PTS - Chapter 2

2 Random Variables

We roll 2 dice: the sum $X = D_1 + D_2$,

$$X : S \longrightarrow \mathbb{R}$$

is a random variable

Idea: We are not interested in arbitrary events but only events that can be described by X having certain values

E.g. weight ≥ 100 kg, height < 1.60 m, ...

Back to dice

$$P([\mathcal{X} = 2]) = \frac{1}{36}$$

$$P([\mathcal{X} = 3]) = \frac{2}{36}$$

$$P([\mathcal{X} = 4]) = \frac{3}{36}$$

$$P([\mathcal{X} = 5]) = \frac{4}{36}$$

$$P([\mathcal{X} = 6]) = \frac{5}{36}$$

$$P([\mathcal{X} = 7]) = \frac{6}{36}$$

$$P([\mathcal{X} = 8]) = \frac{5}{36}$$

$$P([\mathcal{X} = 9]) = \frac{4}{36}$$

$$P([\mathcal{X} = 10]) = \frac{3}{36}$$

$$P([\mathcal{X} = 11]) = \frac{2}{36}$$

$$P([\mathcal{X} = 12]) = \frac{1}{36}$$

Check

$$\sum_{i=2}^{12} P([\mathcal{X} = i]) = 1$$

Other poss. events expressible by $\mathcal{X}$:

$$P[5 \leq \mathcal{X} \leq 9] = \frac{24}{36} = \frac{2}{3}$$

A random variable $X : S \to \mathbb{R}$

is discrete if it has only finitely (or countably) many values $x_1, \dots, x_n, \dots$

X is continuous if it takes a continuum of values (e.g. weight, ...)

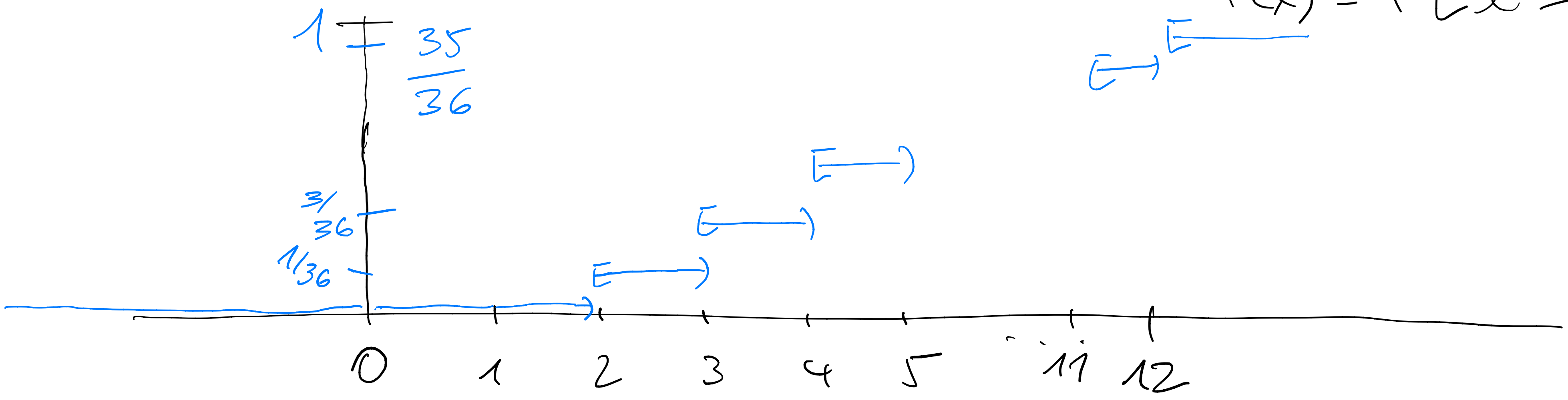
Definition 24: The (cumulative) distribution function of X is

$$F : \mathbb{R} \to [0, 1]$$
$$F(x) = P[X \le x]$$

"$X \sim F$" means "F is distribution of X"

Distribution for $X = D_1 + D_2$

$F(x) = P[X \leq x]$



$F(2) = P[X \leq 2]$

$F(3) = P[X \leq 3] = \frac{1}{36} + \frac{2}{36} = \frac{3}{36}$

F answers all probability questions about X:

Eg $\quad P[a < X \leq b] = ?$

$$[X \leq b] = [X \leq a] \,\dot{\cup}\, [a < X \leq b]$$

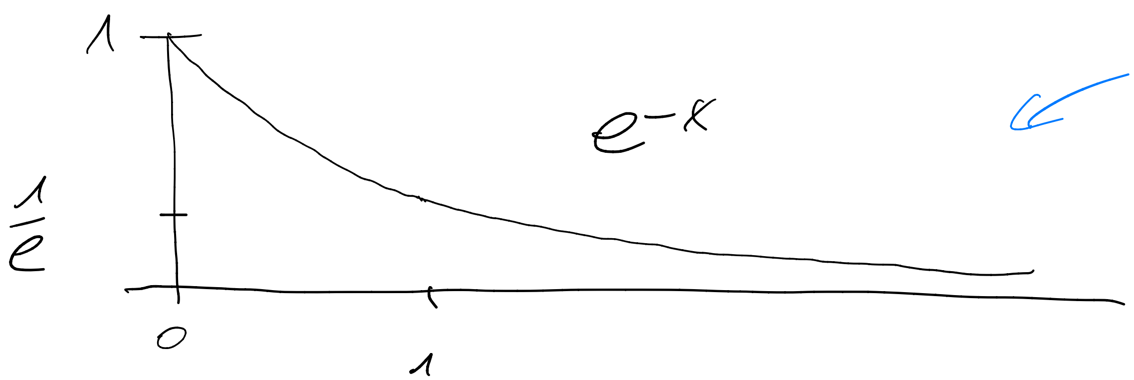
$$P[a < X \leq b] = P[X \leq b] - P[X \leq a]$$

$$= F(b) - F(a)$$

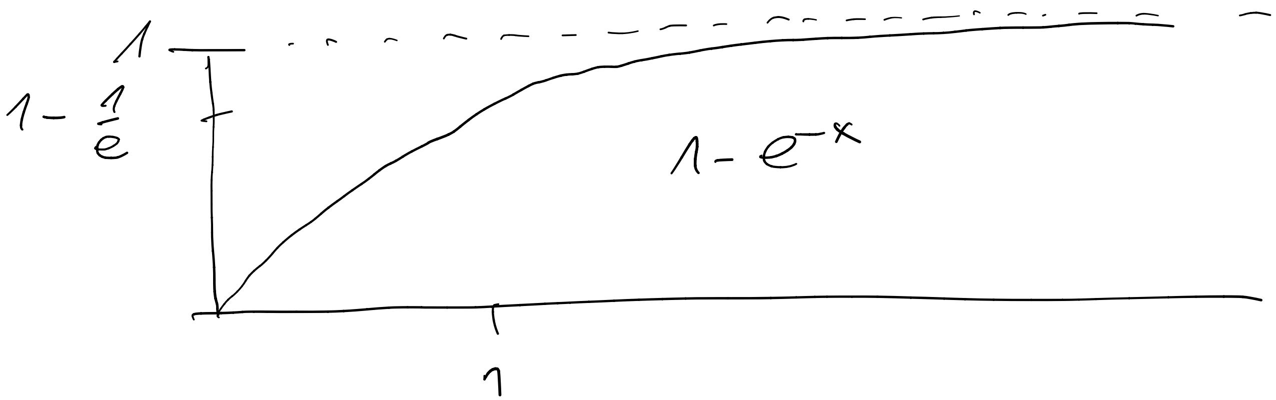
Example 25 a) Suppose $\mathcal{X} \sim F$

e.g., time until a device breaks, an atom decays

$$F(x) = \begin{cases} 0 & x \leq 0 \\ 1 - e^{-x}, & x > 0 \end{cases}$$



density of the exponential distribution

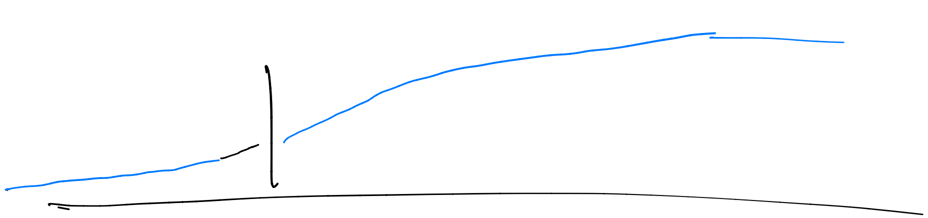


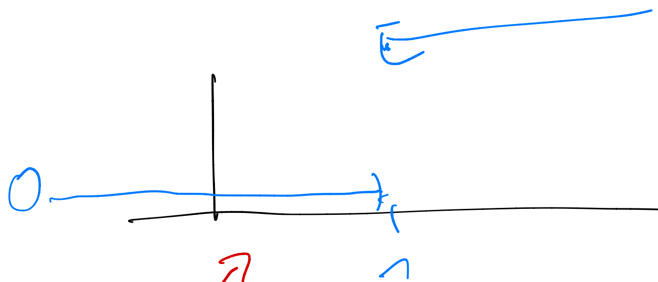
$F(x) = P[\mathcal{X} \leq x]$

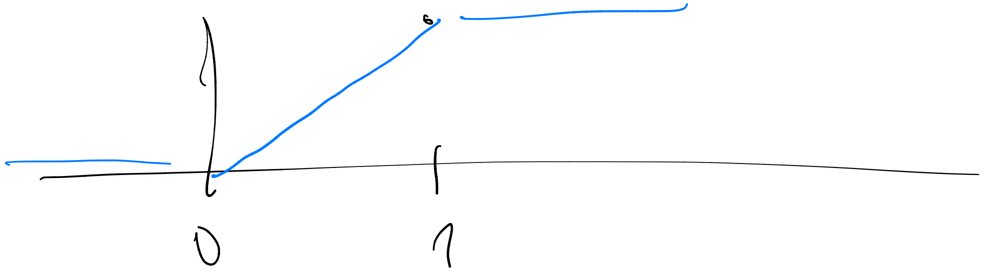
All distribution functions satisfy

- $0 \le F(x) \le 1$ (since $F(x) = P[X \le x]$ is a probability)
- F is monotonically increasing
- $\lim\limits_{x \to -\infty} F(x) = 0$
- $\lim\limits_{x \to +\infty} F(x) = 1$
- $\lim\limits_{\substack{x \to +x_0 \\ x > x_0}} F(x) = F(x_0)$

Possible shapes







X is discrete and the only possible value is 1

2.1 Types of Random Variables

Let X be discrete.

$$p : \mathbb{R} \longrightarrow [0,1]$$

(pmf)

$$p(x) = P[X = x]$$

is the probability mass function of X

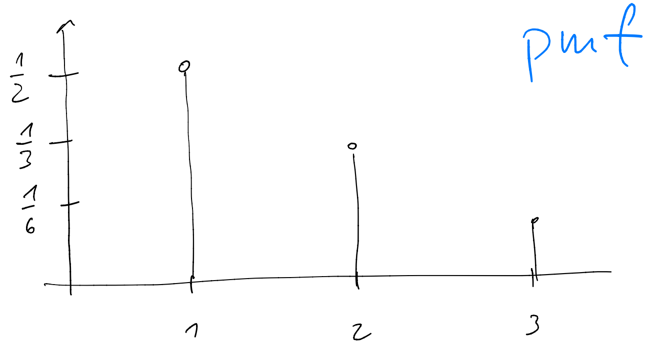
Let $x_1, \dots, x_n, \dots$ be the possible values of X

$$\sum_{i=1}^{\infty} p(x_i) = 1$$

Example 26: Let X takes values $\{1, 2, 3\}$ and $p(2) = \frac{1}{3}$, $p(3) = \frac{1}{6} \Rightarrow p(1) = \frac{3}{6} = \frac{1}{2}$

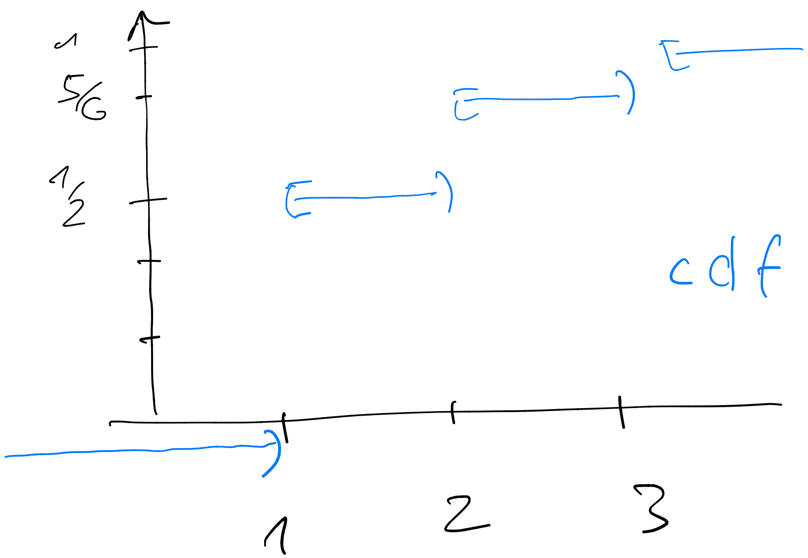
Example 26: Let X takes values $\{1, 2, 3\}$

and $p(2) = \frac{1}{3}$, $p(3) = \frac{1}{6}$ $\Rightarrow$ $p(1) = \frac{3}{6} = \frac{1}{2}$



Cumulative distribution function

$$F(x) = P[X \leq x] = \sum_{y \leq x} p(y)$$



In general:

- F is constant on each interval

$$[x_i, x_{i+1})$$

- F is a step function

Definition 27: X is continuous if there is a fct

$f : \mathbb{R} \to \mathbb{R}$, $f \geq 0$, sth (= such that)

$$P[X \in B] = \int_B f(x)\, dx$$

for all "$\underbrace{\text{reasonable}}_{\text{essentially unions intervals}}$" $B \subseteq R$. We call f the probability density function of X (pdf)

Remark:

