Statistik Zusammenfassung

random experiment

→ process where the possible outcomes can be identified ahead of time

Sample Space

Set Ω of all possible outcomes of an experiment
(Menge)

Type of sets

Countable set

Ω is countable if a bijection exists
↳ a one-to-one function exists → $f: \Omega \to \mathbb{N}$
⇒ otherwise → uncountable

finite set

set is empty or has a finite number of elements
⇒ otherwise → infinite

Event

collection of possible outcomes of an experiment that is a subset of Ω (including Ω itself)

$$A \subseteq \Omega$$

Set Theory

empty set: $\emptyset$

$A \subseteq B$ → A is part of B

$A \cup B$ → A or B (Vereinigung)

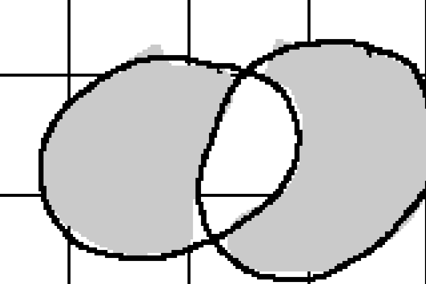
$A \cap B$ → A and B (Gemeinsames)

$A \setminus B$ → A without B $= A - B$

$A^c = A'$ → Complement

$A \triangle B$ → symmetric difference

↳ $(A \setminus B) \cup (B \setminus A)$



Singelton

subset $\{\omega\}$ contains a single outcome ω of Ω

Null set

$\emptyset$ is part of every set (impossible event), $\emptyset \subseteq \Omega$

Universal set

U or Ω (sure event), $\Omega \subseteq \Omega$

Power set

all subsets of a set

↳ $P(A)$ = all possible subsets of A

↳ $A = \{x, y\}$ $P(A) = \{x\}, \{y\}, \{x, y\}, \emptyset$

$|P(A)| = 2^n$

↳ number of Elements of $P(A) = 2^n$ — n ~ number of Elements in A

Properties of set Operations (Theorem)

for any $A, B, C \subset \Omega$:

commutativity:
$$A \cup B = B \cup A$$
$$A \cap B = B \cap A$$

associativity:
$$A \cup (B \cup C) = (A \cup B) \cup C$$
$$A \cap (B \cap C) = (A \cap B) \cap C$$

distributive law:
$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$
$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

De Morgan's law:
$$(A \cup B)^c = A^c \cup B^c$$
$$(A \cap B)^c = A^c \cap B^c$$

Partition

If $A_1, A_2 \dots$ are pairwise disjoint and $\bigcup_i A_i = \Omega$
then the collection $A_1, A_2 \dots$ forms a partition of Ω
↳ collection of unique substets where there union
is the sample space

disjoint

↳ mutually exclusive → $A \cap B = \emptyset$ (nothing in common)

pair wise disjoint → $A_i \cap A_j = \emptyset$ for all $i \neq j$

Field

A field is a set k (körper) with two special elements
$0, 1 \in K$ and two maps $+$ and $\cdot$ ⇒ $(k, \cdot, +)$

$k \times k = k$

⇒ Satisfy 8 axioms (elements written as $\alpha, \beta \dots$)
↳ addition: $\alpha + \beta = \beta + \alpha$

$(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$

$\alpha + 0 = 0 + \alpha = \alpha$

$\forall \alpha \in K \quad \exists -\alpha \in k \rightarrow \alpha + (-\alpha) = 0$

multiplication: $\alpha\beta = \beta\alpha$

$(\alpha\beta)\gamma = \alpha(\beta\gamma)$

$\alpha \cdot 1 = 1 \cdot \alpha = \alpha$

$\forall \alpha \in K \rightarrow \exists \alpha^{-1} \in K \rightarrow \alpha \cdot \alpha^{-1} = 1$

⇒ $(\alpha + \beta) \cdot \gamma = \alpha\gamma + \beta\gamma$

Boolean algebra of sets

collection $\mathcal{G}$ of subsets of Ω with following conditions

1. if $A, B \in \mathcal{G} \Rightarrow A \cup B \in \mathcal{G}$ and $A \cap B \in \mathcal{G}$
   ↳ closed under countable $\cup$ (unions) and $\cap$ (intersections)
2. if $A \in \mathcal{G} \Rightarrow A^c \in \mathcal{G}$
   ↳ closed under complement
3. $\emptyset \in \mathcal{G}$ (same as $\Omega \in \mathcal{G}$ under 2.)

