The Bright Side of Mathematics

The following pages cover the whole Probability Theory course of the Bright Side of Mathematics. Please note that the creator lives from generous supporters and would be very happy about a donation. See more here: https://tbsom.de/support

Have fun learning mathematics!

The Bright Side of Mathematics

Probability Theory – Part 1

(Stochastic, stochastic processes, statistics,...)

Probability measures

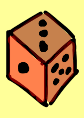
Random variables

Central limit theorem

Probability distributions

Random processes

Statistical tests

Example: 

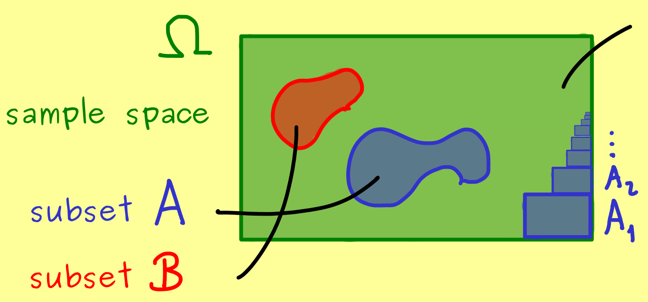
Probability of getting an even number?

$$A = \{2,4,6\} \,, \quad P(A) = \frac{1}{2}$$

$$\frac{\text{number of throws with an even outcome}}{\text{number of total throws}} \longrightarrow \frac{1}{2}$$

Probability Theory – Part 2

Probability measures: measures with total mass $= 1$



$\mathbb{P}: \mathcal{A} \longrightarrow \mathbb{R}$ (collection of subsets)

We want:
- $\mathbb{P}(\Omega) = 1$, $\mathbb{P}(\emptyset) = 0$
- $\mathbb{P}(A) \in [0,1]$
- $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ if A, B are disjoint $\hookrightarrow A \cap B = \emptyset$
- $\mathbb{P}\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} \mathbb{P}(A_j)$ if we have pairwise disjoint sets $\hookrightarrow A_i \cap A_j = \emptyset$ for $i \neq j$

Definition: Let Ω be a set. A collection of subsets $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ (power set) is called a sigma algebra if:

(a) $\emptyset, \Omega \in \mathcal{A}$

(b) If $A \in \mathcal{A}$, then $A^c := \Omega \setminus A \in \mathcal{A}$

(c) If $A_1, A_2, \ldots \in \mathcal{A}$, then $\bigcup_{j=1}^{\infty} A_j \in \mathcal{A}$

$\boxed{\sigma\text{-algebra}}$

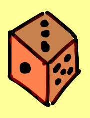
elements $A \in \mathcal{A}$ are called events

Definition: Let $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ be a σ-algebra. A map $\mathbb{P}: \mathcal{A} \longrightarrow [0,1]$ is called a probability measure if:

(a) $\mathbb{P}(\Omega) = 1$, $\mathbb{P}(\emptyset) = 0$

(b) $\mathbb{P}\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} \mathbb{P}(A_j)$

if we have pairwise disjoint sets ($A_i \cap A_j = \emptyset$ for $i \neq j$)

Example:  1 throw: $\Omega = \{1,2,3,4,5,6\}$

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

$\mathbb{P}: \mathcal{A} \longrightarrow [0,1]$, $\mathbb{P}(A) := \dfrac{\# A}{\# \Omega}$ (number of elements in a set)

For example: $\mathbb{P}(\{2\}) = \frac{1}{6}$, $\mathbb{P}(\{2,4,6\}) = \frac{3}{6} = \frac{1}{2}$

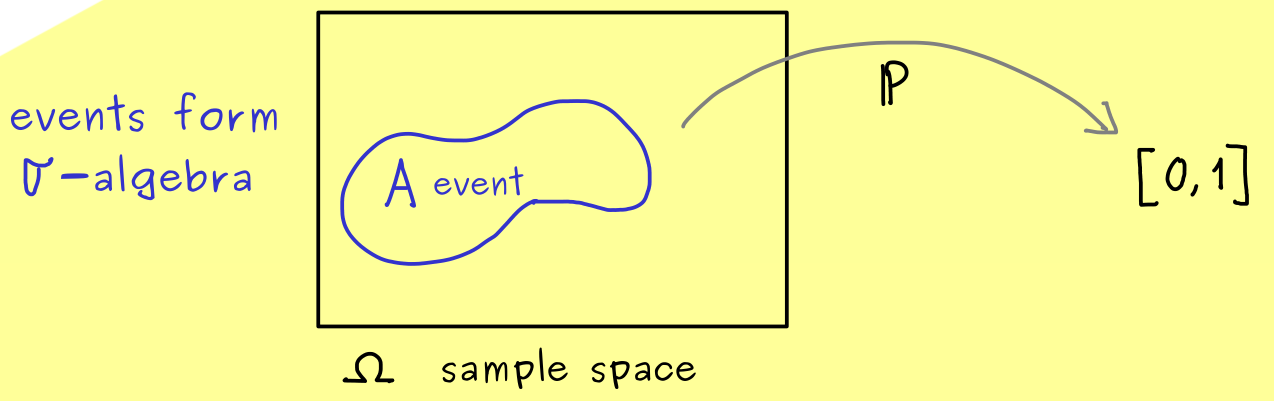
Exercise: Prove: $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$

BECOME A MEMBER

ON STEADY

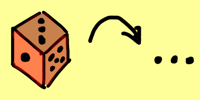
The Bright Side of Mathematics

Probability Theory – Part 3



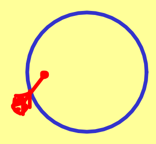
discrete case

"finitely many outcomes"
"countably many outcomes"



(absolutely) continuous case

"uncountably many outcomes"



mixed and other cases

σ-additivity: $\mathbb{P}\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} \mathbb{P}(A_j)$ if we have pairwise disjoint sets

discrete

sample space Ω finite or countable set

(Example: $\Omega = \{\text{Heads}, \text{Tails}\}$, $\Omega = \mathbb{N}$)

σ-algebra $\mathcal{A} = \mathcal{P}(\Omega)$

probability measure $\mathbb{P}: \mathcal{A} \longrightarrow [0,1]$

is completely determined by $\mathbb{P}(\{\omega\})$ for all $\omega \in \Omega$

probability mass function: $(p_\omega)_{\omega \in \Omega}$ with $p_\omega \geq 0$, $\sum_{\omega \in \Omega} p_\omega = 1$

Define: $\mathbb{P}(A) := \sum_{\omega \in A} p_\omega$

Example: $\Omega = \{1,2,3,4,5,6\}$ unfair die

$p_1 = \frac{1}{10}$ $p_2 = \frac{1}{10}$ $p_3 = \frac{1}{10}$ $p_4 = \frac{1}{10}$ $p_5 = \frac{1}{10}$ $p_6 = \frac{1}{2}$

$$\mathbb{P}(\{1,2,3,4,5\}) = \sum_{\omega=1}^{5} p_\omega = 5 \cdot \frac{1}{10} = \frac{1}{2}$$

(abs.) continuous

sample space $\Omega \subseteq \mathbb{R}^n$ uncountable , $\Omega \in \mathcal{B}(\mathbb{R}^n)$ (Borel set)

(Example: $\Omega = [0,1]$)

σ-algebra $\mathcal{A} = \mathcal{B}(\Omega)$

probability measure $\mathbb{P}: \mathcal{A} \longrightarrow [0,1]$

can be described by

probability density function: $f: \Omega \longrightarrow \mathbb{R}$ with $f(x) \geq 0$, $\int_\Omega f(x)\,dx = 1$

measurable!

Define: $\mathbb{P}(A) := \int_A f(x)\,dx$

Example: $\Omega = [0,2]$ throw point into interval

$f: \Omega \longrightarrow \mathbb{R}$ with $f(x) = \frac{1}{2}$

Hence: $\int_0^2 f(x)\,dx = \frac{1}{2} \cdot 2 = 1$

$$\mathbb{P}(A) = \int_A f(x)\,dx = \frac{1}{2} \int_A 1\,dx = \frac{1}{2} \text{Lebesgue measure}(A)$$

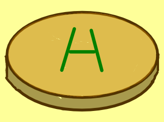
$$\mathbb{P}([a,b]) = \frac{1}{2}(b-a)$$

BECOME A MEMBER

ON STEADY

The Bright Side of Mathematics

Probability Theory – Part 4

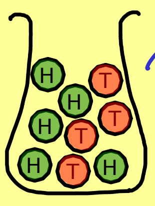


Coin tossing: H, T

Probability for H: $p \in \mathbb{Q} \cap [0,1]$

$p = \frac{a}{a+b}\ , \quad a,b \in \{0,1,2,\dots\}$

(Fair coin: $p = \frac{1}{2}$)



a times H
b times T

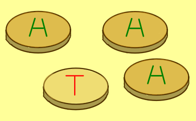
Drawing a ball

Probability for H : $p = \frac{a}{a+b}$

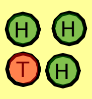
In both cases: $\Omega = \{H, T\}\ , \quad \mathbb{P}(\{H\}) = \underbrace{\frac{a}{a+b}}_{p}\ , \quad \mathbb{P}(\{T\}) = \underbrace{\frac{b}{a+b}}_{1-p}$

Binomial distribution:

- n tosses of the same coin and counting the heads
- draw n balls with replacement and count the heads
- size n , unordered , with replacement



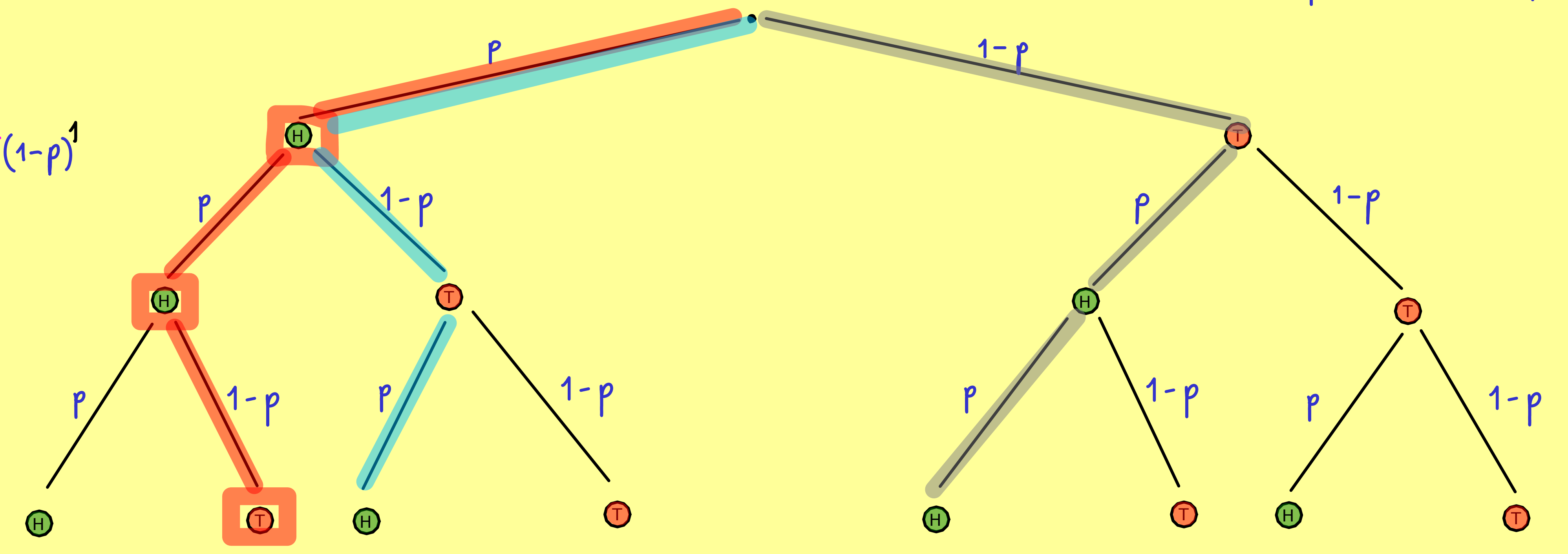
(no order!)



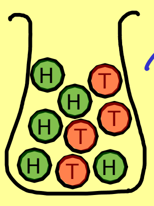
$\Omega = \{0,1,2,\dots,n\}\ , \quad \mathbb{P}(\{k\}) = \binom{n}{k} p^k (1-p)^{n-k}$ two parameters (n, p)

$\mathbb{P} = \mathcal{B}(n,p) = \mathcal{B}in(n,p)$

$\binom{3}{2} p^2 (1-p)^1$



In R:



a times H
b times T

The Bright Side of Mathematics

Probability Theory − Part 5

Probability space $(\Omega, \mathcal{A}, \mathbb{P})$

sample space Ω, σ-algebra $\mathcal{A} \subseteq \mathcal{P}(\Omega)$, probability measure $\mathbb{P}: \mathcal{A} \to [0,1]$

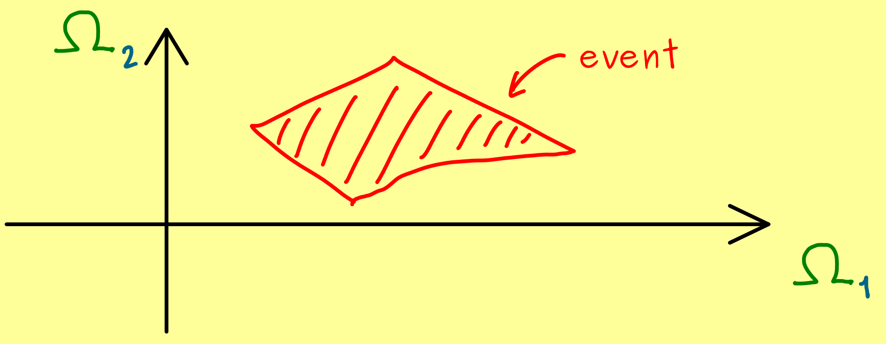
$\rightsquigarrow (\Omega_n, \mathcal{A}_n, \mathbb{P}_n)$, $n \in \{1,2,\dots\}$

Example: first throw a die, then throw a point into the interval $[-1, 1]$

possible outcome: $(3, \frac{1}{4})$ probability?