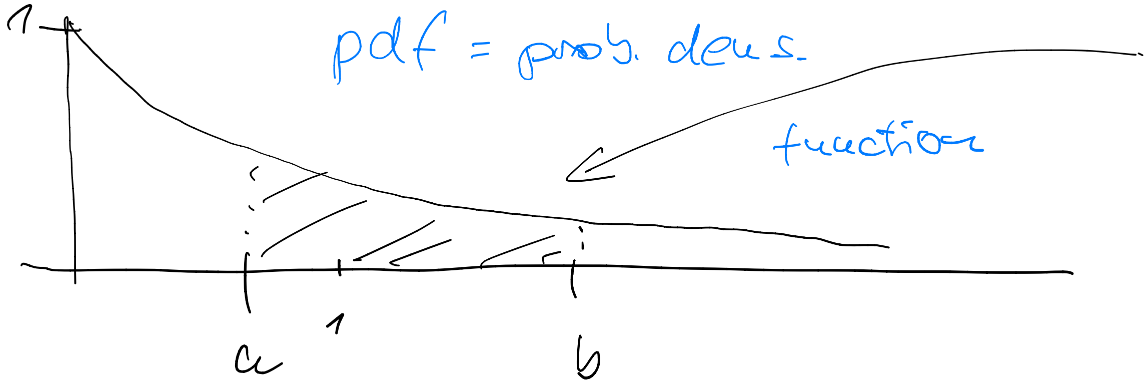
$$1 = P[X \in (-\infty, \infty)]$$

$$= \int_{-\infty}^{\infty} f(x)\, dx$$

$$P[a \leq X \leq b] = \int_a^b f(x)\, dx$$

Example:

$$f(x) = \begin{cases} e^{-x} & x \geq 0 \\ 0 & x < 0 \end{cases} \qquad F(x) = \Bigg\{$$



$P[a < X < b]$

$=$ area under the graph of f between a and b

For a discrete X we would take

$$\sum_{a < x < b} p(x)$$

Connection between cdf and pdf here:

$$F(x) = P[X \leq x] = \int_{-\infty}^{x} f(y)\,dy$$

$$\Longleftrightarrow \quad F'(x) = f(x) \quad \left(\wedge \lim_{x \to -\infty} F(x) = 0 \right)$$

Connection between cdf and pdf here:

$$F(x) = P[X \leq x] = \int_{-\infty}^{x} f(y)\, dy$$

$$\Longleftrightarrow \quad F'(x) = f(x) \quad \left(\wedge \lim_{x \to -\infty} F(x) = 0 \right)$$

What is F?

$$f(y) = \begin{cases} e^{-y} & y \geq 0 \\ 0 & y < 0 \end{cases}$$

$$\boxed{F(x) = \int_{-\infty}^{x} f(y)\, dy} = \begin{cases} \int_{-\infty}^{x} 0 \, dy = \boxed{0} & \text{if } x < 0 \\ \int_{-\infty}^{x} f(y)\, dy & \text{if } x \geq 0 \end{cases}$$

$$\int_{-\infty}^{x} f(y)\, dy = \int_{-\infty}^{0} f(y)\, dy + \int_{0}^{x} f(y)\, dy = 0 + \int_{0}^{x} e^{-y}\, dy$$

$$= \left[-e^{-y} \right]_{0}^{x} = -e^{-x} - (-e^{-0}) = -e^{-x} + 1 = \boxed{1 - e^{-x}}$$

This is the cdf of the exp. distribution

Example 30 Assume the following statistical figures for a community of families

no children	15 %	$= \frac{3}{20}$
1 child	20 %	$= \frac{4}{20}$
2 children	35 %	$= \frac{7}{20}$
3 children	30 %	$= \frac{6}{20}$

The probability of boys and girls is 50 % each.
What is the joint probability mass function for boys and girls?
(i.e.: $X = \#\text{boys}$, $Y = \#\text{girls}$)

Answer using conditioning on # of children!

$$P[\mathcal{X}=0,\ y=0] = P[\text{no children}] = \frac{3}{20} = \frac{12}{80}$$

$$P[\mathcal{X}=0,\ y=1] = P[\text{"1 child"} \cap \text{"child is girl"}]$$

$$= P[\text{"child is girl"} \mid \text{"1 child"}] \cdot P[\text{"1 child"}]$$

$$= \frac{1}{2} \cdot \frac{4}{20} = \frac{8}{80}$$

$$P[\mathcal{X}=0,\ y=2] = P[\text{"2 children are girls"} \mid \text{"2 children"}]$$

$$\cdot\, P[\text{"2 children"}]$$

$$= \frac{1}{2\cdot 2} \cdot \frac{7}{20} = \frac{7}{80}$$

$$P[\mathcal{X}=0,\ y=3] = P[\text{"3 children are girls"} \mid \text{"3 children"}]$$

$$\cdot\, P[\text{"3 children"}]$$

$$= \frac{1}{2^3} \cdot \frac{\cancel{6}\ 3}{\cancel{20}\ 10} = \frac{3}{80}$$

$$P[X=1, Y=0] = P[X=0, Y=1] = \frac{1}{2} \cdot \frac{4}{20} = \frac{8}{80}$$

$$P[X=1, Y=1] = P[\text{"1 boy, 1 girl"} \mid \text{"2 children"}] \cdot P[\text{"2 children"}]$$

$$= \frac{1}{2} \cdot \frac{7}{20} = \frac{14}{80}$$

$$P[X=1, Y=2] = P[\text{"1 boy, 2 girls"} \mid \text{"3 children"}] \cdot P[\text{"3 children"}]$$

$$= 3 \cdot \frac{1}{2^3} \cdot \frac{\overset{3}{\cancel{6}}}{\underset{10}{\cancel{20}}} = \frac{9}{80}$$

$$P[X=2, Y=0] = P[X=0, Y=2]$$

$$P[X=2, Y=1] = P[X=1, Y=2]$$

$$P[X=3, Y=0] = P[X=0, Y=3]$$

This is the summary table (multiples of $\frac{1}{80}$)

$x \backslash y$	0	1	2	3	Sum
0	12	8	7	3	30
1	8	14	9	0	31
2	7	9	0	0	16
3	3	0	0	0	3
Sum	30	31	16	3	80

marginal probabilities of x

marginal probabilities of y

This is the summary table (multiples of $\frac{1}{80}$)

$x \backslash y$	0	1	2	3	Sum
0	12/80	8/80	7/80	3 ...	$\frac{30}{80}$
1	8	14	9	0	$\frac{31}{80}$
2	7	9	0	0	$\frac{16}{80}$
3	3	0	0	0	$\frac{3}{80}$
Sum	30	31	16	3	$\frac{80}{80}$

marginal probabilities of x

marginal probabilities of y

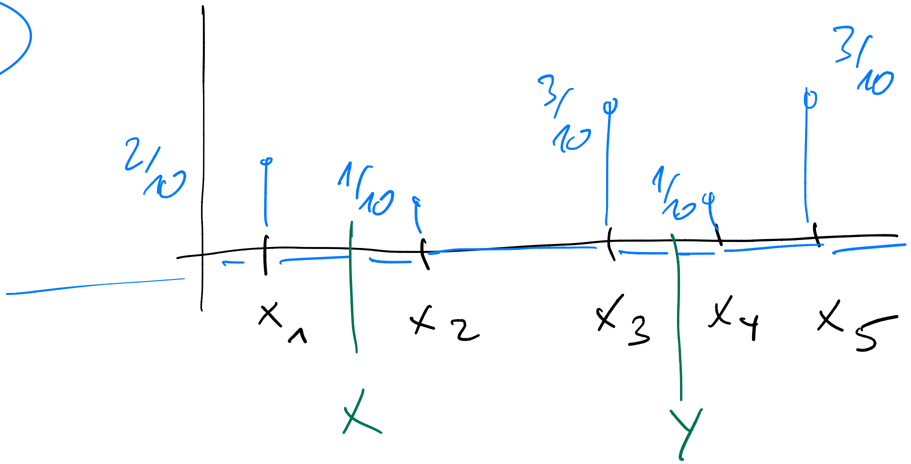
We want a table of the form

$x \backslash y$	0	1	2	3	Sum
0					
1				0	
2			0	0	
3		0	0	0	
Sum					

Revision: Random Variables $\mathcal{X} : S \to \mathbb{R}$

Distinguish 2 kinds

discrete



$$P[x < \mathcal{X} \leq y] = \sum_{x < x_i \leq y} p(x_i)$$

Probabilities $p(x_i)$ for each single value x_i that $\mathcal{X}$ can take

$0 < p(x_i) < 1$

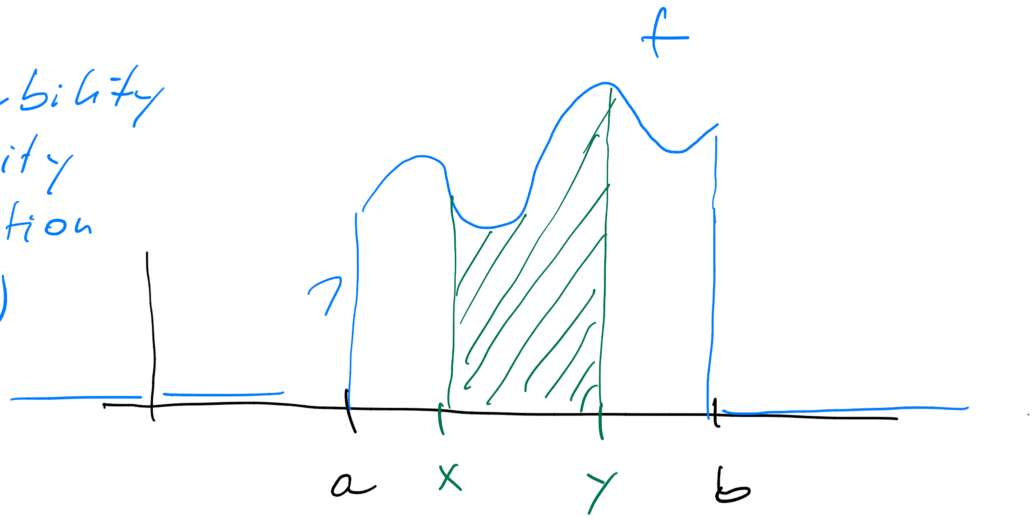
$\sum_i p(x_i) = 1$

p is the probability mass function (pmf) continuous

$$\int_a^b f(z)\,dz = 1 \quad , \quad f(x) \geq 0$$

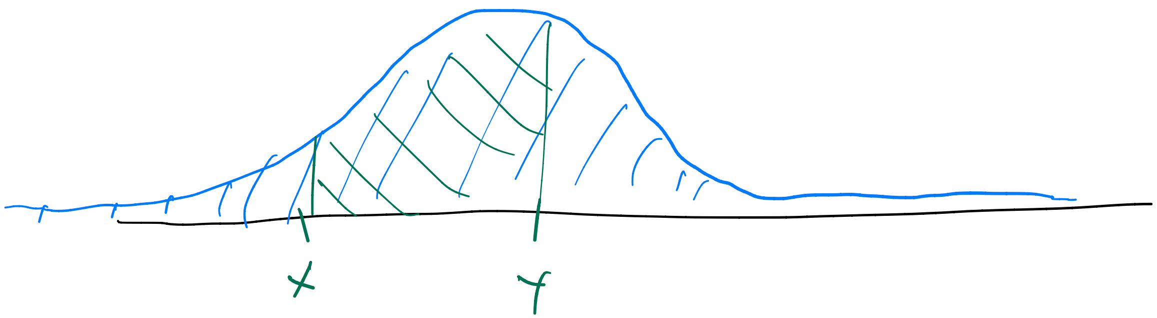
$$\int_{-\infty}^{\infty} f(z)\,dz = 1$$

probability density function (pdf)



$$P[x < \mathcal{X} \leq y] = P[x \leq \mathcal{X} \leq y] = \int_x^y f(z)\,dz$$

$$\int_{-\infty}^{\infty} f(z)\,dz = 1$$



In continuous case

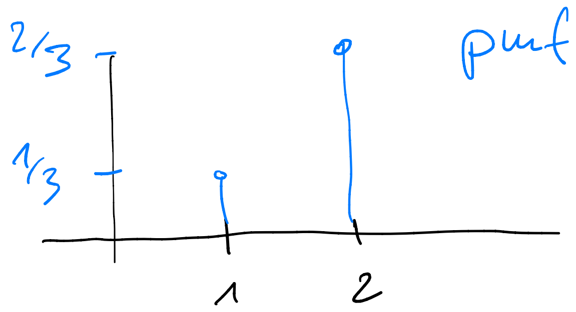
$$P[\mathcal{X} = x] = \int_x^x f(z)\,dz = 0$$

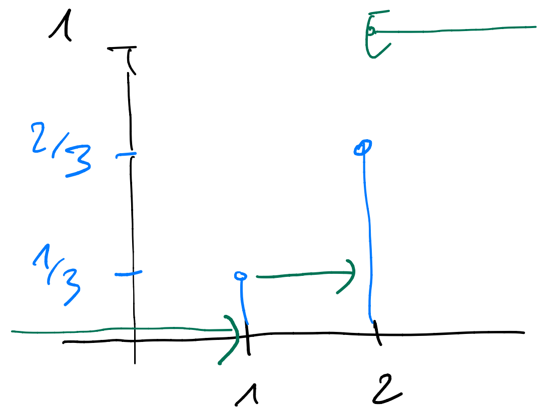
i.e., a single value has probability 0

Distribution Function

$$F(x) = P[X \le x]$$

discrete case





$$F(x) = \sum_{x_i \le x} p(x_i)$$

$$F(x) = \int_{-\infty}^{x} f(z)\,dz$$

continuous case

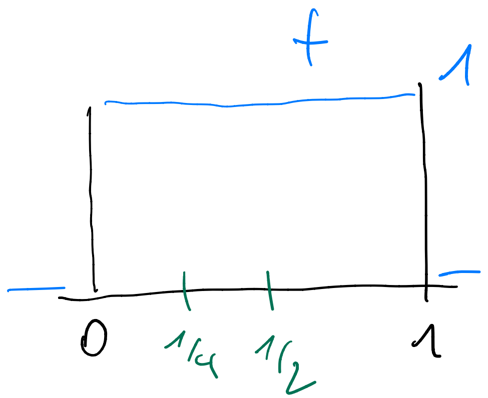
Uniform distribution $U[0,1]$

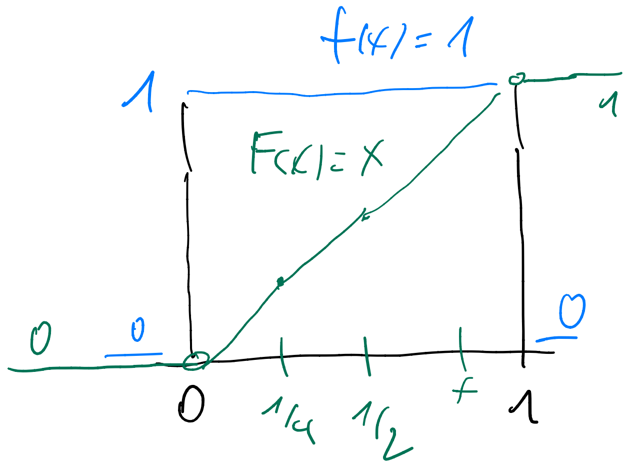
(special case of $U[a,b]$)

$$f(x) = \begin{cases} 1 & \text{if } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Always: $F' = f$

