_n\big)$$
$$= \mathbb{P}\big(Y_{n+1} = i_{n+1} \mid Y_n = i_n\big) \qquad \square$$

In the example, this shows that

- $(X_n, n\geq 0)$ is a MC using $f(y,x) = x$
  since X_{n+1} is indep. of $(X_0, \dots, X_n)$
- $(S_n, n\geq 0)$ is a MC using $f(y,x) = y+x$
  then $S_{n+1} = S_n + X_{n+1}$
  since X_{n+1} is indep. of $(X_0, \dots, X_n)$
  hence of $(S_0, \dots, S_n)$.

$$p_{i,i+1} = p$$
$$p_{i,i-1} = 1-p$$

For (X_n): $P = \begin{array}{c|cc} & -1 & 1 \\ \hline -1 & 1-p & p \\ 1 & 1-p & p \end{array}$

$I = \{-1, 1\}$

Let $i, j \in I$. "i leads to j" or "$i \to j$"

if $\mathbb{P}_i(X_n = j) = p_{ij}^{(n)} > 0$ for some $n \geq 0$

not necessarily in one step!

$(i) \to (j)$ or $(i) \to \circ \to \circ \to \circ \to (j)$

If $i \to j$ and $j \to i$ then we say

"i communicates with j"

This is an equivalence relation.

It partitions I into <u>communicating classes</u>

A chain (or transition matrix) for which I

is a single communicating class is called

<u>irreducible</u>. Equivalently, $i \to j \;\; \forall i, j \in I$

A class C is called <u>closed</u> if

$$p_{ij} = 0 \quad \text{whenever} \quad i \in C \text{ and } j \notin C$$

i.e. $i \not\to j$ if $i \in C$, $j \notin C$

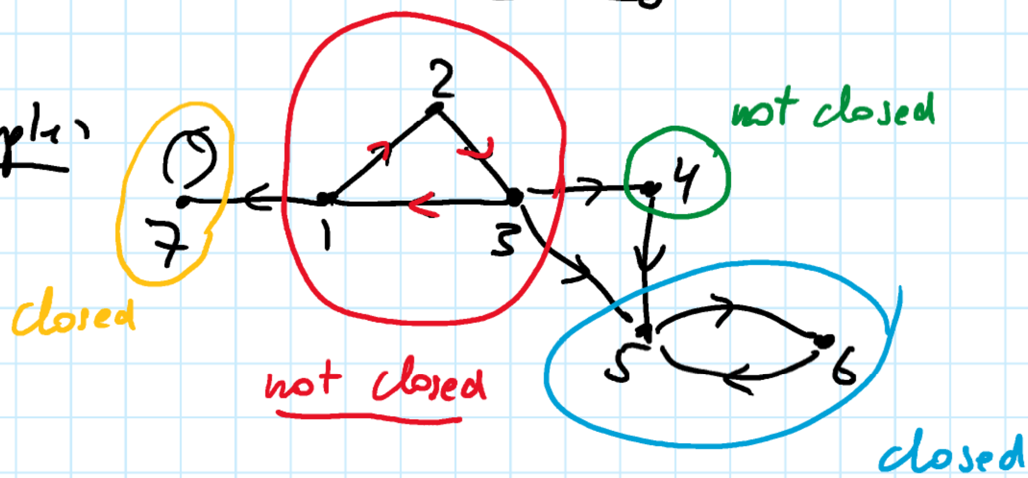
not even in several steps 8

There is no escape from a closed class.

Singleton $C = \{i\}$ closed $\Rightarrow p_{ii} = 1$

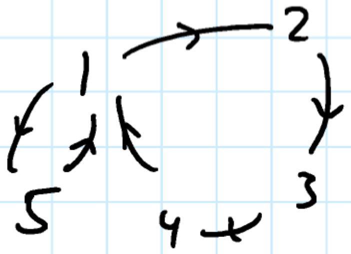
i <u>absorbing</u> state.

<u>Example</u>:



Classes $\{1,2,3\}$, $\{4\}$, $\{5,6\}$, $\{7\}$

$$P = \begin{pmatrix} 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ \frac{1}{3} & 0 & 0 & \frac{1}{3} & \frac{1}{3} & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$



$p_{11}^{(n)} = 0$ whenever n odd

$p_{11}^{(n)} > 0$ whenever n even

$p_{33}^{(n)} = 0$ whenever n odd

$\boxed{p_{33}^{(2)} = 0}$, $p_{33}^{(4)} > 0$, $p_{33}^{(6)} > 0$, ...

Defⁿ: The period of i is $\gcd(\{n \geq 1 : p_{ii}^{(n)} > 0\})$

if $\exists_{n \geq 1} : p_{ii}^{(n)} > 0$

o/w period undefined.

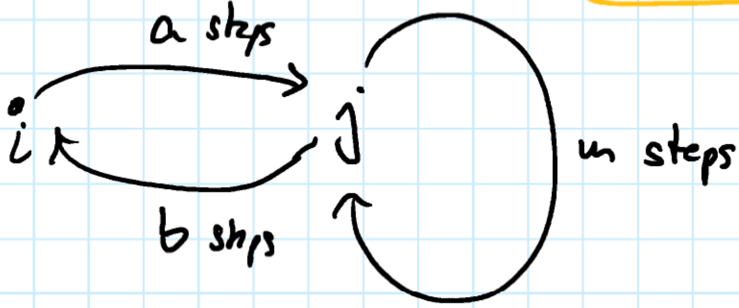
i aperiodic if period is 1.

In the example, the period is 2 for $i = 1$

and $\boxed{\text{for } i = 3}$

Proposition: All states in a communicating class have the same period.

Proof: Suppose $i \longleftrightarrow j$ and $\boxed{d \mid n \text{ whenever } p_{ii}^{(n)} > 0.}$



Find a and b with $p_{ij}^{(a)} > 0$ and $p_{ji}^{(b)} > 0$

Suppose m: $p_{jj}^{(m)} > 0$. Then also

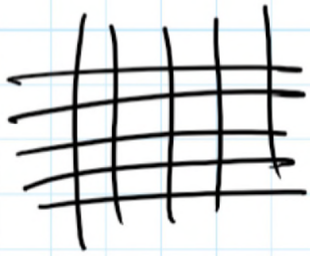
$$p_{ii}^{(a+m+b)} \geq p_{ij}^{(a)} \, p_{jj}^{(m)} \, p_{ji}^{(b)} > 0$$

$$\left.\begin{array}{l} \text{Hence } d \mid (a+m+b) \\ \text{Similarly, } d \mid (a+b) \end{array}\right\} \Rightarrow d \mid m \text{ whenever } p_{jj}^{(m)} > 0 \,!$$

From this and from swapping i and j, $\gcd(\{n : p_{ii}^{(n)} > 0\}) = \gcd(\{n : p_{jj}^{(n)} > 0\})$ $\square$