First probability space: $(\Omega_1, \mathcal{A}_1, \mathbb{P}_1)$, $\Omega_1 = \{1,\dots,6\}$, $\mathcal{A}_1 = \mathcal{P}(\Omega)$, $\mathbb{P}_1(A) = \sum_{k \in A} \frac{1}{6}$

Second probability space: $(\Omega_2, \mathcal{A}_2, \mathbb{P}_2)$, $\Omega_2 = [-1,1]$, $\mathcal{A}_2 = \mathcal{B}(\Omega)$, $\mathbb{P}_2(A) = \int_A \frac{1}{2}\, dx$



new probability space

$(\Omega_1 \times \Omega_2, \sigma(\mathcal{A}_1 \times \mathcal{A}_2), \mathbb{P})$

product σ-algebra, product measure

$\mathbb{P}$ satisfies for $A_1 \in \mathcal{A}_1$, $A_2 \in \mathcal{A}_2$

$$\mathbb{P}(A_1 \times A_2) = \mathbb{P}_1(A_1) \cdot \mathbb{P}_2(A_2)$$

$$\mathbb{P}(\{2,3\} \times [-1,0]) = \mathbb{P}_1(\{2,3\}) \cdot \mathbb{P}_2([-1,0]) = \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{6}$$

Definition: Probability spaces: $(\Omega_n, \mathcal{A}_n, \mathbb{P}_n)$, $n \in \{1,2,\dots\}$

Product space: $(\Omega, \mathcal{A}, \mathbb{P})$ defined by:

- $\Omega = \Omega_1 \times \Omega_2 \times \cdots = \prod_{j \in \mathbb{N}} \Omega_j$ (elements: $(\omega_1, \omega_2, \omega_3, \dots)$)
- $\mathcal{A} = \sigma($"cylinder sets"$)$ product σ-algebra
  $\Omega_1 \times \Omega_2 \times A_3 \times \Omega_4 \times \cdots$
  $A_1 \times \Omega_2 \times \Omega_3 \times \Omega_4 \times \cdots$
- $\mathbb{P}$ product measure

$$\mathbb{P}(A_1 \times A_2 \times \cdots \times A_m \times \Omega_{m+1} \times \Omega_{m+2} \times \cdots) = \mathbb{P}_1(A_1) \cdot \mathbb{P}_2(A_2) \cdot \ldots \cdot \mathbb{P}_m(A_m)$$

Example: throw a die infinitely many times: $(\Omega_0, \mathcal{A}_0, \mathbb{P}_0)$, $\Omega_0 = \{1,\dots,6\}$, $\mathcal{A}_0 = \mathcal{P}(\Omega)$, $\mathbb{P}_0(A) = \sum_{k \in A} \frac{1}{6}$

Product space: $\Omega = \Omega_0 \times \Omega_0 \times \cdots$, $\mathcal{A}$ = product σ-algebra , $\mathbb{P}$ product measure

$A \in \mathcal{A}$ event: "At the 100th throw, we get a six for the first time"

$$A = \underbrace{\{6\}^c \times \{6\}^c \times \cdots \times \{6\}^c}_{99 \text{ times}} \times \{6\} \times \Omega_0 \times \Omega_0 \times \cdots$$

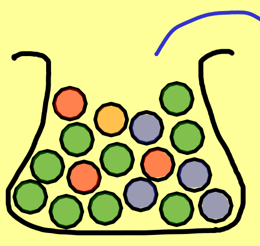
$$\mathbb{P}(A) = \mathbb{P}_0(\{6\}^c) \cdot \ldots \cdot \mathbb{P}_0(\{6\}^c) \cdot \mathbb{P}_0(\{6\}) = \mathbb{P}_0(\{6\}^c)^{99} \cdot \mathbb{P}_0(\{6\}) = \left(\frac{5}{6}\right)^{99} \cdot \frac{1}{6}$$

Probability Theory – Part 6

Hypergeometric distribution (multivariant)

size n , unordered , without replacement

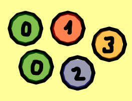
urn model



draw n balls at once

colours: finite set $\mathcal{C}$

(for example: $\mathcal{C} = \{0,1,2,3\}$)

(one possible outcome: 
$n = 5$)

function $\mathcal{C} \longrightarrow \mathbb{N}_0$ or $(2,1,1,1)$

Sample space:

$$\Omega = \Big\{ (k_c)_{c\in\mathcal{C}} \in \mathbb{N}_0^{\mathcal{C}} \;\Big|\; \sum_{c\in\mathcal{C}} k_c = n \Big\}$$

For our example: $\Omega = \Big\{ (k_0, k_1, k_2, k_3) \in \mathbb{N}_0^4 \;\Big|\; k_0 + k_1 + k_2 + k_3 = n \Big\}$

N_c = number of balls for colour c in the urn

$N := \sum_{c\in\mathcal{C}} N_c$ total number of balls

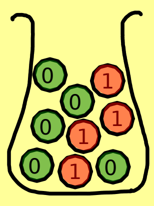
$$\mathbb{P}\big(\{(k_0, k_1, k_2, k_3)\}\big) = \frac{\binom{N_0}{k_0}\cdot\binom{N_1}{k_1}\cdot\binom{N_2}{k_2}\cdot\binom{N_3}{k_3}}{\binom{N}{n}}$$

(multivariant) hypergeometric distribution:

$$\mathbb{P}\big(\{(k_c)_{c\in\mathcal{C}}\}\big) = \frac{\prod_{c\in\mathcal{C}} \binom{N_c}{k_c}}{\binom{N}{n}}$$

Hypergeometric distribution for two colours:

$\mathcal{C} = \{0,1\}$, $N_0 + N_1 = N$



count the 1s : $\Omega = \{0, 1, 2, \dots, n\}$

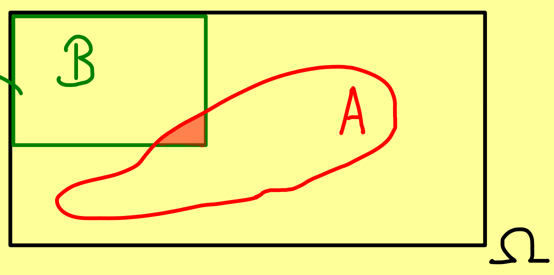
$$\mathbb{P} : \mathcal{P}(\Omega) \longrightarrow [0,1] \;, \quad \mathbb{P}(\{k\}) = \frac{\binom{N_1}{k}\cdot\binom{N-N_1}{n-k}}{\binom{N}{n}}$$

The Bright Side of Mathematics

Probability Theory – Part 7

Conditional probability: $(\Omega, \mathcal{A}, \mathbb{P})$ probability space

subset $B \in \mathcal{A}$
with $\mathbb{P}(B) \neq 0$



$\Longrightarrow$ new probability space: $(B, \tilde{\mathcal{A}}, \tilde{\mathbb{P}})$ $\quad \tilde{\mathbb{P}}(A) = \dfrac{\mathbb{P}(A)}{\mathbb{P}(B)}$

only $A \in \mathcal{A}$ with $A \subseteq B$

$\Longrightarrow$ new probability space: $(\Omega, \mathcal{A}, \mathbb{P}_B)$ $\quad \mathbb{P}_B(A) = \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$

Definition: $(\Omega, \mathcal{A}, \mathbb{P})$ probability space, $B \in \mathcal{A}$ with $\mathbb{P}(B) \neq 0$.

$$\mathbb{P}(A \mid B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$ is called the conditional probability of A under B

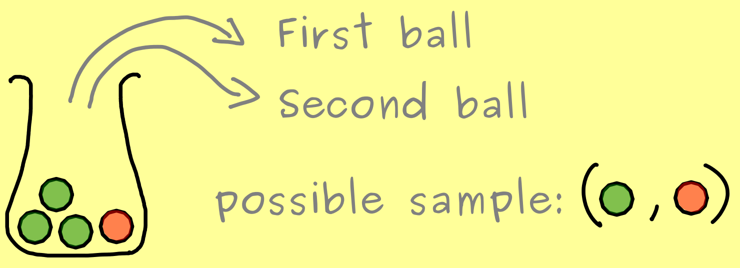
$\mathbb{P}(\cdot \mid B) : \mathcal{A} \longrightarrow [0,1]$ is called the conditional probability measure given B

Property: $\mathbb{P}(B \mid B) = 1$ $\quad$ (For $\mathbb{P}(B) = 0$, set $\mathbb{P}(A \mid B) := 0$)

Example: urn model: ordered, without replacement

$C := \{g, r\}$, $\quad \Omega = C \times C$

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



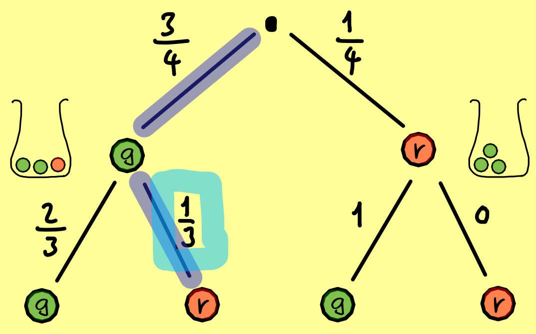
$\mathbb{P}$ given by probability mass function

$\mathbb{P}(\{(g,g)\}) = \frac{1}{2}$

$\mathbb{P}(\{(g,r)\}) = \frac{1}{4}$

$\mathbb{P}(\{(r,g)\}) = \frac{1}{4}$

$\mathbb{P}(\{(r,r)\}) = 0$



event: B = "first ball is green" $= \{(g,g), (g,r)\}$

$$\mathbb{P}(\{(g,r)\} \mid B) = \frac{\mathbb{P}(\{(g,r)\} \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(\{(g,r)\})}{\mathbb{P}(B)} = \frac{\frac{1}{4}}{\frac{3}{4}} = \boxed{\frac{1}{3}}$$

BECOME A MEMBER

ON STEADY

The Bright Side of
Mathematics

Probability Theory – Part 8

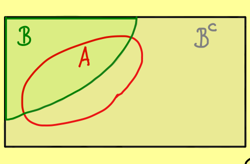
Bayes's theorem: $\mathbb{P}(A|B) = \dfrac{\boxed{\mathbb{P}(A \cap B)}}{\mathbb{P}(B)}$, $\mathbb{P}(B|A) = \dfrac{\boxed{\mathbb{P}(B \cap A)}}{\mathbb{P}(A)}$

$$\Longrightarrow \boxed{\mathbb{P}(A|B) \cdot \mathbb{P}(B) = \mathbb{P}(B|A) \cdot \mathbb{P}(A)}$$

Law of total probability: $(\Omega, \mathcal{A}, \mathbb{P})$ probability space

$A \in \mathcal{A}$

$\mathbb{P}(A)$



Ω

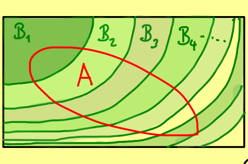
$B \in \mathcal{A}$

$\underbrace{B \cup B^c}_{\text{disjoint union}} = \Omega$

$$\mathbb{P}(A) = \mathbb{P}(\underbrace{(A \cap B) \cup (A \cap B^c)}_{\text{disjoint union}}) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c)$$

$$= \mathbb{P}(A|B) \cdot \mathbb{P}(B) + \mathbb{P}(A|B^c) \cdot \mathbb{P}(B^c)$$

Case with countably many sets: $B_i \in \mathcal{A}$ for $i \in I \subseteq \mathbb{N}$ with $\underbrace{\bigcup_{i \in I} B_i}_{\text{disjoint union}} = \Omega$



Ω

$$\mathbb{P}(A) = \mathbb{P}\Big(\underbrace{\bigcup_{i \in I} (A \cap B_i)}_{\text{disjoint union}}\Big) = \sum_{i \in I} \mathbb{P}(A \cap B_i) = \sum_{i \in I} \mathbb{P}(A|B_i) \cdot \mathbb{P}(B_i)$$

Example: Monty Hall problem

1	2	3

3 doors:
- 1 car
- 2 goats

First: You pick a door (1)

Second: Show master opens a door with a goat (never the door you picked)

Third: Stay or switch

$C_j :=$ car is behind door j , $S_j :=$ show master opens door j (in the second step)

We know: $\mathbb{P}(S_3|C_3) = 0$, $\mathbb{P}(S_3|C_2) = 1$, $\mathbb{P}(S_3|C_1) = \frac{1}{2}$

$$\mathbb{P}(C_2|S_3) \underset{\substack{\uparrow \\ \text{Bayes's theorem}}}{=} \frac{\mathbb{P}(S_3|C_2) \cdot \mathbb{P}(C_2)}{\mathbb{P}(S_3)} \underset{\substack{\uparrow \\ \text{Law of total probability}}}{=} \frac{\mathbb{P}(S_3|C_2) \cdot \mathbb{P}(C_2)}{\sum_{j=1}^{3} \mathbb{P}(S_3|C_j) \cdot \mathbb{P}(C_j)}$$

$$= \frac{\overbrace{\mathbb{P}(S_3|C_2)}^{=1} \cdot \overbrace{\mathbb{P}(C_2)}^{\frac{1}{3}}}{\underbrace{\mathbb{P}(S_3|C_1)}_{=\frac{1}{2}} \cdot \underbrace{\mathbb{P}(C_1)}_{=\frac{1}{3}} + \underbrace{\mathbb{P}(S_3|C_2)}_{=1} \cdot \underbrace{\mathbb{P}(C_2)}_{=\frac{1}{3}} + \underbrace{\mathbb{P}(S_3|C_3)}_{=0} \cdot \underbrace{\mathbb{P}(C_3)}_{=\frac{1}{3}}} = \frac{2}{3}$$

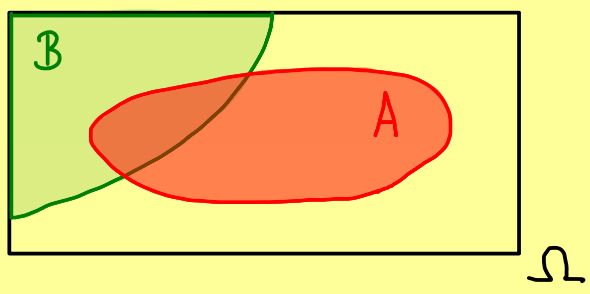
BECOME A MEMBER

ON STEADY

The Bright Side of Mathematics

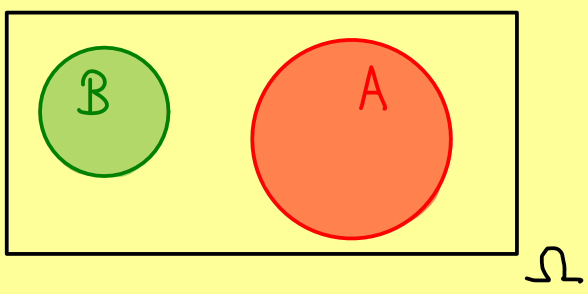
Probability Theory – Part 9

Independence (for events)



$A, \mathcal{B} \subseteq \Omega$ events

independent?

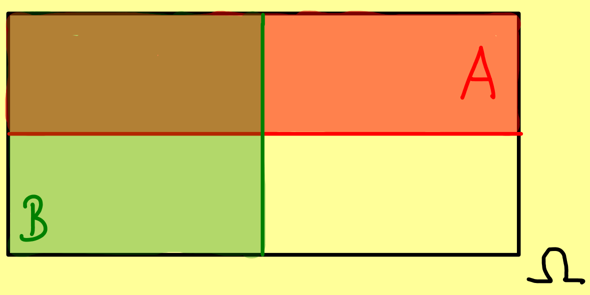


$A, \mathcal{B} \subseteq \Omega$ events

independent!