σ - algebra

→ subset of algebra of sets were it only has to be closed under countable unions

1. if $A_1, A_2, \dots, A_n \in \mathcal{F} \Rightarrow \bigcup_{i \in \mathbb{N}} A_i \in \mathcal{F}$
   ↳ closed under countable unions
2. if $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$
   ↳ closed under complement
3. $\emptyset \in \mathcal{F}$ (Same as $\Omega \in \mathcal{F}$ under 2.)

$\Rightarrow$ Smallest σ-algebra $\mathcal{F} = \{\emptyset, \Omega\}$ → trivial algebra

$\Rightarrow$ easiest way to construct with countable Ω
↳ $\mathcal{F} = P(\Omega)$ $\Rightarrow$ with 2^n elements

Borel σ-algebra

→ a σ-algebra with open sets

→ Borel σ-algebra of $\mathbb{R}$ $]-\infty, a]$, $a \in \mathbb{Q}$

Probability measure (Kolmogorov Axioms)

$\mathbb{P}$ on $(\Omega, \mathcal{F})$ function $\mathbb{P}: \mathcal{F} \to [0,1]$

1. $\mathbb{P}(A) \geq 0$ for any $A \in \mathcal{F}$ (any outcome has a possibility)
2. if $A_1, A_2 \ldots \in \mathcal{F}$ are pairwise disjoint, then $\mathbb{P}(\bigcup_{i \in \mathbb{N}} A_i) = \sum_{i \in \mathbb{N}} \mathbb{P}(A_i)$

   ↳ all outcome possibilities sum to 1 → one of the outcomes will always happen
3. $\mathbb{P}(\Omega) = 1$ | $\mathbb{P}(\emptyset) = 0$

   ↳ one of the possible outcomes will always happen

Probability space

→ triple of $(\Omega, \mathcal{F}, \mathbb{P})$

sample space ; all outcomes ; posibility of all outcomes

uniform Probability

↳ all outcomes are eaqually likely

Probability of an Event

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} \Rightarrow \frac{\text{nr of Events}}{\text{nr of possible Events}}$$

counting rules

	order	no order
replacement	n^k	$\binom{n+k-1}{k}$
no replacement	$\frac{n!}{(n-k)!}$	$\binom{n}{k} = \frac{n!}{(n-k)! \cdot k!}$

$\binom{n}{k}$ → combination ⇒ nCr auf TR

$\frac{n!}{(n-k)!}$ → permutation ⇒ nPr auf TR

n = Elements

r = k = samples / unique values

counting rules for groups

total objects n with n_i: number of objects in each

⇒ number of orderings:

$$\frac{n!}{n_1! \, n_2! \cdots n_k!}$$

Conditional Probability

$(A \mid B) \rightarrow$ Probability of A if $\mathbb{P}(B) > 0$

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

Bayes Rule

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

$$\hookrightarrow \mathbb{P}(A) = \sum_{i \in I} \mathbb{P}(A \mid A_i)\,\mathbb{P}(A_i)$$

$$\hookrightarrow \mathbb{P}(A_i \mid B) = \frac{\mathbb{P}(B \mid A_i)\,\mathbb{P}(A_i)}{\sum_{j \in I} \mathbb{P}(B \mid A_j)\,\mathbb{P}(A_j)}$$

Independence of Events

Events A, B are statistically independent if

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

generally: if the intersection of possibilities of events is the same as the product

$\Rightarrow$

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

$\hookrightarrow$

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

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

mutually independent

→ independence of a collection

$\Rightarrow$ collection $\{A_i \in I\}$ is mutually independent if

$\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\,\mathbb{P}(A_j)$ for all $i, j \in I$, $i \neq j$

conditionally independent

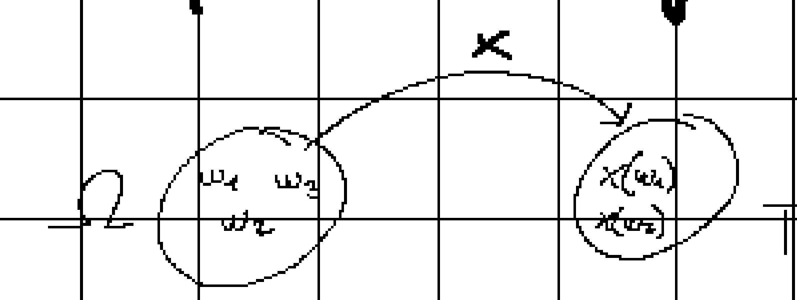
C is an Even with $\mathbb{P}(C) > 0$

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

Random Variable

X can be all the possible values of an experiment

$X : \Omega \to T$



(function that projects the sample space to a meswable space)

