Disclaimer: Diese Zusammenfassung wurde im Rahmen der Vorlesung "Wahrscheinlichkeitstheorie & Statistik" von Prof. B. Acciaio im FS22 geschrieben.
Es besteht keine Garantie für Vollständigkeit & Korrektheit dieses Dokuments. Jedoch bin ich froh über Fehler informiert zu werden oder bei Fragen zu helfen :)

ldewindt@ethz.ch

Zürich, der 07.12.2022

Lina De Windt

WuS Summary

VL: Prof. B. Acciaio FS22 Lina De Windt ldewindt@ethz.ch

1.1 Sample spaces and events

Probability space: $(\Omega, \mathcal{F}, \mathbb{P})$ where
- Ω: Sample Space
- $\mathcal{F}$: Set of events
- $\mathbb{P}$: Probability measure

Def. 1.1 (Sample Space): The set Ω is called the sample space. An element $\omega \in \Omega$ is called an outcome.

Def. 1.2 ($\mathcal{F}$ = Set of all the events) A set $\mathcal{F}$ of all the events is given by a subset of $\mathcal{P}(\Omega)$ satisfying the following hypotheses:
- H1) $\Omega \in \mathcal{F}$
- H2) $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$
- H3) $A_1, A_2, \ldots \in \mathcal{F} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$

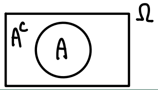
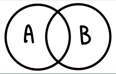
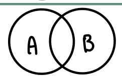
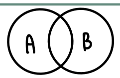
↪ A set satisfying H1) ~ H3) is called a σ-Algebra.

Remark: $\mathcal{P}(\Omega)$ = the set of all possible subsets of Ω, including the empty set and Ω itself.
Power set

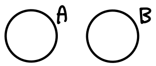
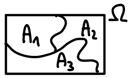
Terminology: Let $\omega \in \Omega$ (ω is a possible outcome), A an event.
We say the event A occurs (for ω) if $\omega \in A$.
We say the event A does not occur (for ω) if $\omega \notin A$.

Remark:
- The event $A = \emptyset$ never occurs.
- The event $A = \Omega$ always occurs.

⚠ Events are defined as subsets of $\Omega \Rightarrow$ use operations from set theory!

Event	Graphical representation	Probabilistic Interpretation
A^c		A does not occur
$A \cap B$		A and B occur
$A \cup B$		A or B occurs
$A \Delta B$		One and only one of A or B occurs

Relations betw. events & interpretations:

Relation	Graphical	Probabilistic Interpretation
$A \subset B$		If A occurs, then B occurs
$A \cap B = \emptyset$		A and B cannot occur at the same time
$\Omega = A_1 \cup A_2 \cup A_3$ with A_1, A_2, A_3 pairwise disjoint		For each outcome ω, one and only one of the events A_1, A_2, A_3 is satisfied.

1.2 Mathematical definition of probability spaces

Def. 1.6. (Probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$)
Let Ω be a sample space, $\mathcal{F}$ a set of events. $\mathbb{P}$ on $(\Omega, \mathcal{F})$ is a map:

$$\mathbb{P}: \mathcal{F} \to [0,1]$$
associates to each event a number in $[0,1]$!
$$A \mapsto \mathbb{P}[A]$$
that satisfies:
- P1) $\mathbb{P}[\Omega] = 1$
- P2) (Countable additivity): $\mathbb{P}[A] = \sum_{i=1}^{\infty} \mathbb{P}[A_i]$ if $A = \bigcup_{i=1}^{\infty} A_i$ (disjoint union)

for example: n: how many times the experiment is repeated
n_A: the number of occurances of the event A $\Rightarrow \mathbb{P}[A] = \frac{n_A}{n}$

How to define $\mathbb{P}[A]$ for a given $(\Omega, \mathcal{F})$?
If the sample space Ω is finite / countable (⚠ does not work otherwise).
1. associate to every ω the prob. p_ω that the output of the experiment is ω.
2. Then for an event $A \subset \Omega$: $\boxed{\mathbb{P}[A] = \sum_{\omega \in A} p_\omega}$ ⚠ does not work for Ω uncountable ⚠ $\mathbb{P}[\Omega] = \sum_{\omega \in \Omega} p(\omega) = 1$

Borel σ-Algebra: contains all $A = [x_1, x_2] \times [y_1, y_2]$ with $0 \leq x_1 \leq x_2 \leq 1$, $0 \leq y_1 \leq y_2 \leq 1$ and it is the smallest collection ob subsets of Ω which satisfies H1, H2 and H3 in Def. 1.2

Proposition 1.8: Direct consequences of the Def. of $\mathbb{P}$:
Let $\mathbb{P}$ be a prob. measure on $(\Omega, \mathcal{F})$:
- i) $\mathbb{P}[\emptyset] = 0$
- ii) Additivity: Let $k \geq 1$, $A_1, \ldots, A_k$ pairwise disjoint events, then
$$\boxed{\mathbb{P}[A_1 \cup A_2 \cup \ldots \cup A_k] = \mathbb{P}[A_1] + \mathbb{P}[A_2] + \ldots + \mathbb{P}[A_k]}$$
- iii) Let A be an event. Then: $\boxed{\mathbb{P}[A^c] = 1 - \mathbb{P}[A]}$
- iv) If A and B are 2 events (not necessarily disjoint!) then:
$$\boxed{\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \cap B]}$$

Useful Inequalities:

Proposition 1.9: Monotonicity
Let $A, B \in \mathcal{F}$. Then $\boxed{A \subset B \Rightarrow \mathbb{P}[A] \leq \mathbb{P}[B]}$

Proposition 1.10: Union Bound
Let $A_1, A_2, \ldots$ be a sequence of events (not necessarily disjoint). Then
$$\boxed{\mathbb{P}\Big[\bigcup_{i=1}^{\infty} A_i\Big] \leq \sum_{i=1}^{\infty} \mathbb{P}[A_i]}$$
→ this also applies to a finite collection of events!

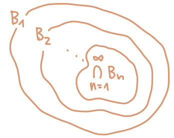
How to construct a probabilistic model: Give:
- → A sample space Ω ("all the possible outcomes of the experiment")
- → A set of events $\mathcal{F} \subset \mathcal{P}(\Omega)$ ("set of all possible observations")
- → A probabilistic measure $\mathbb{P}$ ("gives a number in $[0,1]$ to every event")

Continuity properties of probability measures

Proposition 1.12
Let (A_n) be an increasing sequence of events ($A_n \subset A_{n+1}$ $\forall n$). Then:

$$\boxed{\lim_{n\to\infty} \mathbb{P}[A_n] = \mathbb{P}\Big[\bigcup_{n=1}^{\infty} A_n\Big]}$$ increasing limit

Let (B_n) be a decreasing sequence of events ($B_n \supset B_{n+1}$ $\forall n$). Then:



$$\boxed{\lim_{n\to\infty} \mathbb{P}[B_n] = \mathbb{P}\Big[\bigcap_{n=1}^{\infty} B_n\Big]}$$ decreasing limit

By Monotonicity, we have $\mathbb{P}[A_n] \leq \mathbb{P}[A_{n+1}]$ and $\mathbb{P}[B_n] \geq \mathbb{P}[B_{n+1}]$ $\forall n$. Hence the limits in the prop. are well defined as monotone limits.

1.3 Laplace Models and Counting

Laplace Model: (Intuition) The Sample set Ω is an arbitrary finite set, and all the outcomes have the same probability $\boxed{p_\omega = \frac{1}{|\Omega|}}$

Definition 1.14: Laplace Model
Let Ω be a finite sample space. The Laplace Model on Ω is the Triplet $(\Omega, \mathcal{F}, \mathbb{P})$ where:
- → $\mathcal{F} = \mathcal{P}(\Omega)$
- → $\mathbb{P}: \mathcal{F} \to [0,1]$ is defined by $\forall A \in \mathcal{F}$: $\boxed{\mathbb{P}[A] = \frac{|A|}{|\Omega|}}$

1.4 Random Variables and distribution functions

Definition 1.15: Random Variable
Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A random variable (r.v.) is a map
$$\boxed{X: \Omega \to \mathbb{R} \text{ s.t. } \forall a \in \mathbb{R}: \{\omega \in \Omega : X(\omega) \leq a\} \in \mathcal{F}}$$
this part is needed for $\mathbb{P}[\{\omega \in \Omega : X(\omega) \leq a\}]$ to be well-defined.

Notation: for $a \leq b$ we write:
- → $\{X \leq a\} = \{\omega \in \Omega : X(\omega) \leq a\}$
- → $\{a < X < b\} = \{\omega \in \Omega : a < X(\omega) < b\}$
- → $\{X \in \mathbb{Z}\} = \{\omega \in \Omega : X(\omega) \in \mathbb{Z}\}$

Same for probabilities:
- → $\mathbb{P}[X \leq a] = \mathbb{P}[\{X \leq a\}] = \mathbb{P}[\{\omega \in \Omega : X(\omega) \leq a\}]$

Example of a r.v.: Gambling with one die:
We roll a fair die. The sample space is $\Omega = \{1,2,3,4,5,6\}$ and the associated probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Suppose that we gamble on the outcome in such a way that our profit is:
- -1 if the outcome is $1,2,3$
- 0 if the outcome is 4
- 2 if the outcome is $5,6$

Where a negative profit corresponds to a loss. Our profit can be represented by the r.v. X defined by:
$$X(\omega) = \begin{cases} -1 & \text{if } \omega = 1,2,3 \\ 0 & \text{if } \omega = 4 \\ 2 & \text{if } \omega = 5,6 \end{cases}$$

Indicator function of an event: Let $A \in \mathcal{F}$.
$$\boxed{\mathbb{1}_A(\omega) = \begin{cases} 0 & \text{if } \omega \notin A \\ 1 & \text{if } \omega \in A \end{cases}}$$

⚠ $\mathbb{1}_A$ is a random variable! Indeed: $\{\mathbb{1}_A \leq a\} = \begin{cases} \emptyset & \text{if } a < 0 \\ A^c & \text{if } 0 \leq a < 1 \\ \Omega & \text{if } a \geq 1 \end{cases} \in \mathcal{F}$

Definition 1.16: Distribution function of X "Verteilungsfunktion"

Let X be a r.v. on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The distribution function of X is the function $F_X: \mathbb{R} \to [0,1]$ defined in

$$\boxed{\forall a \in \mathbb{R}: F_X(a) = \mathbb{P}[X \le a]}$$

Proposition 1.17: Basic Identity of F_X:

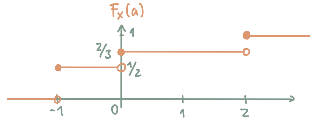
Let $a<b$ be 2 real numbers. Then $\boxed{\mathbb{P}[a < X \le b] = F(b) - F(a)}$

Example:

Let $X(\omega) = \begin{cases} -1 & \text{if } \omega = 1,2,3 \\ 0 & \text{if } \omega = 4 \\ 2 & \text{if } \omega = 5,6 \end{cases}$

Then: $\{X \le a\} = \{\omega: X(\omega) \le a\} = \begin{cases} \emptyset & \text{if } a < -1 \\ \{1,2,3\} & \text{if } -1 \le a < 0 \\ \{1,2,3,4\} & \text{if } 0 \le a < 2 \\ \{1,2,3,4,5,6\} & \text{if } a \ge 2 \end{cases}$

Hence, $F_X(a) = \begin{cases} 0 & \text{if } a < -1 \\ 1/2 & \text{if } -1 \le a < 0 \\ 2/3 & \text{if } 0 \le a < 2 \\ 1 & \text{if } a \ge 2 \end{cases}$



Theorem 1.18: Properties of distribution functions:

Let X be a r.v. on some prob. space $(\Omega, \mathcal{F}, \mathbb{P})$. The distribution function $F = F_X: \mathbb{R} \to [0,1]$ of X satisfies the following properties:

i) F is nondecreasing
ii) F is right continuous
iii) $\lim_{a \to -\infty} F(a) = 0$ and $\lim_{a \to +\infty} F(a) = 1$

} characterize distribution fct's!