(from Fundamental Theorem of Calculus)





Here, $F(x) = \int_0^x f(z)\,dz$

$= \int_0^x 1\,dz = x$

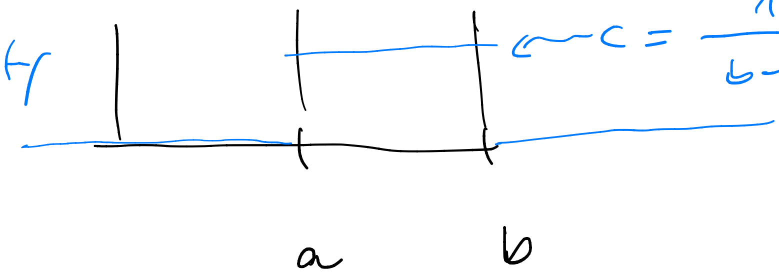
since f is 0 on $[-\infty, 0]$

Examples of Continuous Distributions

1) <u>Uniform Distribution:</u> "Waiting for bus"

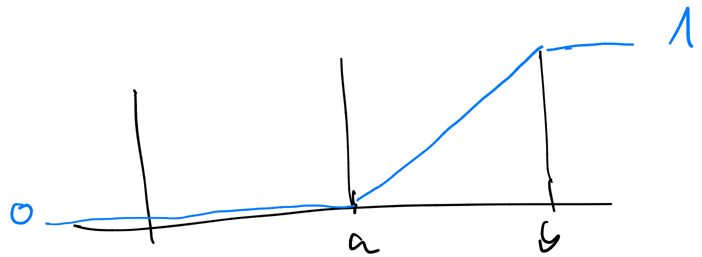
$U[a,b]$: uniform on $[a,b]$

density



$c = \frac{1}{b-a} = f(x)$, $a \leq x \leq b$

distribution



2 Exponential Distribution

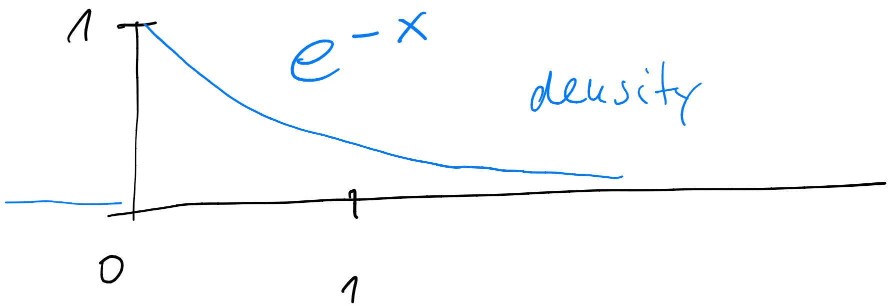
Waiting for: a customer, an email, an atom to decay

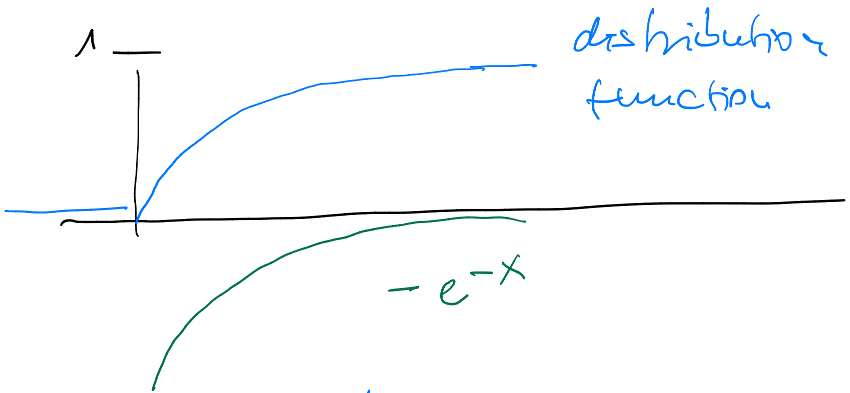
"Process w/o memory":

The probability to wait at least a minutes does not depend on having waited already b minutes

$$P[X > a] = P[X > a+b \mid X > b]$$

This leads to the exponential distribution,





$1 - e^{-x}, \; x \geq 0;$

0 otherwise

Check: the derivative of the distribution is the density

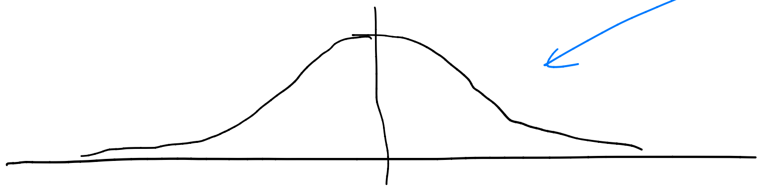
$$\frac{d}{dx}\, 1 - e^{-x} = 0 - e^{-x}(-1) = e^{-x}$$

3. Normal Distribution $\mathcal{N}(0, \frac{1}{2})$

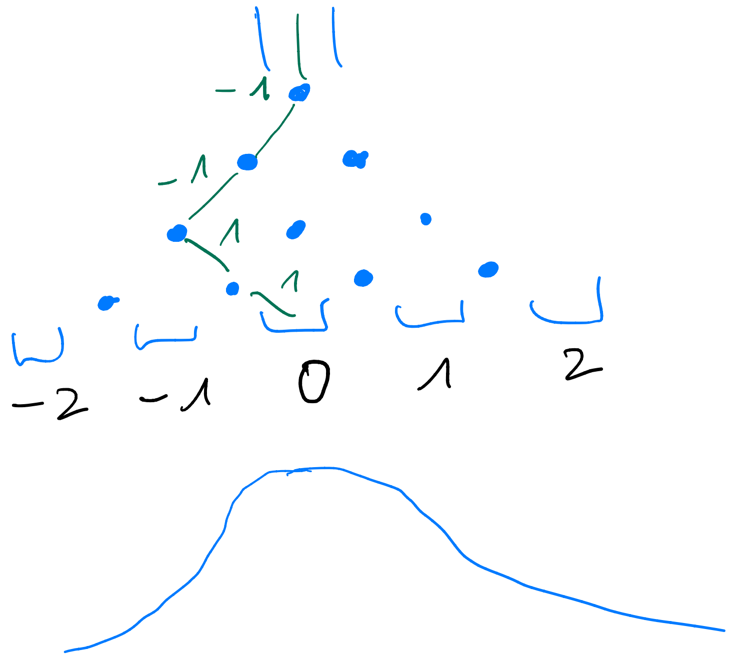
mean — variance

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

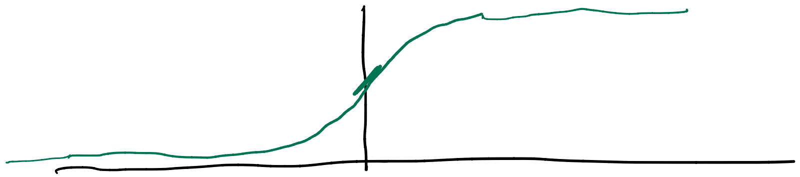
density of $\mathcal{N}(0, \frac{1}{2})$



Sum of independent RVs, with same distribution, e.g. Galton Board



$\mathcal{N}(0, \frac{1}{2})$ distribution

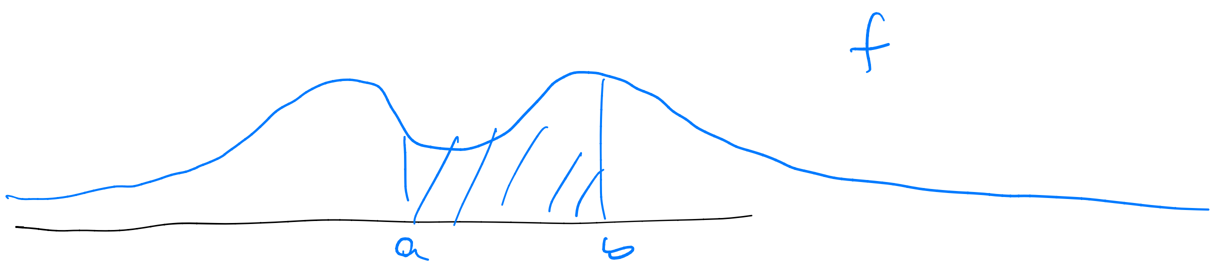


Distribution function of the normal distribution is written Φ for $\mathcal{N}(0,1)$

Density vs. Distribution (Student Question)

$F(x) = P[X \le x]$ is the definition of the distribution function of X

Suppose f is the density of X (X continuous)



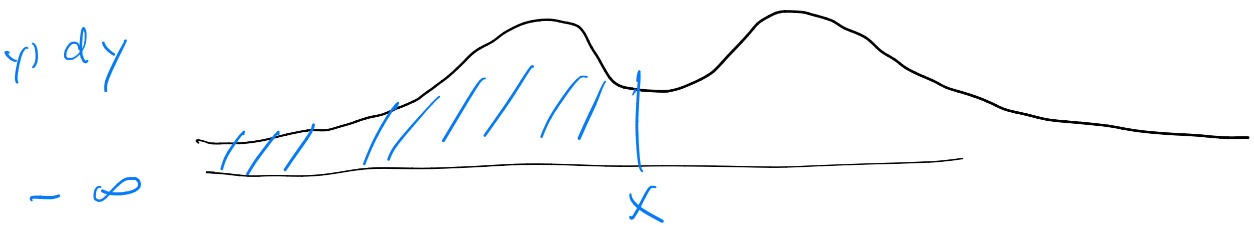
Then $f \ge 0$,

$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

"f is density of" means "$P[a \le X \le b] = \int_a^b f(x)\,dx$"

What is F in this case?

$$F(x) = P[X \le x] = \int_{-\infty}^{x} f(y)\,dy$$

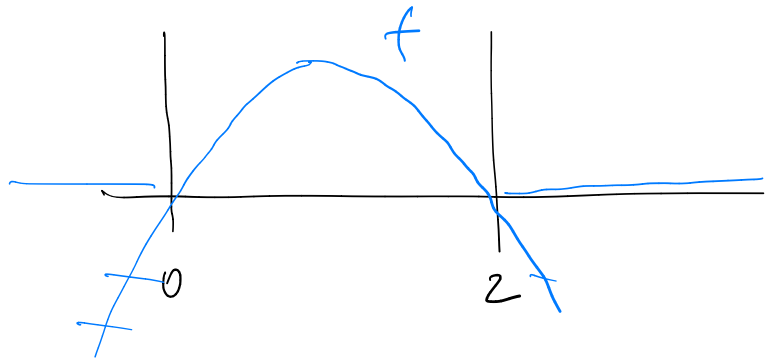


Exercice (Quiz)

$$f(x) = \begin{cases} c\,(2x - x^2), & 0 \le x \le 2 \\ 0 & \text{otherwise} \end{cases}$$

$$f(x) = c \quad x\,(2-x)$$
$$= -c \quad (x-0)(x-2)$$

Which c turns f into a density?



We want

$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

Here,
$$\int_{-\infty}^{\infty} f(x)\,dx = \int_0^2 f(x)\,dx = \int_0^2 c\,(2x - x^2)\,dx$$

$$= c\int_0^2 2x - x^2\,dx = c\left(\int_0^2 2x\,dx - \int_0^2 x^2\,dx\right)$$

$$= c\left(\left[x^2\right]_0^2 - \left[\frac{x^3}{3}\right]_0^2\right) = c\left((2^2 - 0^2) - \left(\frac{2^3}{3} - \frac{0^3}{3}\right)\right)$$

$$= c\left(4 - \frac{8}{3}\right) = c\left(\frac{12}{3} - \frac{8}{3}\right) = c\,\frac{4}{3} = 1 \Rightarrow c = \frac{3}{4}$$

2.4 Expectation

Throwing a die: let X be the value of the die. On "average" we see

$$1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{7}{2}$$

Tossing a coin n times: Expected # of heads:

$$1 \cdot \frac{1}{2} + 0 \cdot \frac{1}{2} = \frac{1}{2} \quad \text{if we toss once}$$

$n = 2$

$$0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} = 0 + \frac{1}{2} + \frac{1}{2} = 1$$

In general: $\quad \dfrac{n}{2} = n \cdot \dfrac{1}{2}$

Tossing the coin until first head:

$$1 \cdot \frac{1}{2} + 2 \cdot \left(\frac{1}{2}\right)^2 + 3 \cdot \left(\frac{1}{2}\right)^3 + \dots$$

head 1st time

T, H

T, T, H

$\sum \frac{1}{n}$

not converging..

$$= \sum_{k=1}^{\infty} k \cdot \left(\frac{1}{2}\right)^k =$$

Geometric series

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x} \qquad |x|<1$$

$$\sum_{k=0}^{n} x^k = \frac{1-x^{n+1}}{1-x}$$

We want to calculate:

$$\sum_{k=1}^{\infty} k \cdot \left(\frac{1}{2}\right)^k$$

Reminder: Geometric series

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}, \quad |x|<1$$

It turns out that a more general sum is easier to calculate:

$$\sum_{k=1}^{\infty} k \cdot x^k, \quad |x|<1$$

Note that

$$k \cdot x^{k-1} = \frac{d}{dx} x^k$$

Consequently,

$$k x^k = x \cdot k \cdot x^{k-1} = x \cdot \frac{d}{dx} x^k$$

Therefore, we can rewrite our sum as follows:

$$\sum_{k=1}^{\infty} k \cdot x^k = \sum_{k=1}^{\infty} x \cdot k \cdot x^{k-1} = x \cdot \sum_{k=1}^{\infty} k \cdot x^{k-1}$$

$$= x \cdot \sum_{k=1}^{\infty} \frac{d}{dx} x^k \overset{f'+g'=(f+g)'}{=} x \frac{d}{dx} \sum_{k=1}^{\infty} x^k \overset{\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}}{=} x \cdot \frac{d}{dx} \frac{1}{1-x}$$

$$= x \cdot \frac{d}{dx} (1-x)^{-1} \overset{\frac{d}{dx} x^{\alpha} = \alpha x^{\alpha-1}}{=} x \cdot (-1)(1-x)^{-2} = \frac{x}{(1-x)^2}$$

Now, we can plug in $\frac{1}{2}$ for x:

$$\sum_{k=1}^{\infty} k \left(\frac{1}{2}\right)^k = \frac{\frac{1}{2}}{\left(1-\frac{1}{2}\right)^2} = \frac{\frac{1}{2}}{\left(\frac{1}{2}\right)^2} = \frac{1}{\frac{1}{2}} = 2$$

So, the expected number of coin tosses until the first head is 2.

