Def. 4.1.3 : let X, Y be continuous r.v.s

1) The joint p.d.f. of X and Y is a function $f_{X,Y}: \mathbb{R}\times\mathbb{R} \to [0,\infty)$ such that

(i) $$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy\,dx = 1.$$

(ii) For any $a, b \in \mathbb{R}$

$$\underbrace{P(X \le a, Y \le b)}_{= F_{X,Y}(a,b)} = \int_{-\infty}^{a}\int_{-\infty}^{b} f_{X,Y}(x,y)\,dy\,dx$$

$= F_{X,Y}(a,b)$ joint c.d.f. of X and Y.

region of integration

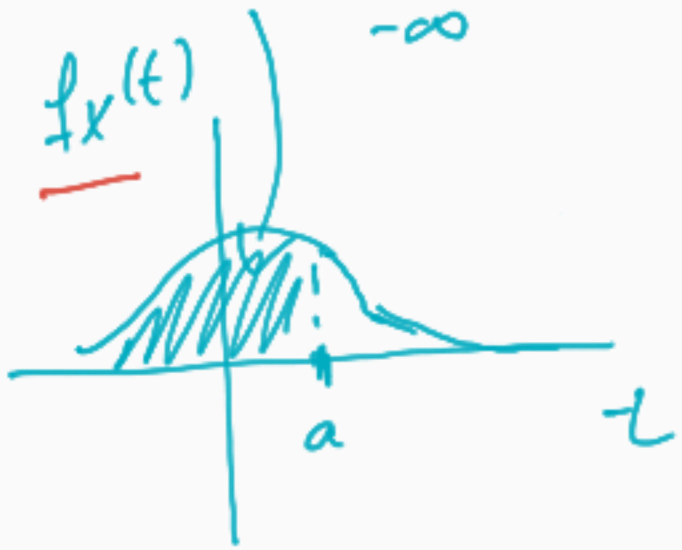
X r.v.

p.d.f. $f_X: \mathbb{R} \to [0,\infty)$

1) $\int_{-\infty}^{\infty} f_X(t)\,dt = 1$

2) $P(a \le X \le b) = \int_a^b f_X(t)\,dt$

$P(X \le a) = \int_{-\infty}^{a} f_X(t)\,dt$



not same!!

2) The marginal pdf of X is

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

[The marginal p.d.f. of Y is

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)\, dx$$

]

X,Y discrete

p_{XY} joint pmf

- marginal pmf of X

$$p_X(k) = \sum_{\substack{l \text{ possible} \\ \text{value of} \\ Y}} p_{XY}(k,l)$$

⚠ $\mathbb{E}[X,Y]$ makes no sense!! But for any function $g: \mathbb{R}\times\mathbb{R} \to \mathbb{R}$

$$\mathbb{E}[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\, f_{XY}(x,y)\, dy\, dx.$$

Example 4.1.4 :

$X \sim Unif\,(0,10)$

$f_X(t) = \begin{cases} \frac{1}{10} & t \in (0,10) \\ 0 & \text{other} \end{cases}$

T_1 = time waited until placing our order (in min.)

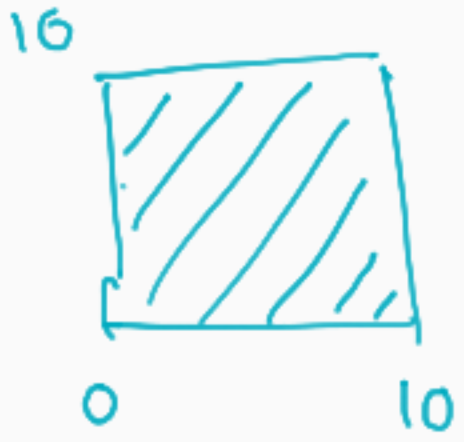
T_2 = " " order is ready (in min.)

Assume T_1, T_2 are jointly distributed uniformly over $(0,10) \times (0,10)$

? What is the prob. that in total you wait more than 6 minutes?

$T_1 + T_2$

1) joint p.d.f. of T_1, T_2 :



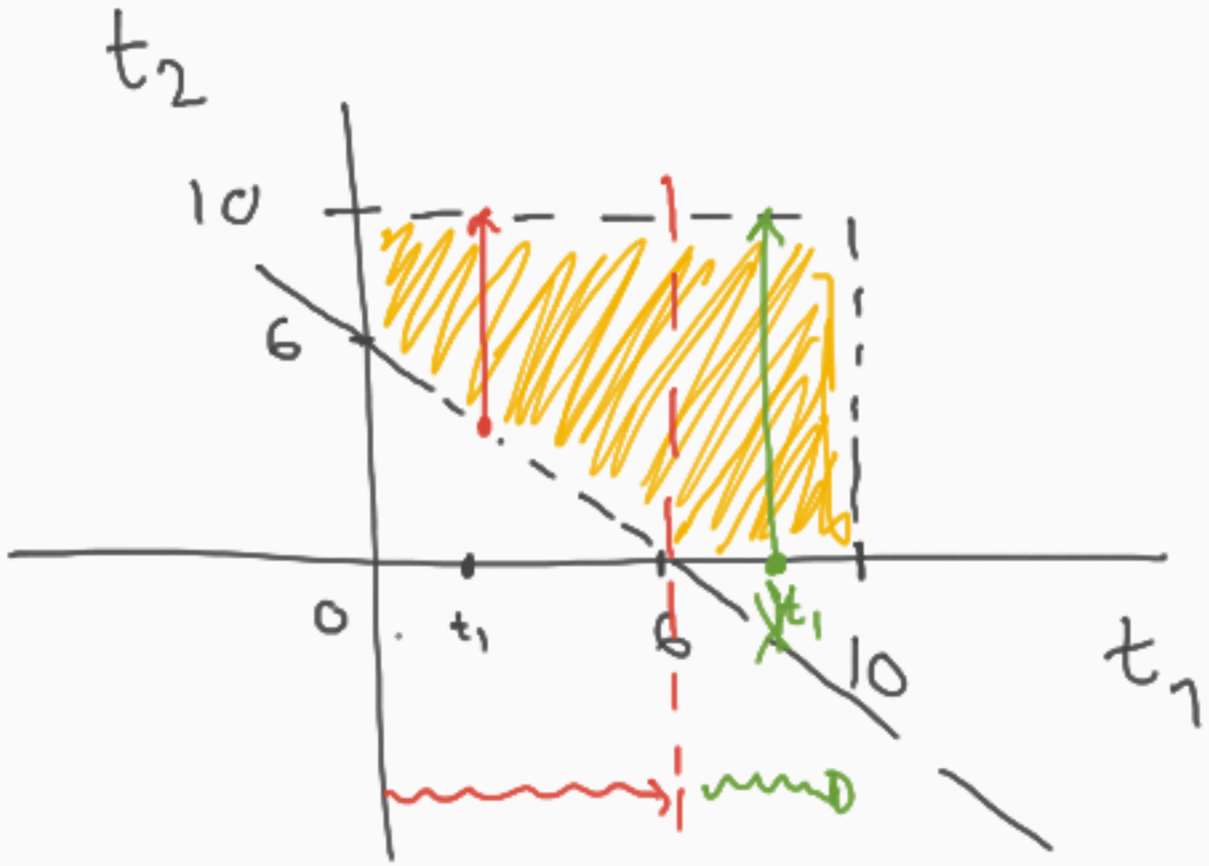
$$f_{T_1T_2}(t_1,t_2) = \begin{cases} \frac{1}{100} & t_1 \in (0,10)\ ,\ t_2 \in (0,10) \\ 0 & \text{otherwise} \end{cases}$$

2) $\mathbb{P}(T_1+T_2>6)$

$$= \mathbb{P}(\underbrace{T_2 > 6-T_1}_{})$$

$t_2 = 6 - t_1$

area of integration



$$= \int_{\underset{t_1}{0}}^{6} \int_{\underset{t_2}{6-t_1}}^{10} \frac{1}{100}\, dt_2\, dt_1$$

$$+ \int_6^{10} \int_0^{10} \frac{1}{100}\, dt_2\, dt_1 = \frac{1}{100}\int_0^6 4+t_1\, dt_1 + \frac{4\cdot 10}{100}$$

$$= \frac{1}{100}\left[4t_1 + \frac{t_1^2}{2}\right]_0^6 + \frac{40}{100} = \frac{24+18}{100} + \frac{40}{100} = \frac{82}{100}.$$