Theorem 1.19:

Let $F: \mathbb{R} \to [0,1]$ satisfying i) ~ iii). Then there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a r.v. $X: \Omega \to \mathbb{R}$ s.t. $F = F_X$.

this means: One can define a r.v. via its distribution function!

⚠ the presice choice of the prob. space is often not important.

Discontinuity / Continuity points of F:

F is always right continuous, but not always left continuous!

i.e.: $\boxed{F(a^-) = \lim_{h \downarrow 0} F(a-h)}$

Proposition 1.20: probability of a given value:

Let $X: \Omega \to \mathbb{R}$ be a r.v. with distribution function F. Then $\forall a \in \mathbb{R}$:

$$\boxed{\mathbb{P}[X=a] = F(a) - F(a^-)}$$

Interpretation:

→ If F is not continuous @ a point $a \in \mathbb{R}$, then the "jump size" $F(a) - F(a^-)$ is equal to $\mathbb{P}[X=a]$

→ If F is continuous @ a point $a \in \mathbb{R}$, then $\mathbb{P}[X=a] = 0$

Discrete & Continuous random variables:

Definition 1.21: Discrete random variables:

A r.v. $X: \Omega \to \mathbb{R}$ is said to be discrete if its image

$\boxed{X(\Omega) = \{x \in \mathbb{R}: \exists \omega \in \Omega, X(\omega) = x\}}$ is at most countable.

This means: a r.v. is discrete if it only takes finitely or countably many values $x_1, x_2, \dots$. In this case:

$p(x_1) = \mathbb{P}[X = x_1]$, $p(x_2) = \mathbb{P}[X = x_2]$

→ the probabilistic properties of X are fully described!

Definition 1.22: Continuous Random Variables:

A r.v. $X: \Omega \to \mathbb{R}$ is said to be continuous if its distribution function F_X can be written as: $\boxed{F_X(a) = \int_{-\infty}^{a} f(x)\, dx}$ $\forall a \in \mathbb{R}$

for some nonnegative function $f: \mathbb{R} \to \mathbb{R}_+$, called the density of X

$f(x)$ = "Verteilungsdichtefunktion"

$f(x)$ represents the probability that X takes a value in the infinitesimal interval $[x, x+dx]$

1.5 Conditional Probabilities

Definition 1.23: Conditional Probability

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be some probability space. Let A, B be 2 events with $\mathbb{P}[B] > 0$. The conditional probability of A given B is:

$$\boxed{\mathbb{P}[A \mid B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}}$$

Proposition 1.25:

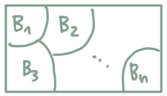
Let $(\Omega, \mathcal{F}, \mathbb{P})$ be some probability space. Let B be an event w. positive prob. Then $\mathbb{P}[\cdot \mid B]$ is a probability measure on Ω.

Proposition 1.26: Formula of total probability

Let $B_1, \dots, B_n$ be a partition (d.h. $B_1 \cup \dots \cup B_n$ and the events B_i are pairwise disjoint) of the sample space Ω with $\mathbb{P}[B_i] > 0$ $\forall 1 \le i \le n$. Then:

$$\boxed{\forall A \in \mathcal{F}: \mathbb{P}[A] = \sum_{i=1}^{n} \mathbb{P}[A \mid B_i] \cdot \mathbb{P}[B_i]}$$

*partition of Ω:



B_i pairwise disjoint!

Proposition 1.27: Bayes Formula:

Let $B_1, \dots, B_n \in \mathcal{F}$ be a partition of Ω with $\mathbb{P}[B_i] > 0$ $\forall i$. Then for all events A with $\mathbb{P}[A] > 0$ we have $\forall i = 1, \dots, n$:

$$\boxed{\mathbb{P}[B_i \mid A] = \frac{\mathbb{P}[A \mid B_i] \cdot \mathbb{P}[B_i]}{\sum_{j=1}^{n} \mathbb{P}[A \mid B_j] \cdot \mathbb{P}[B_j]}}$$

1.6 Independence

Definition 1.28: Independence of events

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Two events A and B are independent if

$$\boxed{\mathbb{P}[A \cap B] = \mathbb{P}[A] \cdot \mathbb{P}[B]}$$

Remarks:

1) If $\mathbb{P}[A] \in \{0,1\}$, then A is independent of every event, i.e. $\forall B \in \mathcal{F}: \mathbb{P}[A \cap B] = \mathbb{P}[A] \cdot \mathbb{P}[B]$
2) If an event A is independent with itself (e.g. $\mathbb{P}[A \cap A] = \mathbb{P}[A]^2$) then $\mathbb{P}[A] \in \{0,1\}$
3) A is independent of B if and only if A is independent of B^c

$$\boxed{A, B \text{ independent} \Leftrightarrow A, B^c \text{ independent}}$$

Concept of independence ≙ idea that 2 events do not influence each other.

Proposition 1.30

Let $A, B \in \mathcal{F}$ be 2 events with $\mathbb{P}[A], \mathbb{P}[B] > 0$. Then the following are equivalent:

i) $\mathbb{P}[A \cap B] = \mathbb{P}[A] \cdot \mathbb{P}[B]$
⟺
ii) $\mathbb{P}[A \mid B] = \mathbb{P}[A]$
⟺
iii) $\mathbb{P}[B \mid A] = \mathbb{P}[B]$

Definition 1.31: Independence of n events

n events $A_1, \dots, A_n$ are independent if:

$$\boxed{\forall J \subset \{1, \dots, n\}: \mathbb{P}\Big[\bigcap_{j \in J} A_j\Big] = \prod_{j \in J} \mathbb{P}[A_j]}$$

d.h.: for $n=3$: 3 events A, B, C are independent if these 4 equations are satisfied:

1) $\mathbb{P}[A \cap B] = \mathbb{P}[A] \cdot \mathbb{P}[B]$
2) $\mathbb{P}[A \cap C] = \mathbb{P}[A] \cdot \mathbb{P}[C]$
3) $\mathbb{P}[B \cap C] = \mathbb{P}[B] \cdot \mathbb{P}[C]$
4) $\mathbb{P}[A \cap B \cap C] = \mathbb{P}[A] \cdot \mathbb{P}[B] \cdot \mathbb{P}[C]$

⚠ all of these have to be satisfied!

Independence of random variables:

Definition 1.32: Let $X_1, \dots, X_n$ be n r.v. on some prob. space $(\Omega, \mathcal{F}, \mathbb{P})$. We say that $X_1, \dots, X_n$ are independent if: $\forall a_1, \dots, a_n \in \mathbb{R}$:

$$\boxed{\mathbb{P}[X_1 \le a_1, \dots, X_n \le a_n] = \mathbb{P}[X_1 \le a_1] \cdot \dots \cdot \mathbb{P}[X_n \le a_n]}$$

Definition 1.33: Independence of an infinite sequence of r.v:

Let $X_1, X_2, \dots$ be an infinite sequence of r.v. We say that $X_1, X_2, \dots$ are independent if $X_1, \dots, X_n$ are independent $\forall n$.

Definition 1.35: Independent and identically distributed (i.i.d.)

Some r.v.'s $X_1, X_2, \dots$ are said to be independent and identically distributed (i.i.d) if they are independent and have the same distribution function, $\forall i,j: F_{X_i} = F_{X_j}$

Theorem 1.34:

Let $F_1, \dots, F_n$ be n distribution functions. Then $\exists$ a prob. space $(\Omega, \mathcal{F}, \mathbb{P})$ and n r.v. $X_1, \dots, X_n$ on this prob. space such that:

→ $\forall i$, X_i has the distribution function F_i (i.e. $\forall a: \mathbb{P}[X_i \le a] = F_i(a)$)

→ $X_1, \dots, X_n$ are independent.

Chapter 2: Discrete Distributions

Definition 2.1: Sum of nonnegative numbers

Let $(a_x)_{x \in E}$ be a sequence of nonnegative numbers (i.e. $\forall x, a_x \ge 0$). The sum of the a_x is defined by:

$$\boxed{\sum_{x \in E} a_x := \sup_{\substack{F \subset E \\ F \text{ finite}}} \sum_{x \in F} a_x}$$

→ the sum can be infinite, e.g. if E is infinite and $a_x = 1$

→ $\forall$ sequences $F_n \uparrow E$ we can check that the limit $\lim_{n \to \infty} \big(\sum_{x \in F_n} a_x\big)$ makes sense and $\lim_{n \to \infty} \big(\sum_{x \in F_n} a_x\big) = \sum_{x \in E} a_x$

Definition 2.3: Sum of an integrable sequence:
A sequence $(a_x)_{x\in E}$ of real numbers is integrable if:

$$\boxed{\sum_{x\in E} |a_x| < \infty}$$

In this case, we define: $\boxed{\sum_{x\in E} a_x = \sum_{x\in E} a_x^+ - \sum_{x\in E} a_x^-}$ where

$a_x^+ = \max(0, a_x)$ and $a_x^- = \max(0, -a_x)$ represent the positive and negative parts of the sequence.

→ a_x integrable ⟹ sum is finite

→ $\forall$ sequence $F_n \uparrow E$, $\lim_{n\to\infty}(\sum_{x\in F_n} a_x)$ makes sense and $\lim_{n\to\infty}(\sum_{x\in F_n} a_x) = \sum_{x\in E} a_x$

⚠ lim does not depend on the chosen sequence $F_n \uparrow E$!

Theorem 2.5: Fubini for integrable sequences:
Let E, F be 2 finite or countable sets. Let $(u_{x,y})_{(x,y)\in E\times F}$ be a family of real numbers.
Assume that $\sum_{x\in E}(\sum_{y\in F}|u_{x,y}|) < \infty$. Then:

$$\boxed{\sum_{x\in E}\left(\sum_{y\in F} u_{x,y}\right) = \sum_{y\in F}\left(\sum_{x\in E} u_{x,y}\right) = \sum_{(x,y)\in E\times F} u_{x,y}}$$

Theorem 2.6: Fubini for nonnegative sequences:
Let E, F be 2 finite or countable sets. Let $(u_{x,y})_{(x,y)\in E\times F}$ be a family of nonnegative numbers. Then:

$$\boxed{\sum_{(x,y)\in E\times F} u_{x,y} = \sum_{x\in E}\left(\sum_{y\in F} u_{x,y}\right) = \sum_{y\in F}\left(\sum_{x\in E} u_{x,y}\right)}$$

2.2 Definitions and Examples

Definition 2.7: Discrete random variables:
A r.v. $X: \Omega \to \mathbb{R}$ is discrete if $\exists$ some set $E \subset \mathbb{R}$ finite or countable s.t.:

$$\boxed{\forall \omega \in \Omega: X(\omega) \in E}$$

Definition 2.8: Distribution of a discrete r.v.:
Let X be a discrete r.v. taking some values in some finite or countable set $E \subset \mathbb{R}$. The distribution of X is the sequence of numbers:

$$\boxed{\forall x \in E: \ p_x = \mathbb{P}(X = x)}$$

Proposition 2.9:
The distribution $(p_x)_{x\in E}$ of a discrete r.v. satisfies:

$$\boxed{\sum_{x\in E} p_x = 1}$$

⚠ if we are given a sequence of numbers p_x with values $[0,1]$ such that $\sum_{x\in E} p_x = 1$, then $\exists$ a prob. space $(\Omega, \mathcal{F}, \mathbb{P})$ and a r.v. X w. associated distribution (p_x) !