We want: $\mathbb{P}(A \mid \mathcal{B}) = \mathbb{P}(A)$ and $\mathbb{P}(\mathcal{B} \mid A) = \mathbb{P}(\mathcal{B})$

Example:



$\mathbb{P}(A) = \frac{1}{2}$, $\mathbb{P}(A \mid \mathcal{B}) = \frac{1}{2}$

$\mathbb{P}(\mathcal{B}) = \frac{1}{2}$, $\mathbb{P}(\mathcal{B} \mid A) = \frac{1}{2}$

$\Longrightarrow$ independent!

Recall: $\mathbb{P}(A) \overset{!}{=} \mathbb{P}(A \mid \mathcal{B}) = \frac{\mathbb{P}(A \cap \mathcal{B})}{\mathbb{P}(\mathcal{B})}$, $\mathbb{P}(\mathcal{B}) \overset{!}{=} \mathbb{P}(\mathcal{B} \mid A) = \frac{\mathbb{P}(A \cap \mathcal{B})}{\mathbb{P}(A)}$

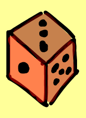
$$\Longleftrightarrow \quad \mathbb{P}(A \cap \mathcal{B}) \overset{!}{=} \mathbb{P}(A) \cdot \mathbb{P}(\mathcal{B})$$

Definition: Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space.

Two events $A, \mathcal{B} \in \mathcal{A}$ are called independent if $\mathbb{P}(A \cap \mathcal{B}) = \mathbb{P}(A) \cdot \mathbb{P}(\mathcal{B})$.

A family $(A_i)_{i \in I}$ with $A_i \in \mathcal{A}$ is called independent if

$$\mathbb{P}\Big(\bigcap_{j \in J} A_j\Big) = \prod_{j \in J} \mathbb{P}(A_j) \quad \text{for all finite } \underset{\emptyset \neq}{J} \subseteq I .$$

Example:  2 throws with order: $(\Omega, \mathcal{A}, \mathbb{P})$ with $\Omega = \{1,2,3,4,5,6\}^2$, $\mathcal{A} = \mathcal{P}(\Omega)$, $\mathbb{P}$ = uniform distribution

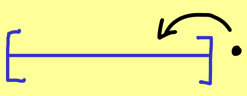
$$\mathbb{P}(\{(\omega_1, \omega_2)\}) = \frac{1}{36}$$

A = "first throw gives 6" $= \{(\omega_1, \omega_2) \in \Omega \mid \omega_1 = 6\}$

$\mathcal{B}$ = "sum of both throws is 7" $= \{(\omega_1, \omega_2) \in \Omega \mid \omega_1 + \omega_2 = 7\}$

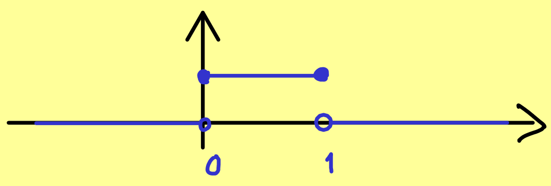
$$\mathbb{P}(A) = \frac{1}{6} \quad , \quad \mathbb{P}(\mathcal{B}) = \mathbb{P}(\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}) = \frac{6}{36} = \frac{1}{6}$$

$$\mathbb{P}(A \cap \mathcal{B}) = \mathbb{P}(\{(6,1)\}) = \frac{1}{36} = \mathbb{P}(A) \cdot \mathbb{P}(\mathcal{B}) \quad \Longrightarrow \quad A, \mathcal{B} \text{ are independent}$$

Example:  throw a point into unit interval $(\Omega, \mathcal{A}, \mathbb{P})$ with $\Omega = [0,1]$, $\mathcal{A} = \mathcal{B}(\Omega)$, $\mathbb{P}$ = uniform distribution density function

$$\mathbb{P}([a,b]) = \int_{[a,b]} 1 \, dx = b - a \quad \boxed{\text{for } b > a \text{ and } a, b \in \Omega}$$

$f: \Omega \to \mathbb{R}$ with $f(x) = 1$



indicator function: $\mathbb{1}_{[0,1]}(x) := \begin{cases} 1 , & x \in [0,1] \\ 0 , & \text{else} \end{cases}$

For two independent events $A, \mathcal{B} \in \mathcal{A}$, we have:

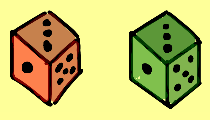
$$\int_{A \cap \mathcal{B}} \mathbb{1}_{[0,1]}(x) \, dx = \mathbb{P}(A \cap \mathcal{B}) = \mathbb{P}(A) \cdot \mathbb{P}(\mathcal{B}) = \int_A \mathbb{1}_{[0,1]}(x) \, dx \cdot \int_{\mathcal{B}} \mathbb{1}_{[0,1]}(x) \, dx$$

$$\int_{[0,1]} \mathbb{1}_{A \cap \mathcal{B}}(x) \, dx = \int_{[0,1]} \mathbb{1}_A(x) \, dx \cdot \int_{[0,1]} \mathbb{1}_{\mathcal{B}}(x) \, dx$$

The Bright Side of Mathematics

Probability Theory − Part 10

Random variables $X: \Omega \longrightarrow \mathbb{R}$ with some properties.

Example: Throwing two dice 

$$(\underset{\{1,2,3,4,5,6\}^2}{\underset{\parallel}{\Omega}}, \underset{\mathcal{P}(\Omega)}{\underset{\parallel}{\mathcal{A}}}, \mathbb{P})$$ $\mathbb{P}$ = uniform distribution

$$X: \Omega \longrightarrow \mathbb{R}, \quad (\omega_1, \omega_2) \longmapsto \omega_1 + \omega_2$$

random variable gives sum of the numbers the dice show

Definition: Let $(\Omega, \mathcal{A})$ and $(\tilde{\Omega}, \tilde{\mathcal{A}})$ be measurable spaces (= event spaces).

A map $X: \Omega \longrightarrow \tilde{\Omega}$ is called a random variable if

$$X^{-1}(\tilde{A}) \in \mathcal{A} \quad \text{for all } \tilde{A} \in \tilde{\mathcal{A}}.$$

Examples: (a) $(\underset{\{1,2,3,4,5,6\}^2}{\underset{\parallel}{\Omega}}, \underset{\mathcal{P}(\Omega)}{\underset{\parallel}{\mathcal{A}}})$ and $(\underset{\mathbb{R}}{\underset{\parallel}{\tilde{\Omega}}}, \underset{\mathcal{B}(\mathbb{R})}{\underset{\parallel}{\tilde{\mathcal{A}}}})$, $X: \Omega \longrightarrow \mathbb{R}, \; (\omega_1, \omega_2) \longmapsto \omega_1 + \omega_2$

$$X^{-1}(\tilde{A}) \in \mathcal{P}(\Omega) \quad \text{for all } \tilde{A} \in \tilde{\mathcal{A}}. \Rightarrow$$ X is a random variable

(b) $(\underset{\{1,2,3,4,5,6\}^2}{\underset{\parallel}{\Omega}}, \underset{\{\emptyset, \Omega\}}{\underset{\parallel}{\mathcal{A}}})$ and $(\underset{\mathbb{R}}{\underset{\parallel}{\tilde{\Omega}}}, \underset{\mathcal{B}(\mathbb{R})}{\underset{\parallel}{\tilde{\mathcal{A}}}})$, $X: \Omega \longrightarrow \mathbb{R}, \; (\omega_1, \omega_2) \longmapsto \omega_1 + \omega_2$

$$X^{-1}(\{2\}) = \{(1,1)\} \notin \mathcal{A} \Rightarrow$$ X is not a random variable

Notation: Let $(\Omega, \mathcal{A})$ and $(\tilde{\Omega}, \tilde{\mathcal{A}})$ be measurable spaces (= event spaces).

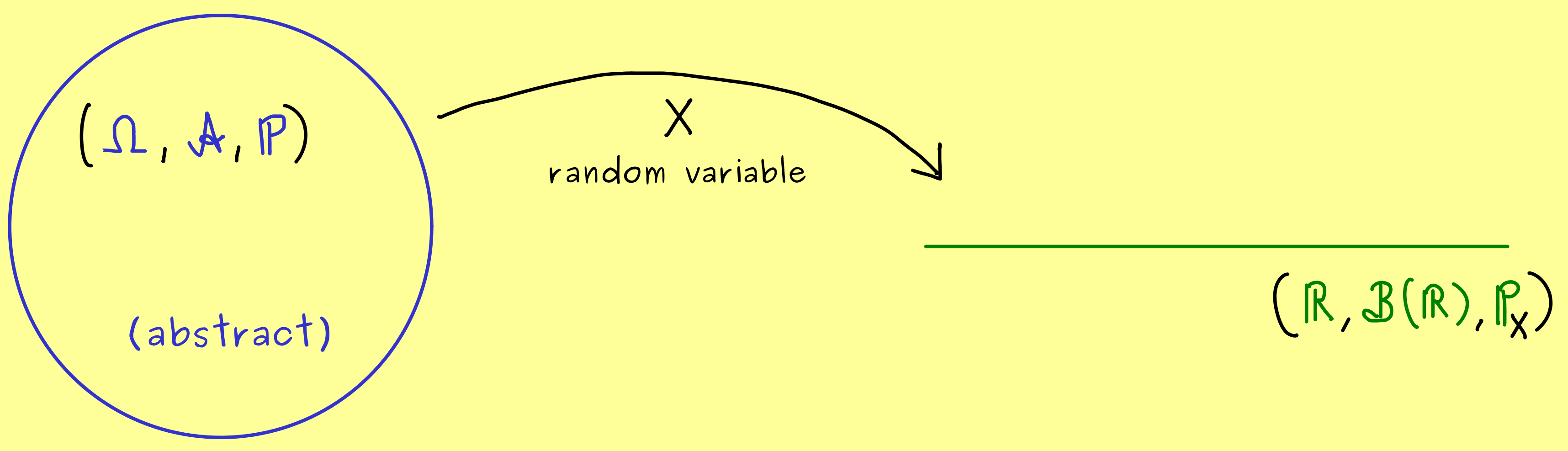
probability measure $\mathbb{P}: \mathcal{A} \rightarrow [0,1]$, $X: \Omega \longrightarrow \tilde{\Omega}$ random variable

$$\mathbb{P}(X \in \tilde{A}) := \mathbb{P}\big(X^{-1}(\tilde{A})\big) = \mathbb{P}\big(\{\omega \in \Omega \mid X(\omega) \in \tilde{A}\}\big)$$

The Bright Side of Mathematics

Probability Theory – Part 11

$(\Omega, \mathcal{A}), (\tilde{\Omega}, \tilde{\mathcal{A}})$ event spaces, $X : \Omega \longrightarrow \tilde{\Omega}$, where $\tilde{\Omega} = \mathbb{R}$ and $\tilde{\mathcal{A}} = \mathcal{B}(\mathbb{R})$



Definition: Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $X : \Omega \longrightarrow \mathbb{R}$ be a random variable.
(with Borel sigma algebra)

Then $\mathbb{P}_X : \mathcal{B}(\mathbb{R}) \longrightarrow [0,1]$ defined by

$$\mathbb{P}_X(B) := \mathbb{P}(X^{-1}(B)) = \mathbb{P}(X \in B)$$

is called probability distribution of X.

Proposition: $\mathbb{P}_X$ is a probability measure.

Proof:

$\mathbb{P}$ is a probability measure

$$X^{-1}(\mathbb{R}) = \Omega \quad \Rightarrow \quad \mathbb{P}_X(\mathbb{R}) = \mathbb{P}(X^{-1}(\mathbb{R})) = \mathbb{P}(\Omega) = 1$$

$$X^{-1}(\emptyset) = \emptyset \quad \Rightarrow \quad \mathbb{P}_X(\emptyset) = \mathbb{P}(X^{-1}(\emptyset)) = \mathbb{P}(\emptyset) = 0$$

For σ-additivity: Choose $B_1, B_2, B_3, \ldots \in \mathcal{B}(\mathbb{R})$ pairwise disjoint.

Then: $i \neq j \Rightarrow X^{-1}(B_i) \cap X^{-1}(B_j) = X^{-1}(\underbrace{B_i \cap B_j}_{=\emptyset}) = \emptyset$

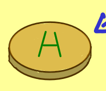
So: $X^{-1}(B_1), X^{-1}(B_2), X^{-1}(B_3) \ldots \in \mathcal{A}$ pairwise disjoint.

And: $\mathbb{P}_X\left(\bigcup_{j=1}^{\infty} B_j\right) = \mathbb{P}\left(X^{-1}\left(\bigcup_{j=1}^{\infty} B_j\right)\right) = \mathbb{P}\left(\bigcup_{j=1}^{\infty} X^{-1}(B_j)\right)$

$\mathbb{P}$ is a probability measure

$$= \sum_{j=1}^{\infty} \mathbb{P}(X^{-1}(B_j)) = \sum_{j=1}^{\infty} \mathbb{P}_X(B_j) \qquad \square$$

Notation: If $\tilde{\mathbb{P}}$ probability measure and $\mathbb{P}_X = \tilde{\mathbb{P}}$, then $X \sim \tilde{\mathbb{P}}$.

Example:  $\leftarrow p$ n tosses of the same coin $(\Omega, \mathcal{A}, \mathbb{P})$, where $\Omega = \{0,1\}^n$, $\mathcal{A} = \mathcal{P}(\Omega)$, $\mathbb{P}$ BERNOULLI:

$$\mathbb{P}(\{\omega\}) = p^{\#1s} \cdot (1-p)^{\#0s}$$

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

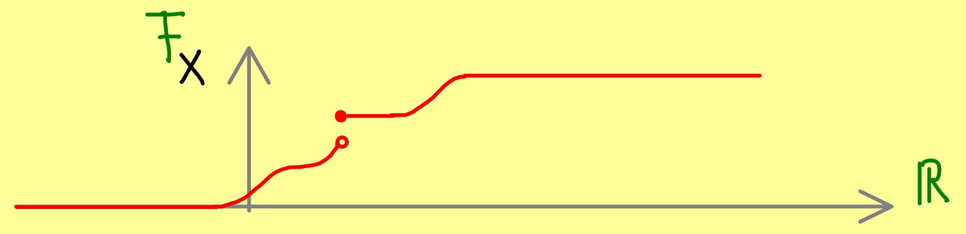
$X(\omega) :=$ number of 1s in ω $\overset{\text{part 4}}{\Rightarrow}$ $X \sim \mathrm{Bin}(n, p)$

The Bright Side of Mathematics

Probability Theory – Part 12

Cumulative distribution function (cdf)

$(\Omega, \mathcal{A}, \mathbb{P})$ (abstract) $\xrightarrow[\text{random variable}]{X}$ $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_X)$



Definition: Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $X: \Omega \to \mathbb{R}$ be a random variable.
(with Borel sigma algebra)

$$F_X : \mathbb{R} \to [0,1] \quad , \quad F_X(x) := \mathbb{P}_X((-\infty, x]) = \mathbb{P}(X \leq x)$$

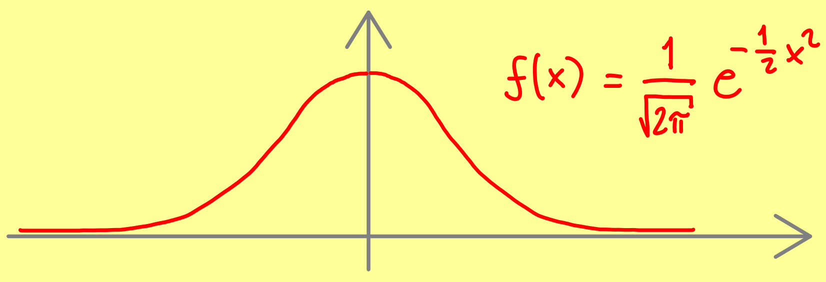
is called the cumulative distribution function of X.