Definition Let X be a **discrete** R.V., with values $x_1, \dots, x_k, \dots$

Then $$E[X] := \sum_{i=1}^{n} x_i \, P[X = x_i], \quad \text{if } X \text{ has } n \text{ values}$$

$$E[X] := \sum_{i=1}^{\infty} x_i \, P[X = x_i], \quad \text{if } X \text{ has } \infty \text{ many values,}$$

$E[X]$ is the expected value of X.

Definition Let X be a continuous R.V. with density f.

Then $$E[X] := \int_{-\infty}^{\infty} x \, f(x) \, dx$$

is the expected value of X

(if the integral exists)

$f(x) \cdot \Delta$
prob. of
$P[x \le X \le x + \Delta]$

Waiting time for alarm :

$\mathcal{X}$ waiting time RV, takes value $[0,2]$

Density $f(x) = \begin{cases} 1/2 & x \in [0,2] \\ 0 & \text{otherwise} \end{cases}$

$$E[\mathcal{X}] = \int_{-\infty}^{\infty} x f(x)\, dx = \int_0^2 x \frac{1}{2}\, dx$$

$$= \left[\frac{1}{\cancel{2}}\frac{x^2}{2}\right]_0^2 = \left[\frac{x^2}{4}\right]_0^2 = \frac{2^2}{4} - \frac{0^2}{4} = \frac{4}{4} = 1$$

uniform probability

$c \cdot 2 = 1$

2.2 Joint Distributions

Consider two RUs X, Y together. Study probabilities

$$P[X = x, Y = y]$$

or

$$P[a < X \leq b, c < Y \leq d]$$

Example 29: 9 batteries, 2 new, 3 part. charged, 4 empty.

Randomly select 3 out 9 batteries.

X # new batteries $\qquad X \in \{0,1,2\}$

Y # partially charged $\qquad Y \in \{0,1,2,3\}$

let $p(x,y) = P[X=x, Y=y]$ — joint pmf of X and Y

$$p(0,0) = \frac{\binom{4}{3}}{\binom{9}{3}} = \frac{4}{84} \qquad p(0,1) = \frac{\binom{3}{1}\binom{4}{2}}{\binom{9}{3}} = \frac{18}{84}$$

$$p(0,2) = \frac{\binom{3}{2}\binom{4}{1}}{\binom{9}{3}} = \frac{12}{84} \qquad p(0,3) = \frac{\binom{3}{3}}{\binom{9}{3}} = \frac{1}{84}$$

$$p(1,0) = \frac{\binom{2}{1}\binom{4}{2}}{\binom{9}{3}} = \frac{12}{84}$$

$$\binom{9}{3} = \frac{\cancel{9}^{3}\cdot\cancel{8}^{4}\cdot 7}{\cancel{3}\cdot\cancel{2}\cdot 1} = 12\cdot 7 = 84$$

$$p(1,1) = \frac{\binom{2}{1}\binom{3}{1}\binom{4}{1}}{\binom{9}{3}} = \frac{24}{84}$$

$$p(1,2) = \frac{\binom{2}{1}\binom{3}{2}}{\binom{9}{3}} = \frac{6}{84}$$

$$p(2,0) = \frac{\binom{2}{2}\binom{4}{1}}{\binom{9}{3}} = \frac{4}{84}$$

$$p(2,1) = \frac{\binom{2}{2}\binom{3}{1}}{\binom{9}{3}} = \frac{3}{84}$$

We have computed the joint pmf of X and Y.
Summarize in table (by multiples of $\frac{1}{84}$)

$X \backslash Y$	0	1	2	3	Sum
0	4	18	12	1	35
1	12	24	6	0	42
2	4	3	0	0	7
Sum	20	45	18	1	84

joint pmf

probability of $X = i$, i.e. $P[X = i]$

marginal probabilities

prob. of $Y = j$, i.e.
$P[Y = j]$

Joint cumulative probability distrib.

$$F(x,y) = P[X \leq x, Y \leq y]$$

Distribution of X

$$F_X(x) = P[X \leq x] = P[X \leq x, Y < \infty]$$
$$= F(x, \infty)$$

How we get the marginal pmf of $\mathcal{X}$ out of the joint pmf $p(x,y) = P[\mathcal{X}=x, \mathcal{Y}=y]$?

$$p_{\mathcal{X}}(x) = P[\mathcal{X}=x] = \sum_{j=1}^{n} P[\mathcal{X}=x, \mathcal{Y}=y_j]$$

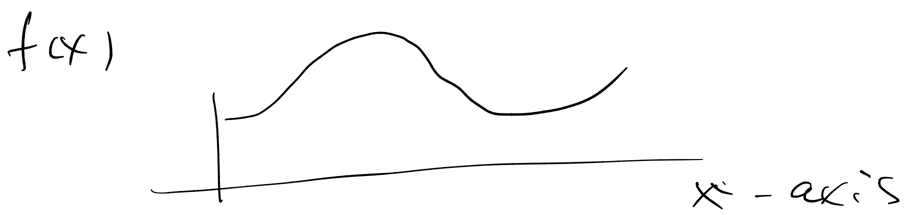
$$= \sum_{j=1}^{n} p(x, y_j)$$

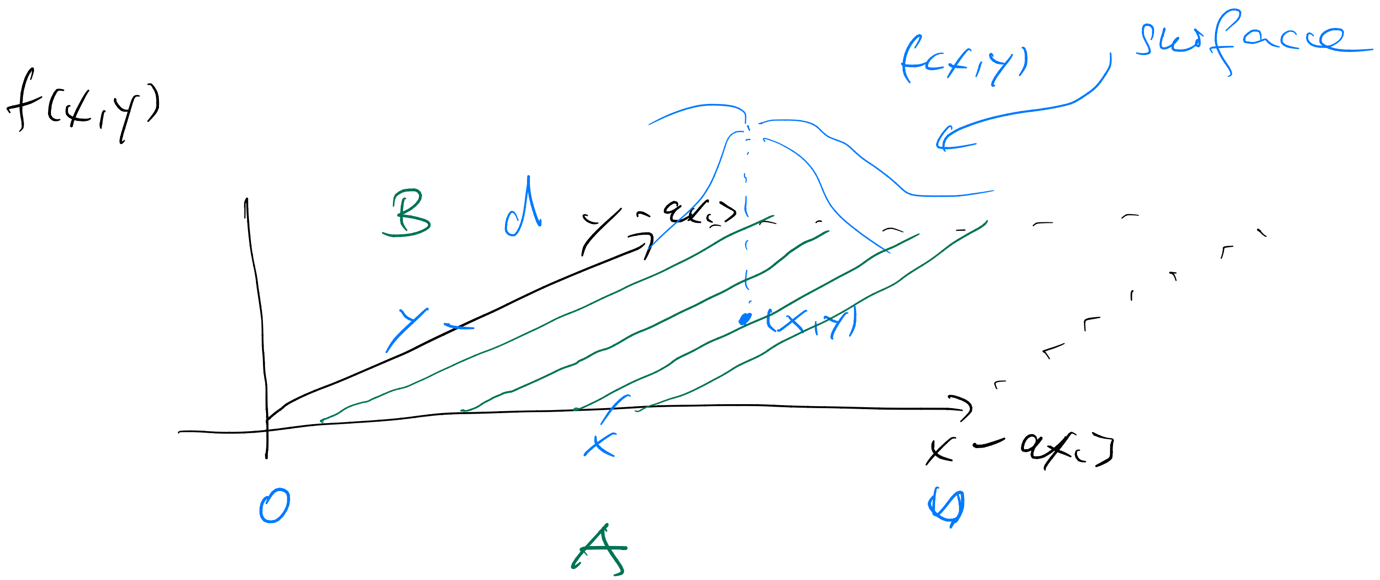
$$p_{\mathcal{Y}}(y) = P[\mathcal{Y}=y] = \sum_{i=1}^{m} p(x_i, y)$$

How can we model joint probabilities in cont. case?

discrete case: joint pmf $p(x,y)$

continuous case: joint pdf $f(x,y)$
(p. density f.)



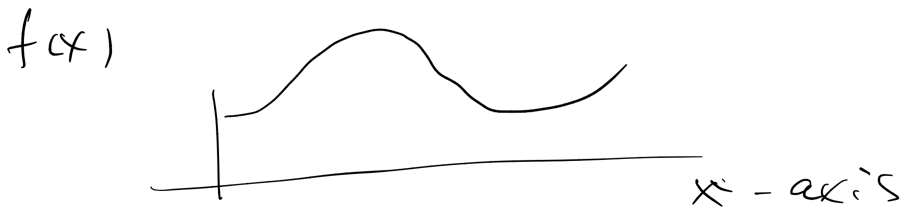


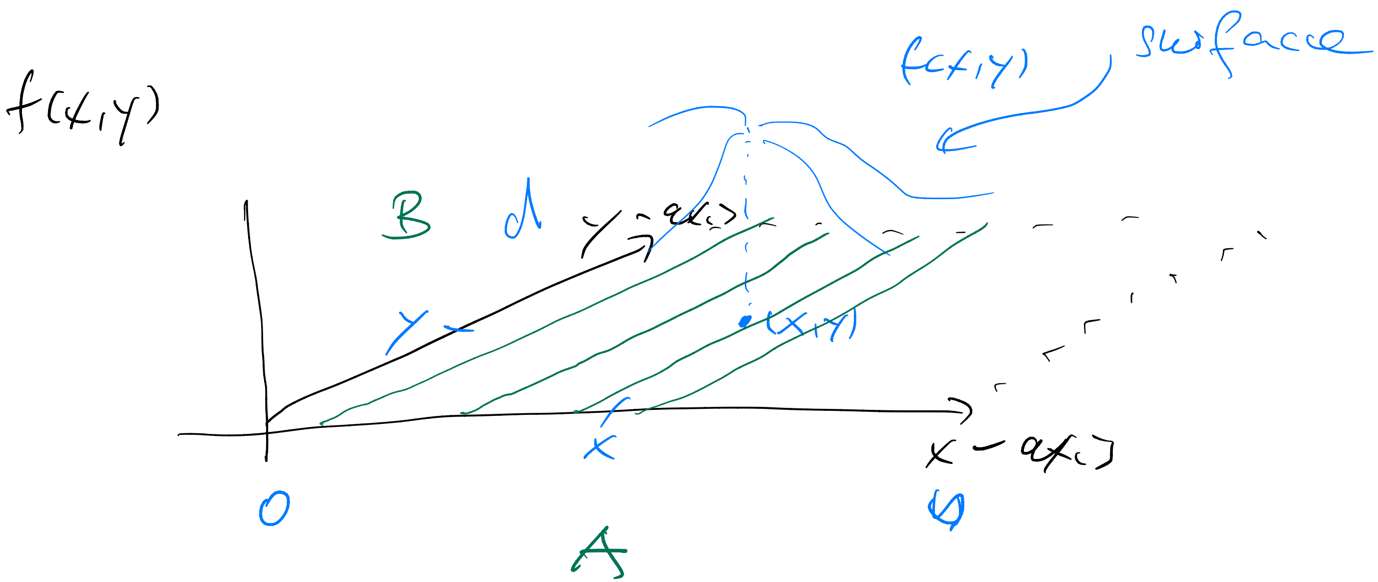
Integrals = areas under surface

How can we model joint probabilities in the continuous case?