Bernoulli random variable

Let $0 \le p \le 1$. A r.v. X is a Bernoulli r.v. with parameter p if
→ it takes values in $E = \{0,1\}$ and
→ $\boxed{\mathbb{P}[X=0] = 1-p}$ and $\boxed{\mathbb{P}[X=1] = p}$

Notation: $\boxed{X \sim Ber(p)}$

Binomial random variable

Let $0 \le p \le 1$, $n \in \mathbb{N}$. A r.v. X is a Binomial r.v. w. parameters n and p if:
→ it takes values in $E = \{0, 1, \dots, n\}$
→ $\forall k \in \{0, 1, \dots, n\}$: $\boxed{\mathbb{P}[X=k] = \binom{n}{k} p^k (1-p)^{n-k}}$

Notation: $\boxed{X \sim Bin(n,p)}$

Proposition 2.12: Sum of independent Ber and Bin:
Let $0 \le p \le 1$, $n \in \mathbb{N}$. Let $X_1, \dots, X_n$ be independent Ber(p). Then:

$$\boxed{S_n := X_1 + \dots + X_n \text{ is } \sim Bin(n,p).}$$

→ In particular, the distribution Bin(1, p) is the same as Ber(p) !
→ If $X \sim Bin(m,p)$ and $Y \sim Bin(n,p)$ and X, Y are independent:

$$\boxed{X + Y = Bin(m+n, p)}$$

Geometric random variable:

Let $0 \le p \le 1$. A r.v. X is a geometric r.v. w. parameter p if:
→ it takes values in $E = \mathbb{N}\setminus\{0\}$ and
→ $\forall k \in \mathbb{N}\setminus\{0\}$: $\boxed{\mathbb{P}[X=k] = (1-p)^{k-1} \cdot p}$ ⚠ here, $0^0 = 1$

Notation: $\boxed{X \sim Geom(p)}$

Proposition 2.16:
Let $X_1, X_2, \dots$ be a sequence of ∞ many independent Ber(p). Then:

$$\boxed{T := \min\{n \ge 1: X_n = 1\} \sim Geom(p)}$$

Proposition 2.18: Absence of memory of the geometric distribution:
Let $T \sim Geom(p)$ for some $0 \le p \le 1$. Then:

$$\boxed{\forall n \ge 0, \forall k \ge 1: \ \mathbb{P}[T \ge n+k \mid T > n] = \mathbb{P}[T \ge k]}$$

Poisson random variable

Let $\lambda > 0$ be a positive real number. A r.v. X is a Poisson r.v. with parameter λ if:
→ it takes values in $E = \mathbb{N}$
→ $\forall k \in \mathbb{N}$: $\boxed{\mathbb{P}[X=k] = \frac{\lambda^k}{k!} e^{-\lambda}}$

$\boxed{E[X] = \lambda}$

Notation: $\boxed{X \sim Poisson(\lambda)} = Poiss(\lambda)$

⚠ the poiss. distribution is an approx. of a bin. dist. Bin(n,p) if param. n is large and param. p is small!

Proposition 2.20: Poisson approximation of the Bin.
Let $\lambda > 0$. $\forall n \ge 1$, consider a r.v. $X_n \sim Bin(n, \frac{\lambda}{n})$. Then:

$$\boxed{\forall k \in \mathbb{N}: \lim_{n\to\infty} \mathbb{P}[X_n = k] = \mathbb{P}[N = k] \text{ with } N \sim Poiss(\lambda)}$$

Intuitively, this says that X_n and N have very similar prob. properties for n large.

2.3 Joint distribution & image of r.v.'s

Definition 2.22: Joint distribution:
Let $X_1: \Omega \to E_1$, $X_2: \Omega \to E_2$, ..., $X_n: \Omega \to E_n$ be n discrete r.v.'s on $(\Omega, \mathcal{F}, \mathbb{P})$ with values in some finite or countable sets $E_1, \dots, E_n$. The joint distribution of the r.v.'s $X_1, \dots, X_n$ is the family $(p_{x_1,\dots,x_n})_{x_1\in E_1,\dots,x_n\in E_n}$ with

$$\boxed{p_{x_1,\dots,x_n} = \mathbb{P}[X_1 = x_1, \dots, X_n = x_n]}$$

→ this characterises the prob. properties of the random vector $(X_1, \dots, X_n)$
⚠ If the r.v.'s $(X_1, \dots, X_n)$ are independent, the joint dist. is:

$$\boxed{p_{x_1,\dots,x_n} = \mathbb{P}[X_1 = x_1] \cdot \dots \cdot \mathbb{P}[X_n = x_n]}$$

Proposition 2.23: Consider r.v. as Image of discrete r.v.:
Let $n \ge 1$, $\phi: \mathbb{R}^n \to \mathbb{R}$, $X_1: \Omega \to E_1, \dots, X_n: \Omega \to E_n$ n discrete r.v.'s on $(\Omega, \mathcal{F}, \mathbb{P})$ with values in some finite/countable sets $E_1, \dots, E_n$.
Then $Z = \phi(X_1, \dots, X_n)$ is a discrete r.v. with values in the discrete set $F = \phi(E_1 \times \dots \times E_n)$ and with distribution given by:

$$\forall z \in F: \boxed{\mathbb{P}[Z = z] = \sum_{\substack{x_1\in E_1,\dots,x_n\in E_n \\ \phi(x_1,\dots,x_n) = z}} \mathbb{P}[X_1 = x_1, \dots, X_n = x_n]}$$

2.3.1 Almost sure events

Definition 2.24: Let $A \in \mathcal{F}$ be an event. A occurs almost surely (a.s.) if:

$$\boxed{\mathbb{P}[A] = 1}$$

→ Can be extended to say any set A (not necessarily an event!) occurs almost surely if $\exists$ an event $A' \in \mathcal{F}$ s.t. $A' \subset A$ and $\mathbb{P}[A'] = 1$

2.4 Expectation → Intuitively: Average

Definition 2.26: Expectation of a r.v.:
Let $X: \Omega \to E$ be a discrete r.v., assume $X \ge 0$ a.s. Then the expectation of X is:

$$\boxed{\mathbb{E}[X] = \sum_{x\in E} x \cdot \mathbb{P}[X = x]}$$

But what when a r.v. does not have a constant sign?

Definition 2.28: Integrability
Let $X: \Omega \to E$ be a discrete r.v.. X is integrable if: $\boxed{\mathbb{E}[|X|] < \infty}$
In this case, the expectation of X is: $\boxed{\mathbb{E}(X) = \sum_{x\in E} x \cdot \mathbb{P}[X = x]}$

Expectation value of a Bernoulli r.v.

$X \sim Ber(p)$. Then $\boxed{\mathbb{E}[X]=p}$

Expectation value of a Poisson r.v.

$X \sim Poiss(\lambda)$, $\lambda>0$. Then $\boxed{\mathbb{E}[X]=\lambda}$

Expectation value of an Indicator of an event:

Let A be an event. Then $\mathbb{1}_A \sim Ber(\mathbb{P}[A])$. Hence

$$\boxed{\mathbb{E}[\mathbb{1}_A]=\mathbb{P}[A]}$$

Expectation value of a Geometric r.v.:

$T \sim Geom(p)$, $0<p\leq 1$. Then $\boxed{\mathbb{E}[T]=\frac{1}{p}}$

Proposition 2.30: Tailsum formula:

Let X be a discrete r.v. taking values in $\mathbb{N}=\{0,1,\dots\}$. Then:

$$\boxed{\mathbb{E}[X]=\sum_{n=1}^{\infty}\mathbb{P}[X\geq n]}$$

Theorem 2.31: Image of random variables:

Let $X_1,\dots,X_n: \Omega\to E$ be n r.v.'s, $\phi: \mathbb{R}^n\to\mathbb{R}$. Then $Z=\phi(X_1,\dots,X_n)$ defines a discrete r.v. and if

$\sum_{x_1,\dots,x_n\in E}|\phi(x_1,\dots,x_n)|\cdot\mathbb{P}[X_1=x_1,\dots,X_n=x_n]<\infty$, Z is integrable and

$$\boxed{\mathbb{E}[\phi(X_1,\dots,X_n)]=\sum_{x_1,\dots,x_n\in E}\phi(x_1,\dots,x_n)\cdot\mathbb{P}[X_1=x_1,\dots,X_n=x_n]}$$

→ Formula for $n=1$: Let $X:\Omega\to E$ be a discrete r.v. and $\phi:\mathbb{R}\to\mathbb{R}$.
Assume $\sum_{x\in E}|\phi(x)|\cdot\mathbb{P}[X=x]<\infty$. Then the discrete r.v.
$Z=\phi(X)$ is integrable and $\boxed{\mathbb{E}[Z]=\mathbb{E}[\phi(X)]=\sum_{x\in E}\phi(x)\mathbb{P}[X=x]}$

→ In the case ϕ takes values in the set $F\subset[0,\infty)$ ($\Rightarrow Z\geq$ a.s.) then $\mathbb{E}[\phi(X_1,\dots,X_n)]=\sum\phi(x_1,\dots,x_n)\mathbb{P}[X_1=x_1,\dots,X_n=x_n]$ is always true, even without the integrability assumption :)

Theorem 2.34: Linearity of the Expectation:

Let $X,Y:\Omega\to\mathbb{R}$ be 2 integrable discrete r.v.'s, $\lambda\in\mathbb{R}$.
Then λX and $X+Y$ are also integrable discrete r.v.'s and:

1) $\mathbb{E}[\lambda\cdot X]=\lambda\cdot\mathbb{E}[X]$
2) $\mathbb{E}[X+Y]=\mathbb{E}[X]+\mathbb{E}[Y]$

$\Rightarrow \mathbb{E}[\alpha X+\beta Y]=\alpha\cdot\mathbb{E}[X]+\beta\mathbb{E}[Y]$

→ r.v.'s X and Y do not need to be independent for this!
→ this implies: $\forall n\geq 1$ ($n\in\mathbb{N}$):
$\mathbb{E}[\lambda_1 X_1+\dots+\lambda_n X_n]=\lambda_1\mathbb{E}[X_1]+\dots+\lambda_n\mathbb{E}[X_n]$
$\forall$ integrable discrete r.v.'s $X_1,\dots,X_n:\Omega\to E$ and $\forall\lambda_1,\dots,\lambda_n\in\mathbb{R}$

Expectation value of a Binomial r.v.

Let $n\geq 1$, $0\leq p\leq 1$, $S\sim Bin(n,p)$. Then

$$\mathbb{E}[S]=\sum_{k=0}^{n}k\cdot\binom{n}{k}p^k(1-p)^{n-k}$$

Use linearity of $\mathbb{E}$ and that here, $\mathbb{E}[X_i]=p\ \forall i$ $\Big\{$ S has the same distribution as $S_n=X_1+\dots+X_n$ w. $X_1,\dots,X_n$ iid $\sim Ber(p)$

$$\boxed{\mathbb{E}[S]=\mathbb{E}[S_n]=n\cdot p}$$

Theorem 2.37: Jensen's Inequality:

Let X be a discrete r.v., $\phi:\mathbb{R}\to\mathbb{R}$ a convex function.
If $\mathbb{E}[\phi(X)]$ and $\mathbb{E}[X]$ are well defined, then

$$\boxed{\phi(\mathbb{E}[X])\leq\mathbb{E}[\phi(X)]}$$

→ apply this to $\phi(x)=|x|\Rightarrow\forall$ integrable discrete r.v. X, $\boxed{|\mathbb{E}[X]|\leq\mathbb{E}[|X|]}$

→ apply this to $\phi(x)=x^2\Rightarrow\forall$ discrete r.v. X, $\boxed{\mathbb{E}[|X|]\leq\sqrt{\mathbb{E}[X^2]}}$

2.5 Independence of discrete r.v.'s

Theorem 2.38:

Let X,Y be discrete r.v.'s. Then the following are equivalent:

i) X,Y are independent $\Leftrightarrow$
ii) $\forall a,b\in\mathbb{R}$: $\mathbb{P}[X=a,Y=b]=\mathbb{P}[X=a]\cdot\mathbb{P}[Y=b]$ $\Leftrightarrow$
iii) $\forall f:\mathbb{R}\to\mathbb{R}, g:\mathbb{R}\to\mathbb{R}$: $\mathbb{E}[f(X)g(Y)]=\mathbb{E}[f(X)]\cdot\mathbb{E}[g(X)]$ whenever $\mathbb{E}$'s are well-defined.