4.2. Independence, covariance & correlation

Recall: A, B events are independent when $\mathbb{P}(\underset{\{X\le a\}}{A}\cap \underset{\{Y\le b\}}{B}) = \mathbb{P}(A)\cdot\mathbb{P}(B)$

Def 4.2.1: Let X, Y be r.v.s. They are called independent if for any $a,b\in\mathbb{R}$

$$\underbrace{\mathbb{P}(X\le a, Y\le b)}_{F_{X,Y}(a,b)} = \underbrace{\mathbb{P}(X\le a)}_{F_X(a)}\cdot\underbrace{\mathbb{P}(Y\le b)}_{F_Y(b)}$$

joint c.d.f of X,Y = marginal c.d.f. of X · marginal c.d.f of Y

Also: X, Y independent means

joint $\left\{\begin{array}{l}\text{p.m.f.}\\ \text{p.d.f.}\end{array}\right.$ of X, Y = product of marginal p.m.f$_s$ / p.d.f

__Observation / examples:__ ($\nearrow$ central limit theorem)

(a) $X \sim Pois(\lambda_1)$, $Y \sim Pois(\lambda_2)$ $\boxed{\text{are indep.}}$ then $X+Y \sim Pois(\lambda_1+\lambda_2)$

(b) $X \sim Bin(n_1, p)$, $Y \sim Bin(n_2, p)$ " " " $X+Y \sim Bin(n_1+n_2, p)$ think about!

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

__Def. 4.2.2__: Let X, Y be r.v.s $\mathbb{E}[X^2] < \infty$, $\mathbb{E}[Y^2] < \infty$. Their covariance is defined as

$$\mathrm{cov}(X,Y) = \mathbb{E}[X \cdot Y] - \mathbb{E}[X] \cdot \mathbb{E}[Y] \overset{!}{=} \mathbb{E}\big[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])\big]$$

$$\mathrm{cov}(X,X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \mathrm{Var}(X) = \mathbb{E}\big[(X-\mathbb{E}[X])^2\big]$$

and their correlation

$$\rho(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} \in [-1, 1]$$

$\rho(X,Y) > 0$: $X \uparrow \to Y \uparrow$, $X \downarrow \to Y \downarrow$

$\rho(X,Y) < 0$: $X \uparrow \to Y \downarrow$, $X \downarrow \to Y \uparrow$

We say that X, Y are uncorrelated if $\operatorname{cov}(X,Y) = 0$.

$\rho(X,Y) = 0$

⚠ uncorrelated $\neq$ independent

however, independent $\Rightarrow$ uncorrelated. (but uncorrelated $\not\Rightarrow$ indep.)

! check!!

$$\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y] \Rightarrow \operatorname{cov}(X,Y) = \mathbb{E}[X \cdot Y] - \mathbb{E}[X] \cdot \mathbb{E}[Y] = 0$$

Recall: X, Y independent, then $\operatorname{Var}(X+Y) = \operatorname{Var}(X) + \operatorname{Var}(Y)$

Why?

$$\begin{aligned}
\operatorname{Var}(X+Y) &= \mathbb{E}[(X+Y)^2] - (\mathbb{E}[X] + \mathbb{E}[Y])^2 \\
&= \mathbb{E}[X^2 + 2XY + Y^2] - ((\mathbb{E}[X])^2 + 2\mathbb{E}[X] \cdot \mathbb{E}[Y] + (\mathbb{E}[Y])^2) \\
&= \mathbb{E}[X^2] + 2\mathbb{E}[X \cdot Y] + \mathbb{E}[Y^2] - (\mathbb{E}[X])^2 - 2\mathbb{E}[X]\mathbb{E}[Y] - (\mathbb{E}[Y])^2 \\
&= \operatorname{Var}(X) + \operatorname{Var}(Y) + \underbrace{2\operatorname{cov}(X,Y)}_{=0 \text{ if } X, Y \text{ independent}}
\end{aligned}$$

☺

Example 4.2.3 :

$Y \backslash X$	-1	0	1	marginal pmf of Y
1	0.1 +	0.2 +	0.3	0.6
2	0.2 +	0.1 +	0.1	0.4
marginal p.m.f. X	0.3	0.3	0.4	

1) $\operatorname{cov}(X,Y)$?

2) X, Y indep ? NO b/c $\operatorname{cov}(X,Y) \neq 0$.

$$\operatorname{cov}(X,Y) = \mathbb{E}[X \cdot Y] - \mathbb{E}[X] \cdot \mathbb{E}[Y] = 0 - 0.1 \cdot 1.4 = -0.14$$

$$\mathbb{E}[X \cdot Y] = (-1) \cdot (1) \cdot 0.1 + (-1) \cdot 2 \cdot 0.2 + \cancel{0 \cdot 1 \cdot 0.2} + \cancel{0 \cdot 2 \cdot 0.1} + 1 \cdot 1 \cdot 0.3 + 1 \cdot 2 \cdot 0.1$$

$$= -0.1 - 0.4 + 0.3 + 0.2 = 0$$

$$\mathbb{E}[X] = -1 \cdot 0.3 + 0 \cdot 0.3 + 1 \cdot 0.4 = 0.1$$

$$\mathbb{E}[Y] = 1 \cdot 0.6 + 2 \cdot 0.4 = 1.4$$

4.3 Conditional probability

• Discrete case.

Example 4.3.1:

X = # e-devices w/ mobile data access

Y = # unlimited data plans.

$Y \backslash X$	0	1	2	marginal pmf of Y
0	$\frac{1}{15}$ +	$\frac{1}{15}$ +	$\frac{2}{15}$	$\frac{4}{15}$
1	$\frac{1}{15}$	$\frac{1}{5}$	$\frac{1}{15}$	
2	$\frac{2}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	

? Given that a randomly chosen household has no unlimited data plans, what is the prob. that it has one e-device w/ mobile data access?

$$\mathbb{P}(X=1 \mid Y=0) = \frac{\mathbb{P}(X=1, Y=0)}{\mathbb{P}(Y=0)} = \frac{1/15}{4/15} = \frac{1}{4}.$$

joint pmf of X,Y at (1,0)

marginal pmf of Y at zero

Def 4.3.2 : let X, Y be discrete r.v. with joint pmf $p_{X,Y}(k,l)$.

(i) The conditional p.m.f. of X given $Y = l$ (assuming $P(Y=l)>0$)

is defined as

$$p_{X|Y}(k \mid l) = \frac{p_{X,Y}(k,l)}{p_Y(l)}$$

fixed

← joint pmf

← marginal pmf of Y

(ii) The conditional c.d.f of X given $Y = l$ is

$$F_{X|Y}(a \mid l) = P(X \le a \mid Y = l) = \sum_{\substack{\text{possible} \\ k \le a \text{ for } X}} p_{X|Y}(k \mid l)$$

$Y \le a \quad | X = k$

conditional pmf of Y given $X = k$

$$p_{Y|X}(l \mid k) = \frac{p_{X,Y}(k,l)}{p_X(k)}$$

; conditional c.d.f of Y given $X = k$ $\quad F_{Y|X}(a \mid k)$

Def. 4.3.3 : Let X, Y be continuous r.v.s with joint pdf $f_{X,Y}(x,y)$.