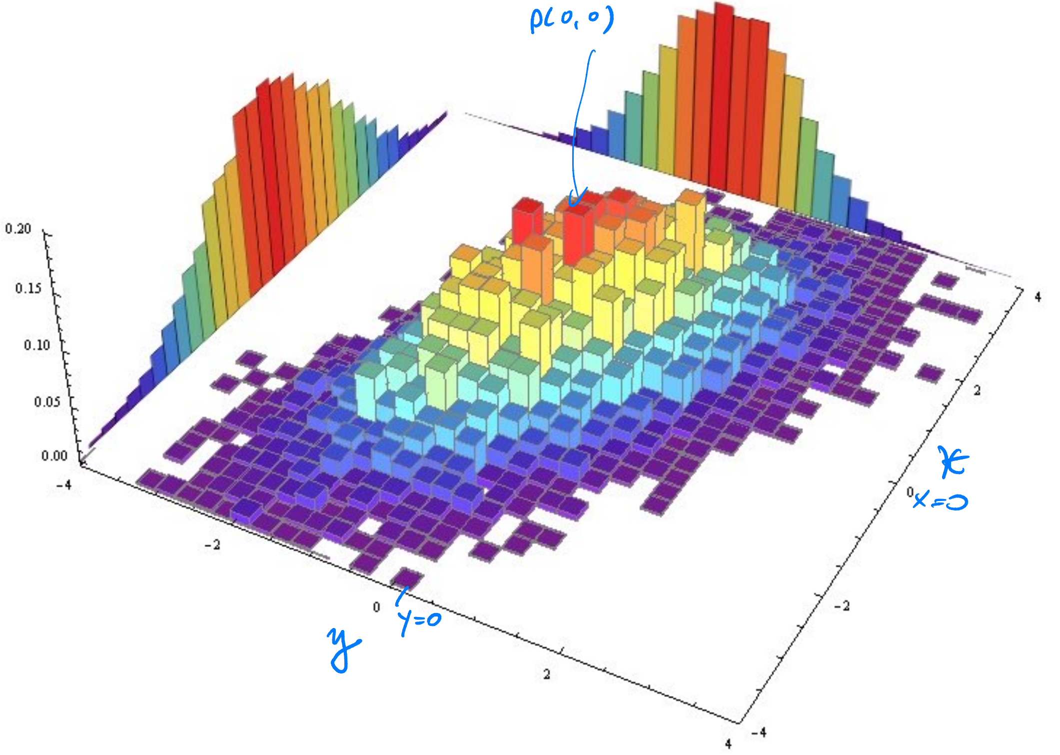
Discrete case: joint pmf $p(x,y)$ "discrete"

Continuous case: joint pdf $f(x,y)$





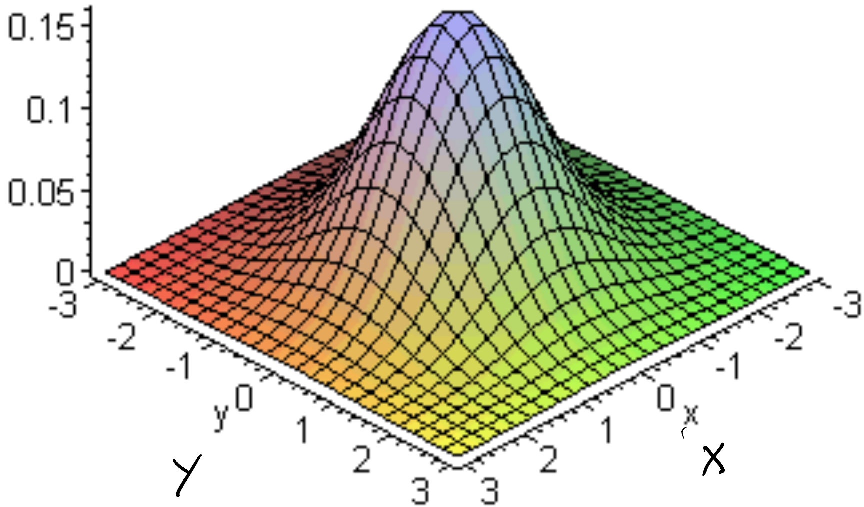
Integrals = volume under surface

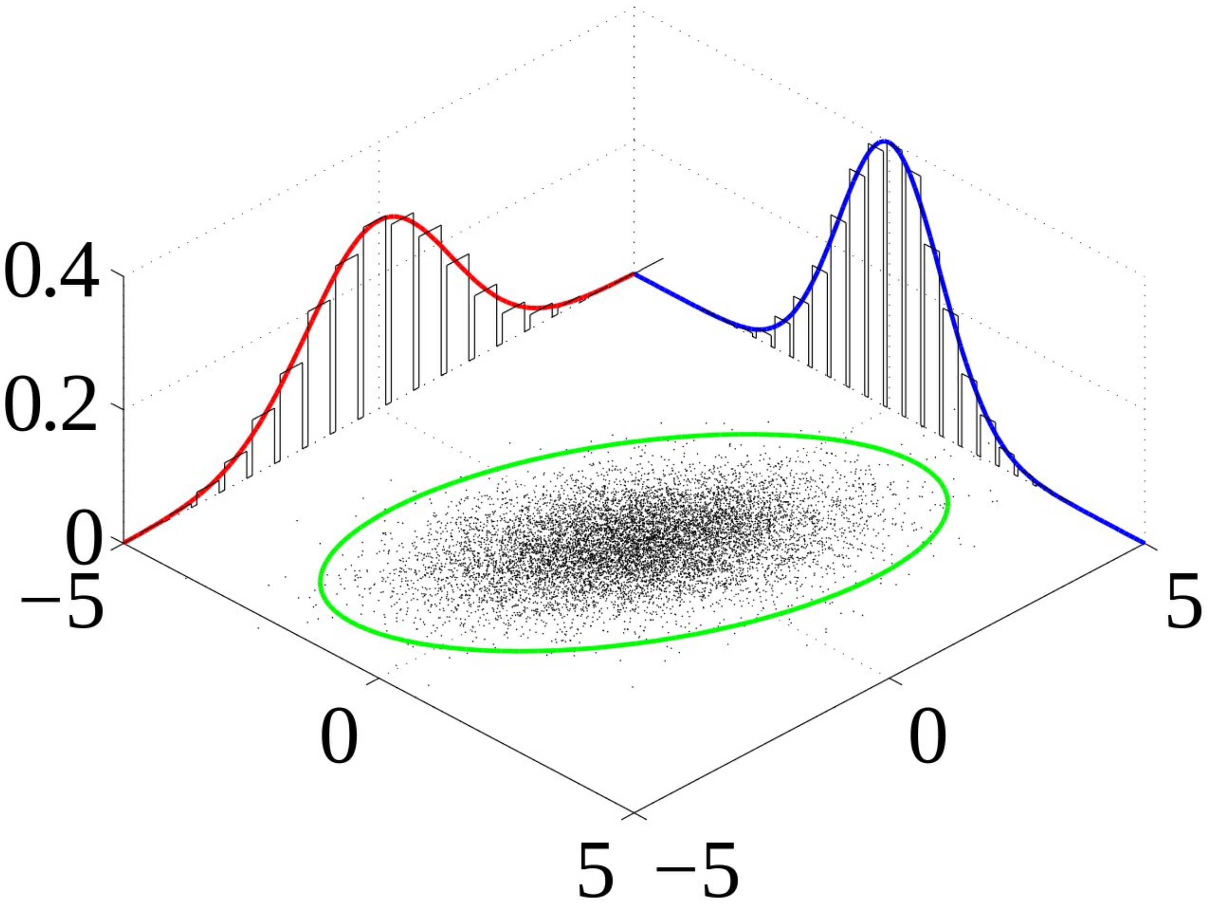


ρ(0, 0)
0.20
0.15
0.10
0.05
0.00
-4
-2
0
2
4
y
y=0
x
x=0

Bivariate Normal

Cavalieri's Principle





0.4
0.2
0
−5
0
5
−5
0
5

Let X, Y continuous, joint pdf is a fct

$$f : \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{R},$$

For $C \subseteq \mathbb{R} \times \mathbb{R}$, a reasonable subset, we have

$$P[(X, Y) \in C] = \iint_{(x,y) \in C} f(x,y)\, dx\, dy$$

Requirements of f:

$$f(x,y) \geq 0, \quad \iint_{\mathbb{R} \times \mathbb{R}} f(x,y)\, dx\, dy = 1$$

If $A, B \subseteq \mathbb{R}$, then

$$P[X \in A, Y \in B] = \int_A \left(\int_B f(x,y)\, dy \right) dx$$

$$= \int_B \left(\int_A f(x,y)\, dx \right) dy$$

Change of integration order is similar to change of summation order:

Consider

$$\begin{matrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{matrix}$$

Then

$$\sum_{i=1}^{2} \sum_{j=1}^{2} a_{ij} = (a_{11} + a_{12}) + (a_{21} + a_{22})$$

$$\sum_{j=1}^{2} \sum_{i=1}^{2} a_{ij} = (a_{11} + a_{21}) + (a_{12} + a_{22})$$

The two sums are identical due to associativity and commutativity of addition.

Similarly, we have $\int_A \int_B f(x,y)\, dx\, dy = \int_B \int_A f(x,y)\, dy\, dx$

Example 32: Let the joint pdf of X, Y be

$$f(x,y) = \begin{cases} 2e^{-x}e^{-2y} & x,y > 0 \\ 0 & \text{otherwise} \end{cases}$$

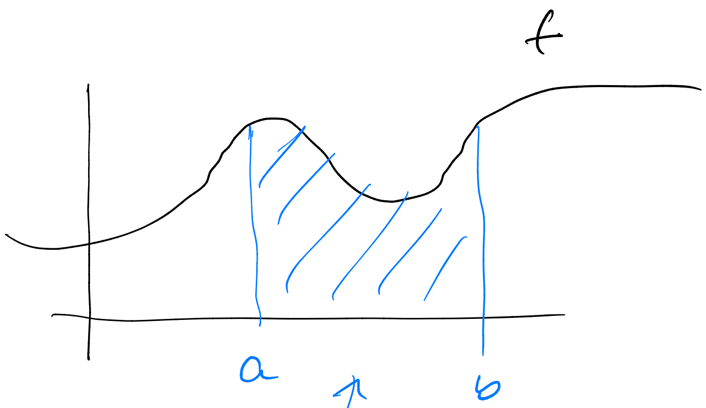
$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = \int_0^{\infty}\int_0^{\infty} 2e^{-x}e^{-2y}\,dx\,dy$$

$$= \int_0^{\infty} 2e^{-2y}\left(\int_0^{\infty} e^{-x}\,dx\right)dy = \int_0^{\infty} 2e^{-2y}\left[-e^{-x}\right]_0^{\infty} dy$$

$$\int_0^{\infty} 2e^{-2y}(0-(-1))\,dy = \int_0^{\infty} 2e^{-2y}\,dy = \left[-e^{-2y}\right]_0^{\infty}$$

$$= 0 - (-e^{-2\cdot 0}) = 0 - (-1) = 1$$

Student Question: "When evaluating an integral of f with an antiderivative F, why do we plug the upper bound first into F?"



F is an antiderivative of f if

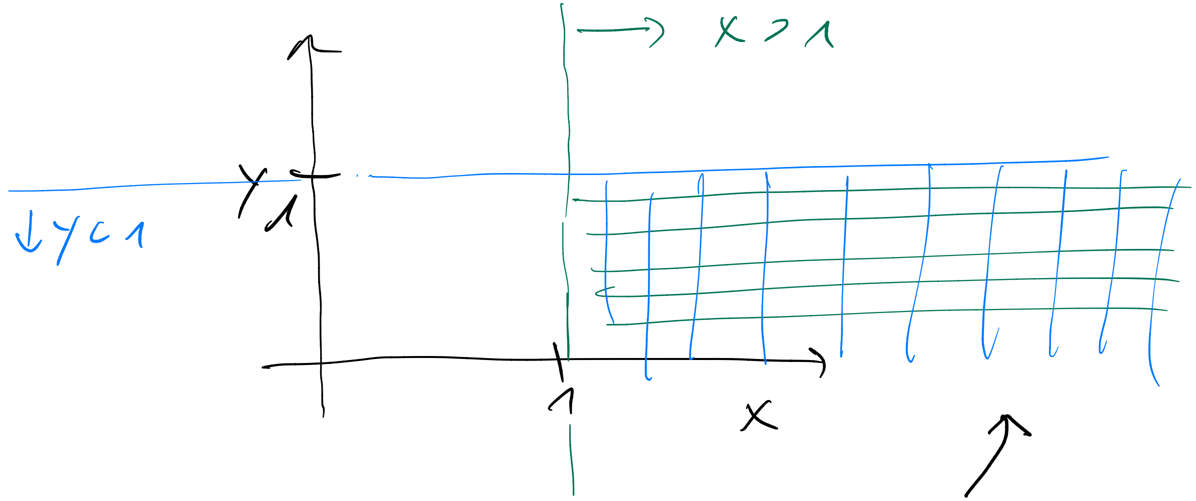
$F' = f$

$$\int_a^b f(x)\,dx = F(b) - F(a) = \Big[F(x) \Big]_a^b$$

$$\int_a^\infty f(x)\,dx = \lim_{b\to\infty} \int_a^b f(x)\,dx =$$

$$\lim_{b\to\infty} \big(F(b) - F(a)\big) = \left(\lim_{b\to\infty} F(b)\right) - F(a)$$

$$P[X>1,\ Y<1]$$



$$= \int_1^\infty \int_0^1 f(x,y)\, dy\, dx$$

$$= \int_1^\infty \int_0^1 e^{-x} 2e^{-2y}\, dy\, dx$$

$$= \int_1^\infty e^{-x} \int_0^1 2e^{-2y}\, dy\, dx = \int_1^\infty e^{-x} \left[-e^{-2y}\right]_0^1 dx$$

$$= \int_1^\infty e^{-x} \left(-e^{-2} - (-e^0)\right) dx = \int_1^\infty e^{-x}(1-e^{-2})\, dx$$

$$= (1-e^{-2}) \int_1^\infty e^{-x}\, dx = (1-e^{-2}) \left[-e^{-x}\right]_1^\infty$$

$$= (1-e^{-2})\, e^{-1} = e^{-1} - e^{-3}$$

$$\int_1^\infty \int_0^1 e^{-x}\, 2e^{-2y}\, dy\, dx$$

Special case of:

This is a constant that can be pulled out of the integral.

$$\int_A \boxed{\int_B g(x)\cdot h(y)\, dy}\, dx = \int_A g(x) \boxed{\int_B h(y)\, dy}\, dx$$

$$= \int_B h(y)\, dy \cdot \int_A g(x)\, dx = \boxed{\int_A g(x)\, dx \cdot \int_B h(y)\, dy}$$

If

1) $f(x) = g(x)\, h(x)$

2) Integration area has form $A \times B$

Then

$$\iint_{A\times B} f\cdot g = \int_A f \cdot \int_B g$$

$$\int_1^\infty \int_0^1 e^{-x} 2e^{-2y} \, dy \, dx$$

$$= \int_1^\infty e^{-x} \, dx \cdot \int_0^1 2e^{-y} \, dy = \Big[-e^{-x} \Big]_1^\infty \cdot \Big[-e^{-2y} \Big]_0^1$$

$$= e^{-1} \left(1 - e^{-2} \right)$$

$P[X < a]$ $\qquad a > 0$

$$= \int_0^a \int_0^\infty e^{-x} \cdot 2e^{-2y} \, dy \, dx$$

$$= \int_0^a e^{-x} \, dx \cdot \int_0^\infty 2e^{-2y} \, dy$$

Density of Exp(2)