Example(s) 1) SRW on $\mathbb{Z}^2$

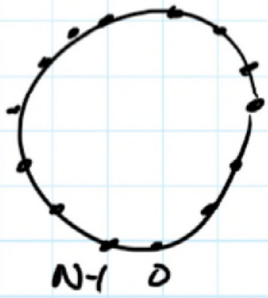
period 2



2) RW on a cycle

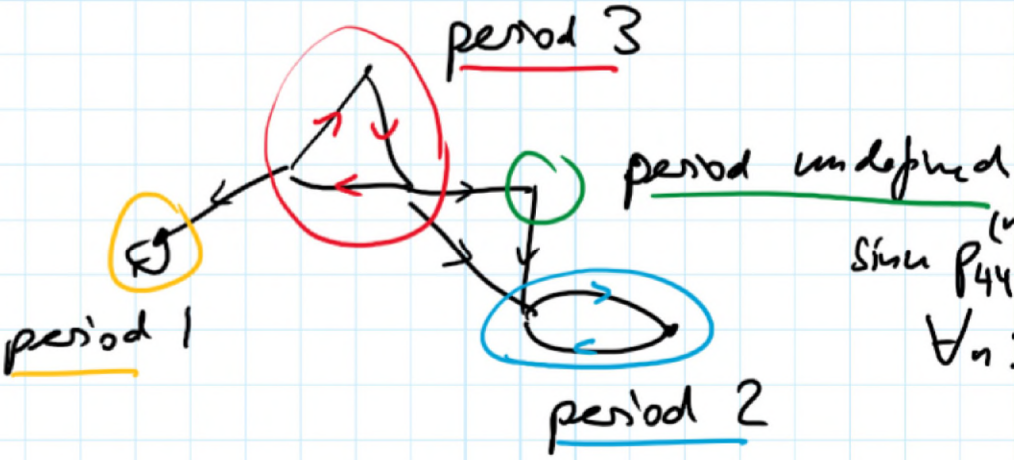
If N odd, period 1

If N even, period 2



$p \in (0,1)$

3)



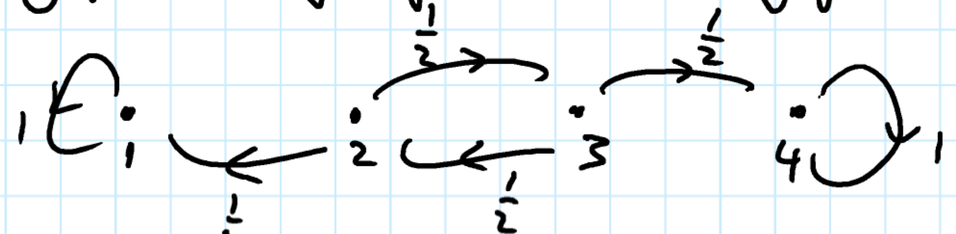
since $p_{44}^{(n)} = 0$ $\forall n \geq 1$.

Let (X_n) be a MC. Let $A \subseteq I$. Define

$$h_i^A = \mathbb{P}_i\left(X_n \in A \ \boxed{\text{for some } n \geq 0}\right)$$

the "hitting probability" of set A starting from i.

Examples:



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

Starting from 2, what is the probability of absorption in 4?

Let $h_i = \mathbb{P}_i(\text{reach } 4)$. Then $\boxed{h_1 = 0}$, $h_4 = 1$.

Also $\boxed{\text{informally}}$,

$$\boxed{\begin{cases} h_2 = \frac{1}{2} h_1 + \frac{1}{2} h_3 \\ h_3 = \frac{1}{2} h_2 + \frac{1}{2} h_4 \end{cases}}$$

Solve $h_2 = \frac{1}{2} h_3 = \frac{1}{4} h_2 + \frac{1}{4} \Rightarrow h_2 = \frac{1}{3}$

(also $h_3 = \frac{2}{3}$)

__Theorem:__ The vector of hitting probabilities $(h_i^A, i\in I)$ is the minimal non-negative solution to

$$h_i^A = \begin{cases} 1 & \text{if } i\in A \\ \sum_{j\in I} p_{ij} h_j^A & \text{if } i\notin A \end{cases} \qquad (*)$$

"Minimal" means that if $(x_i, i\in I)$ is another non-negative sol^n of $(*)$, then $x_i \geq h_i^A$ $\forall i\in I$.

__Proof:__ Certainly, if $i\in A$, then $h_i^A = 1$.

If $\boxed{i\notin A}$, then

$$\begin{aligned} h_i^A &= \mathbb{P}_i(X_n\in A \text{ for some } n\geq 0) \\ \text{condition on the first step } &= \mathbb{P}_i(X_n\in A \text{ for some } n\geq 1) \\ &= \sum_{j\in I} \mathbb{P}_i(\underline{X_1}=j)\, \mathbb{P}(X_n\in A \text{ for some } n\geq 1 \mid \cancel{X_0=i}, \underline{X_1=j}) \\ \text{Markov prop. } &= \sum_{j\in I} p_{ij}\, \mathbb{P}_j(X_n\in A \text{ for some } n\geq 0) \end{aligned}$$

$$= \sum_{j \in I} p_{ij} h_j^A \,.$$ So indeed, (*) holds.

To prove minimality, suppose $(x_i, i \in I)$ is any non-negative sol^{n} to (*). We want $h_i^A \le x_i \ \forall i \in I$.

<u>Claim:</u> for all $M \in \mathbb{N}$, and all $i \in I$, (**)

$$x_i \ge \mathbb{P}_i \left(X_n \in A \text{ for all } n \le M \right)$$

Proof by induction on M

Base case: $M=0$: for $i \in A$, LHS $=1$, for $i \notin A$, RHS $=0$

Induction step: Suppose for all $j \in I$

$$x_j \ge \mathbb{P}_j \left(X_n \in A \text{ for some } n \le M-1 \right)$$

Now we want (**).

If $i \in A$, then $x_i = 1$ and (**) is true

If $i \notin A$,

$$\mathbb{P}_i(X_n \in A \text{ for some } n \le M)$$

cond on 1st step + MP

$$= \sum_{j \in I} p_{ij} \, \mathbb{P}_j(X_n \in A \text{ for some } n \le M-1)$$

ind hyp.

$$\le \sum_{j \in I} p_{ij} \, x_j = x_i \quad \text{and } (**) \text{ holds}$$

Now

$$x_i \ge \lim_{M \to \infty} \mathbb{P}_i(X_n \in A \text{ for some } n \le M)$$

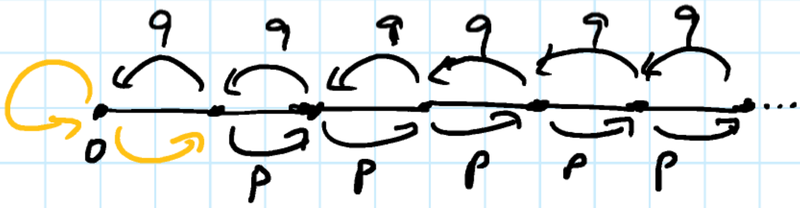
Lemma on unions of increasing seq. of events (p.13)

$$= \mathbb{P}_i\left(\bigcup_{M \ge 1} \{X_n \in A \text{ for some } n \le M\}\right)$$

$$= \mathbb{P}_i(X_n \in A \text{ for some } n \ge 0)$$

$$= h_i^A$$

$I = \{0, 1, 2, \dots\}$



$p_{00} = 1$

$$\left.\begin{array}{l} p_{i,i+1} = p \\ p_{i,i-1} = q = 1-p \end{array}\right\} i \geq 1, \text{ for } p \in (0,1)$$

What is the probability of hitting 0 starting from i?

Let $h_i = \mathbb{P}_i(\text{hit } 0)$.