Types of random variables

Discrete random variable

if X can only take a finite number or a countable infinite number of different values

Continuous random variables

if X can take uncountable number of different values

→ uncountable many values in T'

(ex: water in a glass → unlimited possibilities)

Comulative Distribution Function (CDF)

function that calculate the probability that the random variable will take a value less then or equal to a cetain value

$\Rightarrow$ the CDF is the sum of all probabilities up to the given point

$$F : \mathbb{R} \to [0,1]$$

$F(x) = \mathbb{P}(X \leq x) = \mathbb{P}(\{\omega \mid X(\omega) \leq x\})$ for all $x \in \mathbb{R}$

(Intervals: $\mathbb{P}(]a,b]) = F(b) - F(a) \Rightarrow \mathbb{P}(X \leq b) - \mathbb{P}(X \leq a)$)

Properties of CDF

- We can compute $\mathbb{P}$ if we know F

F is a CDF of a unique Probability if the following properties hold:

1. right continuity : $\lim\limits_{\Delta x \to 0} F(x + \Delta x) = F(x)$ at every x
   ↳ as Δx gets smaller we approach $F(x)$ more and more

2. Nondecreasing : $F(a) \leq F(b)$ for any $a < b$
   ↳ the probability gets higher, as "past events" get more likely

3. Limits : $\lim\limits_{x \to -\infty} F(x) = 0$ and $\lim\limits_{x \to +\infty} F(x) = 1$
   ↳ 0 → event can't happen | 1 → event will happen

Example CDF

The absolute difference between two rolled dice

$X = |i-j| \quad i,j \in \{1,2,\dots,6\}$

$\Rightarrow$ X can be 0-5

$\hookrightarrow$ calculate possibilities of the outcomes of X

$X=0 \rightarrow 6$ pairs $\rightarrow \frac{6}{36} = \frac{1}{6}$

$X=1 \rightarrow 10$ pairs $\rightarrow \frac{10}{36} = \frac{5}{18}$

$X=2 \rightarrow 8$ pairs $\rightarrow \frac{8}{36} = \frac{2}{9}$

$X=3 \rightarrow 6$ pairs $\rightarrow \frac{6}{36} = \frac{1}{6}$

$X=4 \rightarrow 4$ pairs $\rightarrow \frac{4}{36} = \frac{1}{9}$

$X=5 \rightarrow 2$ pairs $\rightarrow \frac{2}{36} = \frac{1}{18}$

$$\Rightarrow F(x) = \begin{cases} 0 & x < 0 \\ 0 + \frac{1}{6} = \frac{1}{6} & 0 < x < 1 \\ 0 + \frac{1}{6} + \frac{5}{18} = \frac{4}{9} & 1 < x < 2 \\ 0 + \frac{1}{6} + \frac{5}{18} + \frac{2}{9} = \frac{6}{9} & 2 < x < 3 \\ 0 + \frac{1}{6} + \frac{5}{18} + \frac{2}{9} + \frac{1}{6} = \frac{5}{6} & 3 < x < 4 \\ 0 + \frac{1}{6} + \frac{5}{18} + \frac{2}{9} + \frac{1}{6} + \frac{1}{9} = \frac{17}{18} & 4 < x < 5 \\ 0 + \frac{1}{6} + \frac{5}{18} + \frac{2}{9} + \frac{1}{6} + \frac{1}{9} + \frac{1}{18} = 1 & x \geq 5 \end{cases}$$

Probability Mass function (PMF)

for a discrete random Variable

if there exists a non-negative function

$f : \mathbb{R} \to [0,1]$ any $x \in \mathbb{R}$

$$F(x) = \mathbb{P}(X \leq x) = \sum_{x_i \leq x} \mathbb{P}(X = x_i) = \sum_{x_i \leq x} f(x_i)$$

$\Rightarrow$ PMF gives us a list of the possibilities of all possible outcomes

$$f(x) = \mathbb{P}(X = x)$$

$\Rightarrow$ $p_i = \mathbb{P}(X = x_i)$ only positive at $x = x_i$ belonging to T' of X

$$f(x) = \mathbb{P}(X = x) = \begin{cases} p_i, & x = x_i \in T' \\ 0, & \text{else.} \end{cases}$$