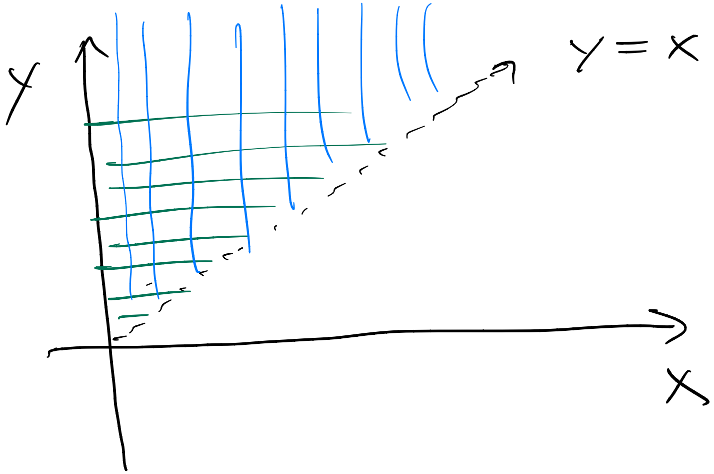
$$= \left[-e^{-x}\right]_0^a \cdot \left[-e^{-2y}\right]_0^\infty$$

$$= \left(e^0 - e^{-a}\right) \cdot \left(e^0 - 0\right)$$

$$= \left(1 - e^{-a}\right) \cdot 1$$

$$P[X < Y]$$

$$= \int_0^\infty \int_x^\infty f(x,y)\, dy\, dx \quad (*)$$



$$= \int_0^\infty \int_0^y f(x,y)\, dx\, dy$$

$$\overset{(*)}{=} \int_0^\infty \int_x^\infty e^{-x}\, 2 e^{-2y}\, dy\, dx = \int_0^\infty e^{-x} \left[-e^{-2y} \right]_x^\infty dx$$

$$= \int_0^\infty e^{-x} e^{-2x}\, dx = \int_0^\infty e^{-3x}\, dx = \left[-\frac{1}{3} e^{-3x} \right]_0^\infty$$

$$= \frac{1}{3} e^{-3 \cdot 0} = \frac{1}{3}$$

2.3 Independent Random Variables

$\mathcal{E}, \mathcal{F}$ ind. $\iff$ $P[\mathcal{E}\mathcal{F}] = P(\mathcal{E}) \cdot P(\mathcal{F})$

$\iff$ $P(\mathcal{E} \mid \mathcal{F}) = P(\mathcal{E})$

($\mathcal{E}$ = "$X \in A$",

$\mathcal{F}$ = "$Y \in B$")

X, Y are independent iff

$$P[X \in A, Y \in B] = P[X \in A] \cdot P[Y \in B]$$

for all $A, B \subseteq \mathbb{R}$

Equivalent: $P[X \le a, Y \le b] = P[X \le a] \cdot P[Y \le b]$,

f.a. $a, b \in \mathbb{R}$

that is

$$F(a, b) = F_X(a) \cdot F_Y(b)$$

Equivalent for discrete RVs:

$$p(x,y) = p_X(x) \cdot p_Y(y) \qquad \text{f.a. } x,y \in \mathbb{R}$$

For cont. RVs

$$f(x,y) = f_X(x) \cdot f_Y(y) .$$

Example 33: Let X, Y be independent, each with

density

$$f(x) = \begin{cases} e^{-x}, & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

What can we say about the quotient of two independent and exponentially distributed RVs?

What is the density of $\frac{X}{Y}$?

Two steps:

1) cdf of $\frac{X}{Y}$

2) pdf is derivative of cdf

1) $$F(a) = P\left[\frac{X}{Y} \le a\right] = P[X \le aY]$$

$$= \int_0^\infty \int_0^{aY} e^{-x} e^{-Y}\, dx\, dy$$

$$= \int_0^\infty e^{-Y} \left[-e^{-x}\right]_0^{aY} dy = \int_0^\infty e^{-Y}(1 - e^{-aY})\, dy$$

$$= \int_0^\infty e^{-Y} dy - \int_0^\infty e^{-(1+a)Y}\, dy$$

$$= 1 - \left[-\frac{1}{1+a} e^{-(1+a)Y}\right]_0^\infty = 1 + [\dots]_0^\infty$$

$$= 1 + \left(-\frac{1}{1+a}\right) = 1 - \frac{1}{1+a}$$ cdf

2) $$f(a) = \frac{d}{da} F(a) = -\frac{d}{da}(1+a)^{-1} = -(-1)(1+a)^{-2}$$

$$= \frac{1}{(1+a)^2}$$ pdf

Remark: Generalization to n RVs $X_1, \dots, X_n$ is possible:

- joint pmf $\quad p(x_1, \dots, x_n)$
- joint pdf $\quad f(x_1, \dots, x_n)$
- independence

$$p(x_1, \dots, x_n) = p_{X_1}(x_1) \cdot \dots \cdot p_{X_n}(x_n)$$

$$f(x_1, \dots, x_n) = f_{X_1}(x_1) \cdot \dots \cdot f_{X_n}(x_n)$$

Joint distribution of RVs X, Y: Revision

Individual RVs can be described by

pmfs (discrete)

pdfs (continuous)

Joint pmfs: $p(x_i, y_j)$ if x_i, y_j are the possible values of X, Y

Joint pdfs: $f(x, y)$

$$\sum_{i,j} p(x_i, y_j) = 1$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\, dy\, dx = 1$$

$$= \sum_i \sum_j p(x_i, y_j) = \sum_j \sum_i p(x_i, y_j)$$

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

Marginal pmfs, pdfs

$$p_X(x_i) = \sum_j p(x_i, y_j) \qquad p_Y(y_j) = \sum_i p(x_i, y_j)$$

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

Independence of events

$\mathcal{E}, \mathcal{F}$ indep. if

- "inf about $\mathcal{E}$ does not provide information about $\mathcal{F}$"

- $P(\mathcal{F} \mid \mathcal{E}) = P(\mathcal{F}) \quad (\Leftrightarrow P(\mathcal{F} \mid \mathcal{E}) = P(\mathcal{E})$

$\Leftrightarrow \quad P(\mathcal{E}\mathcal{F}) = P(\mathcal{E})\, P(\mathcal{F})$

Independence of RVs

X, Y are indep. if

all events that can be described in terms of X

"$5 < X < 9$", "$X > 2$", "$X < 0$ or $X > 2$"

are independent of events that can be described in terms of Y.

"$Y \geq 4$"

<u>Proposition</u> X, Y are independent iff

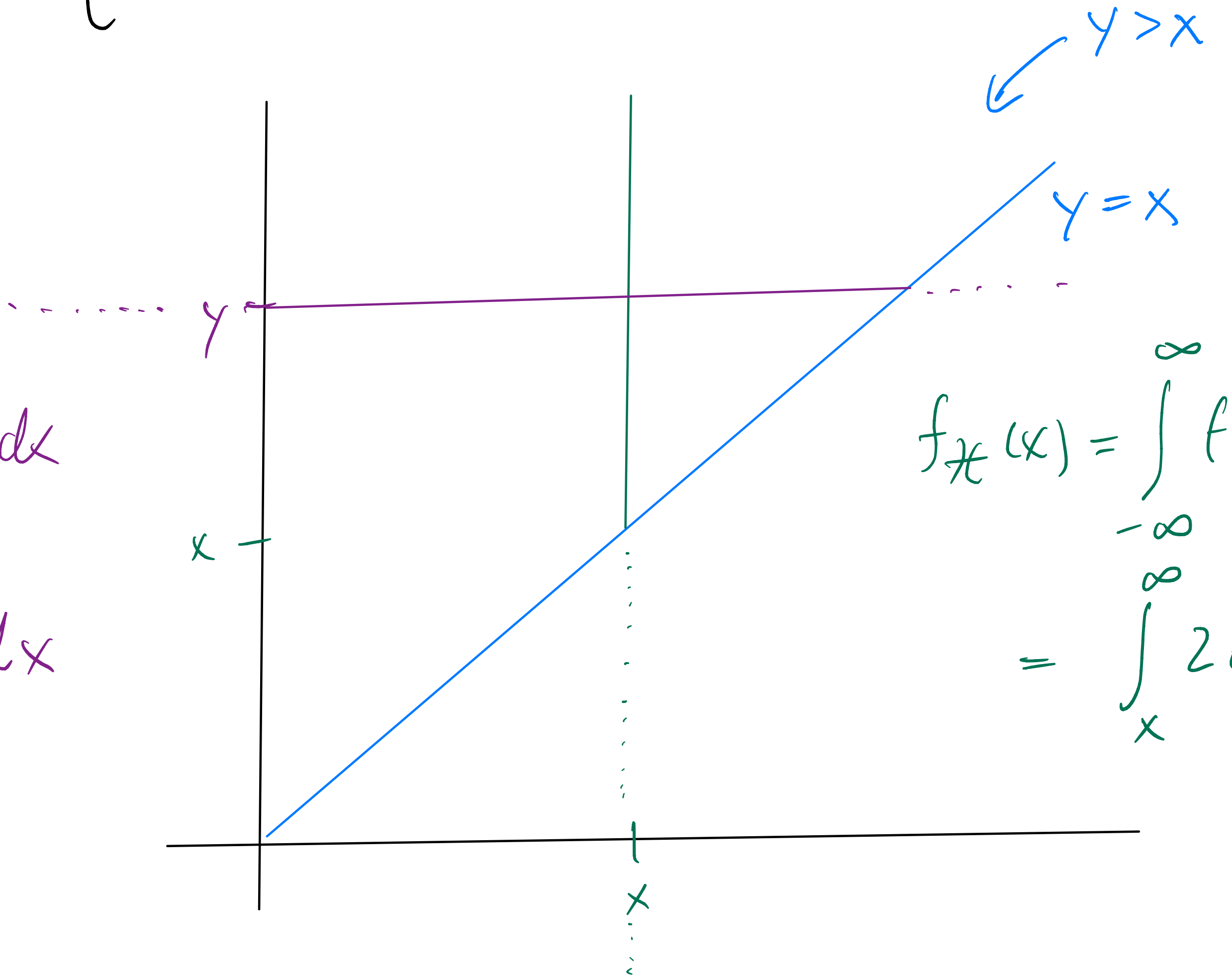
- $p(x_i, y_j) = p_X(x_i) \cdot p_Y(y_j)$, f.a. x_i, y_j

- $f(x,y) = f_X(x) \cdot f_Y(y)$, f.a. x, y

Marginal Densities and Independence

Suppose X, Y have the joint distribution

$$f(x,y) = \begin{cases} 2e^{-x}e^{-y} & 0 < x < y \\ 0 & \text{otherwise} \end{cases}$$



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

$$= \int_0^y 2e^{-x}e^{-y}\,dx$$

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

$$= \int_x^{\infty} 2e^{-x}e^{-y}\,dy$$

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \int_x^{\infty} 2e^{-x}e^{-y}\,dy$$

$$= 2e^{-x} \int_x^{\infty} e^{-y}\,dy = 2e^{-x}\left[-e^{-y}\right]_x^{\infty}$$

$$= 2e^{-x}e^{-x} = 2e^{-2x}$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \int_0^{y} 2e^{-x}e^{-y}\,dx$$

$$= 2e^{-y} \int_0^{y} e^{-x}\,dx = 2e^{-y}\left[-e^{-x}\right]_0^{y}$$

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

Expected Values of RVs

Die: X has values $1, \dots, 6$, all with $p(x_i) = \frac{1}{6}$

$$E[X] = 1 \cdot p(1) + 2 \cdot p(2) + \dots + 6 \cdot p(6)$$

$$= (1 + 2 + \dots + 6) \frac{1}{6} = \frac{35}{6}$$

Discrete RV X

$$E[X] = \sum_i x_i \cdot p(x_i)$$

Continuous RV

$$E[X] = \int_{-\infty}^{\infty} x \cdot f(x)\, dx$$

2.5 Properties of Expectation

X RV, $g(x)$ function $\Rightarrow$ $g(X)$ is a RV

X points of die, $g(x) = x^2$ $\Rightarrow$ $g(X) = X^2$, squares of points

X^2 is a new RV

values	probabilities
1	1/6
4	1/6
9	1/6
16	1/6
25	1/6
36	1/6

$$E[X^2] = 1 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + \dots \dots 36 \cdot \frac{1}{6}$$

$$= \frac{1+4+9+16+25+36}{6}$$

$$= \frac{91}{6}$$

Imagine: a die with numbers $-3, -2, -1, 1, 2, 3$

X is the number on top of the die: $\quad E[X] = 0$

Let $Z := X^2$

1) Find the pmf of Z and compute $E[Z]$

$$E[Z] = 1 \cdot \frac{1}{3} + 4 \cdot \frac{1}{3} + 9 \cdot \frac{1}{3}$$

pmf of Z

values	probabilities
9	$\frac{1}{6} + \frac{1}{6} = \frac{1}{3}$
4	—— , ——
1	—— , ——

2) Take a weighted average of the $g(x_i)$

$$E[X^2] = (-3)^2 \cdot \frac{1}{6} + (-2)^2 \cdot \frac{1}{6} + (-1)^2 \cdot \frac{1}{6}$$

$$3^2 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{6} + 1^2 \cdot \frac{1}{6}$$

$$= \sum_i x_i^2 \cdot p(x_i)$$

We found:

$$\boxed{E[g(X)] = \sum_i g(x_i) \cdot p(x_i)}$$

Proposition 39 ("Law of the Unconscious Statistician" LOTUS)

X RV, $g: \mathbb{R} \to \mathbb{R}$

- $$E[g(X)] = \sum_i g(x_i)\, p(x_i)$$
- $$E[g(X)] = \int_{-\infty}^{\infty} g(x) \cdot f(x)\, dx$$

Applications of LOTUS

X RV, $a, b \in \mathbb{R}$, $g(x) = ax + b$

$$E[aX + b] \stackrel{?}{=} aE[X] + b$$

$$E[g(X)] = \int_{-\infty}^{\infty} (ax+b) f(x)\, dx = \int_{-\infty}^{\infty} ax f(x) + b f(x)\, dx$$

$$= a \int_{-\infty}^{\infty} x f(x)\, dx + \int_{-\infty}^{\infty} b \cdot 1 f(x)\, dx$$

$$= a E[X] + b \cdot \int_{-\infty}^{\infty} 1 f(x)\, dx$$

$$= a E[X] + b \cdot 1$$

Let X, Y be RVs. What about $E[X+Y]$?

$$E[X+Y] \stackrel{?}{=} E[X] + E[Y]$$

(Do we need independence?)