~~By~~ Thus, we need the minimal non-negative sol^{n} of

$$\begin{cases} h_0 = 1 & (1) \\ h_i = p h_{i+1} + q h_{i-1}, \quad i \geq 1 & (2) \end{cases}$$

If $p \neq q$, (2) has the general solution

$$h_i = A + B\left(\frac{q}{p}\right)^i \qquad \text{check!}$$

If $p = q$, then $h_i = A + Bi$

Cases: $\underline{p<q}$. From (1), $A+B=1$

Since $\frac{q}{p}>1$, $B\neq 0$ is impossible for $h_i\in[0,1]$,

so $B=0$, $A=1$, so $h_i=1$.

$\underline{p>q}$. Again $A+B=1$

Now $\left(\frac{q}{p}\right)^i \xrightarrow[i\to\infty]{} 0$, so $A\geq 0$

for non-negativity

Then we want $A=0$, $B=1$ for $\boxed{\text{minimality}}$

(since $1\geq \left(\frac{q}{p}\right)^i$)

There are plenty of other sol$^{\underline{ns}}$, for any $A\in[0,1]$
$B=1-A$

Here $h_i=\left(\frac{q}{p}\right)^i \neq 1$!

$\underline{p=q}$: From (1), $A=1$

Since $B<0 \Rightarrow A+Bi \xrightarrow[i\to\infty]{} -\infty$, $B=0$ minimal $\Bigg\}\ h_i=1$

Starting from i, what is the chance of returning to i?

Two possibilities:

(1) $\mathbb{P}_i(X_n = i \boxed{\text{for some } n \geq 1}) = p < 1.$

Using the Markov property at n s.th. $X_n = i$, # visits to i ~ geom$(1-p)$, since each time we visit, we have a probab. of $1-p$ of never returning. Then $\mathbb{P}_i(\text{hit } i \text{ infinitely often}) = 0.$

State i is called <u>transient</u>.

(2) $\mathbb{P}_i(X_n = i \boxed{\text{for some } n \geq 1}) = 1.$ Then

$\mathbb{P}_i(\text{hit } i \text{ infinitely often}) = 1.$

State i is called <u>recurrent</u>.

Typically, $\mathbb{P}_i(X_n = i) < 1$ $\boxed{\text{for all } n \geq 1}$, even if i recurrent

<u>Theorem:</u> i is recurrent $\iff \sum_{n=0}^{\infty} p_{ii}^{(n)} = \infty$

<u>Proof:</u> The total number of visits to i is

$$\sum_{n=0}^{\infty} \mathbb{1}_{\{X_n = i\}}$$

which has expectation

$$\sum_{n=0}^{\infty} \mathbb{E}\left(\mathbb{1}_{\{X_n=i\}}\right) = \sum_{n=0}^{\infty} p_{ii}^{(n)}.$$

If i is transient, # visits to i is geom(r_p) with mean $\frac{1}{1-p} < \infty$.

If i is recurrent, # visits to i is infinite w.p. 1, so has mean ∞.

Hence,

$$i \text{ recurrent} \iff \sum_{n=0}^{\infty} p_{ii}^{(n)} = \infty$$

$\square$

Proposition: (1) Let C be a communicating class. Either all states in C are recurrent, or all states in C are transient, i.e. recurrence and transience are class properties (and we may call the class recurrent/transient).

(2) Every recurrent class is closed.

(3) Every finite closed class is recurrent.

The proof is an exercise.

Simple random walk $I = \mathbb{Z}$, $p_{i,i+1} = p$, $p_{i,i-1} = q = 1-p$

Look at $\sum_{n=0}^{\infty} p_{00}^{(n)}$. First consider $p = q = \frac{1}{2}$.

We will use Stirling's formula $n! \sim \sqrt{2\pi}\, n^{n+\frac{1}{2}} e^{-n}$

where $a_n \sim b_n$ means $\frac{a_n}{b_n} \to 1$ as $n \to \infty$

If n odd, $p_{00}^{(n)} = 0$ (by periodicity)

If $n = 2m$, to return to 0 in $2m$ steps,

we need m up-steps and m down-steps.

$$p_{00}^{(2m)} = \binom{2m}{m} \cancel{\left(\tfrac{1}{2}\right)^{2m}} (pq)^m$$

more generally for $p \neq q$

$$= \frac{(2m)!}{m!\,m!} \cancel{\left(\tfrac{1}{2}\right)^{2m}} (pq)^m$$

$$\sim \frac{1}{\sqrt{\pi}} \frac{1}{m^{1/2}} (4pq)^m$$

So $\sum_{n=0}^{\infty} p_{00}^{(n)} = \sum_{m=0}^{\infty} p_{00}^{(2m)} = \infty$ if $p = q = \frac{1}{2}$

$< \infty$ if $p \neq q$

By the theorem on p. 94, SRW recurrent if $p=q=\frac{1}{2}$

transient if $p\neq q$.

<u>d-dimensional RW</u>: Cf. p. 73

<u>either</u>: $2d$ directions ... notes

<u>or</u>: 2^d directions: $X^{(1)},\dots,X^{(d)}$ indep. SSRW

$\underline{0}=(0,\dots,0)$ $\qquad p^{(2n+1)}_{\underline{0},\underline{0}}=0$

$$p^{(2n)}_{\underline{0}\underline{0}}=\left(p^{(2n)}_{00}\right)^d\sim\left(\frac{1}{\sqrt{\pi}}\frac{1}{\sqrt{n}}\right)^d=\frac{1}{\pi^{d/2}}\frac{1}{n^{d/2}}$$

Hence $\displaystyle\sum_{n=0}^{\infty}p^{(n)}_{\underline{0},\underline{0}}=\infty$ for $d=2$ $\left(\sum\frac{1}{n}=\infty\right)$

recurrent δ

$\displaystyle\sum_{n=0}^{\infty}p^{(n)}_{\underline{0},\underline{0}}<\infty$ for $d\geq3$ $\left(\sum\frac{1}{n^{d/2}}<\infty\right)$

transient δ

Let $H^A = \inf\{n \geq 0 : X_n \in A\}$ hitting time of $A \subseteq I$

Note $H^A = \infty$ is possible. In fact $h_i^A = \mathbb{P}_i(H^A < \infty)$.

Let $k_i^A = \mathbb{E}_i(H^A)$ mean hitting time of A from i

If $h_i^A < 1$, then $\mathbb{P}_i(H^A = \infty) > 0$, so $k_i^A = \infty$.

But also maybe $k_i^A = \infty$ when $h_i^A = 1$!

Theorem: The vector of mean hitting times $(k_i^A, i \in I)$ is the minimal nonnegative solution to

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

Sketch proof: For $i \notin A$:

$$k_i^A = \mathbb{E}_i(H^A) = \sum_{j \in I} \mathbb{E}_i(H^A \mid X_1 = j) \mathbb{P}_i(X_1 = j)$$

$$= \sum_{j\in I} p_{ij}\left(1+k_j^A\right) = 1+\sum_{j\in I} p_{ij} k_j^A$$

Minimality is also seen similarly.

$\square$

<u>Mean return time to i</u>

$$m_i = \mathbb{E}_i\left(\inf\{n \geq 1 : X_n = i\}\right)$$
$$= 1 + \sum_{j\in I} p_{ij}\, k_j^{\{i\}} \qquad \left(\neq k_i^{\{i\}} = 0\right)$$

Then i transient $\Rightarrow m_i = \infty$.