(i) The conditional p.d.f. of X given $Y=y$ (assume $f_Y(y) > 0$)

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} \begin{array}{l} \leftarrow \text{joint pdf} \\ \leftarrow \text{marginal pdf} \end{array}$$

fixed

(ii) The conditional c.d.f. of X given $Y=y$ is

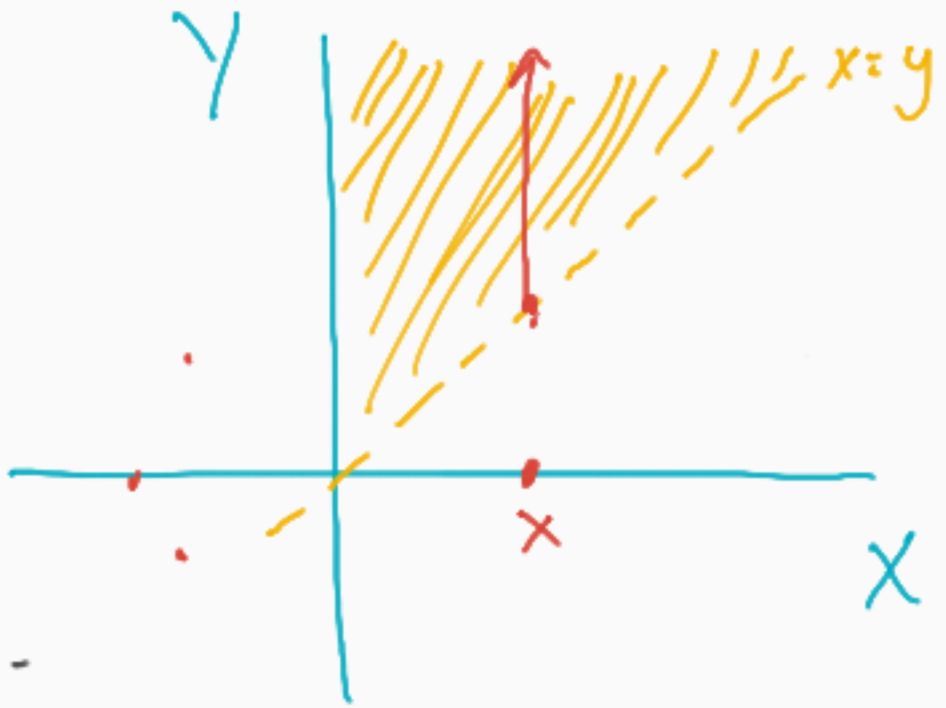
$$F_{X|Y}(a|y) = \mathbb{P}(X \leq a \mid Y=y) = \int_{-\infty}^{a} f_{X|Y}(x|y)\,dx$$

[conditional pdf of Y given $X=x$ $\quad f_{Y|X}(y|x) = \dfrac{f_{X,Y}(x,y)}{f_X(x)}$]

Example 4.3.4 :

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{18} e^{-\frac{x+y}{6}} & \boxed{0<x<y} \\ 0 & \text{otherwise} \end{cases}$$

region of integration



? Find the conditional pdf of Y given X=2.

$$f_{Y|X}(y|2) = \frac{f_{X,Y}(2,y)}{f_X(2)} = \begin{cases} \dfrac{\frac{1}{18} e^{-\frac{2+y}{6}}}{\frac{1}{3} e^{-2/3}} = \frac{1}{6} e^{\frac{2-y}{6}} & y>2 \\ 0 & \text{otherwise.} \end{cases}$$

marginal pdf of X

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy = \int_x^{\infty} \frac{1}{18} e^{-\frac{\cancel{x}+y}{6}}\,dy = \frac{1}{18} e^{-\frac{x}{6}} \cdot \int_x^{\infty} e^{-\frac{y}{6}}\,dy$$

fixed

fixed

$$= \frac{1}{18} e^{-\frac{x}{6}} \left[-6 e^{-\frac{y}{6}}\right]_x^{\infty} = \frac{1}{18} e^{-\frac{x}{6}} \cdot 6 e^{-\frac{x}{6}} = \frac{1}{3} e^{-\frac{x}{3}}$$

if $x \le 0$, $f_{X,Y}(x,y) = 0 \Rightarrow f_X(x) = 0$

if $x > 0$

4.4. Conditional expectation

Def. 4.4.1: let X, Y be r.v.s

(i) If X and Y are discrete, the conditional expectation of X given $Y = \ell$ (assuming $\mathbb{P}(Y=\ell)>0$) is

$$\mathbb{E}[X \mid Y=\ell] = \sum_{\substack{k \text{ possible} \\ \text{value of } X}} k \cdot p_{X|Y}(k|\ell)$$

[If $g: \mathbb{R} \to \mathbb{R}$ is a transformation, then $\mathbb{E}[g(X) \mid Y=\ell] = \sum_k g(k) \cdot p_{X|Y}(k|\ell)$.]

(ii) If X, Y are continuous, the conditional expectation of X given $Y = y$ (assuming $f_Y(y) > 0$) is

$$\mathbb{E}[X \mid Y=y] = \int_{-\infty}^{\infty} x \, f_{X|Y}(x|y) \, dx$$

$g: \mathbb{R} \to \mathbb{R}$ transformation

$$\mathbb{E}[g(X) \mid Y=y] = \int_{-\infty}^{\infty} g(x) f_{X|Y}(x|y) \, dx$$

⊛ In particular: $\mathrm{Var}(X \mid Y=y) = \mathbb{E}[X^2 \mid Y=y] - (\mathbb{E}[X \mid Y=y])^2$

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

⊛ Further general properties:

(a) Linearity : $a, b \in \mathbb{R}$ $\quad \mathbb{E}[aX + b \mid Y=y] = a \cdot \mathbb{E}[X \mid Y=y] + b$

$\mathbb{E}[aX+b] = \dots$

(b) X_1, X_2 : $\mathbb{E}[X_1 + X_2 \mid Y=y] = \mathbb{E}[X_1 \mid Y=y] + \mathbb{E}[X_2 \mid Y=y]$

$\mathbb{E}[X_1 + X_2] = \mathbb{E}[X_1] + \mathbb{E}[X_2]$

(c) X, Y independent : $\mathbb{E}[X \mid Y=y] = \mathbb{E}[X]$.

!

Application (Wald's identity): let $X_1, X_2, \dots X_n$ be i.i.d. r.v.s.

and let N be a r.v. also indepent of all others. independent identically distributed

Then,

$$\mathbb{E}\left[\sum_{i=1}^{N} X_i\right] = \sum_{k=0}^{\infty} \mathbb{E}\left[\sum_{i=1}^{N} X_i \,\middle|\, N=k\right] \cdot \mathbb{P}(N=k)$$

$$= \sum_{k=0}^{\infty} \mathbb{E}\left[\sum_{i=1}^{k} X_i \,\middle|\, N=k\right] \cdot \mathbb{P}(N=k)$$

indep.

$$= \sum_{k=0}^{\infty} \mathbb{E}\left[\sum_{i=1}^{k} X_i\right] \cdot \mathbb{P}(N=k)$$

X_i same distribution

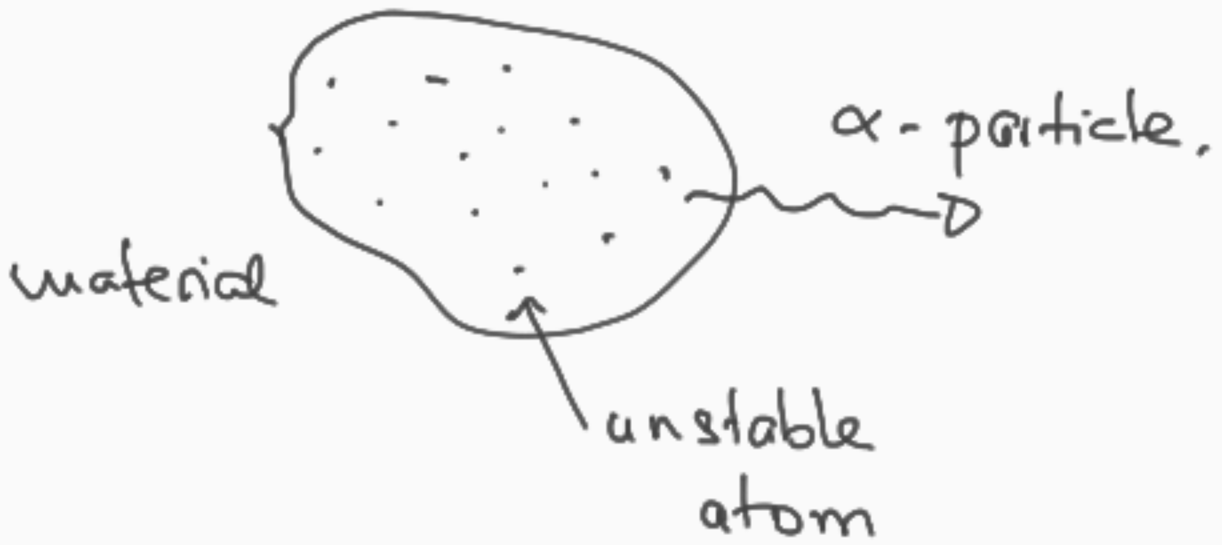
$$= \sum_{k=0}^{\infty} \underbrace{\sum_{i=1}^{k} \mathbb{E}[X_i]}_{} \cdot \mathbb{P}(N=k)$$

so $\mathbb{E}[X_i]$ same for all!

$$= \sum_{k=0}^{\infty} k \cdot \mathbb{E}[X_1] \cdot \mathbb{P}(N=k) = \mathbb{E}[X_1] \boxed{\sum_{k=0}^{\infty} k \cdot \mathbb{P}(N=k)} = \mathbb{E}[X_1] \cdot \mathbb{E}[N]$$

$= \mathbb{E}[N]$

Example 4.4.2 : Radioactive decay



$\mathbb{P}(\text{unstable atom sends } \alpha) = 0.001$

$N = \#\text{ unstable atoms} \sim \text{Pois}(60)$.