Diffeence CDF PMF

→ The CDF gives us the cumulative probablities up to a point while the PMF gives us the probability of each specific outcome

Uniform distribution

→ every outcome has the same probability

$\Rightarrow$ PMF $f(x;N) = \mathbb{P}(X=x \mid N) = \begin{cases} \frac{1}{N}, & x = 1,2,\dots,N \\ 0, & \text{else} \end{cases}$

Hypergeometric distribution

→ calculate the probabilities when picking elements from a set without replacing them

PMF: $f(x;N,M,n) = \mathbb{P}(X=x \mid N,M,n) = \dfrac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$

Example Hypergeometric distribution

balls with two different colors in a bag. calculate the probability of a certain outcome when taking a number of balls at once

$|\Omega| = \underset{\text{all}}{\binom{N}{n}}$, $\underset{\text{red}}{\binom{M}{x}}$, $\underset{\text{blue}}{\binom{N-M}{n-x}}$

$\Rightarrow$ 10 balls, 6 red, 4 blue

pick 3; what is the probability of drawing 2 red?

$\Rightarrow \left.\begin{matrix} N=10 & n=3 \\ M=6 & x=2 \end{matrix}\right\} \mathbb{P}(X=2) = \dfrac{\binom{6}{2}\binom{10-6}{3-2}}{\binom{10}{3}}$

$= 0.5 = \underline{\underline{50\%}}$

Bernoulli distribution

→ simple distribution with only two possible autcome success and failure

$$f(x;p) = \begin{cases} 1-p, & x=0 \\ p, & x=1 \\ 0, & \text{else} \end{cases}$$

p: probability of success

Binomial distribution

→ repetition of an experiment with only two possbe autcomes

PMF: $f(x,p,n) = \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x} \qquad x = 0,1,2,\dots n$

x = number of success
p = Probability of success
n = number of trials

Poisson distribution

→ probability of a certain number of events happening in a fixed interval of time or space

PMF: $$f(x;\lambda) = \begin{cases} \frac{\lambda^x}{x!} \cdot e^{-\lambda}, & \text{for } x = 0,1,2\dots \\ 0, & \text{else} \end{cases}$$

x = number of events we want to find the probability for
λ = Average Rate of event happening

Continuous random variable – CDF & PDF

Probability Density Function (PDF)

→ calculate probabilities of a continuous random variable using Integrals

$$f(x) = \frac{d}{dx} F(x)$$

Comulative Distribution Function (CDF)

→ function to show the probability, that the random variable will take a value less or equal to a certain value.

⇒ can be expressed using the PDF f

$$F(x) = \int_{-\infty}^{x} f(t)\, dt$$

⇒ with $A =]a, b]$

$$\mathbb{P}(X \in A) = \mathbb{P}(a < X \leq b) = F(b) - F(a)$$

$$= \int_{a}^{b} f(x)\, dx = \int_{A} f(x)\, dx$$

Properties of PDF

1. $f(x) \geq 0$, for all $x \in \mathbb{R}$

2. $\int_{-\infty}^{\infty} f(x) \, dx = 1$ → integrates to 1

$\Rightarrow$ any non negative finitely integrable function can be a PDF

A PDF is not able to calculate the probability of a single point as it has no area

$$\Rightarrow \mathbb{P}\left(X = \tfrac{1}{3}\right) = \int_{\frac{1}{3}}^{\frac{1}{3}} f(x) \, dx \doteq 0$$

Uniform Distribution

→ every value in a certain range has the same chance of being picked (ex. random selection)

PDF: $f(x;a,b) = \begin{cases} \frac{1}{b-a} & , \text{if } x \in [a,b] \\ 0 & , \text{else} \end{cases}$

Exponential Distribution

→ calculate the time or distance until some specific event happens. (ex. time for a part to break)

PDF: $f(x;\lambda) = \begin{cases} \lambda \cdot e^{-\lambda x} & , \text{if } x \geq 0 \\ 0, & \text{else} \end{cases}$

→ λ = rate how often an event is expected to happen

⇒ can be "memoryless", if the time passed has no effect on the probability of an event happening

↳ ex: the chance of a bus arriving doesn't change with the amount of time you have been waiting.

Gamma Distribution

→ calculate the time until a multitude of events happen. (ex. how long does it take a call center to receive its 5th call on an average rate of 1 per 10 min?)