Theorem 2.39:

Let $X_1,\dots,X_n$ be discrete r.v.'s. Then the following are equivalent:

i) $X_1,\dots,X_n$ are independent $\Leftrightarrow$
ii) $\forall x_1,\dots,x_n\in\mathbb{R}$: $\mathbb{P}[X_1=x_1,\dots,X_n=x_n]=\mathbb{P}[X_1=x_1]\cdot\dots\cdot\mathbb{P}[X_n=x_n]$
iii) $\forall f_1:\mathbb{R}\to\mathbb{R},\dots,f_n:\mathbb{R}\to\mathbb{R}$ s.t. $f_1(X_1),\dots,f_n(X_n)$ are integrable, $\Leftrightarrow$
$\mathbb{E}[f_1(X_1)\cdot\dots\cdot f_n(X_n)]=\mathbb{E}[f_1(X_1)]\cdot\dots\cdot\mathbb{E}[f_n(X_n)]$

2.6 Variance

Definition 2.40: Variance

Let X be a disc. r.v. s.t. $\mathbb{E}[X^2]<\infty$. Then the Variance of X is:

$$\boxed{\sigma_X^2=\mathbb{E}[(X-m)^2]\quad\text{where}\quad m=\mathbb{E}[X]}$$

Standard Deviation of X: $\boxed{\sigma_X=\sqrt{\sigma_X^2}}$

→ Since $\mathbb{E}[X^2]<\infty\Rightarrow\mathbb{E}[|X|]<\infty$, $m=\mathbb{E}[X]$ is well-defined :)

Interpretation of the standard deviation:
Indicator of how large the fluctuations of X around $m=\mathbb{E}[X]$ are.

Example (Deterministic r.v.): Let $a\in\mathbb{R}$, r.v. $X(\omega)=a\ \forall\omega$.
Then $m=\mathbb{E}[X]=a$ and $\sigma_X^2=\mathbb{E}[(X-m)^2]=0$

Example (Uniform r.v. on 2 points): Let $a<b\in\mathbb{R}$, r.v. X w. distribution $\mathbb{P}[X=a]=\mathbb{P}[Y=b]=\frac{1}{2}$. Then:
$m=\mathbb{E}[X]=\frac{a+b}{2}$ and $\sigma_X=\sqrt{\mathbb{E}[(X-m)^2]}=\frac{a-b}{2}$

Proposition 2.42: Basic Properties of the Variance:

1) Let X be a discrete r.v. with $\mathbb{E}[X^2]<\infty$. Then:

$$\boxed{\sigma_X^2=\mathbb{E}[X^2]-\mathbb{E}[X]^2}$$ Very useful when calculating the Variance!

2) Let $X_1,\dots,X_n$ be pairwise independent r.v.'s, $S:=X_1+\dots+X_n$.
Then: $\boxed{\sigma_S^2=\sigma_{X_1}^2+\dots+\sigma_{X_n}^2}$

Variance of a Binomial r.v.

Let $X_1,\dots,X_n$ iid $\sim Ber(p)$. Then $S:=X_1+\dots+X_n\sim Bin(n,p)$ and

$$\boxed{\sigma_S^2=n\cdot p\cdot(1-p)}$$

→ Important effect of summing iid r.v.'s:
$\mathbb{E}[S]=n\cdot p$ $\quad\sigma_S=\sqrt{n}\cdot\sqrt{p(1-p)}$

⚠ identical distribution means: $\sigma_{X_1}^2+\sigma_{X_2}^2+\dots+\sigma_{X_n}^2=n\cdot\sigma_{X_1}^2$
⚠ independence implicates: $\sigma_S^2=\sigma_{X_1}^2+\dots+\sigma_{X_n}^2$.

Chapter 3: Continuous Distributions

Definition 3.1: Continuous r.v.'s:

A r.v. $X:\Omega\to\mathbb{R}$ is continuous if its distribution function F_X can be written as:

$$\boxed{F_X(a)=\int_{-\infty}^{a}f(x)\,dx}=\mathbb{P}[X\leq a]\quad\forall a\in\mathbb{R}$$

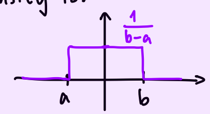
$f(x)=F_X(x)'$

for some nonnegative function $f:\mathbb{R}\to\mathbb{R}_+$ called the Density of X.
⚠ $\mathbb{P}[X=a]=F(a)-F(a-)=0\quad\forall a\in\mathbb{R}$ when X cont. r.v.
⚠ $\mathbb{P}[a<X<b]=\mathbb{P}[a\leq X<b]=\mathbb{P}[a<X\leq b]=\mathbb{P}[a\leq X\leq b]=\int_a^b f(x)\,dx$

Uniform distribution on $[a,b]$, $a<b$:

A cont. r.v. X is uniform in $[a,b]$ if its density is:

$$\boxed{f_{a,b}(x)=\begin{cases}\frac{1}{b-a} & x\in[a,b]\\ 0 & x\notin[a,b]\end{cases}}$$

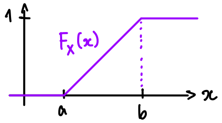


Notation: $\boxed{X\sim\mathcal{U}([a,b])}$

Intuition: X represents an uniformly chosen point in $[a,b]$.
Properties of an uniform r.v. X in $[a,b]$:
→ The probability to fall in an interval $[c,c+l]\subset[a,b]$ depends only on its length l: $\boxed{\mathbb{P}[X\in[c,c+l]]=\frac{l}{b-a}}$

→ The distribution function of X is:

$$F_X(x) = \begin{cases} 0 & , x<a \\ \frac{x-a}{b-a} & , a \le x \le b \\ 1 & , x>b \end{cases}$$



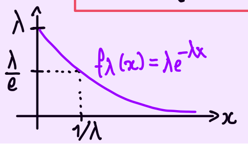
Exponential distribution with parameter $\lambda > 0$

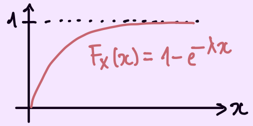
Exp. dist. in cont. $\hat{=}$ geom. dist. in disc.!

Cont. r.v. T is exponential w. param. $\lambda>0$ if its density is:

$$f_\lambda(x) = \begin{cases} \lambda e^{-\lambda x} & , x \ge 0 \\ 0 & , x<0 \end{cases}$$

Notation: $T \sim \exp(\lambda)$





Intuition/Application:

T represents the time of an "alarm clock ring"

Properties of an exp. r.v. T w. param. λ:

→ The waiting prob. is exponentially small: $\forall t \ge 0: \mathbb{P}[T<t] = e^{-\lambda t}$

→ It has the absence of memory property:

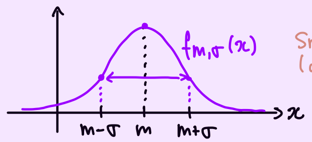
$\forall t,s \ge 0: \mathbb{P}[T>t+s \mid T>t] = \mathbb{P}[T>s]$

Normal distribution with parameter $m \in \mathbb{R}$ and $\sigma^2>0$

A cont. r.v. X is normal w. param. m and $\sigma^2>0$ if its density is:

$$f_{m,\sigma}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

Notation: $X \sim \mathcal{N}(m,\sigma^2)$



Small $\sigma \,\hat{=}$ small fluctuations of X (d.h. X is typically close to m)

Properties of normal r.v.'s:

→ if $X_1, \dots, X_n$ indep. normal r.v.'s w. param $(m_1,\sigma_1^2), \dots, (m_n,\sigma_n^2)$ then

$\boxed{Z := m_0 + \lambda_1 X_1 + \lambda_2 X_2 + \dots + \lambda_n X_n \sim \mathcal{N}(m,\sigma^2)}$ where

$\boxed{m = m_0 + m_1 + \dots + m_n}$, $\boxed{\sigma^2 = \lambda_1^2\sigma_1^2 + \dots + \lambda_n^2\sigma_n^2}$

→ Standard Normal r.v.: $X \sim \mathcal{N}(0,1) \Rightarrow Z := m + \sigma X \sim \mathcal{N}(m,\sigma^2)$

→ if $X \sim \mathcal{N}(m,\sigma^2)$, then all the "probability mass" is certainly in $[m-3\sigma, m+3\sigma]$.

⚠ Every $X \sim \mathcal{N}(m,\sigma^2)$ can be transformed to $Z \sim \mathcal{N}(0,1)$:

Let $X \sim \mathcal{N}(m,\sigma^2)$. Set $\boxed{Z := \frac{X-m}{\sigma}}$. Then $Z = \frac{1}{\sigma}X - \frac{m}{\sigma} \sim \mathcal{N}(0,1)$.

$$\Rightarrow \mathbb{P}[|X-m| \ge 3\sigma] = \mathbb{P}\left[\frac{|X-m|}{\sigma} \ge 3\right] = \underbrace{\mathbb{P}[|Z| \ge 3]}$$

Read off directly from Table ($\exists$ only for $\mathcal{N}(0,1)$).

Theorem 3.2:

Let X be a r.v. Assume the dist. fct. F_X is cont. and piecewise C^1, i.e. $\exists x_0 = -\infty < x_1 < \dots < x_{n-1} < x_n = +\infty$ s.t. F_X is C^1 on every interval (x_i, x_{i+1}). Then X is a cont. r.v. and a density f can be constructed by defining: $\forall x \in (x_i, x_{i+1})$: $\boxed{f(x) = F_X'(x)}$ ← Ableitung!

and setting arbitrary values @ $x_1, \dots, x_{n-1}$

Kochrezept: How to compute the density of $Y = \phi(X)$:

Let X be a cont. r.v. w. density f, $\phi: \mathbb{R} \to \mathbb{R}$ some fct., $Y = \phi(X)$. (X transformed). Assume X and ϕ are sufficiently "nice" (i.e. cont., ...)

1) Compute the distribution function:

$$\boxed{F_Y(x) = \mathbb{P}[\phi(X) \le x] = \dots}$$

2) If F_Y is cont. and piecewise cont. ⇒ Y is cont. and its density is:

$$\boxed{f(x) = F_Y'(x)}$$

on each interval where $F_Y'(x)$ is well-defined.

Example: Density of $Y = X^2$, $X \sim \mathcal{N}(0,1)$

1) $F_Y(x) = 0 \;\; \forall x<0$ (because y is never <0)

$F_Y(x) = \mathbb{P}[X^2 \le x] = \mathbb{P}[-\sqrt{x} \le X \le \sqrt{x}] = \Phi(\sqrt{x}) - \Phi(-\sqrt{x}) \;\; \forall x \ge 0$

where $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2}\, dy$

$F_Y(x)$ is cont. & piecewise C^1 (✓)

2) $f(x) = \begin{cases} F_Y'(x) = 0 & \forall x<0 \\ F_Y'(x) = \frac{1}{2\sqrt{x}}\left(\Phi'(\sqrt{x}) + \Phi'(-\sqrt{x})\right) = \frac{1}{\sqrt{2\pi x}} e^{-x/2} & \forall x \ge 0 \end{cases}$

Example: Density of $Y = X^n$, $X \sim \mathcal{U}([0,1])$

1) $F_Y(x) = \begin{cases} \mathbb{P}[Y \le x] = \mathbb{P}[X^n \le x] = \mathbb{P}[X \le x^{1/n}] = x^{1/n} & , \forall x \in (0,1] \\ 0 & , \forall x \le 0 \\ 1 & , \forall x > 1 \end{cases}$ → cont. & piecewise C^1 (✓)

2) $f(x) = F_Y'(x) = \begin{cases} 0 & , \forall x \le 0 \\ \frac{1}{n} x^{\frac{1}{n}-1} & , \forall x \in (0,1] \\ 0 & , \forall x>1 \end{cases}$

3.2 Expectation and Variance

Definition 3.2: Expectation of a positive r.v.:

Let $X: \Omega \to \mathbb{R}$ be a cont. r.v. w. density f and assume $X \ge 0$ a.s. Then:

Expectation of X: $\boxed{\mathbb{E}[X] = \int_{-\infty}^{\infty} x \cdot f(x)\, dx}$ ⚠ can be finite or infinite!

↳ also If $!X \ge 0$ but $\mathbb{E}[|X|] < \infty$ (X is integrable)

Definition 3.6: Expectation of the image

Let X be a r.v. w. density f, $\phi: \mathbb{R} \to \mathbb{R}$ s.t. $\int_{-\infty}^{\infty} |\phi(x)| f(x)\, dx < \infty$.