? How many α-particles does the material emit on average?

[Use Wald's equality! What is X_i ?]

$X_i = \text{atom } i \text{ sends } \alpha\text{-particle } \begin{cases} 0 \\ 1 \end{cases} \sim \text{Ber}(0.001)$

$\#\alpha\text{-particles} = \sum_{i=1}^{N} X_i$

$$\mathbb{E}\left[\sum_{i=1}^{N} X_i\right] \underset{\text{Wald}}{=} \mathbb{E}[X_1] \cdot \mathbb{E}[N] = 0.001 \cdot 60$$

Chapter 5. Moment generating function

Def. 5.1: The moment generating function (m.g.f.) of a r.v. X is given by

$$m_X(s) = \mathbb{E}[e^{sX}] \qquad s \in \mathbb{R}$$

(as long as it is finite).

⚠ m.g.f. is a function of s.

Q: Why "moment generating"?

↳ First derivative at zero: $\frac{d}{ds} m_X(s) = \frac{d}{ds}\mathbb{E}[e^{sX}] \overset{!}{=} \mathbb{E}\left[\frac{d}{ds} e^{sX}\right] = \mathbb{E}[X e^{sX}]$

$$m_X'(0) = \mathbb{E}[X] \quad (\to \text{1st moment of } X)$$

⇝ Second derivative at zero : $\frac{d}{ds}\left(\frac{d}{ds} m_X(s)\right) = \frac{d}{ds}\mathbb{E}[Xe^{sX}] \overset{!}{=} \mathbb{E}\left[\frac{d}{ds} Xe^{sX}\right] = \mathbb{E}[X^2 e^{sX}]$

"

⋮

$$m''_X(0) = \mathbb{E}[X^2] \quad (\rightsquigarrow \text{2nd moment of } X)$$

⇝ nth derivative at zero : $m_X^{(n)}(0) = \mathbb{E}[X^n]$ ($\rightsquigarrow$ nth moment of X)

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = m''_X(0) - (m'_X(0))^2$$

(using mgf)
moments...

$$\boxed{\mathrm{Var}(X+Y) = m''_{X+Y}(0) - (m'_{X+Y}(0))^2}$$

Prop. 5.2: let X be r.v. with m.g.f. $m_X(s)$.

(i) For $a, b \in \mathbb{R}$, the m.g.f. of $aX+b$ satisfies

$$m_{aX+b}(s) = \mathbb{E}[e^{s(aX+b)}] = \mathbb{E}[e^{s \cdot aX} e^{sb}] = e^{sb}\,\mathbb{E}[e^{s \cdot aX}]$$

non-random!

$$= e^{sb} \cdot m_X(sa)$$

$\rightsquigarrow \mathbb{E}[g(X)h(Y)] = \mathbb{E}[g(X)] \cdot \mathbb{E}[h(Y)]$

(ii) If Y is a r.v. independent of X,

X, Y indep!

$$m_{X+Y}(s) = \mathbb{E}[e^{s(X+Y)}] = \mathbb{E}[e^{sX} \cdot e^{sY}] \overset{!}{=} \mathbb{E}[e^{sX}] \cdot \mathbb{E}[e^{sY}]$$

$$= m_X(s) \cdot m_Y(s).$$

(iii) If X and Y have the same m.g.f., then X and Y have same distribution.

Def. 5.3: let X, Y be r.v.s, their joint m.g.f. is

$$m_{X,Y}(s,t) = \mathbb{E}\left[e^{sX+tY}\right]$$

⚠ not to be confused with $m_{X+Y}(s)$

If X, Y are independent, then

$$m_{X,Y}(s,t) = \mathbb{E}\left[e^{sX} \cdot e^{tY}\right] = \mathbb{E}\left[e^{sX}\right] \cdot \mathbb{E}\left[e^{tY}\right]$$

$$= m_X(s) \cdot m_Y(t)$$

different !!

$$\left[\, m_{X+Y}(s) = m_X(s) \cdot m_Y(s) \,\right]$$

Example 5.4: The power consumption of two cities A and B is modeled by r.v.s X, Y with

$$m_X(s) = \frac{1}{1-s}, \quad s<1$$

$$m_Y(s) = e^{s^2/2}, \quad s\in\mathbb{R}$$

$$m_{X+Y}(s) = m_X(s)\cdot m_Y(s) = \boxed{\frac{e^{s^2/2}}{1-s}} \quad s<1$$

$\leadsto m_{X+Y}(0) = 1$

Assuming X, Y independent, compute the variance of the total power consumption of both cities.

$= X+Y$ (r.v.)

$$\mathrm{Var}(X+Y) = m''_{X+Y}(0) - \left(m'_{X+Y}(0)\right)^2 = (1\cdot 1 + 1\cdot 2) - (1\cdot 1)^2 = 3-1 = 2$$

$$m_{X+Y}'(s) = s\cdot e^{s^2/2}\cdot\frac{1}{1-s} + e^{s^2/2}\cdot\frac{1}{(1-s)^2} = \boxed{\frac{e^{s^2/2}}{1-s}}\left(s+\frac{1}{1-s}\right) = m_{X+Y}(s)\cdot\left(s+\frac{1}{1-s}\right)$$

$$m''_{X+Y}(s) = m'_{X+Y}(s)\left(s+\frac{1}{1-s}\right) + m_{X+Y}(s)\left(1+\frac{1}{(1-s)^2}\right) = m_{X+Y}(s)\left(s+\frac{1}{1-s}\right)^2 + m_{X+Y}(s)\left(1+\frac{1}{(1-s)^2}\right)$$

Chapter 6. Limit theorems

What for? $X_1, X_2, X_3, \dots$ available data (result of an experiment)
sample

Big data!

Quantity of interest: $S_n = \sum_{i=1}^{n} X_i$

Q1: $\frac{1}{n} S_n$ (sample mean) as n becomes larger?

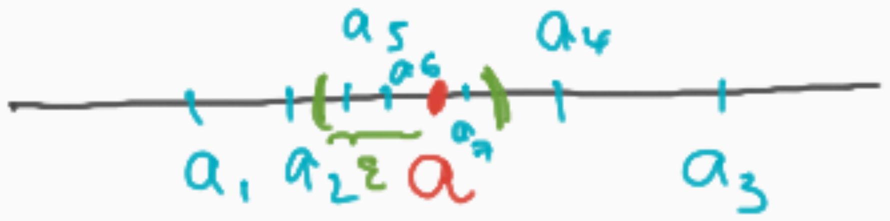
Q2: How is $(\ )S_n$ distributed (approx.) as n becomes large?
e.g. approx. of Bin by Normal

limit as $n \to \infty$.

6.1. Convergence of sequences of r.v.s

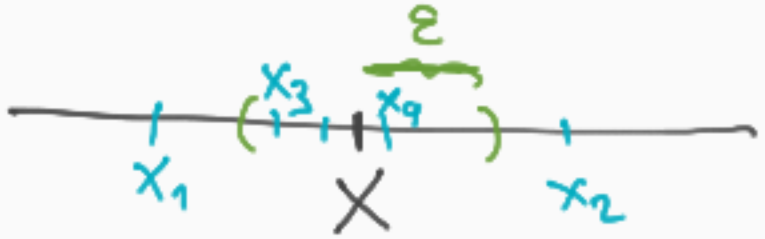
Def 6.1.6: Let $a_1, a_2, a_3, \dots$ be a sequence of real numbers. We say the sequence converges to $a \in \mathbb{R}$ (and write $\lim_{n\to\infty} a_n = a$) if for any $\boxed{\varepsilon > 0}$ there exists $\underset{5}{N} > 0$ such that

$$|a_n - a| < \varepsilon \quad \text{for all } n \geq \underset{5}{N}$$



Def. 6.1.2: Let $X_1, X_2, X_3, \dots$ be a sequence of r.v.s. We say that the sequence converges to a r.v. X (or a number $a \in \mathbb{R}$) $\boxed{\text{in probability}}$ if for any $\varepsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}\left(|X_n - \underset{a}{X}| > \varepsilon\right) = 0 \qquad \left(X_n \xrightarrow{P} X \right)$$



Def 6.1.3 : We say that the sequence $X_1, X_2, X_3, \ldots$ converges to the r.v. X (or the number $a \in \mathbb{R}$) $\boxed{\text{almost surely}}$ if

$$\mathbb{P}\left(\lim_{n\to\infty} X_n = \underset{(a)}{X} \right) = 1 \qquad \left(X_n \xrightarrow{a.s.} X \right)$$

! $X_n \xrightarrow{a.s} X$ implies $X_n \xrightarrow{P} X$

almost surely ⇝ "strong" in probability ⇝ "weak"

Def. 6.1.4 : We say that the sequence $X_1, X_2, X_3, \ldots$ converges to the r.v. X $\boxed{\text{in distribution}}$ if for all $a \in \mathbb{R}$ where the c.d.f of X is continuous we have

$$\lim_{n\to\infty} F_{X_n}(a) = F_X(a) \qquad \left(X_n \xrightarrow{d} X \right)$$

$$\lim_{n\to\infty} \mathbb{P}(X_n \leq a) = \mathbb{P}(X \leq a)$$

Example 6.1.5 : Roll a 4-sided die (each time indep.)