PDF:

$$f(x;\alpha,\beta) = \begin{cases} \frac{1}{\Gamma(\alpha)\beta^{\alpha}} \cdot x^{\alpha-1} \cdot e^{-\frac{x}{\beta}} &, \text{if } 0<x<\infty \\ 0 &, \text{else} \end{cases}$$

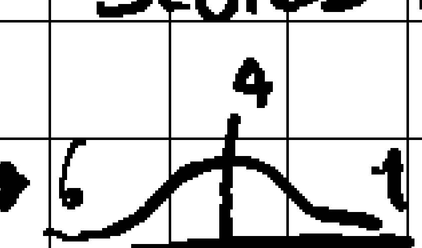
α = shape parameter → number of events we are waiting for (ex. 5 calls → $\alpha = 5$)

β = scale parameter → average time between events (ex. call every 10 min → $\beta = 10$)

$\Gamma(\alpha)$ = Gamma function → $\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1} e^{-t} \, dt$

Normal Distribution

→ describes how a set of continuous data is spread out (ex. test scores of a class → how many were around average etc.

↳ 

)

PDF: $f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{x-\mu}{2\sigma^2}} \quad , -\infty < x < \infty$

μ = Mean → average center of the distribution

σ = standard deviation → how spread out the data is
(Standart Abweichung)
↳ small → more clustered

χ^2 (Chi-Square) Distribution

→ Special case of gamma distribution

PDF: $f(x) = \frac{1}{\Gamma(\frac{r}{2})\, 2^{\frac{r}{2}}} x^{\frac{r}{2}-1} e^{-\frac{x}{2}}$, $x > 0$

r = degrees of freedom → number of independent standard normal variables

↳ number of categories − 1 !

Integral Refresher

$$\int_a^b f(x) = Fx \Big|_a^b = F(b) - F(a)$$

Rules

$$\int_a^b f(x) \pm g(x) = \int_a f(x) \pm \int_a^b g(x)$$

$$\int_a^b c \cdot f(x) = c \cdot \int_a^b f(x)$$

Integration by Parts

$$\int_a^b f(x)\, g'(x)\, dx = \big(f(x)\, g(x)\big)\Big|_a^b - \int_a^b f'(x)\, g(x)$$

Substitution

Derivation (Ableitung)

$$\int_a^b \sin(x^2)\, x\, dx \Rightarrow u = x^2 \Rightarrow \frac{du}{dx} = 2x$$

$$\Rightarrow x\,dx = \frac{1}{2}\, du$$

$$= \int_a^b \sin(u)\, \frac{1}{2}\, du$$ $\Rightarrow$ solve normal, the re-substitute

Integrations

Antiderivation
Stammfunktion

$f(x)$	$F(x)$	$f(x)$	$F(x)$
x^n	$\frac{1}{n+1} x^{n+1}$	x^{-1}	$\ln\lvert x\rvert$
$a\, x^n$	$a \frac{1}{n+1} x^{n+1}$	e^x	e^x
$a + x^n$	$ax + \frac{1}{n+1} x^{n+1}$	e^{-x}	$-1 \cdot e^{-x}$
		a^x	$\frac{a^x}{\ln(a)}$

Expected Value

↳ basically the mean

↳ repeat the experiment ∞ times

↳ average of X is the expected value

discrete random variable:

$$\mathbb{E}[X] = \sum_{x \in X} x \cdot \underbrace{\mathbb{P}(x)}_{\text{PMF}}$$

continuous random variable:

$$\mathbb{E}[X] = \int_{x \in X} x \cdot \underbrace{f(x)}_{\text{PDF}} \, dx$$

Same goes for functions

$$\mathbb{E}[g(X)] = \sum_{x \in X} \cdot g(x) \cdot \mathbb{P}(x)$$

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

Variance

↳ how far from the mean are the values spread?

general:

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

discrete random Variable:

$$Var[X] = \sum_{x \in X} (x - \mathbb{E}[X])^2 \cdot \mathbb{P}(x)$$

continuous random variable:

$$Var[X] = \int_{x \in X} (x - \mathbb{E}[X])^2 \cdot f(x) \, dx$$

Multiple random variables

↳ multiple aspects of the outcome of an expeiment that we want to look at

ex: 2 dice → $X = \text{Sum}$ $\quad Y = |\text{difference}|$

↳ We get a discrete random vector (X, Y)

Joint PMF

↳ PMF of discrete random vector

$$f_{X,Y}(x,y) = f(x,y) = \mathbb{P}(X=x, Y=y)$$

$$\mathbb{P}((X,Y) \in \underset{A \subset \mathbb{R}^2}{A}) = \sum_{(x,y)\in A} f(x,y)$$