Then the expectation of $\phi(X)$ is:

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

Expectation value of a uniform r.v.

Let $X \sim \mathcal{U}([a,b])$, $a<b$. Then $\boxed{\mathbb{E}[X] = \frac{1}{b-a} \int_a^b x\, dx = \frac{a+b}{2}}$

Expectation value of a exp. r.v.

Let $X \sim \exp(\lambda)$, $\lambda>0$. Then $\boxed{\mathbb{E}[X] = \int_0^\infty x \lambda e^{-\lambda x}\, dx = \frac{1}{\lambda}}$

Expectation value of a normal r.v.

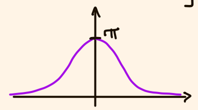
Let $X \sim \mathcal{N}(m,\sigma^2)$. Then X has the same distribution as $m + \sigma Y$, $Y \sim \mathcal{N}(0,1)$. Hence: $\mathbb{E}[X] \overset{\uparrow \text{Def. 3.6}}{=} \mathbb{E}[m+\sigma Y] = m + \sigma\, \mathbb{E}[Y]$

and $\mathbb{E}[Y] = \int_{-\infty}^{\infty} \underbrace{x \cdot f_{0,1}(x)}_{\text{odd}}\, dx = 0 \Rightarrow \boxed{\mathbb{E}[X] = m}$

Expectation value of a Cauchy Distribution

A r.v. X has a Cauchy Dist. if it is cont. and has the density

$$f(x) = \frac{1}{\pi} \cdot \frac{1}{1+x^2} \quad , x \in \mathbb{R}$$



In this case, X is not integrable!

$\Rightarrow \mathbb{E}[|X|] = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{|x|}{1+x^2}\, dx = +\infty$ So, the $\mathbb{E}$ of X is not well defined

Theorem 3.8: Jensen's Inequality:

Let $\phi: \mathbb{R} \to \mathbb{R}$ be a convex fct., X a cont. r.v. s.t. $\mathbb{E}[X]$ and $\mathbb{E}[\phi(X)]$ are well defined. Then:

$$\boxed{\phi(\mathbb{E}[X]) \le \mathbb{E}[\phi(X)]}$$

Definition 3.9: Variance of a cont. r.v.

Let $X: \Omega \to \mathbb{R}$ be a cont. r.v. w. density f. If $\mathbb{E}[X^2] < \infty$, the Variance of X is:

$\boxed{Var(X) = \sigma_X^2 = \mathbb{E}[(X-m)^2] = \int_{-\infty}^{\infty} (x-m)^2 f(x)\, dx}$ where $\boxed{m = \mathbb{E}[X]}$

Standard deviation of X: $\boxed{\sigma_X = \sqrt{\sigma_X^2}}$

Proposition 3.11: Basic properties of the Variance

1) Let X be a cont. r.v. w. $\mathbb{E}[X^2] < \infty$. Then: $\boxed{\sigma_X^2 = \mathbb{E}[X^2] - \mathbb{E}[X]^2}$

2) Let X be a cont. r.v. w. $\mathbb{E}[X^2] < \infty$, $\lambda, \mu \in \mathbb{R}$. Then: $\boxed{\sigma_{\lambda X + \mu}^2 = \lambda^2 \cdot \sigma_X^2}$

3) Let $X_1, \dots, X_n$ be pairwise indep. cont. r.v. w. $\mathbb{E}[X_i^2] < \infty$, $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ and set $S := \lambda_1 X_1 + \dots + \lambda_n X_n$. Then:

$$\boxed{Var(S) = \sigma_S^2 = \lambda_1^2 \sigma_{X_1}^2 + \dots + \lambda_n^2 \sigma_{X_n}^2}$$

Variance of a uniform r.v.

Let $X \sim \mathcal{U}([a,b])$. Then X has the same distribution fct. as $a + (b-a)Y$, $Y \sim \mathcal{U}([0,1])$.

$\mathbb{E}[Y] = \frac{1}{2} = m \Rightarrow \sigma_Y^2 = \int_0^1 x^2\, dx - m^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$

$\overset{\text{prop. 3.11}}{\Longrightarrow} \boxed{\sigma_X^2 = \frac{1}{12}(b-a)^2}$

Variance of a exponential r.v.:

$\mathbb{E}[X^2] = \int_0^\infty x^2 \cdot \lambda e^{-\lambda x}\, dx = \frac{2}{\lambda^2}$ (integration by parts) Use formula $\sigma_X^2 = \mathbb{E}[X^2] - \mathbb{E}[X]^2$ $\boxed{\sigma_X^2 = \frac{1}{\lambda^2}}$

Variance of a normal r.v.:

Let $X \sim \mathcal{N}(m, \sigma^2)$. Use that X has same distribution as $m + \sigma Y$, $Y \sim \mathcal{N}(0,1)$.

Since $\sigma_Y^2 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-x^2/2} dx = 1 \xrightarrow[\text{Prop. 3.11}]{} \boxed{\sigma_X^2 = \sigma^2}$ $\sigma_X^2 = \sigma^2 \cdot \sigma_Y^2$

3.3 Joint Distribution

Definition 3.12: Continuous joint distribution:

Two r.v.'s $X, Y: \Omega \to \mathbb{R}$ have a continuous joint distribution if $\exists$ a fct. $f: \mathbb{R}^2 \to \mathbb{R}_+$ s.t.

$$\boxed{\mathbb{P}[X \in [a,a'], Y \in [b,b']] = \int_a^{a'} \left(\int_b^{b'} f(x,y)\, dy \right) dx}$$

$\forall -\infty < a < a' < +\infty$ and $\forall -\infty < b < b' < +\infty$

$\boxed{f(x,y) = \text{joint density of } (X,Y)}$

Remark 3.13:

A joint density always satisfies: $\boxed{\int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} f(x,y)\, dy \right) dx = 1}$

Also: Given a nonnegative function f satisfying this ↑, one can always construct a prob. space $(\Omega, \mathcal{F}, \mathbb{P})$ and two r.v.'s $X, Y: \Omega \to \mathbb{R}$ with joint density f.

Interpretation: $f(x,y)\, dx\, dy$ represents the prob. that the random point (X,Y) lies in the small rectangle $[x, x+dx] \times [y, y+dy]$.

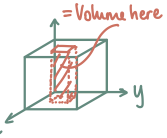
Example: Uniform point in the square:

Consider two r.v.'s X and Y with joint density $f(x,y) = \mathbb{1}_{0 \le x,y \le 1}$, i.e.

$$f(x,y) = \begin{cases} 1 & (x,y) \in [0,1]^2 \\ 0 & (x,y) \notin [0,1]^2 \end{cases}$$

Then we have $\forall$ rectangle $R = [a,a'] \times [b,b'] \subset [0,1]^2$:

$\mathbb{P}[(X,Y) \in R] = (a'-a) \cdot (b'-b) = \text{Area}(R)$



Example: Uniform point in the disk:

Let $D := \{(x,y) : x^2 + y^2 \le 1\}$. Consider two r.v.'s X and Y with joint density $f(x,y) = \frac{1}{\pi} \mathbb{1}_{x^2+y^2 \le 1}$ i.e.

$$f(x,y) = \begin{cases} \frac{1}{\pi} & x^2+y^2 \le 1 \\ 0 & x^2+y^2 > 1 \end{cases}$$

Then $\forall$ rectangle $R = [a,a'] \times [b,b'] \subset D$:

$\mathbb{P}[(X,Y) \in R] = \frac{1}{\pi}(a'-a) \cdot (b'-b) = \frac{\text{Area}(R)}{\text{Area}(D)}$

(X,Y): represents a uniform point in D.

Marginal densities:

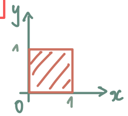
If X, Y possess a joint density f_{XY}, we have:

$$\boxed{\mathbb{P}[X \le a] = \mathbb{P}[X \in [-\infty, a], Y \in [-\infty, \infty]] = \int_{-\infty}^{a} \left(\int_{-\infty}^{\infty} f(x,y)\, dy \right) dx}$$

Hence, X is continuous with density: $\boxed{f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\, dy}$

And Y is continuous with density: $\boxed{f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)\, dx}$

Example: Marginal density of uniform point in square:



If $f_{XY}(x,y) = \mathbb{1}_{0 \le x,y \le 1}$, X has density

$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\, dy = \int_0^1 \mathbb{1}_{0 \le x \le 1} \mathbb{1}_{0 \le y \le 1}\, dy = \mathbb{1}_{0 \le x \le 1}$

and analogously, Y has density: $f_Y(y) = \mathbb{1}_{0 \le y \le 1}$

Example: Marginal density of uniform point in disk

If $f_{XY}(x,y) = \frac{1}{\pi} \mathbb{1}_{x^2+y^2 \le 1}$, X has density

$f_X(x) = \int_{-\infty}^{\infty} \frac{1}{\pi} \mathbb{1}_{x^2+y^2 \le 1}\, dy = \frac{2}{\pi} \sqrt{1-x^2}$ and Y has density

$f_Y(y) = \frac{2}{\pi} \sqrt{1-y^2}$ (aus Symmetriegründen: no need to calculate integral!)

Definition: Expectation of $\phi(X,Y)$:

Let $\phi: \mathbb{R}^2 \to \mathbb{R}$. If X, Y have joint density $f_{X,Y}$, the $\mathbb{E}$ of the r.v. $Z := \phi(X,Y)$ is:

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

Linearity of $\mathbb{E}$ for jointly continuous distributions:

$\forall \lambda, \mu \in \mathbb{R}$; X, Y cont. r.v. w. joint dist. f_{XY}:

$$\boxed{\mathbb{E}[\lambda X + \mu Y] = \lambda \cdot \mathbb{E}[X] + \mu \mathbb{E}[Y]}$$

Theorem 3.14: Independence for continous r.v.'s:

Let X, Y be two cont. r.v.'s w. densities f_X, f_Y. The following are equivalent:

1) X, Y are independent
2) X, Y are jointly continuous with joint density: $\boxed{f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)}$
3) $\forall \phi: \mathbb{R} \to \mathbb{R}, \psi: \mathbb{R} \to \mathbb{R}$: $\boxed{\mathbb{E}[\phi(X)\psi(Y)] = \mathbb{E}[\phi(X)] \cdot \mathbb{E}[\psi(Y)]}$

Whenever the $\mathbb{E}$ are well-defined.

⚠ Important consequence:

two independent continuous r.v.'s are automatically jointly continuous!

Example: Uniform point in the square $[0,1]^2$:

If X, Y have joint density $f_{XY}(x,y) = \mathbb{1}_{0 \le x,y \le 1}$

$\Rightarrow f_{XY}(x,y) = \mathbb{1}_{0 \le x \le 1} \cdot \mathbb{1}_{0 \le y \le 1} = f_X(x) \cdot f_Y(y)$

the 2 coordinates of a uniform random point in $[0,1]^2$ are independent!

So, X, Y are jointly continuous :)

Example: Uniform point in the disk:

X, Y have joint density $f_{XY} = \frac{1}{\pi} \mathbb{1}_D$, and $\begin{cases} f_X(x) = \frac{2}{\pi}\sqrt{1-x^2} \\ f_Y(y) = \frac{2}{\pi}\sqrt{1-y^2} \end{cases}$

$\Rightarrow f_{XY}(x,y) \neq f_X(x) \cdot f_Y(y)$

the 2 coordinates X and Y of a uniform point in D are not independent!

So, X, Y are not jointly continuous.

Asymptotic Results

Fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and an infinite sequence of i.i.d. r.v. $X_1, X_2, \dots$

i.e. let $X_i: \Omega \to \mathbb{R}$: $\forall i_1 < \dots < i_k$, $\forall x_1, \dots, x_k \in \mathbb{R}$: $\mathbb{P}[X_{i_1} \le x_1, \dots, X_{i_k} \le x_k] = F(x_1) \cdots F(x_k)$

w. F = common distribution function.