If i recurrent

if $m_i = \infty$, we say i is <u>null recurrent</u>

if $m_i < \infty$, we say i is <u>positive recurrent</u>

Null recurrence / positive recurrence are class properties

If the chain is irreducible, we call the whole chain transient / null recurrent / positive recurrent.

Example: Gambler's ruin-like chain

$p_{01} = 1$, $p_{i,i+1} = p \in (0,1)$, $p_{i,i-1} = q = 1-p$, $i \geq 1$.

Let k_i = mean time to hit 0 from i

Then $\mathbb{E}_i(\text{time to } 0) = \sum_{j=1}^{i} \underbrace{\mathbb{E}_j(\text{time to } j-1)}_{= k_1}$

$= i k_1$

Need $k_i = 1 + q k_{i-1} + p k_{i+1}$

then $i k_1 = 1 + q(i-1)k_1 + p(i+1)k_1$

$\Rightarrow (q-p) k_1 = 1$ (*) or $k_1 = \infty$

$p > q$: $h_i < \infty \Rightarrow k_i = \infty$ transient

$p < q$: we get, by minimality, $k_i = \frac{1}{q-p} i$

$p = q$: $h_i = 1$, but (*) positive recurrent

has no sol^{n}, so $k_i = \infty$, so null recurrent

<u>Def$^{\underline{n}}$</u>: A dist$^{\underline{n}}$ π on I is called $\boxed{\underline{\text{stationary}}}$ for a MC if $X_0 \sim \pi \Rightarrow \boxed{X_n \sim \pi \quad \forall n \geq 0.}$

Recall, dist$^{\underline{ns}}$ on I are row vectors

<u>Proposition</u>: For a MC with transition matrix P,

$$X_0 \sim \lambda \quad \Rightarrow \quad X_n \sim \lambda P^n$$

<u>Proof</u>: $\mathbb{P}(X_n = j) = \sum_{i \in I} \underbrace{\mathbb{P}(X_0 = i)}_{\lambda_i} \underbrace{\mathbb{P}(X_n = j \mid X_0 = i)}_{p_{ij}^{(n)}}$

$$= \sum_{i \in I} \lambda_i p_{ij}^{(n)}$$

$$= (\lambda P^n)_j \qquad \square$$

invariant dist$^{\underline{n}}$

<u>Corollary</u>: π stationary $\iff \boxed{\pi P = \pi}$

i.e. $\pi_j = \sum_{i \in I} \pi_i p_{ij} \quad \forall j \in I$

i.e. π is a left eigenvector of P with eigenvalue 1.

X MC, $P = (p_{ij})_{i,j\in I}$ transition matrix

IF X, P irreducible $\iff \forall_{i,j\in I}\, \exists_{n\geq 0}\ p_{ij}^{(n)} > 0$

X, P aperiodic $\iff \forall_{i\in I}$ ($\iff \exists i\in I$) $\gcd\{n\geq 1 : p_{ii}^{(n)} > 0\} = 1$

X, P recurrent $\iff \forall_{i\in I}$ ($\iff \exists i\in I$) $\mathbb{P}_i(\exists_{n\geq 1} X_n = i) = 1$

$$= \mathbb{P}_i(\min\{n\geq 1 : X_n = i\} < \infty) = 1$$

X, P positive recurrent $\iff \forall_{i\in I}$ ($\iff \exists i\in I$) $\mathbb{E}_i(\min\{n\geq 1 : X_n = i\}) < \infty$

$$= m_i$$

Theorems: (1) P irred. $\Rightarrow$ (a) $\exists \pi \iff P$ pos. rec.

(b) $\exists \pi \Rightarrow \pi$ unique, $\pi_j = \frac{1}{m_j}$

Convergence Thm

(2) $\left.\begin{array}{l} P \text{ irred} \\ P \text{ aperiodic} \\ \exists \pi \end{array}\right\} \Rightarrow p_{ij}^{(n)} \xrightarrow[n\to\infty]{} \boxed{\pi_j}$

Ergodic Theorem

(3) P irred $\Rightarrow$ $\dfrac{\#\{k=0,\dots,n-1 : X_k = j\}}{n} \xrightarrow[n\to\infty]{\text{a.s.}} \boxed{\dfrac{1}{m_j}}$

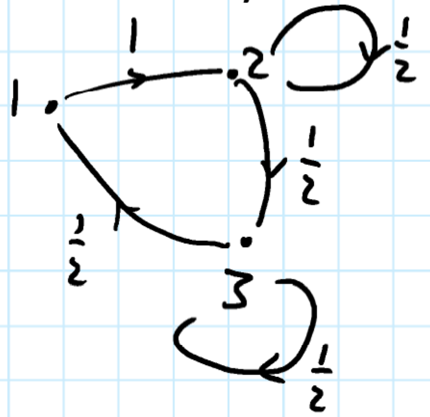
where $\dfrac{1}{m_j} := 0$ if $m_j = \infty$.

Interpretation: (3) long-term proportion of time that X spends in j is $\approx \frac{1}{m_j}$

(2) for large n, $X_n \sim \pi$ approx., i.e. $\mathbb{P}(X_n = j) \simeq \pi_j$ and $(X_m, m \geq n)$ is approx. stationary chain

(1) limits in (2) and (3) are the same.

Example 1:



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

For π to be stationary, we want $\boxed{\pi P = \pi}$, i.e.

$$\begin{cases} \pi_1 = \frac{1}{2}\pi_3 \\ \pi_2 = \pi_1 + \frac{1}{2}\pi_2 \\ \pi_3 = \frac{1}{2}\pi_2 + \frac{1}{2}\pi_3 \end{cases}$$

Any one of them is redundant.

Include $\boxed{\pi_1 + \pi_2 + \pi_3 = 1}$ to normalise

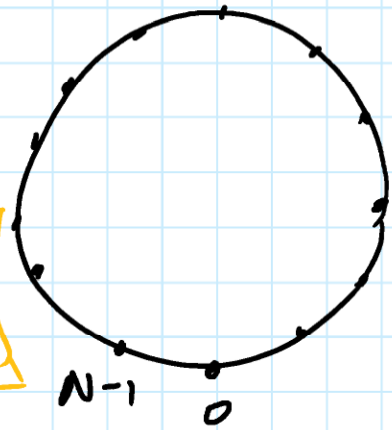
Obtain $(\pi_1, \pi_2, \pi_3) = \left(\frac{1}{5}, \frac{2}{5}, \frac{2}{5}\right)$

So (1) $(m_1, m_2, m_3) = \left(5, \frac{5}{2}, \frac{5}{2}\right)$

(3) $\Rightarrow$ long-term proportion of time in 1 is $\pi_1 = \frac{1}{5}$.

(2) Since MC is irreducible & aperiodic, $p_{11}^{(n)} \xrightarrow[n\to\infty]{} \frac{1}{5}$.

<u>Example 2</u>: $p = \frac{1}{2}$



$p_{ij} = \frac{1}{2}$ if $i-j = \pm 1 \mod N$

Uniform dist$^{\underline{n}}$ $\boxed{\pi_i = \frac{1}{N} \text{ stationary by symmetry}}$

<u>or</u> $\pi_i = \frac{1}{2}\pi_{i+1} + \frac{1}{2}\pi_{i-1}$ for all i

Note also $m_i = N \ \forall i$

Note that just guessing and checking a stat. dist$^{\underline{n}}$, the uniqueness follows from (1) as MC is irreducible.