Properties:

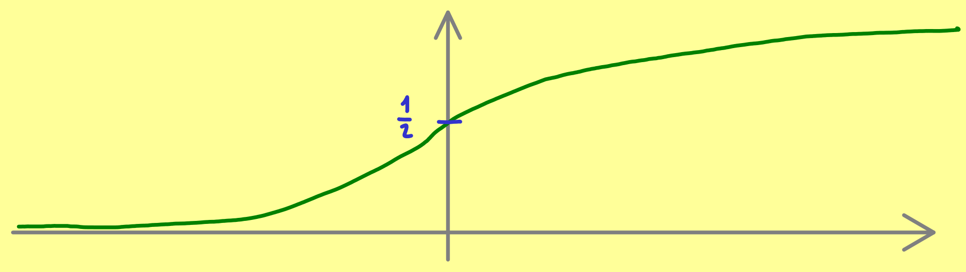
- $F_X(x) \xrightarrow{x \to -\infty} 0 \quad , \quad F_X(x) \xrightarrow{x \to \infty} 1$
- F_X is monotonically increasing $\left(x_1 < x_2 \Rightarrow F_X(x_1) \leq F_X(x_2) \right)$
- F_X is right-continuous $\left(\lim_{x \searrow x_0} F_X(x) = F_X(x_0) \right)$

Example: $X \sim \mathrm{NORMAL}(0, 1^2)$

probability density function

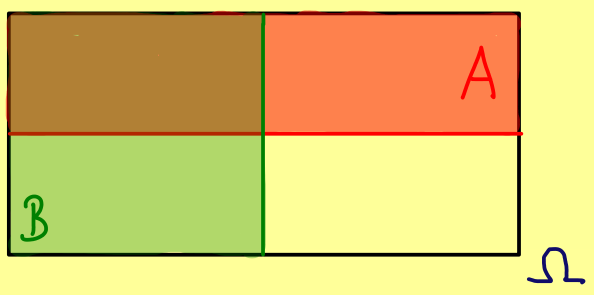


cdf: $F_X(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}t^2} \, dt$



The Bright Side of Mathematics

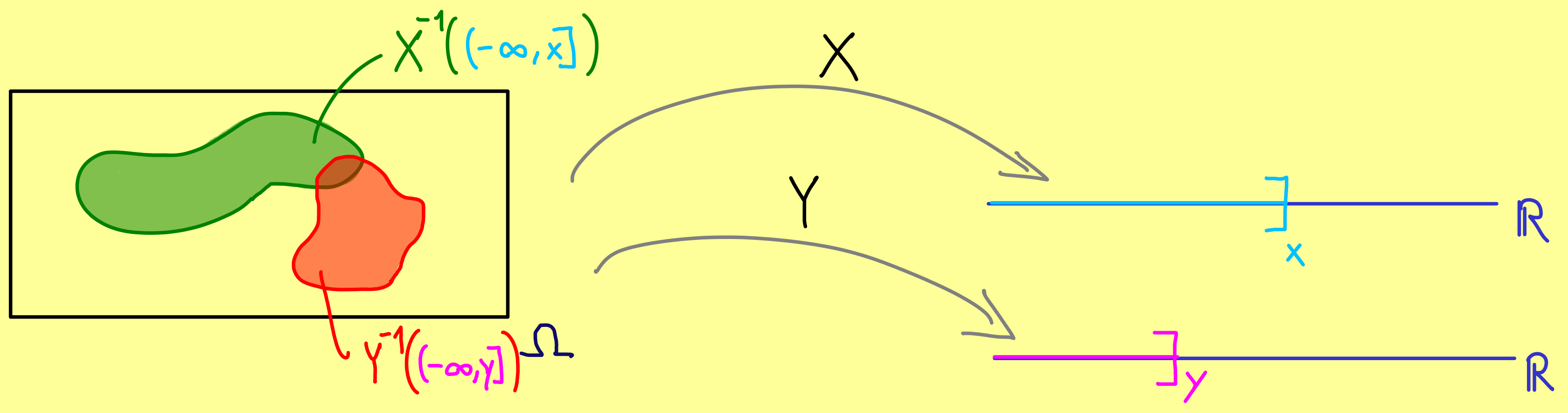
Probability Theory – Part 13



$A, B \subseteq \Omega$

two independents events

$X : \Omega \longrightarrow \mathbb{R}$, $Y : \Omega \longrightarrow \mathbb{R}$ two independent random variables?



Definition: Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and let $X : \Omega \longrightarrow \mathbb{R}$, $Y : \Omega \longrightarrow \mathbb{R}$ be two random variables.

Then X, Y are called independent if for all $x, y \in \mathbb{R}$

$X^{-1}((-\infty, x])$ and $Y^{-1}((-\infty, y])$ are independent events.

$$\Longleftrightarrow \quad \mathbb{P}\big(X^{-1}((-\infty,x]) \cap Y^{-1}((-\infty,y])\big) = \mathbb{P}\big(X^{-1}((-\infty,x])\big) \cdot \mathbb{P}\big(Y^{-1}((-\infty,y])\big)$$

$$\Longleftrightarrow \quad \underbrace{\mathbb{P}(X \leq x, Y \leq y)}_{F_{(X,Y)}(x,y)} = F_X(x) \cdot F_Y(y)$$

$F_{(X,Y)}(x,y)$ ⟵ cdf of random variable $(X,Y) : \Omega \longrightarrow \mathbb{R}^2$

Example: Product space: $\Omega = \Omega_1 \times \Omega_2$, $X : \Omega \longrightarrow \mathbb{R}$, $X(\omega_1, \omega_2) = f(\omega_1)$

$Y : \Omega \longrightarrow \mathbb{R}$, $Y(\omega_1, \omega_2) = g(\omega_2)$

$\Longrightarrow$ X, Y are independent random variables

Definition: A family $(X_i)_{i \in I}$ is called independent if

$$\mathbb{P}\big((X_j \leq x_j)_{j \in J}\big) = \prod_{j \in J} \mathbb{P}(X_j \leq x_j)$$

for all $x_j \in \mathbb{R}$

for all finite $J \subseteq I$, $J \neq \emptyset$

Probability Theory — Part 14

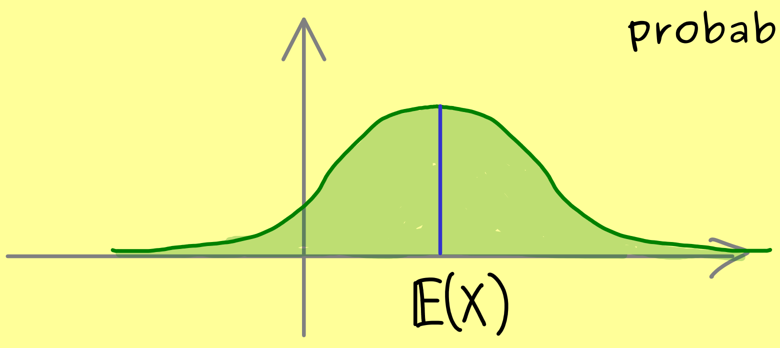
$(\Omega, \mathcal{A}, \mathbb{P})$ probability space

$X: \Omega \longrightarrow \mathbb{R}$ random variable

$\mathbb{E}(X) \in \mathbb{R}$ expectation of X (expected value, mean, expectancy...)

continuous case:

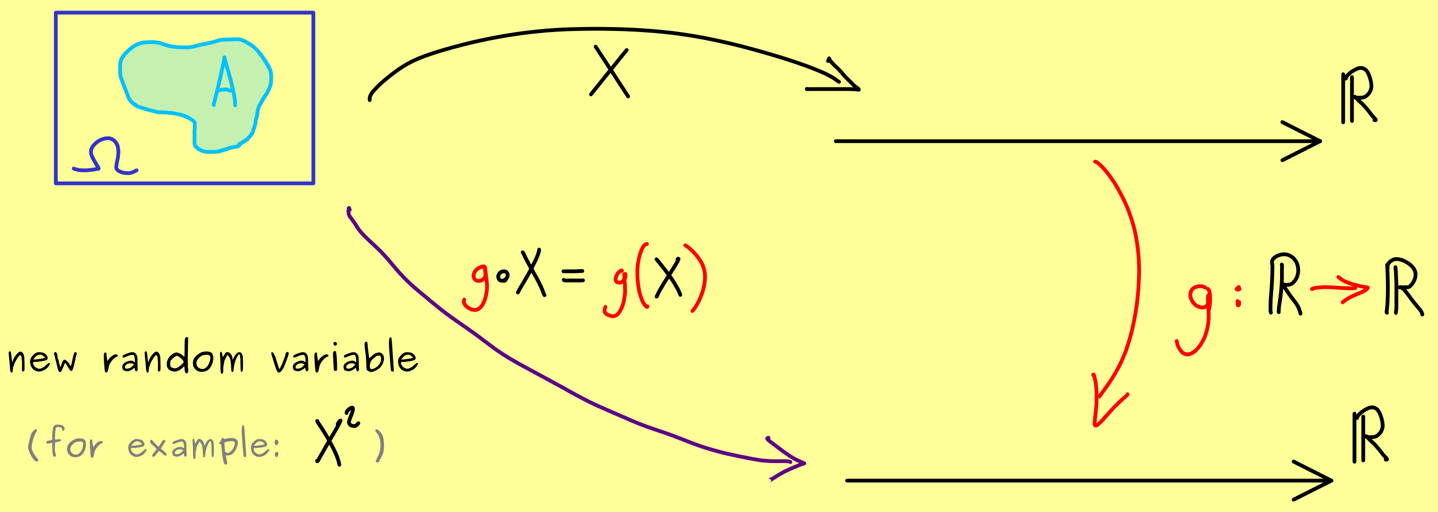
probability density function of $\mathbb{P}_X$



Definition: $(\Omega, \mathcal{A}, \mathbb{P})$ probability space, $X: \Omega \longrightarrow \mathbb{R}$ random variable.

$$\mathbb{E}(X) := \int_\Omega X \, d\mathbb{P} \qquad \text{(abstract integral)}$$

Change of variables:



$$\int_A g(X)\, d\mathbb{P} = \int_A g(\underbrace{X(\omega)}_{=:x})\, d\mathbb{P}(\omega) \quad (\text{"}X^{-1}(x) = \omega\text{"}) \quad = \int_{X(A)} g(x)\, d(\underbrace{\mathbb{P}\circ X^{-1}}_{\mathbb{P}_X})(x)$$

$$= \int_{X(A)} g(x)\, d\mathbb{P}_X(x) = \begin{cases} \displaystyle\int_{X(A)} g(x)\, f_X(x)\, dx & \text{continuous case} \quad (f_X: \text{pdf of } \mathbb{P}_X) \\ \displaystyle\sum_{x\in X(A)} g(x)\cdot p_x & \text{discrete case} \quad (p_x: \text{pmf of } \mathbb{P}_X,\ p_x = \mathbb{P}_X(\{x\})) \end{cases}$$

Remember:

$$\mathbb{E}(X) = \begin{cases} \displaystyle\int_{X(\Omega)} x\cdot f_X(x)\, dx & \text{continuous case} \\ \displaystyle\sum_{x\in X(\Omega)} x\cdot p_x & \text{discrete case} \end{cases}$$

Example: $X: \underbrace{\Omega}_{=\{1,2,3,\dots,6\}} \longrightarrow \mathbb{R}$ throwing a fair die, $X(\omega) = \omega$

$$\mathbb{E}(X) = \sum_{x\in X(\Omega)} x\cdot p_x = 1\cdot\frac{1}{6} + 2\cdot\frac{1}{6} + \dots + 6\cdot\frac{1}{6} = \underline{3.5}$$

Probability Theory – Part 15

$$\mathbb{E}(X) := \int_\Omega X \, d\mathbb{P}$$

Example: $\underbrace{X \sim \text{Exp}(\lambda)}$ (exponential distribution)

$$\mathbb{P}_X(A) = \int_A \underbrace{f_X(x)}_{\text{pdf}} dx \, , \quad f_X(x) = \begin{cases} \lambda e^{-\lambda \cdot x} & , x > 0 \\ 0 & , x \leq 0 \end{cases}$$

$$\mathbb{E}(X) = \int_\Omega X \, d\mathbb{P} = \int_\mathbb{R} x \cdot f_X(x) \, dx = \int_0^\infty x \cdot \lambda e^{-\lambda \cdot x} dx = \frac{1}{\lambda}$$

Properties: $(\Omega, \mathcal{A}, \mathbb{P})$ probability space, $X, Y : \Omega \longrightarrow \mathbb{R}$ random variables, where $\mathbb{E}(X)$ and $\mathbb{E}(Y)$ exist.