Empirical Average: behavior for n large of the r.v.:

$$\boxed{U_n := \frac{X_1(\omega) + \dots + X_n(\omega)}{n}}$$

Law of large numbers:

Theorem 3.16:

Assume $\mathbb{E}[|X_1|]$ well defined & finite, define $m := \mathbb{E}[X_1]$. Then:

$$\boxed{\lim_{n \to \infty} \frac{X_1 + \dots + X_n}{n} = m \text{ a.s.}}$$

Example: Bernoulli r.v.'s:

Let $X_1, X_2, \dots$ infinite sequence of i.i.d. Bernoulli r.v. w. param p.

Then: $\lim_{n \to \infty} \frac{X_1 + \dots + X_n}{n} = p$ a.s.

Example:

Let $T_1, T_2, \dots$ infinite sequence of i.i.d. Exponential r.v.'s w. param λ

Then: $\lim_{n \to \infty} \frac{T_1 + \dots + T_n}{n} = \frac{1}{\lambda}$ a.s.

Central limit theorem:

Answers the question: how far is $\frac{X_1 + \dots + X_n}{n}$ from m typically?

The Gaussian Case:

Let $X_1, X_2, \dots$ a sequence of i.i.d. $\sim \mathcal{N}(m, \sigma^2)$. Then:

$$Z = \frac{X_1 + \dots + X_n}{n} - m \sim \mathcal{N}\left(0, \frac{\sigma^2}{n}\right)$$

Standard deviation $\frac{\sigma}{\sqrt{n}}$ represents the typical fluctuations of Z

$\Rightarrow$ typical distance betw. $\frac{X_1 + \dots + X_n}{n}$ and m is of order $\frac{\sigma}{\sqrt{n}}$

$\Rightarrow$ rescale Z: $\frac{\sqrt{n}}{\sigma} Z$ has fluctuations of order 1, d.h. $\frac{\sqrt{n}}{\sigma} Z \sim \mathcal{N}(0,1)$

So: If we consider i.i.d. r.v.'s $\sim \mathcal{N}(m, \sigma^2)$, then the r.v.

$$\frac{X_1 + \dots + X_n - n \cdot m}{\sqrt{\sigma^2 n}} \sim \mathcal{N}(0,1).$$

General Case:

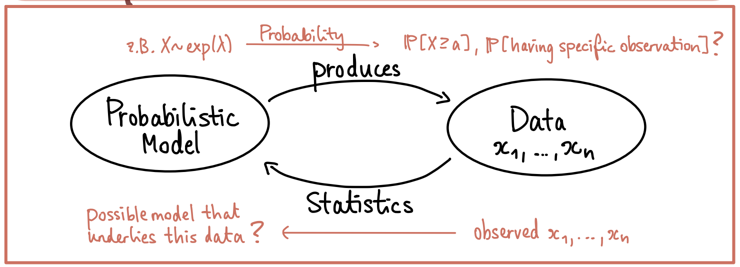
If X_1, X_2 are not normal: $\frac{X_1 + \dots + X_n - n \cdot m}{\sqrt{\sigma^2 n}} \xrightarrow[\text{Th.}]{n \to \infty} \sim \mathcal{N}(0,1)$

Theorem 3.18: Central limit theorem:

Assume $\mathbb{E}[X_1^2]$ well defined and finite, define $m := \mathbb{E}[X_1]$, $\sigma^2 := \text{Var}(X_1)$ and $S_n := X_1 + \dots + X_n$. Then:

$$\boxed{\mathbb{P}\left[\frac{S_n - nm}{\sqrt{\sigma^2 n}} \le a\right] \xrightarrow[n \to \infty]{} \Phi(a) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-x^2/2}\, dx}$$

Chapter 4: Introduction to statistics



4.1 Estimation of parameters

Goal: We want to estimate parameters, given the observed data.

Our framework: Sequence of i.i.d. r.v.'s $X_1, \dots, X_n$ w. distribution depending on one parameter θ or more parameters $\theta_1, \dots, \theta_n$ that we want to estimate.

Examples of discrete models:
- $X_1, \dots, X_n$ i.i.d. $\sim Ber(p)$, $p \in [0,1]$
- $X_1, \dots, X_n$ i.i.d. $\sim Geom(p)$, $p \in [0,1]$
- $X_1, \dots, X_n$ i.i.d. $\sim Poiss(\lambda)$, $\lambda > 0$

Examples of continuous models:
- $X_1, \dots, X_n$ i.i.d. $\sim \mathcal{U}([0,\theta])$, $\theta \geq 0$
- $X_1, \dots, X_n$ i.i.d. $\sim \exp(\lambda)$, $\lambda > 0$
- $X_1, \dots, X_n$ i.i.d. $\sim \mathcal{N}(m, \sigma^2)$, $m \in \mathbb{R}$, $\sigma^2 > 0$

Realization of a model: A vector $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ of possible values for $(X_1, \dots, X_n)$

Main Question of Chapter: We observe a realization $(x_1, \dots, x_n)$ of our model. Can one estimate the underlying parameter(s) of the model?

4.1.1 Maximum likelihood estimator for discrete models

Consider $X_1, \dots, X_n$ i.i.d. discrete r.v.'s w. values in a finite or countable set E, and distribution $(p_\theta(x))$ depending on some parameter $\theta \in \mathbb{R}$.

$$\forall i, \ \forall x \in E, \ \mathbb{P}[X_i = x] = p_\theta(x)$$

Realization (observed) $(x_1, \dots, x_n) \longrightarrow \theta$?

Goal: define $\hat{\theta}(x_1, \dots, x_n)$ = good prediction of θ

Definition 4.1: likelihood function of a realization

Let $x = (x_1, \dots, x_n) \in E^n$ be a possible Realization for $(X_1, \dots, X_n)$.
likelihood function of x := $L(\theta) = L_x(\theta) = \mathbb{P}[X_1 = x_1, \dots, X_n = x_n]$
Likelihood of $x \mathrel{\hat{=}}$ probability to observe the realization $x = (x_1, \dots, x_n)$ when the underlying parameter is θ! $\hat{=}$ in probabilistic terms to the **joint distribution** of $X_1, \dots, X_n$!
for i.i.d. r.v.'s (for us always the case!), likelihood function is:

$$\boxed{L_x(\theta) = \mathbb{P}[X_1 = x_1] \cdots \mathbb{P}[X_n = x_n] = p_\theta(x_1) \cdots p_\theta(x_n)}$$

Definition 4.2: Maximum likelihood estimator

The maximum likelihood estimator for a realization x is the parameter $\hat{\theta} = \hat{\theta}(x_1, \dots, x_n)$ for which the likelihood is maximal:

$$\boxed{L_x(\hat{\theta}) = \max_\theta L_x(\theta)}$$

Example: Bernoulli model

Let $X_1, \dots, X_n \sim Ber(p)$. A possible realization: $x = (x_1, \dots, x_n) \in \{0,1\}^n$

$\Rightarrow$ likelihood fct: $\boxed{L_x(p) = \mathbb{P}[X_1 = x_1, \dots, X_n = x_n] = p^{|x|}(1-p)^{n-|x|}}$

where $|x| = \sum_i x_i$

Now, we want to find $\hat{p}$ s.t. $L_x(\hat{p}) = \max_{0 \leq p \leq 1} L_x(p)$:

$\Rightarrow$ easier to study $\phi(p) = \log(L_x(p)) = |x| \cdot \log(p) + (n - |x|)\log(1-p)$ than $L_x(p)$ (Use Trick of taking the log for $L(\theta)$ in form of a product (always the case for i.i.d. r.v.'s!))

$\Rightarrow \phi'(p) = \frac{|x|}{p} - \frac{n - |x|}{1-p} \overset{!}{=} 0$ (find max. $\to \phi$ has unique max.!)

$\Rightarrow$ Maximum likelihood estimator for p: $\boxed{\hat{p}(x_1, \dots, x_n) = \frac{x_1 + \dots + x_n}{n}}$

4.1.2 Maximum likelihood estimator for continuous models

Consider $X_1, \dots, X_n$ i.i.d. continuous r.v.'s w. density f_θ depending on some parameter $\theta \in \mathbb{R}$. (or more parameters $\theta_1, \theta_2, \dots$)

Goal: define a natural estimator for the parameter θ.

Definition 4.3: likelihood function of a realization

Let $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ be a possible realization for $(X_1, \dots, X_n)$. The likelihood function of x is: $\boxed{L(\theta) = L_x(\theta) = f_\theta(x_1) \cdots f_\theta(x_n)}$

$\rightarrow$ Intuitively: $f_\theta(x_1) \cdots f_\theta(x_n) \mathrel{\hat{=}}$ probability that $(X_1, \dots, X_n) \simeq (x_1, \dots, x_n)$

Definition 4.4: Maximum likelihood estimator:

The maximum likelihood estimator for the realization $x = (x_1, \dots, x_n)$ is the parameter $\hat{\theta} = \hat{\theta}(x_1, \dots, x_n)$ s.t. the likelihood is maximal:

$$\boxed{L_x(\hat{\theta}) = \max_\theta L_x(\theta)}$$

$\rightarrow$ When model involves more parameters $\theta_1, \theta_2, \dots$:
- $\hookrightarrow$ likelihood function of the form $L_x(\theta_1, \theta_2, \dots)$
- $\hookrightarrow$ maximum likelihood estimators for a realization $x = (x_1, \dots, x_n)$ = parameters $\hat{\theta}_1, \hat{\theta}_2, \dots$ s.t. likelihood is maximal:

$$\boxed{L_x(\hat{\theta}_1, \hat{\theta}_2, \dots) = \max_{\theta_1, \theta_2, \dots} L_x(\theta_1, \theta_2, \dots)}$$

Example 1: Normal model

Let $X_1, \dots, X_n$ be i.i.d. $\sim \mathcal{N}(m, \sigma^2)$. Possible realization: $x = (x_1, \dots, x_n) \in \mathbb{R}^n$.

$\Rightarrow$ likelihood function: $L_x(m, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\sum_{i=1}^n \frac{(x_i - m)^2}{2\sigma^2}\right\}$

$\Rightarrow$ maximum likelihood estimator: find $\hat{m}, \hat{\sigma}^2$ s.t.

$$L_x(\hat{m}, \hat{\sigma}^2) = \max_{m \in \mathbb{R}, \sigma^2 > 0} L_x(m, \sigma^2)$$

$\hookrightarrow$ look for the critical points of the function $\log(L)$:

$$\begin{cases} \frac{\partial \log(L)}{\partial m} = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - m) \\ \frac{\partial \log(L)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - m)^2 \end{cases}$$

$\hookrightarrow$ L has a unique maximum!

Maximum likelihood estimators of m and σ^2:

$$\boxed{\hat{m}(x_1, \dots, x_n) = \frac{x_1 + \dots + x_n}{n}}$$

$$\boxed{\hat{\sigma}^2(x_1, \dots, x_n) = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{m}(x_1, \dots, x_n))^2}$$

4.2 Confidence Intervals

$\hookrightarrow$ Goal: find out "how good" the estimators from chapter 4.1 are

Definition 4.7: z%- confidence interval for θ

Consider a probabilistic model w. underlying parameter θ.
For $0 \leq z \leq 100$, a z%- confidence interval for θ associates to each realization $x = (x_1, \dots, x_n)$ an interval $I = [a(x), b(x)] \subset \mathbb{R}$ s.t.

$$\forall \theta, \ \mathbb{P}[\underbrace{a(X_1, \dots, X_n) \leq \overset{\text{not random!}}{\theta} \leq b(X_1, \dots, X_n)}_{\text{random elements}}] \geq \frac{z}{100}$$

Example: Exact confidence interval for the normal model w. $\sigma^2 = 1$

Let $X_1, \dots, X_n$ be i.i.d $\sim \mathcal{N}(m, \sigma^2 = 1)$, m unknown.