(3) $\Rightarrow$ long-run proportion of time spent in state i is $\frac{1}{N}$.

Is $p_{00}^{(n)} \longrightarrow \frac{1}{N}$ as $n \to \infty$?

Yes, if N odd, the MC is aperiodic.

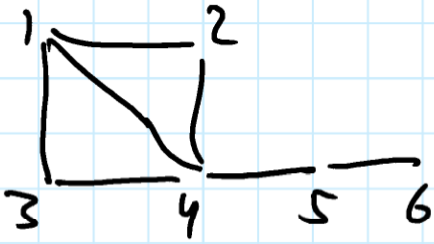
No, if N even, since MC is 2-periodic

Note $p_{00}^{(2m+1)} = 0$, but $p_{00}^{(2m)} \xrightarrow[m\to\infty]{} \frac{2}{N}$

(check!)

Consider 2-step MC $(X_{2m}, m \geq 0)$

<u>Example 3</u>: RW on a graph



degree of vertex

= number of neighbours

Here $(d_i, i \in I) = (3, 2, 2, 4, 2, 1)$

I = set of vertices

$$p_{ij} = \begin{cases} \frac{1}{d_i} & \text{if } \exists\ i - j \\ 0 & \text{o/w} \end{cases}$$

Assume irreducibility (the graph is connected).

Then we know the stat. dist$^{\underline{\text{n}}}$ is unique

<u>Claim</u>: $\pi_i \propto d_i$, i.e. $\exists_A \ \pi_i = \frac{d_i}{A}$.

<u>Proof</u>: Check $dP = d$

$$d_j = \sum_{i \in I} d_i \frac{1}{d_i} \mathbb{1}\{\exists \text{ edge } i - j\}$$

$$= \sum_{i \in I} d_i p_{ij} \qquad \square$$

To get π, normalise: $\pi_i = \frac{d_i}{\sum_{j \in I} d_j}$

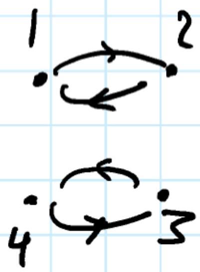
Here $\sum_{j \in I} d_j = 14$, so $\pi = \left(\frac{3}{14}, \frac{1}{7}, \frac{1}{7}, \frac{2}{7}, \frac{1}{7}, \frac{1}{14}\right)$

e.g. $m_1 = \frac{14}{3}$

<u>Example 4</u>: $\tilde{P} = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$

$$\tilde{\pi} = \left(\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta}\right)$$

$$P = \begin{pmatrix} 1-\alpha & \alpha & 0 & 0 \\ \beta & 1-\beta & 0 & 0 \\ 0 & 0 & 1-\gamma & \gamma \\ 0 & 0 & \delta & 1-\delta \end{pmatrix}$$



Not irreducible! $\pi = \left(\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta}, 0, 0\right)$

and $\pi' = \left(0, 0, \frac{\delta}{\gamma+\delta}, \frac{\gamma}{\gamma+\delta}\right)$

Non-uniqueness!

In fact, also mixture of these is also stationary

$$q\pi + (1-q)\pi' = \left(q\frac{\beta}{\alpha+\beta}, q\frac{\alpha}{\alpha+\beta}, (1-q)\frac{\delta}{\gamma+\delta}, (1-q)\frac{\gamma}{\gamma+\delta}\right)$$

for $q \in [0,1]$.

<u>Example 5</u>: 1-dim RW again, $I = \{0,1,2,\dots\}$

$p_{01} = p$, $p_{00} = q$, $p_{i,i+1} = p$, $p_{i,i-1} = q$, $i \geq 1$.

$\pi P = \pi$ leads to a recurrence relation

$$\begin{cases} \pi_i = \pi_{i-1}\, p + \pi_{i+1}\, q, \quad i \geq 1 \\ \pi_0 = \pi_0\, q + \pi_1\, q \end{cases}$$

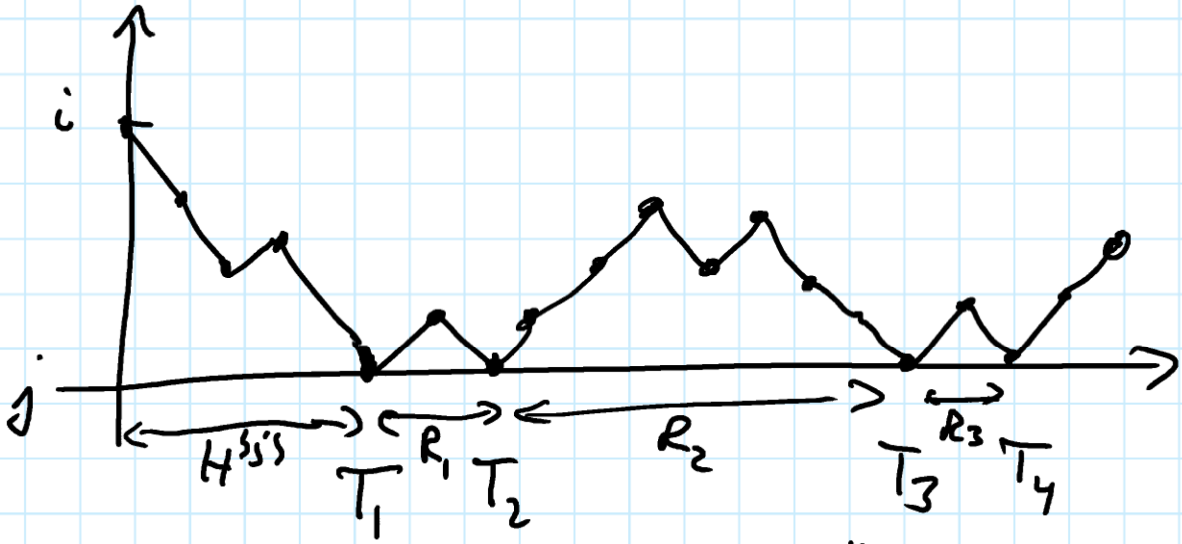
Solve these, normalise: $\pi \sim \text{geom}\left(1 - \frac{q}{p}\right)$

when $p < q$

(req^d^ for pos. rec.)

Ergodic Thm: Starting from i.