(a) $\mathbb{E}(a \cdot X + b \cdot Y) = a \cdot \mathbb{E}(X) + b \cdot \mathbb{E}(Y)$ for all $a, b \in \mathbb{R}$

(b) If X, Y are independent, then: $\mathbb{E}(X \cdot Y) = \mathbb{E}(X) \cdot \mathbb{E}(Y)$

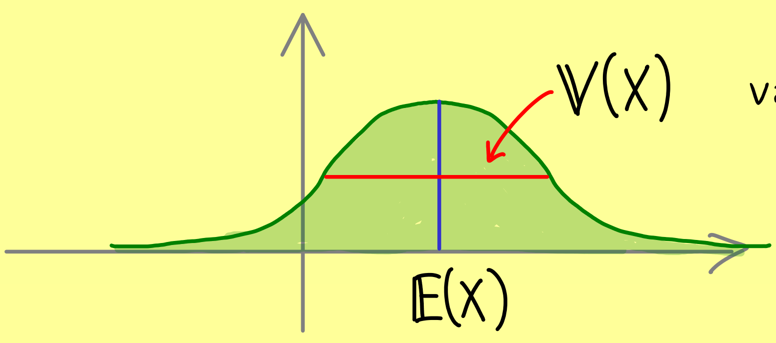
(c) If $\mathbb{P}_X = \mathbb{P}_Y$, then: $\mathbb{E}(X) = \mathbb{E}(Y)$

(d) If $X \leq Y$ almost surely $\mathbb{P}(\{\omega \in \Omega \mid X(\omega) \leq Y(\omega)\}) = 1$,

then: $\mathbb{E}(X) \leq \mathbb{E}(Y)$

The Bright Side of Mathematics

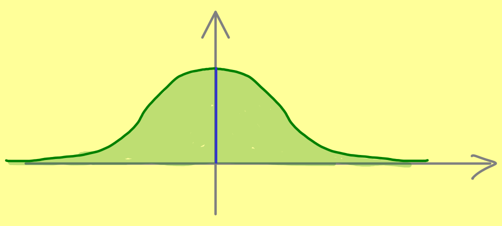
Probability Theory – Part 16

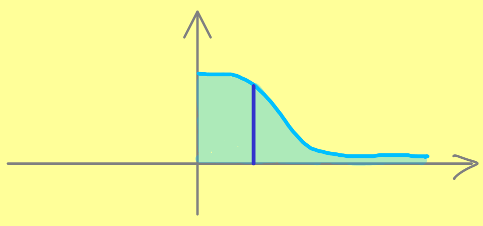


variance?

Definition: $(\Omega, \mathcal{A}, \mathbb{P})$ probability space, $X : \Omega \longrightarrow \mathbb{R}$ random variable.

$$Var(X) := \mathbb{E}\Big(\underbrace{\big(X - \mathbb{E}(X)\big)^2}_{\text{new random variable}}\Big)$$





$$= \mathbb{E}\big(X^2 - 2\cdot\mathbb{E}(X)\cdot X + \mathbb{E}(X)^2\big)$$

$$\overset{\text{linearity}}{=} \mathbb{E}(X^2) - 2\cdot\mathbb{E}(X)\,\mathbb{E}(X) + \underbrace{\mathbb{E}\big(\mathbb{E}(X)^2\big)}_{\mathbb{E}(X)^2 \cdot \underbrace{\int_\Omega 1 \, d\mathbb{P}}_{\mathbb{P}(\Omega) = 1}}$$

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

We need to assume that $\mathbb{E}(X^2) = \int_\Omega X^2 \, d\mathbb{P}$ exists

$$\overset{\text{change-of-variables}}{=} \begin{cases} \int_{X(\Omega)} x^2 \cdot f_X(x)\, dx & \text{continuous case} \\ \sum_{x \in X(\Omega)} x^2 \cdot p_x & \text{discrete case} \end{cases}$$

Examples:

(a) $X \sim \text{Uniform}(\{x_1, x_2, \dots, x_n\})$ discrete case with $\mathbb{P}_X(\{x_i\}) = \frac{1}{n}$

$$\mathbb{E}(X) = \int_\Omega X \, d\mathbb{P} = \sum_{j=1}^{n} x_j \, \mathbb{P}_X(\{x_j\}) = \frac{1}{n} \underbrace{\sum_{j=1}^{n} x_j}_{=\,\bar{x}}$$

arithmetic mean

$$Var(X) = \int_\Omega \big(X - \underbrace{\mathbb{E}(X)}_{\bar{x}}\big)^2 \, d\mathbb{P} = \sum_{j=1}^{n} (x_j - \bar{x})^2 \cdot \mathbb{P}_X(\{x_j\})$$

$$= \frac{1}{n} \cdot \sum_{j=1}^{n} (x_j - \bar{x})^2$$

(b) $X \sim \text{Exp}(\lambda)$ (exponential distribution) $\mathbb{E}(X) = \frac{1}{\lambda}$

$$\mathbb{E}(X^2) = \int_\Omega X^2 \, d\mathbb{P} = \int_{\mathbb{R}} x^2 \cdot f_X(x) \, dx$$

$$\boxed{f_X(x) = \begin{cases} \lambda e^{-\lambda \cdot x} & , x > 0 \\ 0 & , x \le 0 \end{cases}}$$

$$= \int_0^\infty x^2 \cdot \lambda e^{-\lambda \cdot x} \, dx \overset{\text{integration by parts}}{=} \frac{2}{\lambda^2}$$

$$Var(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{1}{\lambda^2}$$

Probability Theory – Part 17

standard deviation $= \sqrt{\text{variance}}$

<u>Definition:</u> $(\Omega, \mathcal{A}, \mathbb{P})$ probability space, $X : \Omega \longrightarrow \mathbb{R}$ random variable, where $\int_\Omega X^2 \, d\mathbb{P}$ exists. Then:

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

is called the standard deviation of X.

$$\sigma(X) = \sqrt{\mathbb{E}(X^2) - \mathbb{E}(X)^2}$$

<u>Examples:</u> (a) $X \sim \mathrm{Uniform}(\{x_1, x_2, \ldots, x_n\})$ discrete case with $\mathbb{P}_X(\{x_i\}) = \frac{1}{n}$

$$\sigma(X) = \sqrt{\frac{1}{n} \cdot \sum_{j=1}^{n} (x_j - \bar{x})^2}$$

(b) $X \sim \mathrm{Normal}(\mu, \sigma^2)$ continuous case with pdf

$$f_X(x) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{-\frac{1}{2} \cdot \left(\frac{x-\mu}{\sigma}\right)^2}$$

$$\mathbb{E}(X) = \mu$$

$$\sigma(X) = \sigma$$

BECOME A MEMBER

ON STEADY

The Bright Side of Mathematics

Probability Theory – Part 18

Properties of variance and standard deviation:

Let X, Y be independent random variables where $\mathbb{E}(X^2)$ and $\mathbb{E}(Y^2)$ exist.

Then: (a) $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$

(b) $\mathrm{Var}(\lambda \cdot X) = \lambda^2 \cdot \mathrm{Var}(X)$ for every $\lambda \in \mathbb{R}$

(c) $\sigma(\lambda X) = |\lambda| \cdot \sigma(X)$ for every $\lambda \in \mathbb{R}$

Proof: (a)

$$\mathrm{Var}(X+Y) = \mathbb{E}\big((X+Y)^2\big) - \mathbb{E}(X+Y)^2$$

$$= \mathbb{E}(X^2 + 2XY + Y^2) - \big(\mathbb{E}(X) + \mathbb{E}(Y)\big)^2$$

$$= \mathbb{E}(X^2) + 2\,\mathbb{E}(XY) + \mathbb{E}(Y^2) - \mathbb{E}(X)^2 - 2\,\mathbb{E}(X)\mathbb{E}(Y) - \mathbb{E}(Y)^2$$

$$= \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\cdot\big(\underbrace{\mathbb{E}(XY)}_{= \mathbb{E}(X)\mathbb{E}(Y)} - \mathbb{E}(X)\mathbb{E}(Y)\big)$$

independence

(b)

$$\mathrm{Var}(\lambda \cdot X) = \mathbb{E}\big((\lambda \cdot X)^2\big) - \mathbb{E}(\lambda \cdot X)^2$$

$$= \lambda^2\, \mathbb{E}\big((X)^2\big) - \lambda^2\, \mathbb{E}(X)^2 = \lambda^2 \cdot \big(\mathbb{E}(X^2) - \mathbb{E}(X)^2\big)$$

$$= \lambda^2 \cdot \mathrm{Var}(X)$$

(c)

$$\sigma(\lambda X) = \sqrt{\mathrm{Var}(\lambda X)} \overset{(b)}{=} |\lambda| \cdot \sigma(X)$$

The Bright Side of
Mathematics

Probability Theory – Part 19

Definition: $(\Omega, \mathcal{A}, \mathbb{P})$ probability space, $X, Y : \Omega \longrightarrow \mathbb{R}$ random variables $\left(\mathbb{E}(X^2), \mathbb{E}(Y^2) \text{ are finite} \right)$

$$\mathrm{Cov}(X,Y) := \mathbb{E}\Big(\big(X - \mathbb{E}(X)\big) \cdot \big(Y - \mathbb{E}(Y)\big) \Big)$$

$$= \mathbb{E}\Big(X Y - X \cdot \mathbb{E}(Y) - Y \,\mathbb{E}(X) + \mathbb{E}(X)\,\mathbb{E}(Y) \Big)$$

$$\overset{\text{linearity}}{=} \mathbb{E}(XY) - 2 \cdot \mathbb{E}(Y)\,\mathbb{E}(X) + \mathbb{E}(X)\,\mathbb{E}(Y)$$

$$= \mathbb{E}(XY) - \mathbb{E}(Y)\,\mathbb{E}(X)$$

is called the covariance of X and Y.

Remember: X, Y independent $\Longrightarrow$ $\mathrm{Cov}(X,Y) = 0$ (X, Y uncorrelated)

only in special situations
(for example: X, Y normally distributed)

Property: $\mathrm{Cov}(X,Y)^2 \leq \mathrm{Cov}(X,X)\,\mathrm{Cov}(Y,Y)$

Definition: $\rho_{X,Y} := \dfrac{\mathrm{Cov}(X,Y)}{\sigma(X)\,\sigma(Y)} \in [-1, 1]$ correlation coefficient

Example: $\Omega = \{a, b, c\}$, $\mathbb{P}$ uniform on Ω $\left(\mathbb{P}(\{a\}) = \mathbb{P}(\{b\}) = \mathbb{P}(\{c\}) = \frac{1}{3} \right)$

$X, Y : \Omega \longrightarrow \mathbb{R}$,

$$X(a) = 1 \qquad X(b) = 0 \qquad X(c) = -1$$
$$Y(a) = 0 \qquad Y(b) = 1 \qquad Y(c) = 0$$

$$\Rightarrow X \cdot Y = 0, \quad \mathbb{E}(X) = 0 \quad \Rightarrow \quad \mathrm{Cov}(X,Y) = 0$$

Independence? $\mathbb{P}(X \leq x, Y \leq y) = \mathbb{P}(X \leq x) \cdot \mathbb{P}(Y \leq y)$ for all x, y

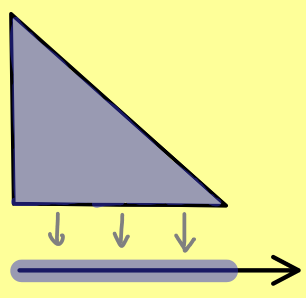
$x = -1$, $y = 0$: $\quad \mathbb{P}(\{c\}) = \mathbb{P}(\{c\}) \cdot \mathbb{P}(\{a, c\})$ ↯

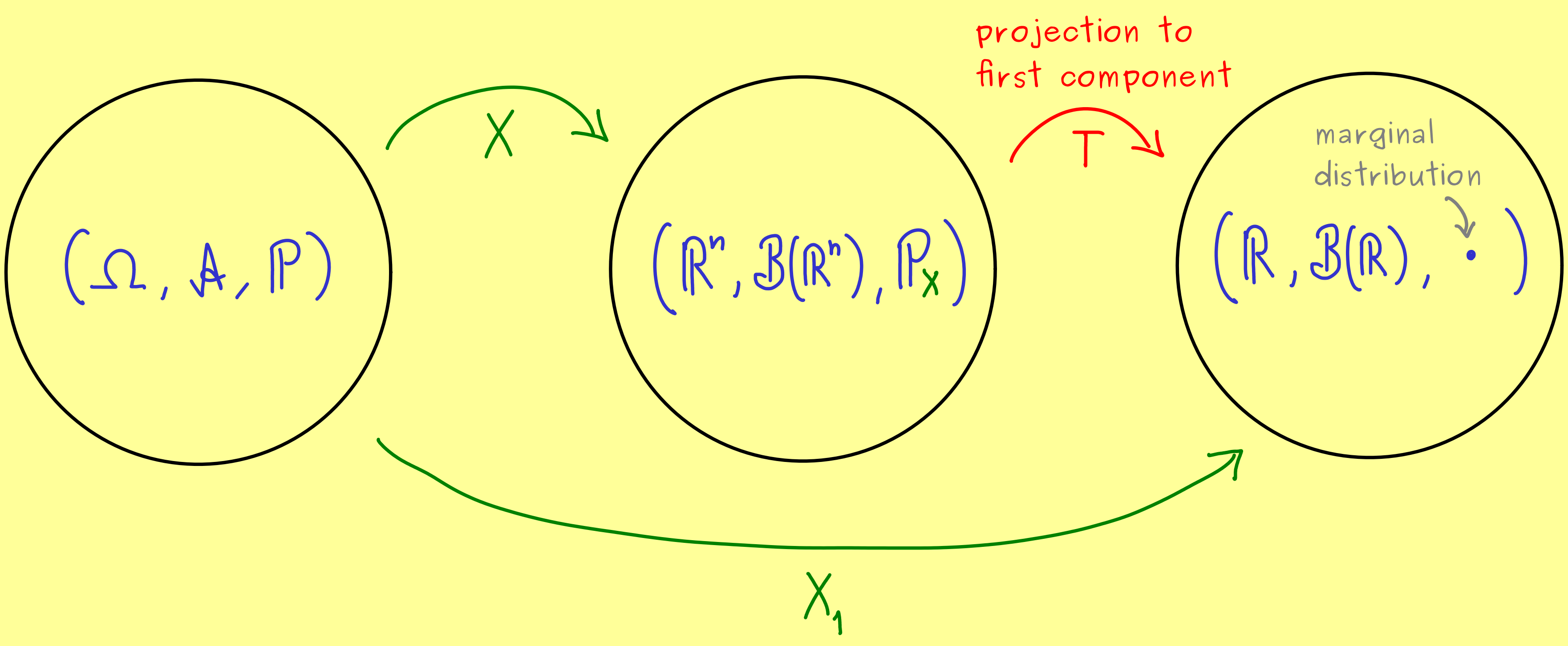
The Bright Side of Mathematics

Probability Theory – Part 20

$X: \Omega \longrightarrow \mathbb{R}^n$ random vector

$\Longrightarrow X_1: \Omega \longrightarrow \mathbb{R}$ random variable





Definition: $\mathbb{P}_{X_1} = (\mathbb{P}_X)_T$ is called the marginal distribution of X with respect to the first component.