LOTUS also holds for 2-dim. densities:

Let X, Y be RVs with joint pdf $f(x,y)$, let $g: \mathbb{R} \times \mathbb{R} \to \mathbb{R}$

Then

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

The summation law follows with $g(x,y) = x+y$

$E[g(X,Y)]$ $g(x,y)$

lotus

$$E[X+Y] = \int_{\mathbb{R}}\int_{\mathbb{R}} (x+y)\, f(x,y)\, dx\, dy$$

$$= \int_{\mathbb{R}}\int_{\mathbb{R}} x \cdot f(x,y)\, dx\, dy + \int_{\mathbb{R}}\int_{\mathbb{R}} y \cdot f(x,y)\, dx\, dy$$

$$= \int_{\mathbb{R}} x \underbrace{\left(\int_{\mathbb{R}} f(x,y)\, dy\right)}_{f_X(x)} dx + \int_{\mathbb{R}} y \underbrace{\left(\int_{\mathbb{R}} f(x,y)\, dx\right)}_{f_Y(y)} dy$$

$$= \int_{\mathbb{R}} x \cdot f_X(x)\, dx + \int_{\mathbb{R}} y \cdot f_Y(y)\, dy$$

$$= E[X] + E[Y]$$

Generalization

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

Throwing 2 dice, adding result $X_1 + X_2$

$$E[X_1 + X_2] = E[X_1] + E[X_2] = \frac{7}{2} + \frac{7}{2} = 7$$

Example 42: Tossing coin n times, $E[\#\text{ heads}] = ?$

$p(H) = p$, $p(T) = 1-p$

Let $X_i = 1$ iff head with i-th toss $\Rightarrow E[X_i] = p$

n times: $E\left[\sum_{i=1}^{n} X_i\right] = n\, E[X_1] = np$

Note: $E[X] = 1 \cdot p + 0 \cdot (1-p) = p$!!

Last topic: Expected value of a RV

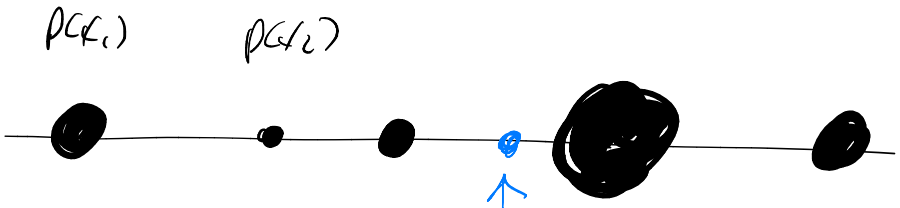
Idea: longterm average

X discr. $\quad \sum_i x_i \, p(x_i)$, x_i: poss. values of X

X cont. $\quad \int_{\mathbb{R}} x \cdot f(x) \, dx$, f pdf

Another interpretation: center of gravity

X discr.: a number of points $x_1, x_2, \ldots$,
with weight $p(x_i)$

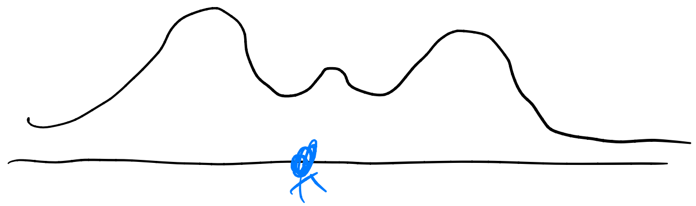


points on a
weightless rod

center of gravity
supports
the construction

X cont

weight proportional to
weight of curve



center of
gravity

Properties of Exp. V.:

$$E[aX] = a\,E[X]$$

$$E[b] = b$$

$$E[X+Y] = E[X] + E[Y]$$

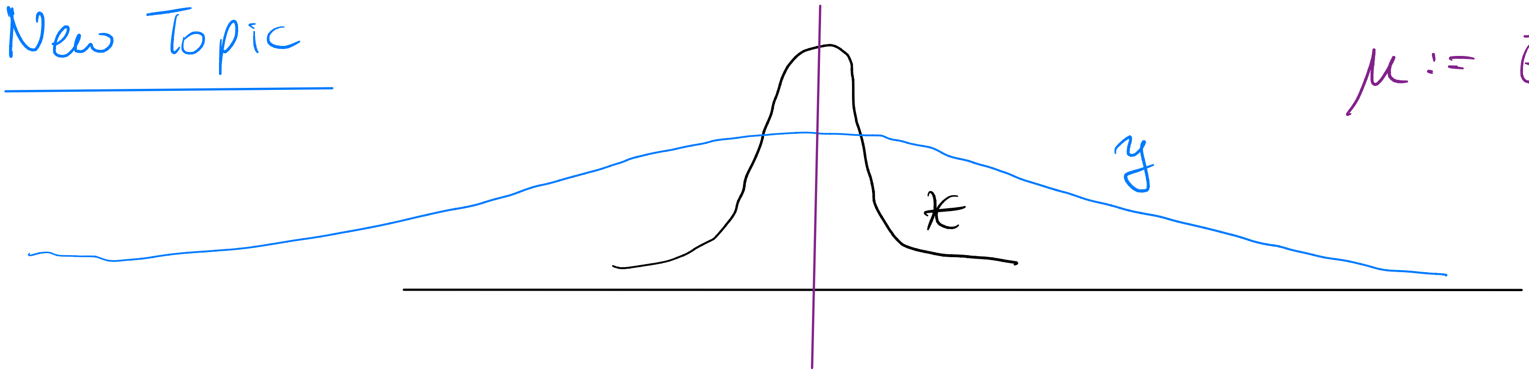
! holds also if X, Y are not independent

$$E[X_1 + \dots + X_n] = \sum_{i=1}^{n} E[X_i]$$

LOTUS

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

New Topic



$\mu := E[X] = E[Y]$

μ mean of X, Y

Values of Y are spread much farther around μ than the values of X.

2.6 Variance

$$\mathrm{Var}(X) := E\left[(X-\mu)^2\right]$$

$$g(x) = (x-\mu)^2$$

is the variance of X.

Why not $E\left[|X-\mu|\right]$?

The definition with the square has better mathematical properties.

How can we calculate $Var(X)$? Suppose $X \sim f$.

1.) Apply LOTUS:

$$Var(X) = \int_{\mathbb{R}} (x-\mu)^2 f(x)\,dx$$

$$= \int_{\mathbb{R}} (x^2 - 2x\mu + \mu^2) f(x)\,dx$$

$$= \int_{\mathbb{R}} x^2 f(x)\,dx - 2\mu \int_{\mathbb{R}} x f(x)\,dx + \mu^2 \int_{\mathbb{R}} f(x)\,dx$$

$\boxed{\text{Shortcut}}$

$$= E[X^2] - 2\mu E[X] + \mu^2 \cdot 1$$

$$= E[X^2] - 2\mu^2 + \mu^2$$

$$= E[X^2] - E[X]^2$$

square of mean

2nd moment of X

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

<u>Example 44</u> $\quad X$ = number on die

$E[X] = \frac{7}{2}$. Also $E[X^2] = \frac{91}{6}$ (Ex. 37)

$$\mathrm{Var}(X) = E[X^2] - E[X]^2 = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{182 - 3\cdot 7^2}{12}$$
$$= \frac{182-147}{12} = \frac{35}{12}$$

Units: unit of measurement of X is metre, sec

$\Rightarrow$ unit of measurement of X^2 is metre^2, sec^2

Get back to original unit: take $\sqrt{\cdot}$:

$$\sigma := \sqrt{Var(X)}$$

is the standard deviation.

One often writes the variance as the square of the standard deviation:

$$Var(X) = \sigma^2$$

$$x = (x_1, x_2, x_3)$$

$$y = (y_1, y_2, y_3)$$

$$dist(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2}$$

Properties of $Var(\cdot)$. Suppose X has $E[X] = \mu$.

- $Var(X + b) = Var(X)$
- $Var(aX) = a^2 Var(X)$
- $Var(X + Y) \overset{?}{=} Var(X) + Var(Y)$ only if X, Y independent

$$
\begin{aligned}
Var(\overbrace{aX + b}^{Y}) &= E[Y^2] - E[Y]^2 \\
&= E[(aX + b)^2] - E[aX + b]^2 \\
&= E[a^2 X^2 + 2abX + b^2] - (aE[X] + b)^2 \\
&= a^2 E[X^2] + 2ab E[X] + E[b^2] \\
&\quad - a^2\mu^2 - 2ab\mu - b^2
\end{aligned}
$$

$$= a^2 E[X^2] + 2ab E[X] + E[b^2] - a^2\mu^2 - 2ab\mu - b^2$$

$$= a^2 E[X^2] + \cancel{2ab}\mu + \cancel{b^2} - a^2\mu^2 - \cancel{2ab}\mu - \cancel{b^2}$$

$$= a^2\left(E[X^2] - \mu^2\right) = a^2 \operatorname{Var}(X)$$

Note: $\sigma_{aX} = a\,\sigma_X$

2.7 Covariance

We note

$$Var(X+X) = Var(2X) = 4\,Var(X)$$
$$\neq Var(X) + Var(X)$$

Definition 45: X, Y RV, $\mu_X = E[X]$, $\mu_Y = E[Y]$

Then

$$Cov(X,Y) = E\left[(X-\mu_X)(Y-\mu_Y)\right]$$

This assumes a joint distribution of X and Y.

Property of Cov

$$\mathrm{Cov}(X, Y) = E\left[(X-\mu_X)(Y-\mu_Y)\right]$$

$$\begin{aligned}
\mathrm{Cov}(X, Y) &= E\left[XY - \mu_X Y - \mu_Y X + \mu_X \mu_Y\right] \\
&= E[XY] - \mu_X E[Y] - \mu_Y E[X] + \mu_X \mu_Y \\
&= E[XY] - \mu_X \mu_Y - \cancel{\mu_Y \mu_X} + \cancel{\mu_X \mu_Y} \\
&= E[XY] - E[X]\,E[Y]
\end{aligned}$$

Observation: X, Y independent $\Rightarrow E[XY] = E[X] \cdot E[Y]$

$\Rightarrow \mathrm{Cov}(X, Y) = 0$

Observation: X, Y ind. $\Rightarrow E[XY] = E[X] \cdot E[Y]$

Suppose $X \sim f$, $Y \sim g$:

LOTUS

$$E[XY] = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} x \cdot y \; f(x) g(y) \, dx \right) dy$$

$$= \int_{\mathbb{R}} y \cdot g(y) \left(\int_{\mathbb{R}} x \cdot f(x) \, dx \right) dy$$

$$= \int_{\mathbb{R}} x \cdot f(x) \, dx \cdot \int_{\mathbb{R}} y \cdot g(y) \, dy$$

$$= E[X] \cdot E[Y]$$

Proposition 46 $\quad Cov(X_1+X_2, Y) = Cov(X_1, Y) + Cov(X_2, Y)$

$$Cov(X_1+X_2, Y) = E[(X_1+X_2)\cdot Y] - E[X_1+X_2]\cdot E[Y]$$

$$= E[X_1 Y] + E[X_2 Y] - E[X_1]\cdot E[Y] - E[X_2]\cdot E[Y]$$

$$= E[X_1 Y] - E[X_1]\cdot E[Y]$$
$$+ E[X_2 Y] - E[X_2]\cdot E[Y]$$

$$= Cov(X_1, Y) + Cov(X_2, Y)$$

Theorem 47

$$Cov\left(\sum_i X_i, \sum_j Y_j\right) = \sum_i \sum_j Cov(X_i, Y_j)$$

Consequence:

$$\mathrm{Var}\left(\sum_i X_i\right) = \mathrm{Cov}\left(\sum_i X_i, \sum_i X_i\right)$$

$$= \sum_i \sum_j \mathrm{Cov}(X_i, X_j)$$

$$= \sum_i \left(\sum_{j \neq i} \mathrm{Cov}(X_i, X_j) + \mathrm{Cov}(X_i, X_i) \right)$$

$$= \left(\sum_i \sum_{j \neq i} \mathrm{Cov}(X_i, X_j) \right) + \sum_i \mathrm{Cov}(X_i, X_i)$$

$$= \sum_i \mathrm{Var}(X_i) + \sum_i \sum_{j \neq i} \mathrm{Cov}(X_i, X_j)$$

For $n=2$:

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X, Y)$$

Meaning of covariance $Cov(X, Y)$

> 0: X, Y grow together in the same direction

< 0: X, Y grow in sync in opposite directions

≈ 0: X, Y vary independently

Normalize RVs X, Y by taking $\frac{X}{\sigma_X}$, $\frac{Y}{\sigma_Y}$,

i.e., $\frac{X}{\sqrt{Var(X)}}$ $\frac{Y}{\sqrt{Var(Y)}}$

$$Corr(X, Y) = \frac{Cov(X, Y)}{\sigma_X \sigma_Y} = \frac{Cov(X, Y)}{\sqrt{Var(X) \cdot Var(Y)}}$$