Joint Expected value (PMF)

$$\mathbb{E}[g(X,Y)] = \sum_{(x,y)\in\mathbb{R}^2} g(x,y) \cdot \mathbb{P}(X=x, Y=y)$$

⇒ The Sum of all possibilities has to be 1

⇒ for any (x,y), $f(x,y) \geq 0$

$$\sum_{(x,y)\in\mathbb{R}^2} f(x,y) = \mathbb{P}((X,Y) \in \mathbb{R}^2) = 1$$

Joint PDF

↳ pdf of continuous random vectors

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

$A \subset \mathbb{R}^2$

integrated over all $(x,y) \in A$

Joint Expected value (PDF)

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

Marginal Distribution (PMF)

↳ we look at only one of the discrete random variables

$$f_X(x) = \sum_{y \in Y} f_{XY}(x,y) \quad \& \quad f_Y(y) = \sum_{x \in X} f_{XY}(x,y)$$

ex: 2 dice → X = sum Y = |difference|

↳ To get the probability of one of them beeing a certain number, we have to sum the probability of the other that gives us the desired value

↳ $y \overset{!}{=} 0 \Rightarrow f_Y(0) = \underbrace{\sum_{x \in X} f_{X,Y}(x,y)}_{\text{Sum of all probabilitys over the other variable}}$

Marginal Distribution PDF

↳ only one of the continuous random variables

$$f_X(x) = \int_{y \in Y} f(x,y) \; dy$$

$$f_Y(y) = \int_{x \in X} f(x,y) \; dx$$

Conditional PMF / PDF

↳ probability of one value if the other is fixed

$$f(y|x) = \mathbb{P}(Y=y \mid X=x) = \frac{f(x,y)}{f_X(x)}$$

↳ probability of y, if x is set

$$= \frac{\text{probability of both happening}}{\text{probability of only } x \text{ happening}}$$

Expected value

PMF:

$$\mathbb{E}[g(Y)|x] = \sum_{y \in Y} g(y)\, f(y|x)$$

PDF:

$$\mathbb{E}[g(Y)|x] = \int_{y \in Y} g(y)\, f(y|x)\, dy$$

$$\text{Bayes } P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

Covariance

↳ relationship between two random variables

↳ how much do they vary together?

$$Cov(X,Y) = \mathbb{E}\left[(X - \mathbb{E}[X]) \cdot (Y - \mathbb{E}[Y])\right]$$

↳ if $Cov(X,Y) > 0$ → X & Y move in the same direction

$Cov(X,Y) < 0$ → X & Y move in oposite direction

$Cov(X,Y) = 0$ → no linear relationship between X & Y

Rules:

- $\mathbb{E}[X,Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y] + Cov(X,Y)$
- X,Y = independent $\Rightarrow Cov(X,Y) = 0$
- $Cov(X,Y) = Cov(Y,X)$
- $Cov(aX+b, cY+d) = a \cdot c \cdot Cov(X,Y)$

Correlation

↳ strength and direction of linear relationships

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X \cdot \sigma_Y} \qquad \left.\begin{matrix}\sigma_X \\ \sigma_Y\end{matrix}\right\} = \text{Standard deviation}$$

↳ if $\rho_{X,Y} > 0$ → X, Y are positively correlated

if $\rho_{X,Y} < 0$ → X, Y are negatively correlated

if $\rho_{X,Y} = 0$ → X, Y are uncorrelated

↳ $\rho(X,Y)$ is between -1 and 1

Central Limit Theorem

↳ distribution of sample means approximates a normal distribution

⇒ Sample means are the means of a sequence of random variables $(X_1 \dots X_n)$

Sum of samples $= S_n = X_1 + X_2 \dots + X_n$

Sample mean $= \overline{X}_n = \frac{S_n}{n}$

$$Z_n = \frac{S_n - n\mathbb{E}[X_i]}{\sqrt{Var[X_i]} \cdot \sqrt{n}} = \frac{\overline{X}_n - \mathbb{E}[X_i]}{\frac{\sqrt{Var[X_i]}}{\sqrt{n}}}$$

⇒ It tells us that the average sample means will be normally distributed if the sample spce is large enough, no matte the distribution of the original data.

ex: Quality controll of a product

↳ take 40 lightbulbs of each production badge and calculate the averege life span $\overline{X}$

↳ with enoug $\overline{X}$s, the distribution of them will approximate a normal distribution.