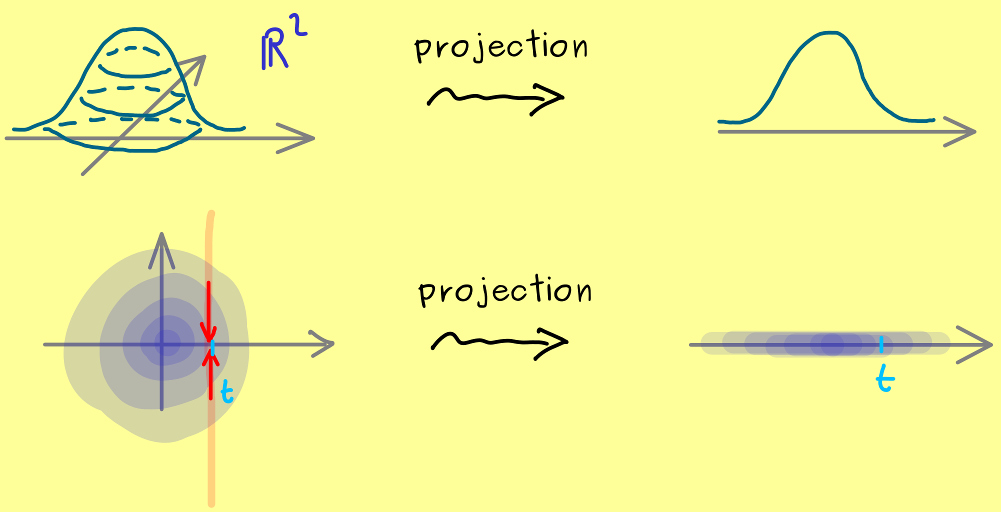
$$F_{X_1}(t) = \mathbb{P}_{X_1}((-\infty, t]) \quad \text{marginal cumulative distribution function}$$

$$= \mathbb{P}_X\big((-\infty, t] \times \mathbb{R} \times \cdots \times \mathbb{R}\big)$$

$$= \mathbb{P}\big(X_1 \leq t,\ X_2 \in \mathbb{R}, \ldots, X_n \in \mathbb{R}\big)$$

Two important cases:

(1) (abs.) continuous: $\mathbb{P}_X$ has a probability density function $f_X: \mathbb{R}^n \to \mathbb{R}$



$$f_{X_1}(t) = \int_{\mathbb{R}^{n-1}} f_X(t, x_2, x_3, \ldots, x_n)\, d(x_2, \ldots, x_n) \quad \text{marginal probability density function}$$

(2) discrete: $\mathbb{P}_X$ has a probability mass function $(p_x)_{x \in \mathbb{R}^n}$

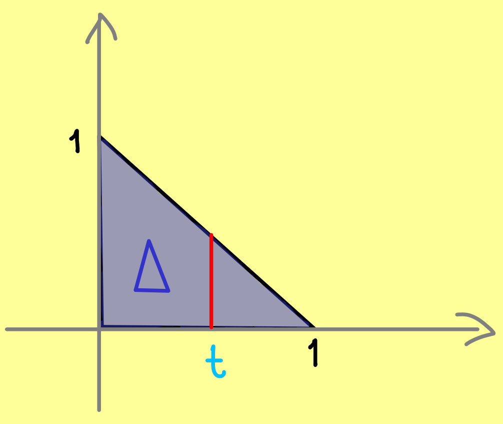
(only countably many are non-zero)

marginal probability mass function $(p_t)_{t \in \mathbb{R}}$ with

$$p_t = \sum_{\substack{x_2, x_3, \ldots \\ \in \mathbb{R}}} p(t, x_2, x_3, \ldots, x_n)$$

Example: $X: \Omega \longrightarrow \mathbb{R}^2$ uniformly distributed on Δ

$$f_X(x_1, x_2) = \begin{cases} 2, & (x_1, x_2) \in \Delta \\ 0, & (x_1, x_2) \notin \Delta \end{cases}$$



marginal probability density function $f_{X_1}(t) = \int_{-\infty}^{\infty} f_X(t, x_2)\, dx_2$

$$= \begin{cases} \int_0^{1-t} 2\, dx_2 & t \in [0,1] \\ 0, & t \notin [0,1] \end{cases}$$

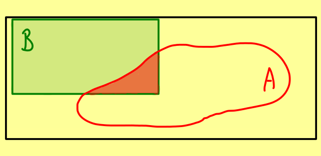
$$= \begin{cases} 2 - 2t, & t \in [0,1] \\ 0, & t \notin [0,1] \end{cases}$$

Probability Theory – Part 21

conditional probability:

$\mathbb{P}(\cdot \mid B) : A \mapsto \mathbb{P}(A \mid B)$

is probability measure $(\mathbb{P}(B) > 0)$



Definition: $(\Omega, \mathcal{A}, \mathbb{P})$ probability space, $B \in \mathcal{A}$ with $\mathbb{P}(B) > 0$

$\left(\Longrightarrow (\Omega, \mathcal{A}, \mathbb{P}(\cdot \mid B)) \right.$ probability space $\left. \right)$

For a random variable $X : \Omega \longrightarrow \mathbb{R}$, we define:

$$\mathbb{E}(X) = \int_\Omega X \, d\mathbb{P} \qquad \text{(expectation of } X\text{)}$$

$$\mathbb{E}(X \mid B) = \int_\Omega X \, d\mathbb{P}(\cdot \mid B) \qquad \text{(conditional expectation of } X \text{ given } B\text{)}$$

Remember:

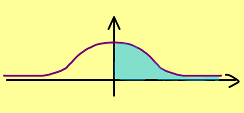
$$\mathbb{E}(X \mid B) = \frac{1}{\mathbb{P}(B)} \int_\Omega X \mathbb{1}_B \, d\mathbb{P} = \frac{1}{\mathbb{P}(B)} \mathbb{E}(\mathbb{1}_B X)$$

$$\boxed{\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{1}{\mathbb{P}(B)} \int_A \mathbb{1}_B \, d\mathbb{P}}$$

indicator function: $\mathbb{1}_B(\omega) = \begin{cases} 1, & \omega \in B \\ 0, & \omega \notin B \end{cases}$

Example: $X \sim \text{NORMAL}(0, 1^2)$, $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$,

$B = \{X > 0\}$



$$\mathbb{E}(X \mid B) = \frac{1}{\mathbb{P}(B)} \int_\Omega \underbrace{X(\omega)}_{x} \mathbb{1}_B(\omega) \, d\mathbb{P}(\omega) = \frac{1}{\mathbb{P}(B)} \cdot \int_\mathbb{R} x \underbrace{\mathbb{1}_B(X^{-1}(x))}_{= \begin{cases} 1, & x > 0 \\ 0, & x \le 0 \end{cases}} f_X(x) \, dx$$

$$= \frac{1}{\mathbb{P}(B)} \cdot \int_0^\infty x f_X(x) \, dx = 2 \cdot \frac{1}{\sqrt{2\pi}} \int_0^\infty x e^{-\frac{1}{2}x^2} dx = \frac{2}{\sqrt{2\pi}} \underbrace{\left(-e^{-\frac{x^2}{2}} \Big|_0^\infty \right)}_{=1}$$

General example: $\mathbb{E}(\mathbb{1}_A \mid B) = \int_\Omega \mathbb{1}_A \, d\mathbb{P}(\cdot \mid B) = \int_A d\mathbb{P}(\cdot \mid B) = \mathbb{P}(A \mid B)$

Example: Throw one die: $X : \Omega \to \mathbb{R}$, $B = \{X = 5, X = 6\}$

$$\mathbb{E}(X \mid B) = \frac{1}{\mathbb{P}(B)} \cdot \int_B X \, d\mathbb{P} = \frac{1}{\mathbb{P}(B)} \sum_{x = 5,6} x \cdot \mathbb{P}(X = x)$$

$$= \frac{1}{\frac{2}{6}} \cdot \left(5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} \right) = \frac{11}{2} = 5.5$$

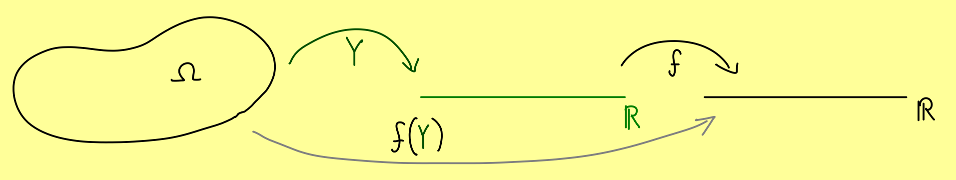
Probability Theory – Part 22

Recall: $X: \Omega \longrightarrow \mathbb{R}$ discrete, B event with $\mathbb{P}(B) > 0$

$$\mathbb{E}(X \mid B) = \int_\Omega X \, d\mathbb{P}(\cdot \mid B) = \sum_x x \cdot \mathbb{P}(X = x \mid B)$$

Consider $Y: \Omega \longrightarrow \mathbb{R}$ discrete, $B = \{Y = y\}$.

Define: $$f(y) := \mathbb{E}(X \mid Y = y) = \sum_x x \, \frac{\overbrace{\mathbb{P}(X = x \text{ and } Y = y)}^{\text{joint pmf of } X \text{ and } Y}}{\mathbb{P}(Y = y)}$$



$f(Y): \Omega \longrightarrow \mathbb{R}$ is called the conditional expectation of X given Y

and denoted by $\mathbb{E}(X \mid Y)$

Example: die throw, $\Omega = \{1, \dots, 6\}$, $X: \Omega \longrightarrow \mathbb{R}$ checks if number is even

$$X(\omega) = \begin{cases} 1, & \omega \in \{2,4,6\} \\ 0, & \text{else} \end{cases}$$

$Y: \Omega \longrightarrow \mathbb{R}$ checks if number is the highest

$$Y(\omega) = \begin{cases} 1, & \omega = 6 \\ 0, & \text{else} \end{cases}$$

$$\mathbb{E}(X \mid Y)(\omega) = \begin{cases} \mathbb{E}(X \mid Y = 0) = \sum_{x=0,1} x \, \frac{\mathbb{P}(X = x \text{ and } Y = 0)}{\mathbb{P}(Y = 0)} = \frac{\frac{2}{6}}{\frac{5}{6}} = \frac{2}{5}, & \omega \in \{1, \dots, 5\} \\ \mathbb{E}(X \mid Y = 1) = \sum_{x=0,1} x \, \frac{\mathbb{P}(X = x \text{ and } Y = 1)}{\mathbb{P}(Y = 1)} = \frac{\frac{1}{6}}{\frac{1}{6}} = 1, & \omega = 6 \end{cases}$$

Definition for (abs.) continuous case: $(X, Y): \Omega \longrightarrow \mathbb{R}^2$ with pdf $f_{(X,Y)}: \mathbb{R}^2 \to \mathbb{R}$

$$g(y) := \mathbb{E}(X \mid Y = y) = \int_{\mathbb{R}} x \cdot \underbrace{\frac{f_{(X,Y)}(x,y)}{f_Y(y)}}_{\text{conditional density}} \, dx$$

$\mathbb{E}(X \mid Y) = g(Y) = g \circ Y$ is called the conditional expectation of X given Y

Properties: (a) X, Y independent $\Longrightarrow$ $\mathbb{E}(X \mid Y) = \mathbb{E}(X)$ and

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

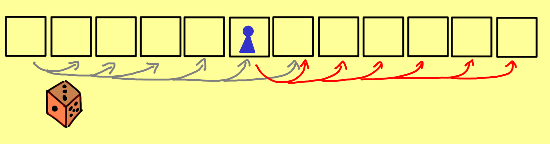
(b) $\mathbb{E}(X \mid X) = X$

(c) $\mathbb{E}\big(\mathbb{E}(X \mid Y)\big) = \mathbb{E}(X)$ (Law of total probability)

Probability Theory – Part 23

Stochastic processes:
- "random variables in a row"
- random experiment with time evolution
  (discrete timesteps, continuous time)

board game:



coin game: toss a coin again and again until two successive heads occur

$$n \in \mathbb{N} \quad X_n : \{H, T\}^{\mathbb{N}} \longrightarrow \{0, 1, 2\}$$

discrete time

no two successive heads in the first n tosses and nth toss is "tails"

no two successive heads in the first n tosses and nth toss is "heads"

two successive heads in the first n tosses

Definition: T set (often $T = \mathbb{N}$, $T = \mathbb{Z}$, $T = \mathbb{R}$).

For each $t \in T$, define: $X_t : \Omega \to \mathbb{R}$ (random variable/ vector)

Then: $(X_t)_{t \in T}$ is called a stochastic process.

For $\omega \in \Omega$: the map $\begin{aligned} T &\to \mathbb{R} \\ t &\mapsto X_t(\omega) \end{aligned}$ is called path.

Example from before:

$$X_n : \{H, T\}^{\mathbb{N}} \longrightarrow \{0, 1, 2\}$$

BECOME A MEMBER

ON STEADY

The Bright Side of
Mathematics

Probability Theory – Part 24

Definition: Let $(X_t)_{t\in T}$ be a stochastic process with $\underbrace{T \subseteq \mathbb{Z}}_{\text{discrete-time}}$ or $\underbrace{T \subseteq \mathbb{R}}_{\text{continuous-time}}$.