Correlation between X and Y

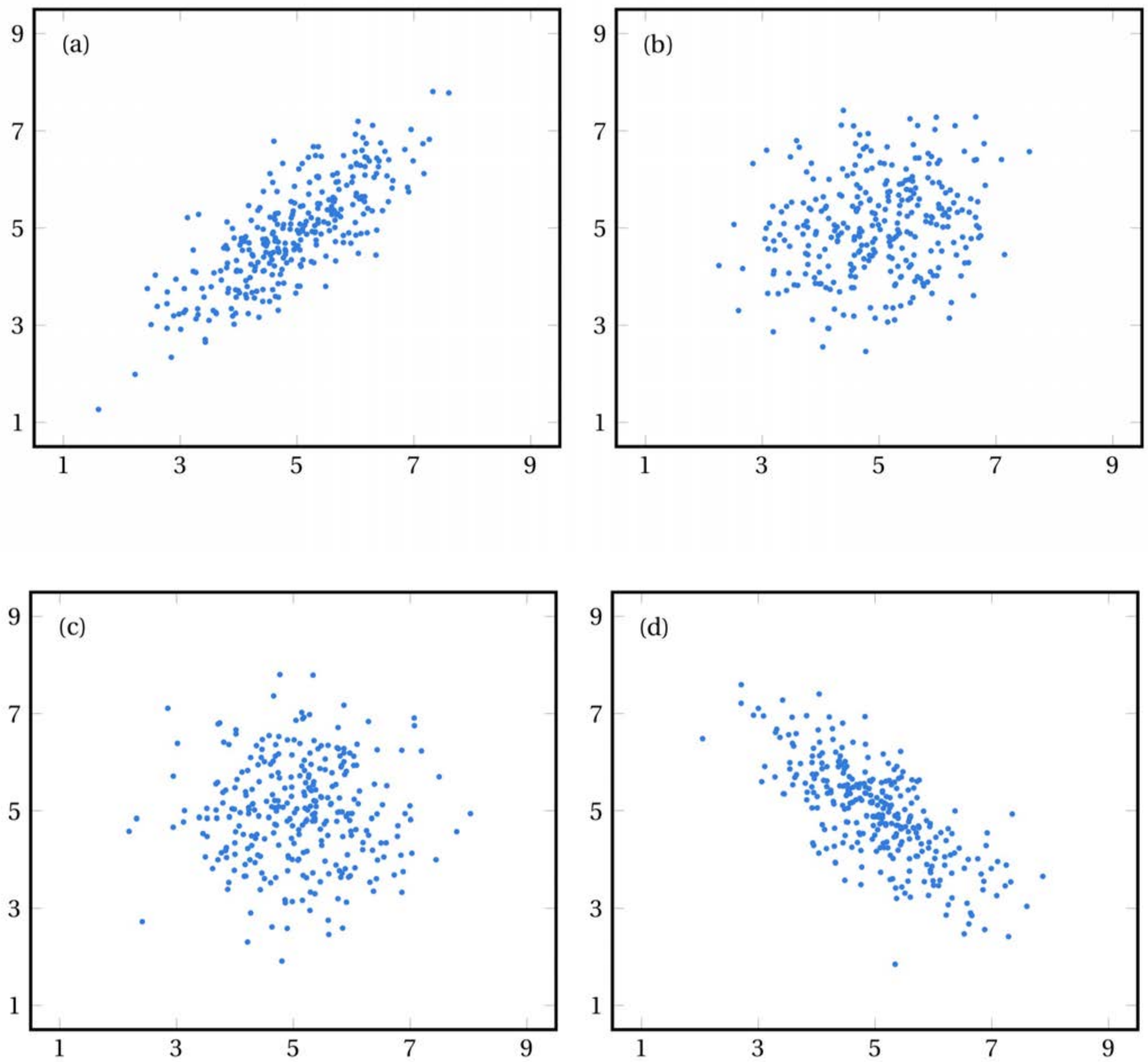


Figure 9: Random variables $\mathscr{X}$ and $\mathscr{Y}$ with correlations (a) 0.75; (b) 0.2; (c) 0; and (d) -0.75.

Example: 10 independent dice rolls: X_i is i-th roll

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) \underset{\text{ind}}{=} \sum_{i=1}^{n} \mathrm{Var}(X_i)$$

$$= \sum_{i=1}^{n} \frac{35}{12} = 10 \cdot \frac{35}{12}$$

What about the standard deviation?

Var has grown by factor 10,

σ grows by factor $\sqrt{10}$!

2.8 The Weak Law of Large Numbers

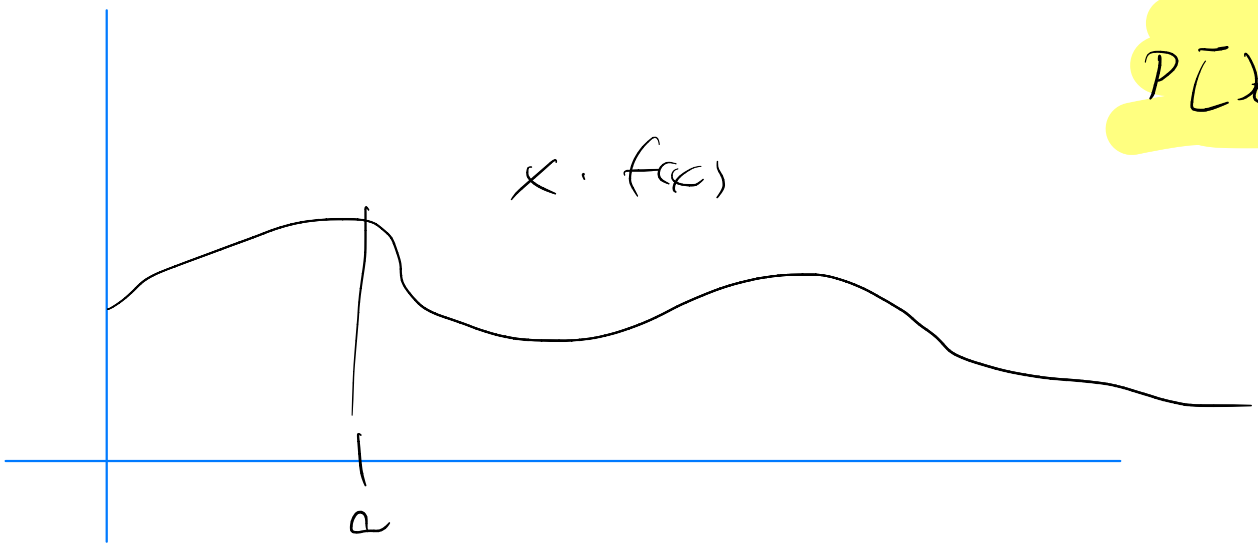
Markov's inequality

Let $X \geq 0$ s.th. $E[X]$ exists. Let also $a > 0$. Then

$$E[X] = \int_0^\infty x\, f(x)\, dx \geq \int_a^\infty x\, f(x)\, dx$$

$$\geq \int_a^\infty a\, f(x)\, dx = a \int_a^\infty f(x)\, dx = a\, P[X \geq a]$$

$$P[X \geq a] \leq \frac{E[X]}{a}$$



Apply Markov's Inequality:

$$Y := (X - \mu)^2 \quad , \quad a = k^2$$

Assume: $Var(X) = \sigma^2 \Rightarrow E[Y] = Var(X) = \sigma^2$

$$P[|X - \mu| \geq k]$$

$$= P[(X-\mu)^2 \geq k^2] = P[Y \geq k^2]$$

$$\leq \frac{E[Y]}{k^2} = \frac{\sigma^2}{k^2}$$

Tchebyshev's Inequality:

$$P[|X - \mu| \geq k] \leq \frac{\sigma^2}{k^2} \quad \text{if } Var(X) = \sigma^2$$

Example 52 Suppose the working time of a person is a RV X with $\mu = 40$ hrs.

1) How propable is it that the person will work more than 60 hrs?

Apply: $P[X \geq a] \leq \frac{E[X]}{a}$

$$P[X \geq 60] \leq \frac{\mu}{60} = \frac{40}{60} = \frac{2}{3}$$

Example 52 Suppose the working time of a person is a RV X with $\mu = 40$ hrs.

2) If $Var(X) = 16$, how probable is it the person will work between 32 and 48 hrs?

Apply: $$P[|X-\mu| \geq k] \leq \frac{\sigma^2}{k^2}$$

$$P[|X-40| \geq 8] \leq \frac{\sigma^2}{8^2} = \frac{16}{8^2} = \frac{1}{4}$$

$$\Rightarrow P[|X-40| \leq 8] \geq 1 - \frac{1}{4} = \frac{3}{4}$$

Example: Small Schools

Educational scientists found that among the schools that fare best in evaluations of teaching success, there are many more small schools than there are small schools among all schools.

(See statistics from North Carolina)

The Gates Foundation decided in the early 2000's to invest heavily in the establishment of small schools (e.g., by splitting larger schools into smaller ones)

Was that a good idea?

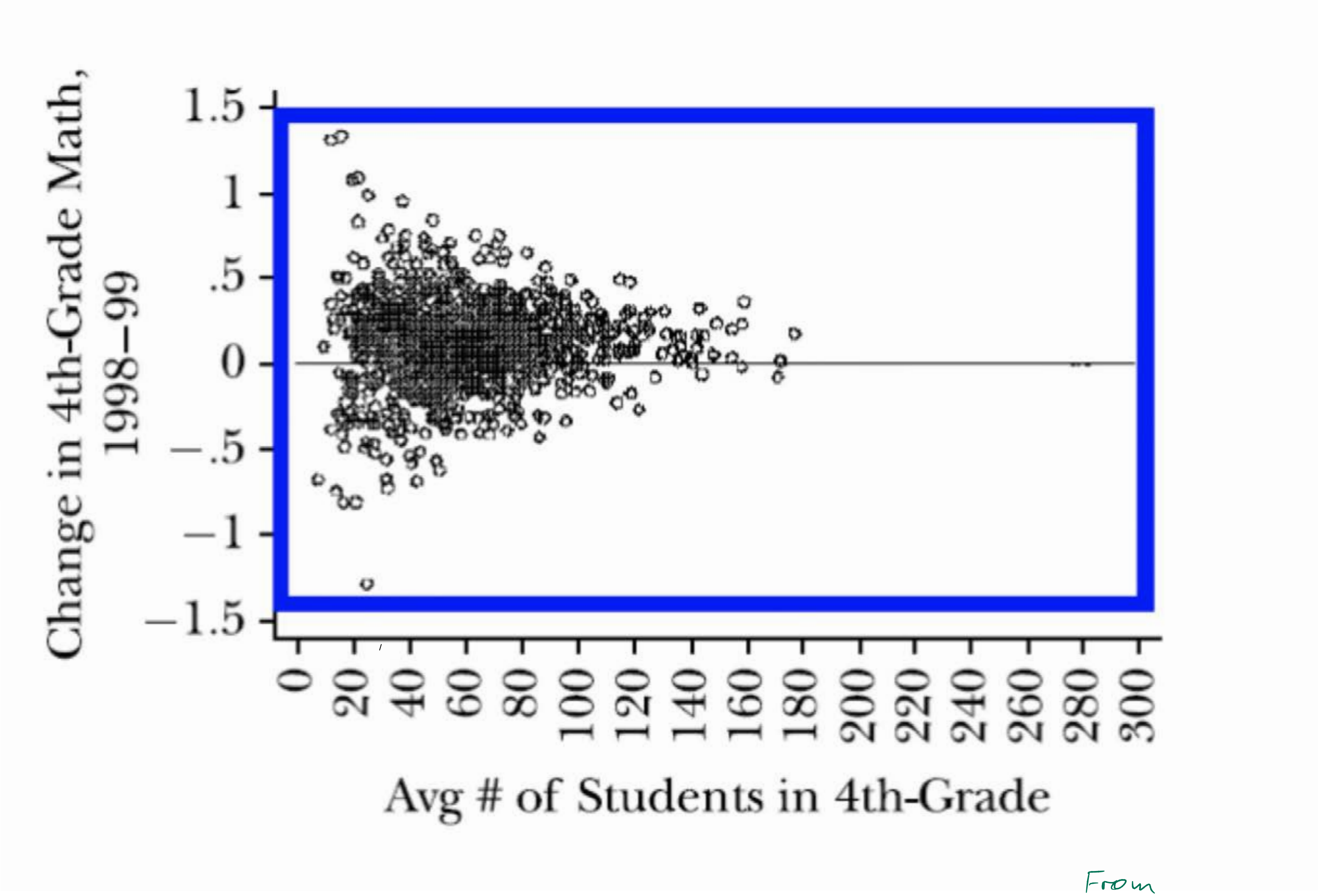
The story is from Daniel Kahnemann, "Thinking, Fast and Slow"

School Size	*Percentage Ever "Top 25" 1997–2000*
Smallest decile	27.7%
2nd	11.8
3rd	8.2
4th	3.6
5th	2.4
6th	3.6
7th	4.8
8th	7.1
9th	0
Largest decile	1.2
Total	7.0

Performance of small schools in North Carolina

From
Alex Tabarrok
"The Small Schools Myth",
September 2, 2010
https://marginalrevolution.com

Distribution of Performance wrt Student Numbers



From
Alex Tabarrok
"The Small Schools Myth",
September 2, 2010
https://marginalrevolution.com

What happens if we execute an experiment many times and take averages of the outcomes?

let X be a RV, let $X_1, \dots, X_n$ be RVs that

1) have the same distribution as X
2) are independent

(i.i.d. RVs, i.e., independent identically distributed RVs)

let $\sigma^2 = Var(X) = Var(X_i)$, $\mu = E[X] = E[X_i]$,

$$\overline{X_n} = \frac{\sum_{i=1}^{n} X_i}{n}$$

We have $\overline{X}_n = \dfrac{\sum_{i=1}^{n} X_i}{n}$. Then

$$E[\overline{X}_n] = \frac{1}{n} \sum_{i=1}^{n} E[X_i] = \frac{1}{n} \cdot (n \cdot \mu) = \mu$$

$$\text{Var}(\overline{X}_n) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} \text{Var}\left(\sum_{i=1}^{n} X_i\right)$$

idependence of the X_i

$$= \frac{1}{n^2} \sum_{i=1}^{n} \text{Var}(X_i) = \frac{1}{n^2} \sum_{i=1}^{n} \sigma^2 = \frac{1}{n^2} \cdot (n \cdot \sigma^2) = \frac{\sigma^2}{n}$$

Chebyshev with $\overline{X}_n$ and $k = \varepsilon$

Probability of ε-outliers

$$P\left[\left|\frac{\sum_{i=1}^{n} X_i}{n} - \mu\right| > \varepsilon\right] = P[|\overline{X}_n - \mu| > \varepsilon]$$

$$\leq \frac{\frac{\sigma^2}{n}}{\varepsilon^2} = \frac{1}{n} \cdot \frac{\sigma^2}{\varepsilon^2} \longrightarrow 0 \qquad (n \to \infty)$$

Theorem (Weak law of Large Numbers)

Let $X_1, \dots, X_n, \dots$ be i.i.d. RVs with mean μ and $Var(X) < \infty$.

Then for every $\varepsilon > 0$

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

i.e., the probability of ε-outliers goes toward 0.

There is also a Strong Law of Large Numbers, which says

$$P\left[\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} X_i = \mu \right] = 1$$

for i.i.d. RVs X_i provided $E[X_i] < \infty$.