Suppose recurrent



Then the times R_k between the k^{th} and $(k+1)^{st}$ visit to j are iid, with $\mathbb{E}(R_k) = m_j \ \forall k \geq 1$

Suppose $m_j < \infty$. First $\frac{T_1}{k} \xrightarrow[k\to\infty]{a.s.} 0$

SLLN $\Rightarrow \frac{T_k}{k} = \frac{T_1}{k} + \frac{R_1 + \cdots + R_{k-1}}{k} \xrightarrow[k\to\infty]{a.s.} m_j$

$$V_j(n) = \#\{l = 0, \dots, n-1 : X_l = j\}$$

But $\frac{T_k}{k} \xrightarrow{a.s.} m_j \iff \frac{V_j(n)}{n} \xrightarrow[n\to\infty]{a.s.} \frac{1}{m_j}$

time per visit — visits per unit time

$\left(\frac{1}{m_i}, i \in I\right)$ stationary

Long-run proportions in $\left\{\begin{matrix} i \\ j \end{matrix}\right.$ is $\frac{1}{m_i}$ $\frac{1}{m_j}$

of jumps from i to j is $\frac{1}{m_i} p_{ij}$

of all jumps $\boxed{\text{into } j}$ is $\sum_{i\in I} \frac{1}{m_i} p_{ij}$

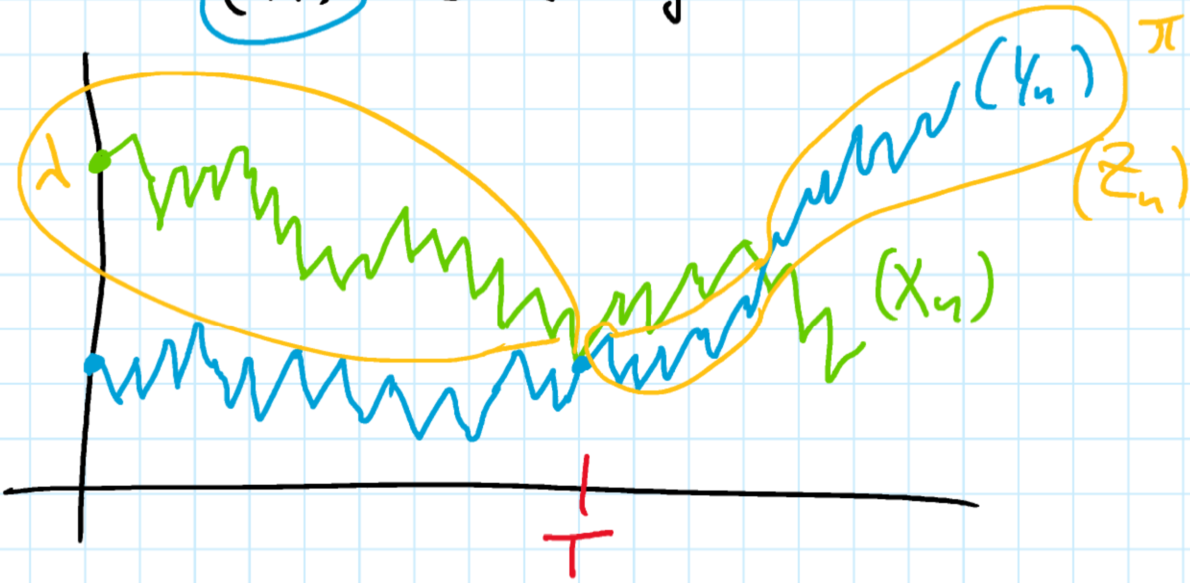
OK if I finite

Hence $\frac{1}{m_j} = \sum_{i \in I} \frac{1}{m_i} p_{ij}$

Convergence Theorem: Assume irreducible, aperiodic, stationary dist^{n} π

λ any initial distⁿ

(X_n) MC starting acc. to λ }
(Y_n) MC starting acc. to π } indep.



Fix $T = \inf\{n \geq 0 : X_n = Y_n\}$

Set $Z_n = \begin{cases} X_n & \text{if } n < T \\ Y_n & \phantom{\text{if }} n \geq T \end{cases}$

Then (Z_n) is a MC starting acc. to λ
and hence dist^d as (X_n)

Suppose $\mathbb{P}(T<\infty) = 1$

thus $\mathbb{P}(T>n) \longrightarrow 0$

so $\left|\mathbb{P}(X_n=j) - \pi_j\right| = \left|\mathbb{P}(Z_n=j) - \mathbb{P}(Y_n=j)\right|$

$$\leq \mathbb{P}(Z_n \neq Y_n)$$

$$= \mathbb{P}(T>n) \xrightarrow[n\to\infty]{} 0$$

Recall: $(\Omega, \mathcal{F}, \mathbb{P})$, $T: \Omega \to [0,\infty)$ r.v.

require $\{T \le t\} = \{\omega \in \Omega : T(\omega) \le t\} \in \mathcal{F}$ $\forall_{t \ge 0}$

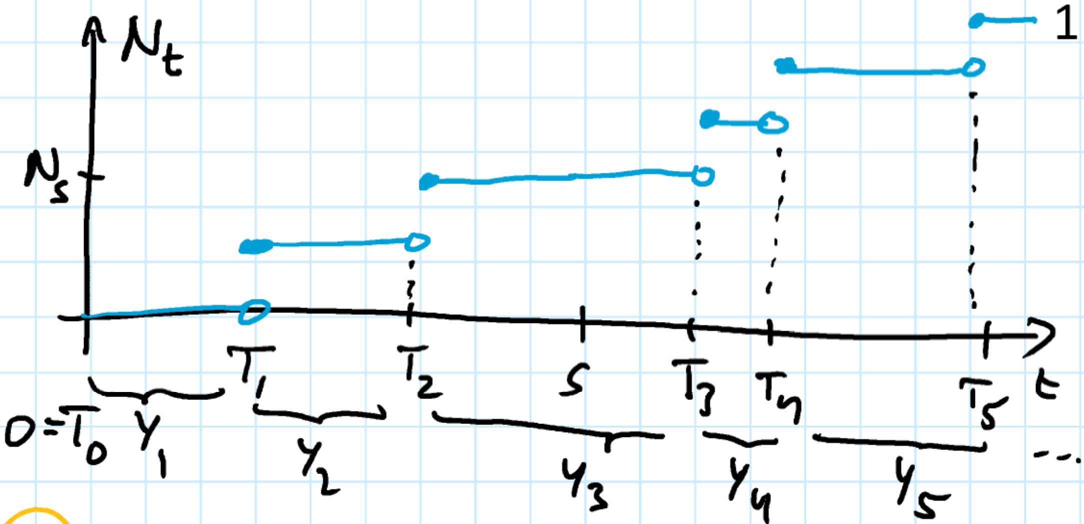
i.e. $\mathbb{1}_{\{T \le t\}}$ discrete r.v. $\forall_{t \ge 0}$.

Defined cdf $F_T(t) = \mathbb{P}(T \le t) = \mathbb{E}(\mathbb{1}_{\{T \le t\}})$

A random process $(N_t,\ t \in [0,\infty))$ ← continuous time!

is a counting process if

- N_t takes values in $\{0,1,2,3,\dots\}$
- $N_s \le N_t$ for $s \le t$
- $t \mapsto N_t$ is right-continuous



$T_k := \inf\{t \geq 0 : N_t \geq k\}$ for $k \geq 0$.

"k^{th} arrival time"

$Y_k = T_k - T_{k-1}$ "k^{th} inter-arrival time".