Descriptive Statistics

↳ ways to summarize data

Scale of random variables

↳ choosing representative data to scale down the data set

3 basic scales

- nominal scale → categorize data without any quantitive value or order
  ex: car brands (BMW, VW, Toyota ...)

- ordinal scale → introduce also an order to the categories
  ex: car performance (luxury, sports ...)
  ↳ they can be ordered

- ratio scale → ordinal scale + zero point to be able to measure the distance between the objects
  ex: Car horsepower
  ↳ car 1 is 20 km/h faster than car 2

Summarize data using numerical techniques

Arithmetic mean (average)

(→ ratio scale)

$$\Rightarrow \quad \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Median

(→ ordinal & ratio scale)

$$\Rightarrow \quad x_{med} = \begin{cases} x_{\frac{n+1}{2}} & \text{if } n \text{ is odd} \\ \frac{1}{2}\left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}\right) & \text{if } n \text{ is even} \end{cases}$$

⇒ the middle value of all the values

Mode

(→ nominal, ordinal, ratio scale)

$$\Rightarrow \quad x_M = x_i$$

⇒ most frequent value

Emperical Quantiles

↳ divide data into equal-sized, ordered segments

$$k = \lfloor \alpha n \rfloor + 1$$

⇒ α = percentile $\left(\begin{array}{l} \text{ex } \alpha = 75 \quad n = 100 \\ \lfloor \alpha n \rfloor + 1 = 76 \end{array} \right)$

$$\alpha\text{-quantile} = \begin{cases} x_k & \text{, if } \alpha n = \text{integer} \\ \frac{1}{2}\left(x_k + x_{k-1}\right) & \text{, if } \alpha n = \text{no integer} \end{cases}$$

Range

↳ biggest - smallest valeue

$$range = x_{max} - x_{min}$$

Mean - quartile

↳ range as dispersion measure

$$MQA = \frac{(Q_3 - Q_2) + (Q_2 - Q_1)}{2} = \frac{IQA}{2}$$

⇒ average Quartile size

Variance

↳ spread of the data from the average

$$s_x^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

Standard deviation

↳ amount of variation from the average (actual value)