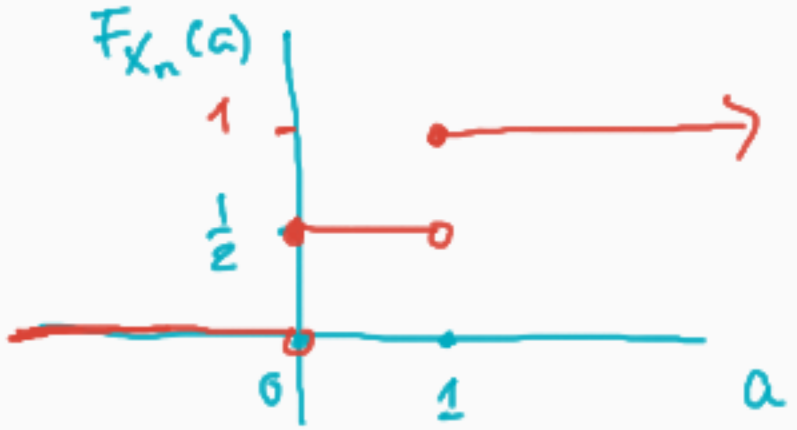
sample space $\Omega = \{1,2,3,4\}$

$$X_n = \begin{cases} 1 & \text{if nth roll is even} \\ 0 & \text{if " " odd} \end{cases} \qquad X = \begin{cases} 1 & \text{if my roll is odd} \\ 0 & \text{" " " even.} \end{cases}$$

(you) (me)

Q: Do we have $X_n \xrightarrow{d} X$? $\boxed{\text{YES}}$ $\left(\lim_{n\to\infty} F_{X_n}(a) = F_X(a) \right)$

① c.d.f. of X_n :



$$F_{X_n}(a) = \begin{cases} 0 & a<0 \\ \frac{1}{2} & 0 \le a < 1 \\ 1 & a \geq 1 \end{cases}$$

$\mathbb{P}(X_n \le a)$

② c.d.f. of X :

$$F_X(a) = \begin{cases} 0 & a<0 \\ \frac{1}{2} & 0 \le a < 1 \\ 1 & a \geq 1 \end{cases}$$

$F_{X_n}(a) = F_X(a)$ for all n!

$\lim_{n\to\infty} F_{X_n}(a) = F_X(a)$ so $X_n \xrightarrow{d} X$

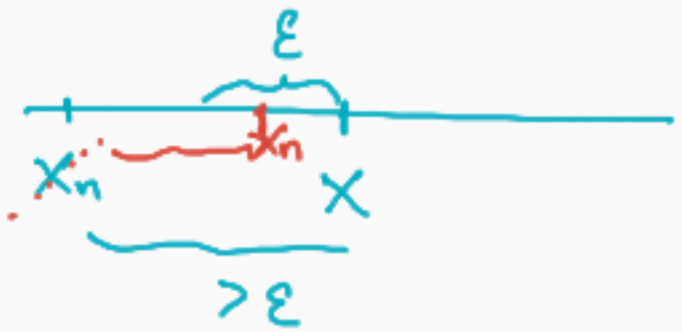
$F_{X_n}(0) = \mathbb{P}(X_n \le 0) = \mathbb{P}(X_n = 0) = \mathbb{P}(\text{odd}) = \frac{1}{2}$

$F_X(0) = \mathbb{P}(X \le 0) = \mathbb{P}(X = 0) = \mathbb{P}(\text{even}) = \frac{1}{2}$

6.2 Markov's and Chebyshev's inequalities

Recall: $X_n \xrightarrow{P} X$ means for any $\varepsilon > 0$ $\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0$

convergence in probability
("weak convergence")



Goal: Give some estimate about $\mathbb{P}(|X_n - X| > \varepsilon)$

(in the hope they will allow to say is small and $\xrightarrow{n\to\infty} 0$)

Lemma 6.2.1: (Markov's inequality) Let X be a non-negative r.v. Then, for any $\varepsilon > 0$

$$\mathbb{P}(X > \varepsilon) \leq \frac{\mathbb{E}[X]}{\varepsilon}$$

Why? ① Look at the r.v. $\mathbb{1}_{\{X>\varepsilon\}}$ (indicator function) $= \begin{cases} 1 & \text{if } X > \varepsilon \\ 0 & \text{otherwise} \end{cases}$

② Look at the r.v. $\varepsilon \cdot \mathbb{1}_{\{X>\varepsilon\}} = \begin{cases} \varepsilon & \text{if } X > \varepsilon \\ 0 & \text{otherwise} \end{cases}$

$\varepsilon \cdot \mathbb{1}_{\{X>\varepsilon\}} \leq X$ i.e. $\underline{\mathbb{1}_{\{X>\varepsilon\}} \leq \frac{X}{\varepsilon}}$

$$\mathbb{P}(X>\varepsilon) = \mathbb{E}[\mathbb{1}_{\{X>\varepsilon\}}] \leq \mathbb{E}\left[\frac{X}{\varepsilon}\right] = \frac{1}{\varepsilon}\cdot\mathbb{E}[X] \qquad \text{"} \smile \text{"} \qquad \varnothing$$

we saw $\mathbb{P}(A) = \mathbb{E}[\mathbb{1}_A]$

Lemma 6.2.2 (Chebyshev's inequality) Let X be a r.v. with $\operatorname{Var}(X) > 0$. Then, for any $\varepsilon > 0$

$$\mathbb{P}(|X - \mathbb{E}[X]| > \varepsilon) \leq \frac{\operatorname{Var}(X)}{\varepsilon^2}$$

how likely it is, that X is "ε-far" from its mean

$$\mathbb{P}(|X - \mathbb{E}[X]| > \varepsilon) = \mathbb{P}(|X - \mathbb{E}[X]|^2 > \boxed{\varepsilon^2}) \leq \frac{\mathbb{E}[|X - \mathbb{E}[X]|^2]}{\boxed{\varepsilon^2}} = \frac{\operatorname{Var}(X)}{\varepsilon^2}$$

non negative r.v. !!

Result is true for any type of distribution!

6.3 The weak law of large numbers (WLLN)

Relate the sample mean with "true" mean (average) of a r.v.

data sample $X_1, X_2, X_3, \ldots, X_n$

sample mean $\frac{1}{n} \sum_{i=1}^{n} X_i$

$E[X]$

r.v. representing "typical data point"

Idea: the more data points we have (the more often an experiment is repeated)

the more confident we are about

the average value of a typical point (average value of outcome)

Theorem 6.3.1 : (wLLN) Let $X_1, X_2, X_3, \dots$ be a sequence of i.i.d. r.v.s with $\mathbb{E}[X_1] < \infty$ and $\mathrm{Var}(X_1) > 0$. Then, for any $\varepsilon > 0$

i.i.d. ↑ independent identically distributed

$$\lim_{n\to\infty} \mathbb{P}\left(\left| \frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}[X_1] \right| > \varepsilon \right) = 0$$

i.e.
$$\underbrace{\frac{1}{n}\sum_{i=1}^{n} X_i}_{\text{sample mean}} \xrightarrow{\mathbb{P}} \underbrace{\mathbb{E}[X_1]}_{\text{"true mean"}} \qquad \text{(weak convergence)}.$$

How ? use Chebyshev's ineq : $\left[\mathbb{P}(|Y - \mathbb{E}[Y]| \geq \varepsilon) \leq \dfrac{\mathrm{Var}(Y)}{\varepsilon^2} \right.$ ⇝ use $\boxed{Y = \frac{1}{n}\sum_{i=1}^{n} X_i}$

??