$\Rightarrow$ Maximum likelihood estimator for m: $\hat{m}(x_1, \dots, x_n) = \frac{x_1 + \dots + x_n}{n}$

$\Rightarrow$ We're looking for a confidence interval for m of the form:

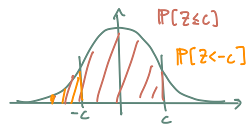
$$I(x) = \left[\hat{m}(x) - \frac{c}{\sqrt{n}}, \hat{m}(x) + \frac{c}{\sqrt{n}}\right], \ c > 0 \text{ constant}.$$

Since $\mathbb{P}[\hat{m}(X_1, \dots, X_n) - \frac{c}{\sqrt{n}} \leq m \leq \hat{m}(X_1, \dots, X_n) + \frac{c}{\sqrt{n}}] = \mathbb{P}[-c \leq Z \leq c]$

w. $Z = \frac{X_1 + \dots + X_n - nm}{\sqrt{n}} \sim \mathcal{N}(0,1)$

$\Rightarrow \mathbb{P}[-c \leq Z \leq c] = \mathbb{P}[Z \leq c] - \mathbb{P}[Z < -c] = 2\Phi(c) - 1$

w. $\Phi(c) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^c e^{-x^2/2} \, dx$



Check in Table: $2\Phi(1.96) - 1 \geq 0.95$

$\Rightarrow$ So, by choosing $c = 1.96$ we get:

$$\mathbb{P}\left[\hat{m}(X_1, \dots, X_n) - \frac{1.96}{\sqrt{n}} \leq m \leq \hat{m}(X_1, \dots, X_n) + \frac{1.96}{\sqrt{n}}\right] \geq \frac{95}{100}$$

$\Rightarrow I(x) = \left[\hat{m}(x) - \frac{1.96}{\sqrt{n}} \leq m \leq \hat{m}(x) + \frac{1.96}{\sqrt{n}}\right]$ is a 95%-confidence interval for m

(d.h. der echte Wert von m befindet sich zu 95% in $I(x)$)

z.B. Temperature at which water boils, when each measurement can be modeled by a $\mathcal{N}(m, 1)$

4.3 Statistical tests

Goal: introduce the notion of test, z.B. used to decide if a coin is fair or not.

Theory of testing: We have general hypothesis on the probabilistic model.

Goal of test: accept/reject these hypothesis by analyzing some realization of the model.

Framework:

n i.i.d. r.v.'s $X_1, \dots, X_n$ whose distribution depends on some parameter θ. Consider two Hypotheses of the form:

Null hypothesis: $H_0: \theta = \theta_0$
Alternative hypothesis: $H_1: \theta = \theta_1$ ← two possible values for the parameter

=> Given a certain realization of our model $x = (x_1, \dots, x_n)$, a test aims to decide whether the hypothesis H_0 is accepted/rejected in favor of H_1.

→ Test: given by a function $\boxed{d = d(x_1, \dots, x_n) \in \{0,1\}}$

$d(x)=0$: H_0 is accepted $d(x)=1$: H_0 is rejected

Errors that occur when setting up a test:

Type I error: Reject H_0 when actually it is true

This type of error is controlled by the relevance level of the test:

$$\alpha := \mathbb{P}_{H_0}[\underbrace{d(X_1, \dots, X_n) = 1}_{\text{reject } H_0}]$$

$\mathbb{P}_{H_0}$ = probability measure when the hypothesis H_0 is satisfied (d.h. $\theta = \theta_0$)

Type II error: Accept H_0 when actually, H_1 occurs:

This type of error is denoted by β and gives rise to the

power of the test $:= 1-\beta = 1 - \mathbb{P}_{H_1}[\underbrace{d(X_1, \dots, X_n) = 0}_{\text{accept } H_0}]$

Principles when choosing the test and the hypothesis:

→ Main Goal: Avoid Type I error (controlled by α)
→ Secondary Goal: Avoid Type II error → look for most powerful test after fixing α.

Example: Eddy's Problem: (Good/Bad Coin?)

$X_1, \dots, X_{10} \sim Ber(p)$

Choose 2 hypotheses for the test:
1. priority: Avoid bringing bad coins to the casino
2. priority: Avoid throwing good coins away

$$\Rightarrow \begin{cases} H_0: & p = 0.7 \\ H_1: & p = 0.5 \end{cases}$$

This means:
H_0 rejected ≙ Eddy concludes that the coin is not bad
H_0 accepted ≙ Eddy concludes that the coin is bad.

Here, we have:
Type I error ≙ bringing a defected coin to the casino
Type II error ≙ throwing a good coin away.

Setting up a test with likelihood ratios:

For us: distribution of $X_1, \dots, X_n$ completely determined under H_0 & H_1.

=> Set up a test using the method of likelihood ratios:

① Fix a threshold $c > 0$

② For each possible realization $x = (x_1, \dots, x_n)$ calculate the ratio

$$r(x) = \frac{\mathbb{P}_{H_0}[X_1 = x_1, \dots, X_n = x_n]}{\mathbb{P}_{H_1}[X_1 = x_1, \dots, X_n = x_n]} = \frac{L_x(\theta_0)}{L_x(\theta_1)}$$

③ Define $d(x) = \begin{cases} 0 & r(x) > c \\ 1 & r(x) \le c \end{cases}$ d.h. we reject H_0 if $r(x) \le c$.

④ Relevance level of the test: $\alpha = \mathbb{P}_{H_0}[r(X_1, \dots, X_n) \le c]$

⑤ Power of the test: $1-\beta = 1 - \mathbb{P}_{H_1}[r(X_1, \dots, X_n) > c] = \mathbb{P}_{H_1}[r(X_1, \dots, X_n) \le c]$

Theorem 4.9: Neyman-Pearson's Lemma:

Consider a statistic framework with two simple hypothesis H_0 and H_1. Among all the tests with relevance level α, the likelihood ratio test is the most powerful.

Using this for Eddy's problem:

For a bad coin ($p=0.7$): likelihood of $x = (x_1, \dots, x_n)$ is:

$$\mathbb{P}_{H_0}(X = x) = L_x(0.7) = 0.7^{|x|}\, 0.3^{10-|x|}$$

For a fair coin ($p = 0.5$): likelihood of $x = (x_1, \dots, x_n)$ is:

$$\mathbb{P}_{H_1}(X = x) = L_x(0.5) = 0.5^{10}$$

The ratio $r(x) = \frac{\mathbb{P}_{H_0}}{\mathbb{P}_{H_1}}$ is equal to: $r(x) = 0.6^{10} \left(\frac{7}{3}\right)^{|x|}$

Represent this in a table:

Tabelle aus Skript

How to "read" this table:

If Eddy chooses $c=1$: he will reject H_0 if $|X| \le 6$. In this case:

Relevance level of test: $\alpha = \mathbb{P}_{H_0}[\text{reject } H_0] = \mathbb{P}_{H_0}[|X| \le 6] \simeq 0.3503$

Power of the test: $1-\beta = \mathbb{P}_{H_1}[\text{reject } H_0] = \mathbb{P}_{H_1}[|X| \le 6] \simeq 0.8228$

↪ D.h. w. choice $c=1$, 35% of the coins Eddy brings to the casino are bad, but 82% of the good coins will be brought to the casino.

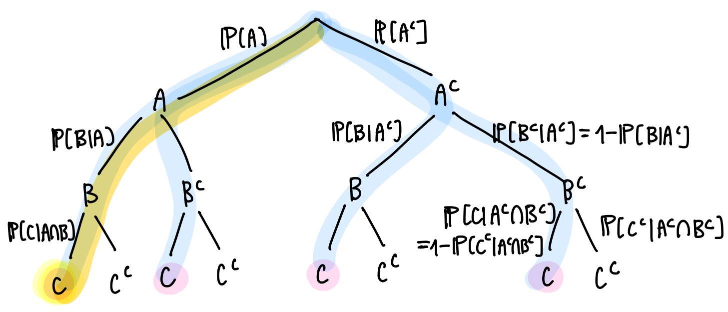
If Eddy chooses $c = 0.1$: he will reject H_0 if $|X| \le 3$. In this case:

Relevance level of test: $\alpha = \mathbb{P}_{H_0}[\text{reject } H_0] = \mathbb{P}_{H_0}[|X| \le 3] \simeq 0.0101$

Power of the test: $1-\beta = \mathbb{P}_{H_1}[\text{reject } H_0] = \mathbb{P}_{H_1}[|X| \le 3] \simeq 0.1719$

↪ D.h. w. choice $c = 0.1$, 1% of the coins Eddy brings to the casino are bad, but only 17% of the good coins will be brought to the casino!

Probability tree:



⇒ $\mathbb{P}[C] = \mathbb{P}[A \cap B \cap C] + \mathbb{P}[A \cap B^c \cap C] + \mathbb{P}[A^c \cap B \cap C] + \mathbb{P}[A^c \cap B^c \cap C]$

⇒ $\mathbb{P}[A \cap B \cap C] = \mathbb{P}[A] \cdot \mathbb{P}[B|A] \cdot \mathbb{P}[C|A \cap B]$

↳ W'keit, da zu landen   ↑ muss durch alle diese!

=> Pfade = $\mathbb{P}$ mit Vorbedingungen

Knoten: $\mathbb{P}[\dots \cap \dots \cap \dots]$

↖ alle schneiden die man durchgeht!

Wichtig bei expectation value:

$$\mathbb{E}[(X+Y)\mathbb{1}_D + (X+Z)\mathbb{1}_{D^c}] = \mathbb{E}[X+Y] \cdot \mathbb{P}[D] + \mathbb{E}[X+Z] \cdot \mathbb{P}[D^c]$$

Wichtig bei Normalverteilung:

Wenn $X \sim \mathcal{N}(m_X, \sigma_X^2)$, $Y \sim \mathcal{N}(m_Y, \sigma_Y^2)$ unabhängig sind, dann ist $X + Y \sim (m_X + m_Y, \sigma_X^2 + \sigma_Y^2)$

Defining a probability space, example:

↳ Roll of die & toss of coin:
$\Omega = \{1,2,3,4,5,6\} \times \{0,1\}$ ($|\Omega| = 12$)
$\mathcal{F} = 2^\Omega$ ($|\mathcal{F}| = 2^{|\Omega|} = 2^{12}$)
$\mathbb{P}: \mathcal{F} \to [0,1]$
$A \mapsto \frac{|A|}{|\Omega|}$

↳ Roll of n die:
$\Omega = \{1, 2, \dots, 6\}^n$ ($|\Omega| = 6^n$)

Wenn nach Verteilung (≙ Distribution) gefragt ist, untersuche:

Diskret: $\mathbb{P}[X = k] = \dots = \text{dist.}$

Continuous: $\mathbb{P}[X \le x] = \dots = F_X(x) = \text{dist.}$

Borel set: Any set in a topological space that can be formed from open sets (or from closed sets) through the operations of countable union, countable intersection, and relative complement.

Borel σ-algebra on $[0,1]^2$:= Die kleinste Menge von Teilmengen von Ω, die:

H1) $\Omega \in \mathcal{F}$
H2) $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$
H3) $A_1, A_2, \dots \in \mathcal{F} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$

erfüllt & alle Mengen $[x_1, x_2] \times [y_1, y_2]$, $0 \le x_1 < x_2 \le 1$, $0 \le y_1 < y_2 \le 1$ enthält.

Anhang

Urnenmodell:

n: # Kugeln, k: # Züge	mit Zurücklegen	ohne Zurücklegen
Reihenfolge relevant (Variation)	n^k	$\frac{n!}{(n-k)!}$
Reihenfolge irrelevant (Kombination)	$\binom{n+k-1}{k}$	$\binom{n}{k}$

Identitäten:

→ $(\alpha+\beta)^n = \sum_{j=0}^{n} \binom{n}{j} \alpha^j \beta^{n-j}$

→ $\sum_{j=0}^{k} \binom{m}{k-j} \cdot \binom{n}{j} = \binom{m+n}{k}$

→ $\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k} = \frac{n!}{k!\,(n-k)!}$

→ $e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$

→ $\int_0^\infty t^n e^{-t}\, dt = n!$

$\partial_x \tan(x) = 1/\cos^2(x)$

→ $\partial_x \ln(x) = \frac{1}{x}$

→ $\int_0^\infty (\lambda t)^n e^{-\lambda t} = \int_0^\infty x^n e^{-x} \frac{dx}{\lambda} = \frac{n!}{\lambda}$

Wichtige Reihen / Summen:

Geometrische Reihe: $\sum_{n=0}^{\infty} q^n = \frac{1}{1-q}$ für $|q|<1$

Geometrische Summe: $\sum_{n=0}^{N} q^n = \frac{1-q^{N+1}}{1-q}$

Arithmetische Summe: $\sum_{k=1}^{n} k = \frac{n\cdot(n+1)}{2}$

Partielle Ableitung: $\int_a^b \underset{\downarrow}{f(x)}\underset{\uparrow}{g(x)}\,dx = [f(x)\cdot G(x)]_a^b - \int_a^b f'(x)\cdot G(x)\,dx$

$f'(x)$ $G(x)$

"Faires Spiel": $\mathbb{E}[X]=0$

Unabhängigkeit: Auch über $\mathbb{E}$ verifizierbar!