$$s_x = + \sqrt{s_x^2}$$

absolute dispersion

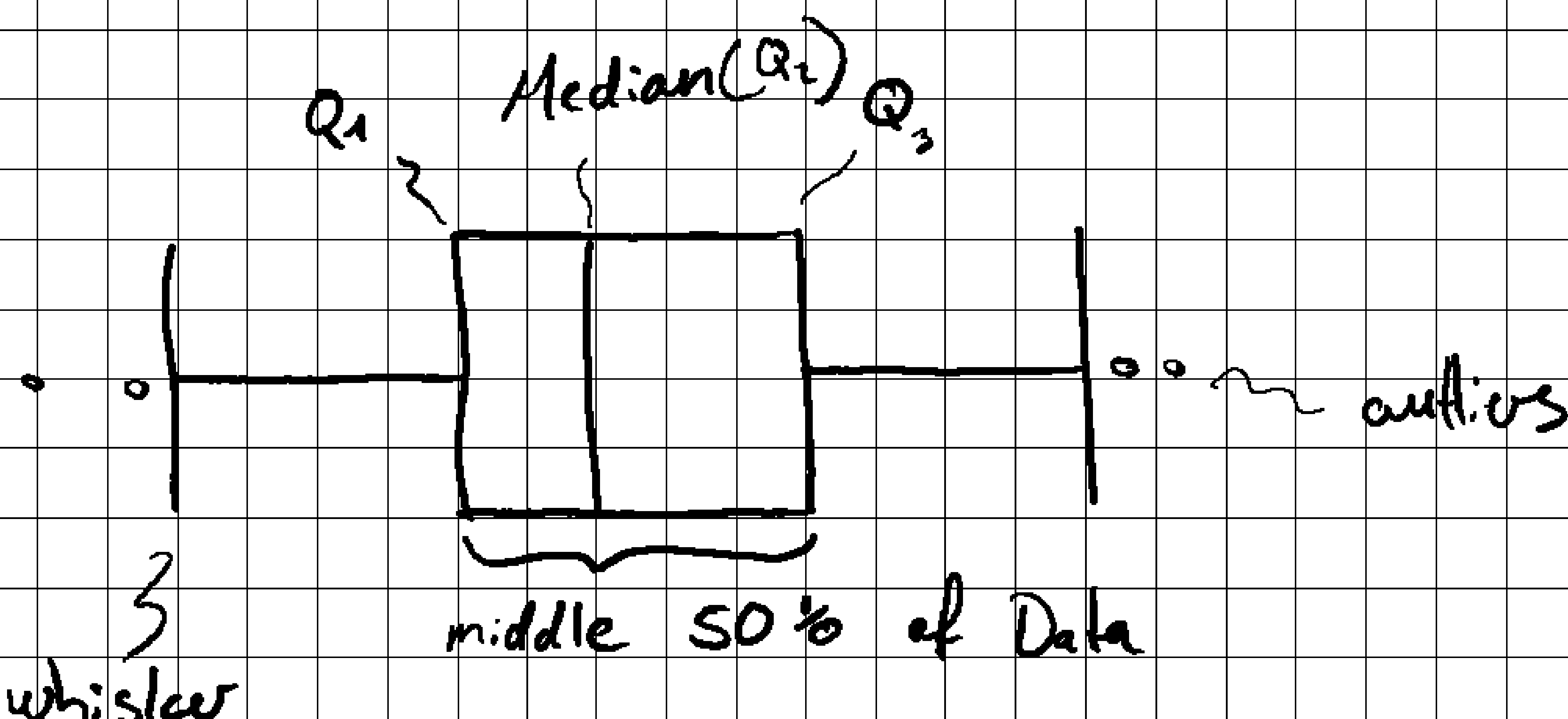
↳ how spread are the values

$$VK_x = \frac{s_x}{|\bar{x}|}$$

Visualize Data

Box Plot

↳ graph to show distribution & outliers of Data



↳ $Q_1 - 1.5 \cdot (IQR)$ } smallest value within this margin

↳ $Q_3 - Q_1$

Interpretation

- length of box and whisker indicate the spread
- if the median is centered, than the distribution is symmetrical
- outliers show possible errors in Data

QQ- Plot

↳ Compare two distributions by plotting their quantiles against each other

⇒ Check how a distribution matches a theoretical distribution

Draw

- calculate quantiles for both distributions
- plot corresponding quantiles

Interpretation

- Depending on the spread you can see how closely one distribution follows the other

Scatter Plot

↳ each value is a point on a line / plane

⇒ see groups and potential relations

⇒ - Single values are plotted on a line
- two dimensional values are plotted on a plane
  ↳ relation between two values
    ↳ e.g. exercise time & blood preasure

Estimation of parameters

↳ estimate population parameter based on sample data

two approaches

- point estimator → calculate single value
- confidence interval → interval in which the true value lies

Point Estimator θ

- Sample mean $(\hat{\mu} = \overline{X})$ → $\frac{\text{Sum of all values}}{\text{Number of values}} = \frac{1}{n}\sum_{i=1}^{n} X_i$

- Sample variance $(\hat{\sigma}^2 = s^2)$ → difference between each sample point and the mean

$$\Rightarrow \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2$$

Bias

↳ the bias is the difference between the point estimator and the expected value

$$\text{Bias}(T) = \mathbb{E}[T] - \theta$$

⇒ if the bias $=0$, then is the estimator unbiased

⇒ if the bias becomes smaller with increasing sample size, and vanishes when size $\to \infty$, then we say that the estimator ; asymtomatically unbiased

$$\lim_{n \to \infty} \mathbb{E}[T] = \theta$$

⇒ if the estimator θ converges to a certain value with an increasing sample size, we say θ is consistent

↳ estimator is consistent, when:

- it is unbiased / asymtomatically unbiased
- variance $\to 0$ when size $\to \infty$

⇒ if the variance is smaller/equal to the unbiased estimator, then it is efficent

Mean squared error (MSE)

↳ criteria to check if an estimator is good

$$\Rightarrow \quad MSE(\theta) = \mathbb{E}[(T-\theta)^2]$$

$$MSE(\theta) = Var(T) + bias^2$$

$\Rightarrow$ if $MSE(\theta) = Var(T)$ → then is θ good

Methods to get estimators

- Method of Moments → Sample mean & variance
- Least-Squares Method →
  ↳ find best fitting line through a set of data points by minimizing the sum of squares of the distances to the points.
- Maximum Likelihood Method
  ↳ the most plausible parameter
    ↳ find the values of the model parameters, that make the observed data most probable

Confidence Intervals

↳ Interval that contains the true value of the estimator with a confidence $\times$