$Y = \frac{1}{n}\sum_{i=1}^{n} X_i$ ⇝ $\mathbb{P}\left(\left| \frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}[X_1] \right| > \varepsilon \right)$

$$\| $$

$$\lim_{n\to\infty} \mathbb{P}(|Y - \mathbb{E}[Y]| > \varepsilon) \leq \lim_{n\to\infty} \frac{\mathrm{Var}(Y)}{\varepsilon^2} = \frac{\frac{1}{n}\mathrm{Var}(X_1)}{\varepsilon^2} = \lim_{n\to\infty} \frac{\mathrm{Var}(X_1)}{n \cdot \varepsilon^2} = 0$$

then, $\boxed{\mathbb{E}[Y] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right]}$

$$= \frac{1}{n}\mathbb{E}\left[\sum_{i=1}^{n} X_i\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[X_i] =$$

$$\overset{\text{indep.}}{=} \frac{1}{n} \cdot n \cdot \mathbb{E}[X_1] = \boxed{\mathbb{E}[X_1]}$$

$$\mathrm{Var}(Y) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) \overset{\text{indep.}}{=} \frac{1}{n^2}\cdot\sum_{i=1}^{n}\mathrm{Var}(X_i) = \frac{1}{n^2}\cdot n \cdot \mathrm{Var}(X_1) = \frac{1}{n}\mathrm{Var}(X_1)$$

Example 6.3.2: The weight of a Tesla S is modeled by a r.v. with mean 4.6 and standard deviation 0.03 (thousands of lbs).

X = weight of "typical" Tesla S, $\mathbb{E}[X] = 4.6$, $\sqrt{Var(X)} = 0.03$

? At least $\boxed{\text{how many}}$ Teslas should be weighted so that the sample mean does not deviate from the true mean by more than two standard deviations with 95% probability or higher?

X_i = weight of ith Tesla S. We know $\mathbb{E}[X_i] = 4.6$, $\sqrt{Var(X_i)} = 0.03$ $\frac{1}{n}\sum_{i=1}^{n} X_i$

Q: $\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - 4.6\right| < 2 \cdot 0.03\right) \geq 0.95$ for which $\boxed{n}$?

Know (WLLN): $\boxed{\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}[X_i]\right| > \underbrace{0.06}_{\varepsilon}\right)} \leq \frac{0.03^2}{n \cdot 0.06^2} \Rightarrow \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - 4.6\right| < 0.06\right) = 1 - \square \geq 1 - \frac{0.03^2}{n \cdot 0.06^2}$

Want: $1 - \frac{0.03^2}{n \cdot 0.06^2} \geq 0.95 \Rightarrow 0.05 \geq \frac{0.03^2}{n \cdot 0.06^2} \Rightarrow n \geq \frac{0.03^2}{0.05 \cdot 0.06^2} = 5$:)

Example 6.2.3 :

$N =$ # million customers per week in Starbucks $\sim N(\underset{E[N]}{60}, \underset{Var(N)}{16})$

? Give a lower bound for the the prob. that this week between 50 and 70 million people visit Starbucks.

[i.e. $\mathbb{P}(\quad\quad) \geq$ lower bound]

$$\mathbb{P}(50 < N < 70) = \mathbb{P}(-10 < N-60 < 10) = \mathbb{P}(|N-60| \boxed{<} 10)$$

$$= 1 - \mathbb{P}(|N-60| > 10)$$

$$\geq 1 - 0.16 = 0.84$$

Chebyshev :

$$\mathbb{P}(|N - E[N]| > \varepsilon) \leq \frac{Var(N)}{\varepsilon^2}$$

$$\mathbb{P}(|N-60| \boxed{>} 10) \leq \frac{16}{100} = 0.16$$

$$\Rightarrow \quad -\mathbb{P}(|N-60|>10) \geq -0.16$$

6.4. The strong law of large numbers (SLLN)

Recall: wLLN $\quad \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \mathbb{E}[X_1] \quad$ i.e. $\quad \lim_{n\to\infty} \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}[X_1]\right| > \varepsilon\right) = 0$ for all $\varepsilon > 0$

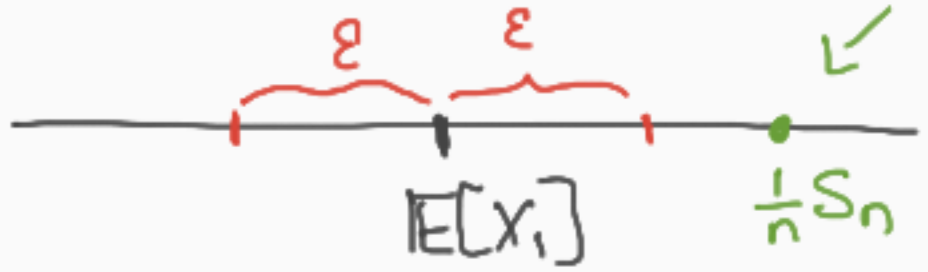
Today: SLLN $\quad \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{a.s.} \mathbb{E}[X_1] \quad$ i.e. $\quad \mathbb{P}\left(\lim_{n\to\infty}\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}[X_1]\right| = 0\right) = 1$

Weak vs strong.

Look at the event

$$E_n = \left\{\left|\underbrace{\frac{1}{n}\sum_{i=1}^{n} X_i}_{=\frac{1}{n}S_n} - \mathbb{E}[X_1]\right| > \varepsilon\right\}$$

wLLN



if E_n is happening with very small prob.

[but it may happen for "too many" ε !]

(countably infinity)

SLLN

if E_n is happening with small prob. if at all, only for finitely many ε !

Theorem 6.4.1 (SLLN): let $X_1, X_2, X_3 \dots$ be a sequence of i.i.d. r.v.s with $\mathbb{E}[X_1] < \infty$ and $Var(X_1) > 0$. Then,

$$\mathbb{P}\left(\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} X_i = \mathbb{E}[X_1] \right) = 1,$$

i.e. $\frac{1}{n} \sum_{i=1}^{n} X_i \xrightarrow{a.s.} \mathbb{E}[X_1]$.

How? long story...

idea (when $\mathbb{E}[X_1^4] < \infty$)

1. Use Markov's ineq to bound $\mathbb{P}\left(\left| \frac{1}{n} \sum_{i=1}^{n} X_i - \mathbb{E}[X_1] \right|^4 > \frac{1}{n^{4\gamma}} \right)$ with some $\gamma \in (0, \frac{1}{4})$.

2. Use previous bound and Borel-Cantelli with the events $\left\{ \left| \frac{1}{n} \sum_{i=1}^{n} X_i - \mathbb{E}[X_1] \right| > \frac{1}{n^{\gamma}} \right\}$

$\varnothing$

<u>Example 6.4.2</u> : E (event) = iPhone 11 weights 6.84 oz.

Based on SLLN, design a way to approx. $\mathbb{P}(E)$.

① Perform the experiment of weighting an iPhone 11.

② Did E occur ? $<$ Yes / No r.v $X_1 = \underbrace{\mathbb{1}_E}_{(\text{indicator})} = \begin{cases} 1 & \text{if } E \text{ occurs} \\ 0 & \text{if not} \end{cases}$

③ Repeat the experiment n times ⤳ $X_1, X_2, X_3, \ldots, X_n$

④ How often did E occur ? Define r.v. $S_n = \underbrace{X_1 + X_2 + \ldots + X_n}_{\#\text{ times } E \text{ occured}}$

$$\underbrace{\frac{1}{n} \boxed{S_n}}_{\text{sample mean}} \overset{\text{SLLN}}{\approx} \mathbb{E}[X_1] = \mathbb{E}[\mathbb{1}_E] = \mathbb{P}(E)$$

☺

What was Tim Cook talking about 2015 ?

$$\frac{1}{n}\sum_{i=1}^{n} X_i \approx \mathbb{E}[X_1] \quad \rightsquigarrow \quad \underbrace{\sum_{i=1}^{n} X_i}_{\$\text{ profit in } n \text{ years}} \approx n \cdot \mathbb{E}[X_1] = \mathbb{E}[X_1] + \mathbb{E}[X_1] + \dots + \mathbb{E}[X_n]$$

each year approx. same profit

X_i = \$ profit in a year

rate of growth of 50% : X_1 X_2 X_3

$10 \quad 10+5 \quad (10+5) + 0.5(10+5) \quad \dots$

each year notably growing profit

6.5 Central limit theorem

Recall: $X_1, X_2, X_3, \ldots\ldots$ data (independent measurments of a quantity of interest X)

Q: What can we say about the distribution $S_n = \sum_{i=1}^{n} X_i$?