Notation: For $s < t$, write $N(s,t] = N_t - N_s$

\# of arrivals in $(s,t]$, "increment of (N_t)"

Note: $T_k = \sum_{j=1}^{k} Y_j$

$N_t = \#\{k \geq 1 : T_k \leq t\} = \sum_{k=1}^{\infty} \mathbb{1}_{\{T_k \leq t\}}$

Poisson process: Two diff^t definitions of what we want to call "Poisson process of rate $\lambda \in (0,\infty)$" : $PP(\lambda)$:

Def^n 1: $(N_t, t\geq 0) \sim PP(\lambda)$ if $Y_k, k\geq 1$, are iid $Exp(\lambda)$ and $N_t = \sum_{k\geq 1} \mathbb{1}_{\{T_k \leq t\}}$, $T_k = \sum_{j=1}^{k} Y_j$

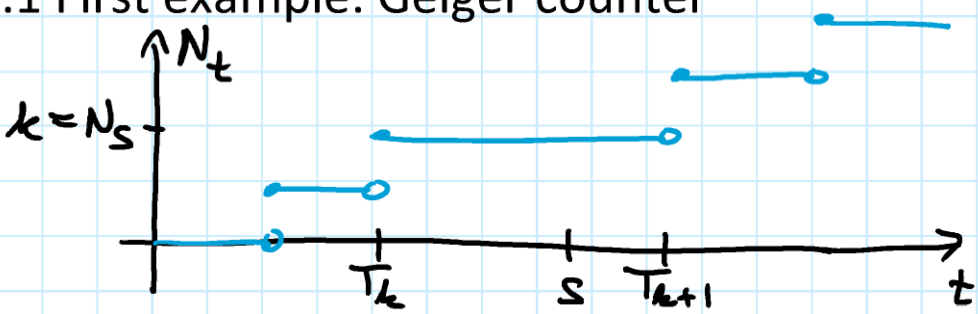
"indep. exponential inter-arrival times"

Def^n 2: $(N_t, t\geq 0) \sim PP(\lambda)$ if

(i) $N_0 = 0$ "independent Poisson increments"

(ii) If $(s_1, t_1], \dots, (s_n, t_n]$ are disjoint intervals in $[0,\infty)$, then increments $N(s_1,t_1], \dots, N(s_n, t_n]$ are independent, where $N(s,t] = N_t - N_s$

(iii) For $s<t$, $N_t - N_s \sim Poisson(\lambda(t-s))$



<u>Example:</u> A Geiger counter near a radioactive source detects particles at an average rate of 1 per 2 seconds.

(a) What is the probab. that no particle is detected for 3 seconds after switching the Geiger counter on?

(b) What is the probab. of detecting at least 3 particles in the first 4 seconds?

<u>Solution:</u> We model this by a $PP(\lambda)$, $\lambda=0.5$

time unit: 1 second

(a) $N_3 \sim \text{Poisson}(3\lambda) = \text{Poisson}(1.5)$

$\mathbb{P}(N_3 = 0) = \underline{e^{-1.5}} = 0.223\ldots$

$T_1 \sim \text{Exp}(\lambda) = \text{Exp}(0.5)$, $\mathbb{P}(T_1 > 3) = e^{-3\lambda}$

$= \underline{e^{-1.5}}$

(b) $N_4 \sim \text{Poisson}(4\lambda) = \text{Poisson}(2)$

$\mathbb{P}(N_4 \geq 3) = 1 - \mathbb{P}(N_4 = 0) - \mathbb{P}(N_4 = 1) - \mathbb{P}(N_4 = 2)$

$= 1 - e^{-2} - 2e^{-2} - \frac{2^2}{2!} e^{-2} = 1 - 5e^{-2}$

$= 0.323\ldots$

$\mathbb{P}(T_3 \leq 4) = \ldots$ $\qquad T_3 = Y_1 + Y_2 + Y_3$

$\sim \text{Gamma}(3, \lambda)$

<u>Claim:</u> $Y_n, n \geq 1$ iid $Exp(\lambda)$ $\iff$ $\begin{cases} N_0 = 0 \\ N_t - N_s \sim Poi(\lambda(t-s)) \\ N_{t_i} - N_{t_{i-1}} \text{ indep}, \ 1 \leq i \leq m \\ \qquad 0 = t_0 < t_1 < \cdots < t_m \end{cases}$

"$\Rightarrow$" $N_0 = 0$ ✓ First $N_t \sim Poi(\lambda t)$

Note $\{N_t = k\} = \{T_k \leq t < T_{k+1}\}$. Hence

$$\mathbb{P}(N_t = k) = \mathbb{P}(T_k \leq t < T_{k+1})$$

$$= \mathbb{P}(T_k \leq t, \underbrace{T_{k+1}}_{= T_k + Y_{k+1}} > t)$$

$\sum_{j=1}^{k} Y_j = T_k \sim Gamma(k, \lambda)$

$Y_{k+1} \sim Exp(\lambda)$

indep.

$$= \mathbb{P}(T_k \leq t, Y_{k+1} > t - T_k)$$

$$= \int_0^t \int_{t-s}^{\infty} \frac{\lambda^k s^{k-1}}{(k-1)!} \cancel{e^{-\lambda s}} \lambda e^{-\lambda y} \, dy \, ds$$

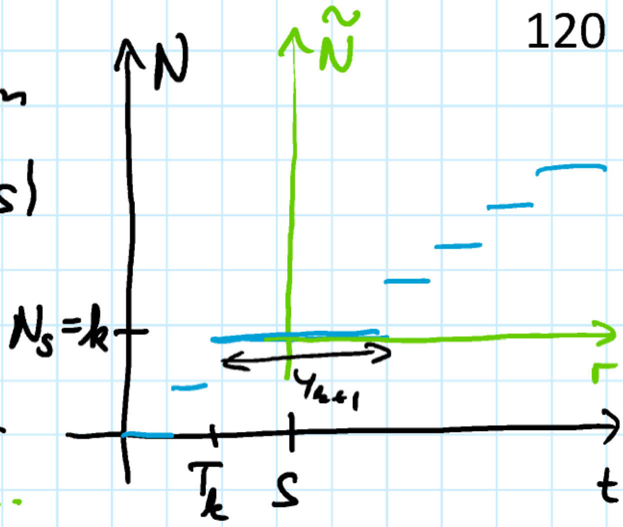
$$\underbrace{\qquad\qquad}_{e^{-\lambda(t-\cancel{s})}}$$

$$= e^{-\lambda t} \frac{\lambda^k}{(k-1)!} \frac{t^k}{k} \quad \Rightarrow N_t \sim Poi(\lambda t)$$

We want to condition on

$\{N_s = k\} = \{T_k \le s, T_{k+1} > s\}$

and show that



$\underbrace{Y_{k+1} - (s - T_k)}_{\tilde{Y}_1}, \underbrace{Y_{k+2}}_{\tilde{Y}_2} \ldots \underbrace{Y_{k+3}}_{\tilde{Y}_3}, \ldots$

are iid $Exp(\lambda)$

We only need to show that $\tilde{Y}_1$ is $Exp(\lambda)$ when when conditioned on $\{N_s = k\}$

$$\mathbb{P}(Y_{k+1} - (s - T_k) > z \mid T_k \le s, T_{k+1} > s)$$

$$= \frac{\mathbb{P}(Y_{k+1} > (s - T_k) + z, T_k \le s)}{\mathbb{P}(Y_{k+1} > (s - T_k), T_k \le s)}$$

$$\mathbb{P}(Y_{k+1} > (s - T_k) + z, T_k \le s) = e^{-\lambda t} \frac{\lambda^k}{(k-1)!} \frac{t^k}{k} e^{-\lambda z}$$

as before

$$= e^{-\lambda z}$$

so cond. dist^{n} of $\tilde{Y}_1$ given $N_s = k$ is $Exp(\lambda)$,

hence so is the uncond. dist^{n}, by LTP

$$\mathbb{P}(\tilde{Y}_1 > z) = \underbrace{\sum_{k=0}^{\infty} \underbrace{\mathbb{P}(\tilde{Y}_1 > z \mid N_s = k)}_{=e^{-\lambda z}} \mathbb{P}(N_s = k)}_{} = e^{-\lambda z}$$

and so $\mathbb{P}(N_s = k, \tilde{Y}_1 > z) = \mathbb{P}(N_s = k)\mathbb{P}(\tilde{Y}_1 > z)$

Hence N_s and $\tilde{Y}_1$ are indep.