Länge eines Konfidenzintervalls: Für X_i, $i=1,2,\dots,n \sim \mathcal{N}(m,\sigma^2)$ ist die Länge des Konfidenzintervalls: $\boxed{|C(x)| = \frac{2\Phi^{-1}(1-\alpha/2)}{\sqrt{n}}\cdot\sigma}$ → falls nach n gesucht: Ergebnis aufrunden!

wobei $1-\alpha$ (%) der gewünschte Konfidenzintervall ist.

(Konfidenzintervall hat die Grenzen $\frac{1}{n}\sum_{i=0}^{n} X_i \pm \sqrt{\frac{\sigma^2}{n}}\,\Phi^{-1}(1-\frac{\alpha}{2})$)

Normalverteilung, Verteilungsfunktionen:

Standardnormalverteilung: $X\sim\mathcal{N}(0,1)$: $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{1}{2}t^2}\,dt$

allgemeine Normalverteilung: $X\sim\mathcal{N}(m,\sigma^2)$: $F(x) = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{1}{2}\left(\frac{t-m}{\sigma}\right)^2}\,dt = \Phi\left(\frac{x-m}{\sigma}\right)$

Subst. $t=\sigma z+m$

neue Integrationsvar. $z=\frac{t-m}{\sigma}$

- Wenn F_X gegeben, und wir wollen $\mathbb{P}[a\le X<b]$:
  $\mathbb{P}[a\le X<b] = \mathbb{P}[X<b]-\mathbb{P}[X<a] = \lim_{t\uparrow b} F_X(t) - \lim_{t\uparrow a} F_X(t)$
- X,Y with joint distribution $f_{X,Y}(x,y)$.
  ↪ $\mathbb{P}[X<3Y] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbb{1}_{\{x<3y\}} f_{X,Y}(x,y)\,dy\,dx$
  ↪ $\mathbb{P}[Y\le \frac{X}{3}] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbb{1}_{\{y\le \frac{x}{3}\}} f_{X,Y}(x,y)\,dy\,dx$
- Es gibt unabhängige sowie nicht unabhängige Zufallsvariablen, für die $\mathbb{E}[\frac{X}{Y}] \ne \frac{\mathbb{E}(X)}{\mathbb{E}(Y)}$ gilt. Es gibt Spezialfälle, in denen $\mathbb{E}[\frac{X}{Y}] = \frac{\mathbb{E}(X)}{\mathbb{E}(Y)}$ gilt.
- Würfel: $\Omega = \{1,2,3,4,5,6\}^n$ n: # Würfel $\Rightarrow |\Omega| = 6^n$

z%-Konfidenzintervall

Sei $Y_1, Y_2, \dots, Y_n$ eine Folge von i.i.d. Messungen einer unbekannten Grösse m. Die Verteilung einer Messung ist $\mathcal{N}(m,\sigma^2)$ mit bekanntem $\sigma=0.1$. Wir betrachten das Konfidenzintervall $I=[\frac{1}{n}\sum_{i=1}^n Y_i - a, \frac{1}{n}\sum_{i=1}^n Y_i + a]$. Wähle das grösste z sodass I ein z%-Konfidenzintervall ist.

Sei $\bar{Y}_n = \sum_{i=1}^n Y_i$. Aufgrund der Eigenschaften von Zufallsvariablen erhalten wir $\bar{Y}_n \sim \mathcal{N}(m, \sigma^2/n)$ und somit $Z := \frac{\sqrt{n}(\bar{Y}_n - m)}{\sigma} \sim \mathcal{N}(0,1)$.

$\mathbb{P}[\bar{Y}_n - a \le m] = \mathbb{P}(\bar{Y}_n - m \le a) = \mathbb{P}\left[\frac{\sqrt{n}(\bar{Y}_n-m)}{\sigma} \le \frac{\sqrt{n}\cdot a}{\sigma}\right] = \mathbb{P}\left[Z\le \frac{\sqrt{n}\cdot a}{\sigma}\right] = \Phi\left(\frac{\sqrt{n}\cdot a}{\sigma}\right)$ und somit

$\mathbb{P}[\bar{Y}_n - a > m] = 1-\Phi\left(\frac{\sqrt{n}\cdot a}{\sigma}\right)$. Analog: $\mathbb{P}[\bar{Y}_n + a \ge m] = \Phi\left(\frac{\sqrt{n}\cdot a}{\sigma}\right) \to \mathbb{P}[\bar{Y}_n + a < m] = 1-\Phi\left(\frac{\sqrt{n}\cdot a}{\sigma}\right)$

$\mathbb{P}[m\in[\bar{Y}_n - a, \bar{Y}_n + a]] = \frac{z}{100}$ ist äquivalent zu $\frac{z}{100} = 1-\mathbb{P}[m\notin[\bar{Y}_n - a, \bar{Y}_n + a]]$. Da die Ereignisse $\{\bar{Y}_n - a > m\}$ und $\{\bar{Y}_n + a < m\}$ disjunkt sind können wir die Wahrscheinlichkeit der Vereinigung als Summe der Wahrscheinlichkeiten schreiben:

$\frac{z}{100} = 1-\mathbb{P}[\{\bar{Y}_n - a > m\}\cup\{\bar{Y}_n + a < m\}] = 1-(\mathbb{P}[\bar{Y}_n - a > m] + \mathbb{P}[\bar{Y}_n + a < m]) =$

$= 1-2\left(1-\Phi\left(\frac{\sqrt{n}\cdot a}{\sigma}\right)\right) = 2\Phi\left(\frac{\sqrt{n}\cdot a}{\sigma}\right) - 1.$

Rote & schwarze Kugeln:

Urne mit 20 roten und 80 schwarzen Kugeln. Ziehe 3 mal mit zurücklegen.

a) Definiere den einfachsten (kleinsten) Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ der dieses Experiment beschreibt, wobei wir nur an den Farben der Kugeln interessiert sind und in welcher Reihenfolge die Farben gezogen werden.

$\Omega := \{0,1\}^3$, wobei 0 für rot und 1 für schwarz steht. $|\Omega| = 2^3 = 8$

$\mathcal{F} := \mathcal{P}(\Omega)$ $|\mathcal{F}| = 2^8 = 256$

3 Kugeln ziehen; 1 or 0 (schwarz or weiss)

$\mathbb{P}: \mathcal{F}\to[0,1]$, $A \mapsto \mathbb{P}(A) := \sum_{w\in A} \prod_{i=1}^{3} 0.2^{1-w_i}\, 0.8^{w_i}$

$w=(w_1,w_2,w_3)$

Medikament

Ein Pharmainstitut behauptet, ein bestimmtes Medikament wirke mit einer W'keit von mindestens 90%. Daraufhin wird das Medikament an 100 Personen verabreicht.

a) In dieser ersten Testreihe zeigt das Mittel bei 97 der 100 Personen Wirkung. Ist damit die Behauptung des Pharmainstituts auf dem Signifikanzniveau von 5% statistisch bewiesen?

Die Nullhypothese wird so gewählt, dass das, was man selbst beweisen will, in der Gegenhypothese steht. Das Pharmainstitut versucht natürlich zu beweisen, dass das Medikament gut ist.

Die Nullhypothese ist dann: $\boxed{H_0: p<0.9}$, was bedeutet, dass das Medikament mit einer geringeren W'keit als 0.9 wirkt.

Die Gegenhypothese ist: $\boxed{H_1: p\ge 0.9}$, also dass das Medikament mit einer Wahrscheinlichkeit von mindestens 90% wirkt. Im folgenden berechnen wir die Wahrscheinlichkeiten unter der Nullhypothese:

$\mathbb{P}_{p=0.9}[X\ge k] \le 0.05 \Leftrightarrow 1-\mathbb{P}_{p=0.9}[X\le k-1] \le 0.05 \Leftrightarrow \mathbb{P}_{p=0.5}[X\le k-1] \ge 0.95$

X ist binomialverteilt mit Parametern $p=0.9$ und $n=100$. Die Binomialverteilung ist für grosses n numerisch schwierig zu berechnen, kann aber gut mit einer Normalverteilung approximiert werden, wie man einfach mit dem Zentralen Grenzwertsatz beweisen kann: $X = X_1 + \dots + X_n$, wobei X_i i.i.d. Bernoulli-Variablen sind mit Parameter 0.9. Somit gilt $\mathbb{E}[X_i] = 0.9$ und $\sigma_{X_i}^2 = 0.9\cdot(1-0.9) = 0.09$. Daher gilt laut dem CLT

$\mathbb{P}_{p=0.9}[X\le k-1] = \mathbb{P}_{p=0.9}\left[\frac{X_1+\dots+X_n - 0.9n}{\sqrt{0.09n}} \le \frac{k-1-0.9n}{\sqrt{0.09n}}\right] \approx \Phi\left(\frac{k-1-0.9n}{\sqrt{0.09n}}\right)$

Als nächstes bringen wir k in der folgenden Ungleichung auf die linke Seite:

$\Phi\left(\frac{k-1-0.9n}{\sqrt{0.09n}}\right) \ge 0.95 \Leftrightarrow \frac{k-1-0.9n}{\sqrt{0.09n}} \ge \Phi^{-1}(0.95)$

$\Leftrightarrow k \ge 95.95$

Die Nullhypothese wird also bei 95 oder weniger Treffern im Test, also Personen bei denen das Medikament wirkt, angenommen und bei 96 oder mehr abgelehnt. Im Test wirkte das Medikament bei 97 Personen, das liegt für das Pharmainstitut im günstigen Ablehnungsbereich. Das Ergebnis der Testreihe ist damit statistisch signifikant bewiesen.

b) Zweifel an Wirkung an Medikament. Neuer Test an 100 Personen. Test Nullhypothese & Gegenhypothese? Wie muss seine Entscheidungsregel lauten, wenn er seine Vermutung auf dem Signifikanzniveau von 5% statistisch belegen will?

Die Nullhypothese wird so gewählt, dass das, was man selbst beweisen will, in der Gegenhypothese steht. Hier wird versucht zu überprüfen, ob das Medikament nicht doch schlechter wirkt, als behauptet.

Die Nullhypothese ist somit $H_0: p\ge 0.9$, also dass das Medikament mit 90% W'keit oder mehr wirkt.

Die Gegenhypothese lautet: $H_1: p<0.9$, also dass das Medikament mit weniger als 90% Wahrscheinlichkeit wirkt.

Wir wollen k finden, sodass $\mathbb{P}_{p=0.9}[X\le k] \le 0.05$. Analog zu a), wende den zentralen Grenzwertsatz an:

$\mathbb{P}_{p=0.9}[X\le k] = \mathbb{P}_{p=0.9}\left[\frac{X_1+\dots+X_n-0.9n}{\sqrt{0.09n}} \le \frac{k-0.09n}{\sqrt{0.09n}}\right] \approx \Phi\left(\frac{k-0.9n}{\sqrt{0.09n}}\right)$ und dann

analog weiterrechnen: $\Phi\left(\frac{k-0.9n}{\sqrt{0.09n}}\right) \le 0.05 \Leftrightarrow \frac{k-0.9n}{\sqrt{0.09n}} \le \Phi^{-1}(0.05) \Leftrightarrow k\le 84.05$

Die Nullhypothese wird bei 84 oder mehr Patienten, bei denen das Medikament wirkt, abgelehnt und demnach bei 85 oder mehr Patienten angenommen.