We call $(X_t)_{t\in T}$ Markov process or Markov chain if

for all $n \in \mathbb{N}$, $t_1, t_2, \dots, t_n, t \in T$, $t_1 < t_2 < \dots < t_n < t$,

and $x_1, x_2, \dots, x_n, x \in \mathbb{R}$, we have:

$$\mathbb{P}(X_t = x \mid X_{t_1} = x_1, X_{t_2} = x_2, \dots, X_{t_n} = x_n)$$
$$= \mathbb{P}(X_t = x \mid X_{t_n} = x_n)$$

for discrete-time Markov chain:

$$\begin{tikzcd} x_n \arrow[r, bend left] & x \end{tikzcd}$$

depends only on x_n, x, t_n

$$p_{x,y}(k,k+1) = \mathbb{P}(X_{k+1} = y \mid X_k = x)$$

transition probability from x to y at time k

$$\begin{tikzcd} x \arrow[r, bend left] & y \end{tikzcd}$$

time $=k$ time $=k+1$

If $p_{x,y}(k,k+1)$ does not depend on k, then we say:

the Markov chain is time-homogeneous

Example: toss a coin again and again until two successive heads occur

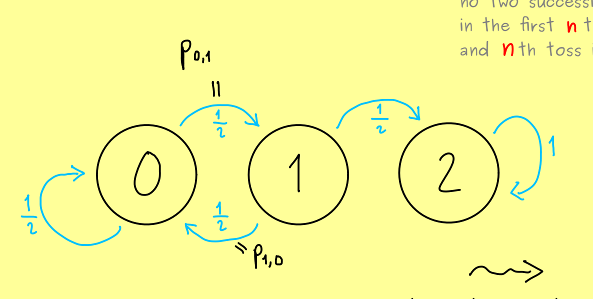
$n \in \mathbb{N}$ discrete time

$$X_n : \{H, T\}^{\mathbb{N}} \longrightarrow \{0,1,2\}$$

0: no two successive heads in the first n tosses and nth toss is "tails"

1: no two successive heads in the first n tosses and nth toss is "heads"

2: two successive heads in the first n tosses



$\rightsquigarrow$ transition matrix $\mathcal{P} = \begin{pmatrix} p_{0,0} & p_{0,1} & p_{0,2} \\ p_{1,0} & p_{1,1} & p_{1,2} \\ p_{2,0} & p_{2,1} & p_{2,2} \end{pmatrix}$

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

Start the game with $q^0 = (1,0,0) \xrightarrow{\text{one time-step}} q^1 = (\frac{1}{2}, \frac{1}{2}, 0)$

$\xrightarrow{\text{one time-step}} q^2 = (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$

$$q^2 = q^1 \mathcal{P} \quad \text{(vector-matrix-multiplication)}$$

Law of total probability

$$\rightsquigarrow \quad q^n = q^0 \mathcal{P}^n \xrightarrow{n\to\infty} (0,0,1)\ ?$$

The Bright Side of
Mathematics

Probability Theory - Part 25

stochastic process: $(X_t)_{t \in T}$ ← subset of $\mathbb{Z}$ or $\mathbb{R}$

discrete-time Markov chains + time-homogeneous:

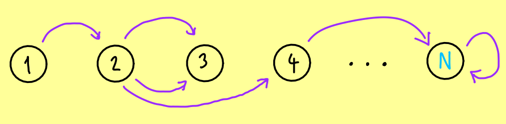
depends only on x and y $\quad x \longrightarrow y$

$$p_{x,y} := \mathbb{P}(X_{k+1} = y \mid X_k = x) \quad \text{independent of } k \in T \subseteq \mathbb{Z}$$

↳ transition matrix $\mathcal{P} = (p_{x,y})_{x,y}$

Important:
- entries of $\mathcal{P}$ lie in $[0,1]$
- $\mathcal{P}$ acts on row vectors from the right

General example: $X_k : \Omega \to \{1,2,\dots,N\}$



start at $k=0$: probability mass function of X_0 (pmf of $\mathbb{P}_{X_0}$)

is given by a row vector $q^0 \in \mathbb{R}^{1 \times N}$

$$(q^0)_m = \mathbb{P}(X_0 = m)$$

at $k=1$:

$$(q^1)_m = \mathbb{P}(X_1 = m) = \sum_{i=1}^{N} \mathbb{P}(X_1 = m \mid B_i) \cdot \mathbb{P}(B_i)$$

law of total probability $\quad \bigcup_{i=1}^{N} B_i = \Omega$ ← disjoint union!

$B_i = \{X_0 = i\}$

$$= \sum_{i=1}^{N} \mathbb{P}(B_i) \cdot \mathbb{P}(X_1 = m \mid B_i)$$

$$= \sum_{i=1}^{N} \underbrace{\mathbb{P}(X_0 = i)}_{(q^0)_i} \cdot \underbrace{\mathbb{P}(X_1 = m \mid X_0 = i)}_{p_{i,m}} = (q^0 \mathcal{P})_m$$

by induction: $q^k = q^0 \cdot \mathcal{P}^k$

Definition: $q \in \mathbb{R}^{1 \times N}$ is called a stationary distribution for the Markov chain if

$$q\mathcal{P} = q \qquad \left(\text{and } q_m \in [0,1],\ \sum_m q_m = 1\right)$$

Note: $q\mathcal{P} = q \iff \mathcal{P}^T q^T = q^T \iff \mathcal{P}^T q^T = 1 \cdot q^T$

column vector $\qquad$ eigenvalue $\qquad$ eigenvector

Example:

$$\mathcal{P} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 1 \end{pmatrix} \rightsquigarrow \mathrm{Ker}(\mathcal{P}^T - 1 \cdot \mathbb{1}) = \mathrm{Ker}\begin{pmatrix} -\frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & -1 & 0 \\ 0 & \frac{1}{2} & 0 \end{pmatrix}$$

row operations

$$= \mathrm{Ker}\begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \mathrm{Span}\left(\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}\right)$$

$\Longrightarrow$ only stationary distribution $q = (0,0,1)$

Probability Theory - Part 26

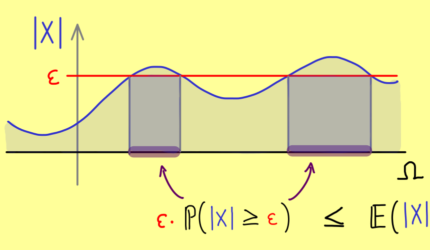
$(\Omega, \mathcal{A}, \mathbb{P})$ probability space

Markov's inequality: $X: \Omega \longrightarrow \mathbb{R}$ random variable.

Then $|X|: \Omega \longrightarrow [0, \infty)$ satisfies:

$$\mathbb{P}(|X| \geq \varepsilon) \leq \frac{\mathbb{E}(|X|^p)}{\varepsilon^p} \qquad \text{for any } \varepsilon > 0, \; p > 0$$

picture for $p = 1$:



$$\varepsilon \cdot \mathbb{P}(|X| \geq \varepsilon) \leq \mathbb{E}(|X|)$$

Proof:

We have: $|X(\omega)| \geq \varepsilon \iff |X(\omega)|^p \geq \varepsilon^p$

indicator function

And: $\varepsilon^p \, \mathbb{P}(|X| \geq \varepsilon) = \varepsilon^p \cdot \mathbb{P}(|X|^p \geq \varepsilon^p) = \varepsilon^p \cdot \mathbb{E}(\mathbb{1}_{\{|X|^p \geq \varepsilon^p\}})$

$$= \mathbb{E}(\varepsilon^p \cdot \mathbb{1}_{\{|X|^p \geq \varepsilon^p\}}) \leq \mathbb{E}(|X|^p) \qquad \square$$

Chebyshev's inequality: $X: \Omega \longrightarrow \mathbb{R}$ random variable where $\mathbb{E}(|X|) < \infty$.

Then: $$\mathbb{P}(|X - \mathbb{E}(X)| \geq \varepsilon) \leq \frac{\mathrm{Var}(X)}{\varepsilon^2} \qquad \text{for any } \varepsilon > 0.$$

Proof: Define: $\tilde{X} := X - \mathbb{E}(X)$. Hence: $\mathrm{Var}(X) = \mathrm{Var}(\tilde{X}) = \mathbb{E}(\tilde{X}^2)$

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq \varepsilon) = \mathbb{P}(|\tilde{X}| \geq \varepsilon) \leq \frac{\mathbb{E}(|\tilde{X}|^2)}{\varepsilon^2} = \frac{\mathrm{Var}(X)}{\varepsilon^2}$$

Markov's inequality for $p = 2$ $\qquad \square$

BECOME A MEMBER

ON STEADY

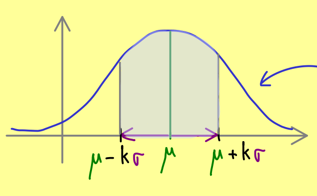
The Bright Side of Mathematics

Probability Theory - Part 27

Assumption: $X : \Omega \to \mathbb{R}$ random variable with

$$\mu := \mathbb{E}(X)$$
$$\sigma := \sqrt{\mathrm{Var}(X)}$$

both should exist!



probability density function

$k = 1, 2, 3, \dots$

Chebyshev's inequality: $$\mathbb{P}\big(|X - \mathbb{E}(X)| \geq \varepsilon\big) \leq \frac{\mathrm{Var}(X)}{\varepsilon^2}$$

$k\sigma$-intervals:
$$\mathbb{P}\big(X \in [\mu - k\sigma, \mu + k\sigma]\big) = \mathbb{P}\big(|X-\mu| \leq k\sigma\big)$$
$$\geq \mathbb{P}\big(|X-\mu| < k\sigma\big)$$
$$= 1 - \mathbb{P}\big(|X-\mu| \geq k\sigma\big)$$

Chebyshev's inequality

$$\geq 1 - \frac{\mathrm{Var}(X)}{k^2\sigma^2} = 1 - \frac{1}{k^2}$$

For $k = 2$: $\mathbb{P}\big(X \in [\mu - 2\sigma, \mu + 2\sigma]\big) \geq 75\%$

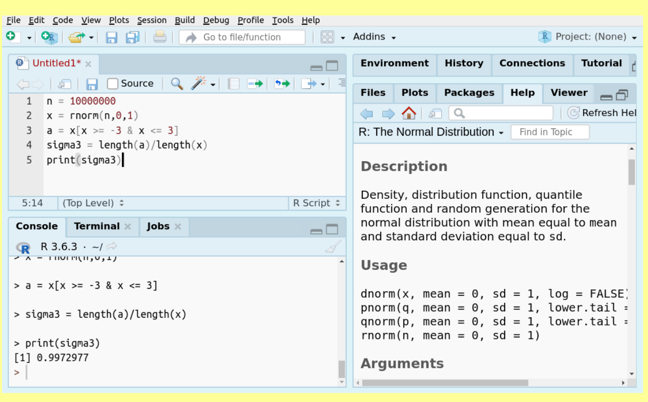
For $k = 3$: $\mathbb{P}\big(X \in [\mu - 3\sigma, \mu + 3\sigma]\big) \geq \frac{8}{9} \geq 88.8\%$

$k\sigma$-intervals for the normal distribution: $\mu = 0$, $\sigma = 1$

$$\mathbb{P}\big(X \in [\mu - 1\sigma, \mu + 1\sigma]\big) \approx 0.682\dots$$

$$\mathbb{P}\big(X \in [\mu - 2\sigma, \mu + 2\sigma]\big) \approx 0.954\dots$$

$$\mathbb{P}\big(X \in [\mu - 3\sigma, \mu + 3\sigma]\big) \approx 0.997\dots$$



The Bright Side of
Mathematics

Probability Theory - Part 28

(theoretical) probability of event $\mathbb{P}(A)$ $\xleftrightarrow{\text{law of large numbers}}$ empirical probability of event $\dfrac{\text{number of outcomes in } A}{\text{total number}}$ = relative frequency of the event

Example: coin toss: $\Omega_0 = \{H, T\}$, $\mathbb{P}_0(\{H\}) = \mathbb{P}_0(\{T\}) = \frac{1}{2}$

repeat random experiment: $\Omega = \Omega_0 \times \Omega_0 \times \cdots$

$\mathbb{P}$ = product measure

(H in kth toss)

define random variables: $X_k : \Omega \longrightarrow \mathbb{R}$, $X_k(\omega) = \begin{cases} 1 , & \omega_k = H \\ 0 , & \omega_k = T \end{cases}$

let's look at n tosses: $\overline{X}_n := \frac{1}{n} \sum_{k=1}^{n} X_k : \Omega \longrightarrow \mathbb{R}$

(relative frequency of heads in the first n tosses)

we expect: $\overline{X}_n \xrightarrow{n \to \infty} \frac{1}{2}$ What does this convergence mean?

Weak law of large numbers: $X_k : \Omega \longrightarrow \mathbb{R}$ random variables.

Let $(X_k)_{k \in \mathbb{N}}$ be independent and identically distributed (= i.i.d.)

$\left[\mathbb{P}\left((X_j \leq x_j)_{j \in J} \right) = \prod_{j \in J} \mathbb{P}(X_j \leq x_j) \quad \text{for all } x_j \in \mathbb{R} \text{, for all finite } \emptyset \neq J \subseteq \mathbb{N} \right]$ $\left[\mathbb{P}_{X_k}(\mathcal{B}) = \mathbb{P}_{X_1}(\mathcal{B}) \quad \text{for all } k \in \mathbb{N} \text{, for all Borel sets } \mathcal{B} \subseteq \mathbb{R} \right]$

and $\mathbb{E}(|X_1|) < \infty$.

Then for $\mu := \mathbb{E}(X_1)$ and for all $\varepsilon > 0$:

$$\mathbb{P}\left(\left| \frac{1}{n} \sum_{k=1}^{n} X_k - \mu \right| \geq \varepsilon \right) \xrightarrow{n \to \infty} 0$$

We say $\overline{X}_n := \frac{1}{n} \sum_{k=1}^{n} X_k$ converges in probability to the expected value μ.

Proof: for the case: $\sigma^2 = \mathrm{Var}(X_1) < \infty$

We have: $\mathbb{E}(\overline{X}_n) = \mathbb{E}\left(\frac{1}{n} \sum_{k=1}^{n} X_k \right) = \frac{1}{n} \sum_{k=1}^{n} \mathbb{E}(X_k) = \mu$

$\mathrm{Var}(\overline{X}_n) = \mathrm{Var}\left(\frac{1}{n} \sum_{k=1}^{n} X_k \right) = \frac{1}{n^2} \sum_{k=1}^{n} \mathrm{Var}(X_k) = \frac{\sigma^2}{n}$

By Chebyshev's inequality:

$$\mathbb{P}\left(|\overline{X}_n - \underbrace{\mathbb{E}(\overline{X}_n)}_{\mu}| \geq \varepsilon \right) \leq \frac{\mathrm{Var}(\overline{X}_n)}{\varepsilon^2} \quad \text{for any } \varepsilon > 0.$$

$$= \frac{\sigma^2}{\varepsilon^2} \cdot \frac{1}{n} \xrightarrow{n \to \infty} 0 \qquad \square$$

BECOME A MEMBER

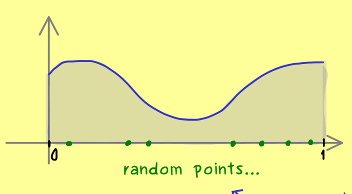
ON STEADY

The Bright Side of Mathematics

Probability Theory - Part 29

law of large numbers: n repetitions $\underbrace{X_1, X_2, \dots, X_n : \frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{n\to\infty} \mathbb{E}(X)}_{\text{"Monte Carlo method"}}$

Monte Carlo integration:



random variables: $X_1, X_2, \dots$

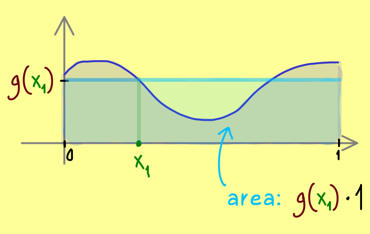
Weak law of large numbers: $Y_k : \Omega \longrightarrow \mathbb{R}$ random variables.

Let $(Y_k)_{k\in\mathbb{N}}$ be i.i.d. and $\mathbb{E}(|Y_1|) < \infty$.

Then for $\mu := \mathbb{E}(Y_1)$ and for all $\varepsilon > 0$: $\mathbb{P}\left(\left| \frac{1}{n}\sum_{k=1}^{n} Y_k - \mu \right| \geq \varepsilon \right) \xrightarrow{n\to\infty} 0$

Monte Carlo integration: Given: $g : [0,1] \longrightarrow [-c,c]$ integrable, $c > 0$.