Conclusion: $\tilde{N}$ is a $PP(\lambda)$ indep. of N_s.

An induction establishes indep. Poisson increments.

"$\Leftarrow$" $(N_t, t \geq 0)$ and $(Y_n, n \geq 1)$ are in 1-1 correspondence

$\Rightarrow$ dist^{n} of $(N_t, t \geq 0)$ determines the dist^{n} of $(Y_n, n \geq 1)$, and vice versa $\square$

<u>Theorem</u>: Superposition theorem.

Let $(L_t, t \geq 0) \sim PP(\lambda)$, $(M_t, t \geq 0) \sim PP(\mu)$ independent. Then $N_t = L_t + M_t$, $t \geq 0$, is a $PP(\lambda + \mu)$.

<u>Proof</u>: (i) $N_0 = L_0 + M_0 = 0$ ✓

(iii) $\left.\begin{array}{l} L(s,t] \sim \text{Poisson}(\lambda(t-s)) \\ M(s,t] \sim \text{Poisson}(\mu(t-s)) \end{array}\right\}$ indep.

$\Rightarrow N(s,t] \sim \text{Poisson}((\lambda+\mu)(t-s))$

(ii) $\left.\begin{array}{l} L(s_1,t_1], \ldots, L(s_n,t_n] \\ M(s_1,t_1], \ldots, M(s_n,t_n] \end{array}\right\}$ all indep.

$\Rightarrow N(s_1,t_1], \ldots, N(s_n,t_n]$ are indep. $\square$

<u>Theorem</u>: Thinning theorem

Let $(N_t, t \geq 0) \sim PP(\lambda)$. Independently mark each point w.p. $p \in (0,1)$. Let

$$M_t = \# \text{ marked points in } [0,t]$$

Then $(M_t, t \geq 0) \sim PP(\lambda p)$

<u>Proof</u>: Properties (i) and (ii) for $(M_t, t \geq 0)$ follow from the corresponding properties for $(N_t, t \geq 0)$, and the fact that the marking in disjoint intervals is independent.

For (iii), if $N \sim \text{Poisson}(\mu)$ and cond. given $N = n$, $M \sim \text{Binomial}(n, p)$, then $M \sim \text{Poisson}(p\mu)$. (Fact from Prelims)

Here, $N(s,t] \sim \text{Poisson}(\lambda(t-s))$, and given $N(s,t]=n$, $M(s,t] \sim \text{Binomial}(n,p)$. So, $M(s,t] \sim \text{Poisson}(\lambda p(t-s))$,

as req^{d} for (iii)

$\square$

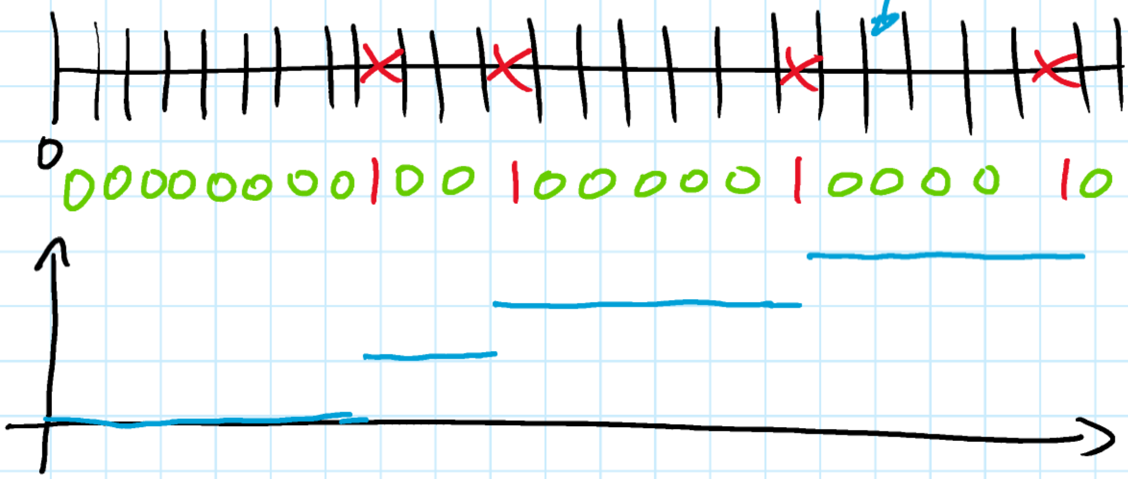
We used Def^{n} 2 to identify Poisson processes. It is instructive to also think about proofs using Def^{n} 1.

Relevant facts:

(1) $X_n \sim \text{Geometric}\left(\frac{\lambda}{n}\right) \Rightarrow \frac{1}{n} X_n \xrightarrow[n\to\infty]{d} \text{Exp}(\lambda)$

(2) $X_n \sim \text{Binomial}\left(n, \frac{\lambda}{n}\right) \Rightarrow X_n \xrightarrow[n\to\infty]{d} \text{Poisson}(\lambda)$

Independent Bernoulli $\left(\frac{\lambda}{n}\right)$ trials slots of width $\frac{1}{n}$



Discrete-time counting process

each time slot contains a cross w.p. $p = \frac{\lambda}{n}$

counting the number of crosses up to the m^{th} trial

(1) Inter-point distances are iid Geometric(p)

(2) Increments during disjoint intervals are indep. Binomial(m,p)

Let $p=\frac{\lambda}{n}$ and rescale time by n (time-steps $\frac{1}{n}$)

Then we get convergence to $PP(\lambda)$.

Example: Call centre, calls from

- existing customers at rate 1 per 20 seconds
- potential new customers at 1 per 30 seconds

$\Big\}$ indep. PP

(1) Dist$^{\underline{n}}$ of number of calls in a given minute?

(2) Suppose potential new customers $\rightsquigarrow$ new contract w.p. $\frac{1}{4}$. Dist$^{\underline{n}}$ of number of new contracts in a given hour?

Solutions: Unit of time: 1 minute, $(E_t) \sim PP(3)$

$(P_t) \sim PP(2)$

(1) By the Superposition Thm $A_t = E_t + P_t$ is such that $(A_t) \sim PP(5)$

$$\Rightarrow A((t, t+1]) \sim \text{Poisson}(5)$$

(2) Thinning w.p. $p = \frac{1}{4}$.

C_t = # new contracts in $[0, t]$

By the Thinning Theorem $(C_t) \sim PP(\underbrace{\tfrac{1}{4} \times 2}_{\frac{1}{2}})$

$\Rightarrow C((t, t+60]) \sim \text{Poisson}(30)$

<u>Example:</u> Genetic recombination model

N_t = # crossover points in $[0, t]$

<u>Model:</u> $(N_t, t \geq 0) \sim PP(\lambda)$

$p = \mathbb{P}(N(a, b] \text{ even}) = \cdots = \frac{1}{2}(1 - e^{-2\lambda x})$

$$\Rightarrow x = -\frac{1}{2\lambda} \log(2p - 1)$$