Know LLN: $\frac{1}{\boxed{n}} S_n \xrightarrow[n\to\infty]{P,\,a.s.} \underbrace{\mathbb{E}[X_1]}$

"true mean" is a number (so no random?!)

- Try a different scaling?
- How? Look at $|S_n - \mathbb{E}[S_n]| = |S_n - n \cdot \mathbb{E}[X_1]|$ ← tells us how much S_n deviates from its mean

Wish: (1) On average, $(S_n - \mathbb{E}[S_n])\frac{1}{\square} = 0$, i.e.

$$\mathbb{E}\left[\frac{1}{\square}(S_n - \mathbb{E}[S_n])\right] = 0$$

(2) Some positive variance, i.e. $\mathrm{Var}\left(\frac{1}{\square}(S_n - \mathbb{E}[S_n])\right) = 1$.

Try: $\square = \sqrt{n\,\mathrm{Var}(X_1)}$

Check wishes: (1) $\mathbb{E}\left[\frac{1}{\sqrt{n\,\mathrm{Var}(X_1)}}(S_n - \mathbb{E}[S_n])\right] = \frac{1}{\sqrt{n\,\mathrm{Var}(X_1)}}(\mathbb{E}[S_n] - \mathbb{E}[S_n]) = 0$ ☺

(2) $\mathrm{Var}\left(\frac{1}{\sqrt{n\,\mathrm{Var}(X_1)}}(S_n - \mathbb{E}[S_n])\right) = \frac{1}{n\,\mathrm{Var}(X_1)}\mathrm{Var}(S_n - \underbrace{\mathbb{E}[S_n]}_{\text{number}})$

$= \frac{1}{n\,\mathrm{Var}(X_1)}\mathrm{Var}(S_n) = 1$ ☺

$(*)$ $\mathrm{Var}(S_n) = \mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = n.\mathrm{Var}(X_1)$

X_i i.i.d. r.v.s

<u>Theorem 6.4.1</u> (Central limit theorem -CLT-) : Let $X_1, X_2, X_3, \ldots$ be a sequence of i.i.d. r.v.s with $\mathbb{E}[X_1] < \infty$ and $\mathrm{Var}(X_1) > 0$. Then,

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{S_n - \overbrace{n \cdot \mathbb{E}[X_1]}^{\mathbb{E}[S_n]}}{\sqrt{\underbrace{n \cdot \mathrm{Var}(X_1)}_{\mathrm{Var}(S_n)}}} \leq a \right) = \underbrace{\mathbb{P}(Z \leq a) = \Phi(a)}_{\text{c.d.f. of } Z \sim N(0,1)} \quad \text{for any } a \in \mathbb{R}$$

i.e. $$\frac{S_n - n \cdot \mathbb{E}[X_1]}{\sqrt{n \mathrm{Var}(X_1)}} \xrightarrow{d} Z .$$ (convergence in distribution)

For any $a < b$

$$\mathbb{P}\left(a \leq \frac{S_n - n\mathbb{E}[X_1]}{\sqrt{n \mathrm{Var}(X_1)}} \leq b \right) = \Phi(b) - \Phi(a) .$$

Example 6.5.2 :

loads 100 packages

each weights independently of the others some quantity uniformly distributed between 5 lb and 55 lb.

? Approx. prob. that the total weight lies between 2500 & 3000 lb.

$$S_{100} = \sum_{i=1}^{100} X_i$$

X_i = weight of ith package

$X_i \sim \text{Unif}\,(5,55)$

$$\mathbb{E}[X_1] = \frac{5+55}{2} = 30$$

$$\text{Var}(X_1) = \frac{(55-5)^2}{12} = \frac{50^2}{12}$$

$$\mathbb{P}(2500 \le S_{100} \le 3000) =$$

(no continuity correction needed b/c X_i continuous!)

$$\mathbb{P}\left(\frac{2500-100\cdot 30}{\sqrt{100\cdot\frac{50^2}{12}}} \le \frac{S_{100}-n\cdot\mathbb{E}[X_1]}{\sqrt{n\,\text{Var}(X_1)}} \le \frac{3000-100\cdot 30}{\sqrt{100\cdot\frac{50^2}{12}}}\right)$$

$$= \mathbb{P}\left(\frac{-\cancel{500}}{10\cdot\cancel{50}\cdot\sqrt{\frac{1}{12}}} \le \qquad \le 0\right) \overset{CLT}{\approx} \Phi(0) - \Phi(\underbrace{-2\sqrt{3}}_{3.46}) = \frac{1}{2} - 1 + \Phi(3.46) =$$

Chapter 7. Markov chains

Basis of mathematical models for dynamics such as the spread/transmission of infectious diseases.

evolution in time (discrete or continuous)

e.g. S I R model

susceptible (not yet immune)

infectious (currently infected)

removed (immune)

S I R

→ 3 possible states

→ Q: How does the group size vary in time?

proportion of population

probability!

X_n = state of typical person at day n

$X_0, X_1, X_2, X_3, \dots, X_n$ ← chain describing the states of the person from day 0 to n.

Def. 7.1: let $\Omega = \{1, 2, 3, \dots, N\}$ and $\{p_{ij}\}_{i,j\in\Omega}$ be numbers such that

1) $p_{ij} \geq 0$ for all $i,j \in \Omega$

2) $\displaystyle\sum_{j=1}^{N} p_{ij} = 1$ for any $i \in \Omega$.

⚠ Only with i fixed, not the other way around! (i.e. j fixed)

A discrete-time Markov chain is a sequence of discrete r.v.s $X_0, X_1, X_2, \dots$ with the property

$$\mathbb{P}(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = p_{ij} \quad \leftarrow \text{transition prob.}$$

for any $n \in \mathbb{N}$ and $j, i, i_{n-1}, i_{n-2}, \dots, i_0 \in \Omega$.

- **What is what?**

Ω = state space

p_{ij} = transition prob.

the current state (j) only depends on the most previous one (i).

current state

most previous state

transition matrix

$$\begin{array}{c|ccccc}
i \backslash j & 1 & 2 & 3 & \dots & N \\
\hline
1 & p_{11} & p_{12} & p_{13} & \dots & p_{1N} \\
2 & p_{21} & p_{22} & & \dots & p_{2N} \\
3 & & & & & \\
\vdots & & \vdots & & & \\
N & p_{N1} & & \dots & \dots & p_{NN}
\end{array}$$

<u>Prop. 7.2</u>: The probability of a path (sequence of states) $i_0, i_1, i_2, \ldots, i_n \in \Omega$ is given by

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

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

Bayes!