We want: $\int_0^1 g(x)\, dx$



Take: X_1 picks a point (randomly = uniformly distributed) from the interval $[0,1]$: $x_1 = X_1(\omega)$

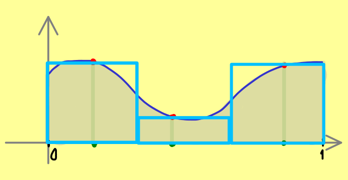
$Y_1 := g(X_1)$ What is $\mathbb{E}(Y_1)$?

$$\mathbb{E}(Y_1) = \mathbb{E}(g(X_1)) \overset{\text{change of variables}}{=} \int_0^1 g(x)\, \underbrace{f_{X_1}(x)}_{=1}\, dx = \int_0^1 g(x)\, dx$$

pdf

procedure: $X_1, X_2, \dots$ i.i.d.+uniformely distributed on $[0,1]$

$\frac{1}{n}\sum_{k=1}^{n} g(X_k)$ approximates $\int_0^1 g(x)\, dx$

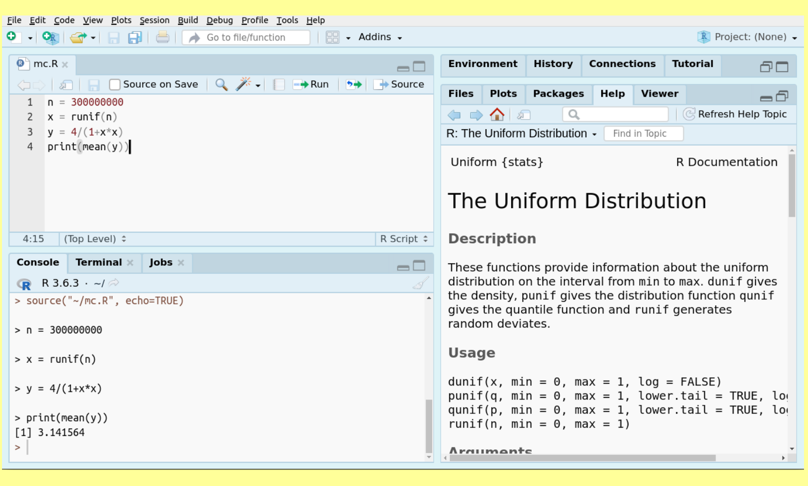


Example:

$$\int_0^1 \frac{4}{1+x^2}\, dx$$

$\parallel$

π



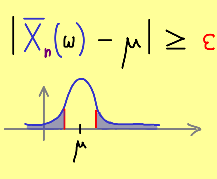
Probability Theory - Part 30

repeating a random experiment: $X_1, X_2, \ldots$ i.i.d. , $\mu := \mathbb{E}(X_1)$

should lead to: $\frac{1}{n}\sum_{k=1}^{n} X_k =: \overline{X}_n \xrightarrow{n\to\infty} \mu$

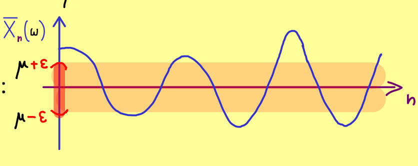
weak law of large numbers: $|\overline{X}_n(\omega) - \mu| \geq \varepsilon$ is unlikely for large n

$$\mathbb{P}\big(\{\omega \in \Omega \mid |\overline{X}_n(\omega) - \mu| \geq \varepsilon\}\big) \xrightarrow{n\to\infty} 0$$



pointwise convergence? $\overline{X}_n(\omega) \xrightarrow{n\to\infty} \mu$?

we could have:



How many $\omega \in \Omega$ have such "bad" behaviour?

Strong law of large numbers: $X_k : \Omega \longrightarrow \mathbb{R}$ random variables.

Let $(X_k)_{k\in\mathbb{N}}$ be i.i.d. and $\mathbb{E}(|X_1|) < \infty$.

Then for $\mu := \mathbb{E}(X_1)$: $\frac{1}{n}\sum_{k=1}^{n} X_k(\omega) =: \overline{X}_n(\omega) \xrightarrow{n\to\infty} \mu$ for $\omega \in \Omega$ almost surely

This means: $\mathbb{P}\big(\{\omega \in \Omega \mid \overline{X}_n(\omega) \xrightarrow{n\to\infty} \mu\}\big) = 1$

(we could have $\overline{X}_n(\omega) \overset{n\to\infty}{\not\longrightarrow} \mu$ but the probability is zero)

Remark: almost sure convergence $\Longrightarrow$ convergence in probability

strong law of large numbers $\Longrightarrow$ weak law of large numbers

The Bright Side of Mathematics

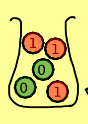
Probability Theory - Part 31

Assumptions of the central limit theorem: $(X_k)_{k \in \mathbb{N}}$ i.i.d. with $\mathrm{Var}(X_1) < \infty$.

part 28

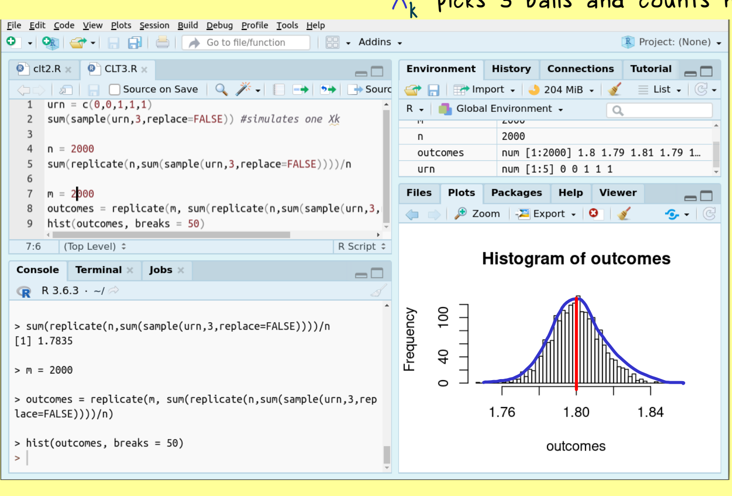
$$\Longrightarrow \quad \overline{X}_n := \frac{1}{n}\sum_{k=1}^{n} X_k \quad \text{satisfies} \quad \mathbb{E}(\overline{X}_n) = \mathbb{E}(X_1), \quad \mathrm{Var}(\overline{X}_n) = \frac{\mathrm{Var}(X_1)}{n}$$

Example:



urn model without replacement = hypergeometric distribution (part 6)

X_k picks 3 balls and counts numbers of 1



$$\mathbb{E}(X_1) = \frac{9}{5} = 1.8$$

$$\overline{X}_n := \frac{1}{n}\sum_{k=1}^{n} X_k$$

What is the distribution?

close to normal distribution!

$$\mathrm{Var}(\overline{X}_n) = \frac{\mathrm{Var}(X_1)}{n}$$

Standardize the random variable: $\mu := \mathbb{E}(X_1)$, $\sigma := \sqrt{\mathrm{Var}(X_1)}$

(1) expectation should be zero: $\overline{X}_n - \mu$

(2) variance should be one: $(\overline{X}_n - \mu) / \left(\frac{\sigma}{\sqrt{n}}\right)$

Central limit theorem: For $(X_k)_{k \in \mathbb{N}}$ i.i.d. with $\mathrm{Var}(X_1) < \infty$, define:

$$Y_n := \left(\frac{1}{n}\sum_{k=1}^{n} X_k - \mu\right) \cdot \left(\frac{\sigma}{\sqrt{n}}\right)^{-1} \quad \text{where } \mu := \mathbb{E}(X_1), \quad \sigma := \sqrt{\mathrm{Var}(X_1)}$$

Then the cdf of Y_n converges to the cdf of $\mathrm{Normal}(0, 1^2)$:

$$\mathbb{P}(Y_n \leq x) \xrightarrow{n \to \infty} \Phi(x) \quad \text{for every } x \in \mathbb{R}$$

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}t^2} \, dt$$

Probability Theory - Part 32

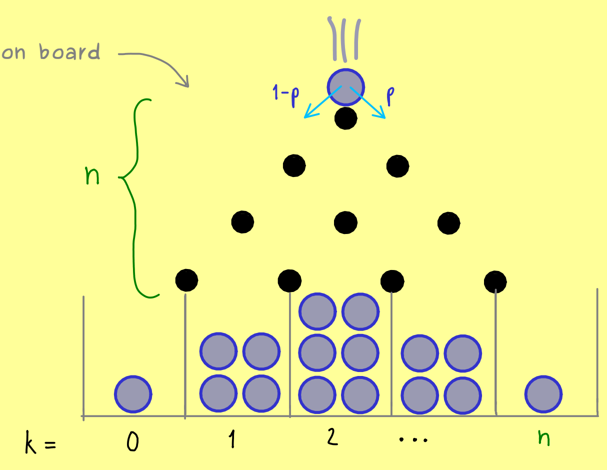
De Moivre-Laplace theorem (special case of central limit theorem)

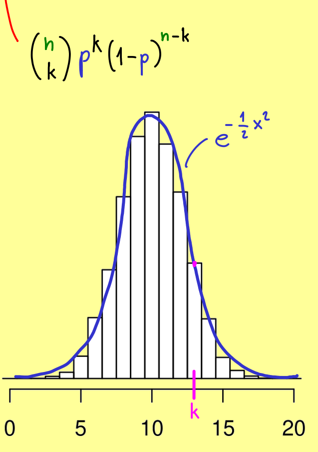
approximation of binomial distribution $\mathrm{Bin}(n, p)$

n times Bernoulli(p)

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

Galton board



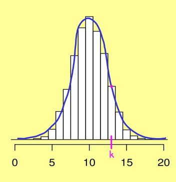


Result: For large n and k close to np, we get:

$$\binom{n}{k} p^k (1-p)^{n-k} \approx \frac{1}{\sqrt{2\pi n p(1-p)}} \, e^{-\frac{(k-np)^2}{2np(1-p)}}$$

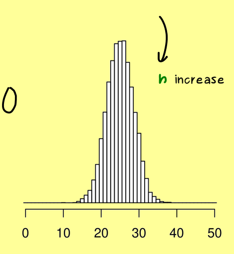
pmf of binomial distribution

pdf of normal distribution



De Moivre-Laplace theorem: Let $p \in (0,1)$. Then for any $C > 0$:

$$\max_{\substack{0 \leq k \leq n \\ \text{with} \\ \frac{(k-np)^2}{2np(1-p)} \leq C}} \left| \frac{\sqrt{np(1-p)} \binom{n}{k} p^k (1-p)^{n-k}}{\frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{(k-np)^2}{2np(1-p)}}} - 1 \right| \xrightarrow{n \to \infty} 0$$

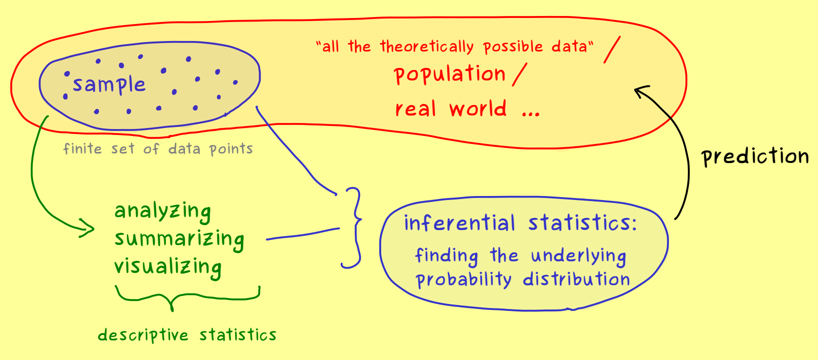


The Bright Side of Mathematics

Probability Theory - Part 33

Statistics | descriptive statistics | inferential statistics

The rough idea:



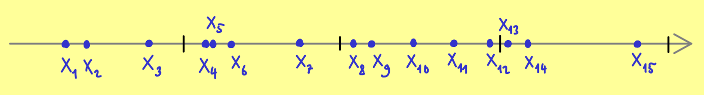
Definition: sample in $\mathbb{R}^d$: $(x_1, x_2, x_3, \dots, x_n)$ with $x_j \in \mathbb{R}^d$

$n \in \mathbb{N}$

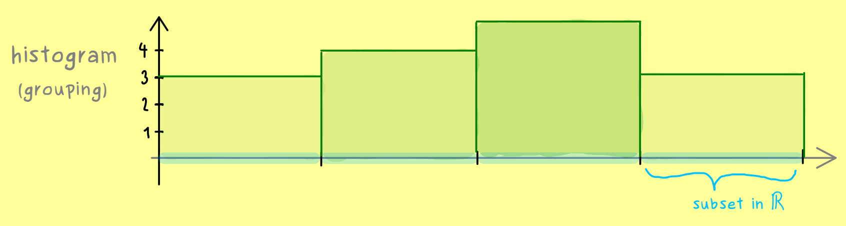
sample in $\mathbb{R}$: $(x_1, x_2, x_3, \dots, x_n)$ with $x_j \in \mathbb{R}$

can be ordered: $x_1 \leq x_2 \leq \dots \leq x_n$

visualization



histogram (grouping)



Definition: For a given sample in $\mathbb{R}$ $(x_1, x_2, x_3, \dots, x_n)$ and subset $A \subseteq \mathbb{R}$,

we define:

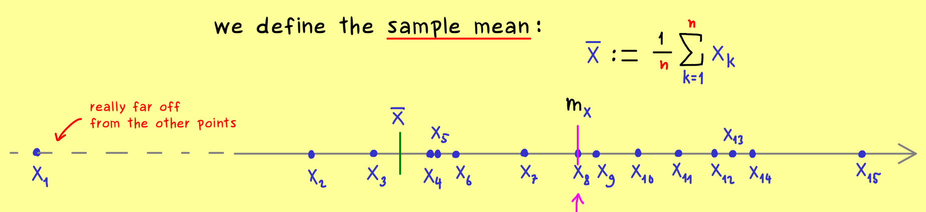
absolute frequency of A $\quad f_{abs}(A) := \#\{k \in \mathbb{N} \mid x_k \in A\}$

relative frequency of A $\quad f_{rel}(A) := \dfrac{\#\{k \in \mathbb{N} \mid x_k \in A\}}{n} \in [0,1]$

Definition: For a given sample in $\mathbb{R}$ $x := (x_1, x_2, x_3, \dots, x_n)$

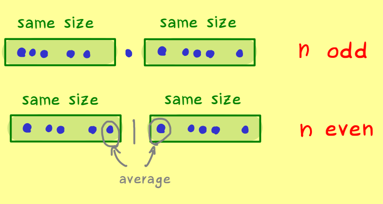
we define the sample mean:

$$\bar{x} := \frac{1}{n} \sum_{k=1}^{n} x_k$$



median: middle point of the sample

$$m_x := \begin{cases} x_{\frac{n+1}{2}}, & n \text{ odd} \\ \frac{1}{2}\left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}\right), & n \text{ even} \end{cases}$$



unbiased sample variance: $\quad s_x^2 := \dfrac{1}{n-1} \sum_{k=1}^{n} (x_k - \bar{x})^2$

makes it unbiased