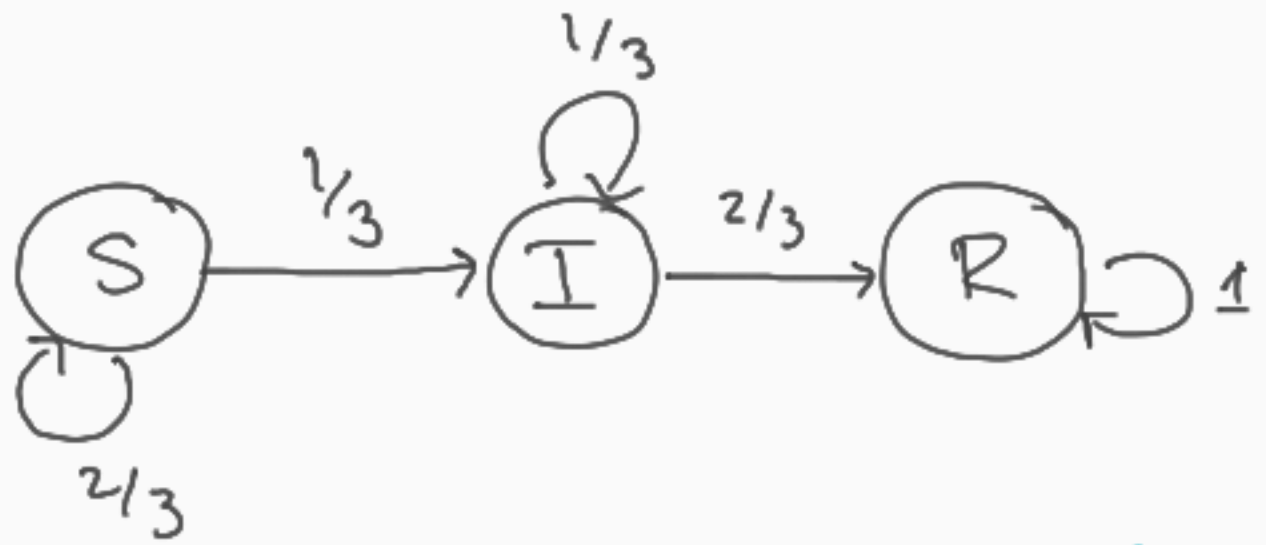
$$= \underbrace{\mathbb{P}(X_n = i_n \mid X_{n-1} = i_{n-1}, \ldots, X_0 = i_0)}_{p_{i_n i_{n-1}}} \cdot \underbrace{\mathbb{P}(X_{n-1} = i_{n-1} \mid X_{n-2} = i_{n-2}, \ldots, X_0 = i_0)}_{p_{i_{n-1} i_{n-2}}} \cdot \mathbb{P}(X_{n-2} = i_{n-2}, \ldots, X_0 = i_0)$$

$$= \ldots\ldots$$

$$= p_{i_n i_{n-1}} \cdot p_{i_{n-1} i_{n-2}} \cdots p_{i_1 i_0} \cdot \boxed{\mathbb{P}(X_0 = i_0)}$$

initial distribution.

Example 7.3: In the model



X_n = state on day n

? Assume nobody is born immune.

What is the prob. to get infected on day 4?

$\Omega = \{S, I, R\}$

X_0	X_1	X_2	X_3	X_4
S	S	S	S	I

$i \backslash j$	S	I	R	
S	2/3	1/3	0	= 1
I	0	1/3	2/3	= 1
R	0	0	1	= 1
	≠ 1	≠ 1	≠ 1	

$$\mathbb{P}(X_4 = I, X_3 = S, X_2 = S, X_1 = S, X_0 = S)$$

$$= p_{SI} \cdot p_{SS} \cdot p_{SS} \cdot p_{SS} \cdot \mathbb{P}(X_0 = S)$$

$$= \frac{1}{3} \cdot \frac{2}{3} \cdot \frac{2}{3} \cdot \frac{2}{3} \cdot 1 = \frac{2^3}{3^4}$$

Recall: $X_1, X_2, X_3, \dots$ Markov chain discrete time

$\{p_{ij}\}_{i,j\in\Omega}$ transition prob. $\rightsquigarrow$ prob. that $\underbrace{\text{today's state}}_{\text{day } n}$ is j given $\underbrace{\text{yesterday's}}_{\text{day } n-1}$ was i

Q: prob. that $\underbrace{\text{today's state}}_{\substack{\text{day } n \\ \text{(or future day } n)}}$ is j given $\underbrace{\text{initial state}}_{\substack{\text{day } 0 \\ \text{(if today counts as 0)}}}$ was i

A: n-step transition probabilities

Def. 7.4: For each $n \geq 1$, the n-step transition prob's of a discrete-time Markov chain with $\Omega = \{1, \dots, N\}$ and $\{p_{ij}\}_{i,j\in\Omega}$ transition prob's are

$$p_{ij}^{(n)} = \mathbb{P}(X_n = j \mid X_0 = i) \qquad i, j \in \Omega$$

? How do we compute $p_{ij}^{(n)}$?

$$p_{ij}^{(1)} = \mathbb{P}(X_1 = j \mid X_0 = i) = p_{ij}$$

$$p_{ij}^{(n)} = \mathbb{P}(X_n = j \mid X_0 = i) = \sum_{k=1}^{N} \underbrace{\mathbb{P}(X_n = j \mid X_{n-1} = k, X_0 = i)}_{= \mathbb{P}(X_n = j \mid X_{n-1} = k, \dots, X_0 = i)} \cdot \mathbb{P}(X_{n-1} = k \mid X_0 = i)$$

law of total prob.

$$= \sum_{k=1}^{N} p_{kj} \cdot p_{ik}^{(n-1)} \quad (*)$$

recursive formula!

Prop. 7.5 (Chapman-Kolmogorov equations): For $1 \le r < n$ and $i, j \in \Omega$

$$p_{ij}^{(n)} = \sum_{k=1}^{N} p_{ik}^{(r)} \cdot p_{kj}^{(n-r)}$$

[use r instead of $n-1$ in the previous computation $(*)$]

Q: How do the dynamics of a chain behave in the long run? ($n \to \infty$)

Theorem 7.6 (Steady-state convergence thm): If there is $n \geq 1$ for which $\boxed{p_{ij}^{(n)} > 0}$ for all $i,j \in \Omega$

such a chain is called ergodic

then for each $j \in \Omega$, $\lim_{n \to \infty} p_{ij}^{(n)} = \pi_j$ (for all $i \in \Omega$).

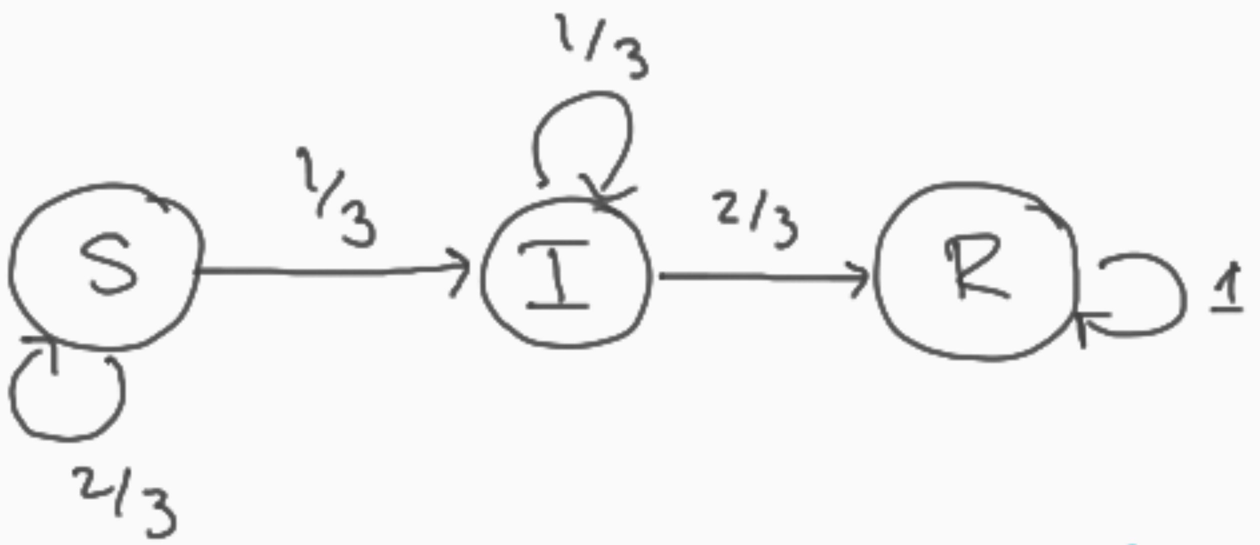
where

① $\pi_j = \sum_{k=1}^{N} \pi_k \cdot p_{kj}$

② $\sum_{k=1}^{N} \pi_k = 1$.

$\{\pi_i\}_{i \in \Omega}$ give the so-called stationary distribution of the chain.

Example 7.3 : In the model



X_n = state on day n

$\Omega = \{S, I, R\}$

Assume nobody is born immune.

? What is the prob. that one is infected by day 2 ? → 2-step transition prob.

maybe on day 1
or on day 2

i \ j	S	I	R
S	2/3	1/3	0
I	0	1/3	2/3
R	0	0	1

$$P^{(2)}_{SI} = P^{(1)}_{SS} \cdot P_{SI} + P^{(1)}_{SI} \cdot P_{II} + P^{(1)}_{SR} \cdot P_{RI}$$

$$= \frac{2}{3} \cdot \frac{1}{3} + \frac{1}{3} \cdot \frac{1}{3} + 0 \cdot 0$$

$$= \frac{1}{3}$$

? ergodic? No!