Probability & statistics

Def. Probability space

A triplet $(\Omega, \mathcal{F}, \mathbb{P})$, where

i). Ω sample space
↳ set of possible outcomes

ii). $\mathcal{F}$ set of events $\mathcal{F} = \mathcal{P}(\Omega)$

iii). $\mathbb{P}$ probability measure $\mathbb{P} : \mathcal{F} \to [0,1]$

Def. Sample space Ω

A set Ω which contains all possible outcomes of an experiment.
$\omega \in \Omega$

Def. Event

Given a sample space Ω, an event is given by a subset $A \subset \Omega$

~ Explicit : $A = \{1, 2, 3\}$

~ Implicit : $A = \{\omega \in \Omega : \omega \leq 3\}$

COMBINATION : Events

→ Given events $A, B \in \mathcal{F}$ with some semantic meaning for each

Using set operations to treat them as a whole

Eg. $\Omega = \{1, 2, 3, 4, 5, 6\}$
$A = \{2, 4, 6\}$ (☆) " the die is even "
$B = \{1, 2, 3\}$ (△) " die is ≤ 3 "

LOGIC	SET	meaning
AND	$A \cap B$	(☆) and (△)
OR	$A \cup B$	(☆) or (△)
NOT	$A^c = \Omega \setminus A$	¬(☆)
	$A \triangle B$	i). (☆) ii). (△) Either i) or ii) is allowed

Def. σ - Algebra

A set of events $\mathcal{F} \subset \mathcal{P}(\Omega)$ is called a σ-algebra when

☑ H1. $\Omega \in \mathcal{F}$

☑ H2. $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$

☑ H3. $A_1, A_2, \ldots \in \mathcal{F} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$

NOTE : In script " $\mathcal{F}$ " denotes "the set of all the events"
$\Rightarrow$ ☑ $\mathcal{F}$ is a σ - algebra

σ-Algebra
i). $\Omega, \emptyset$
ii). for any A, A^c also in it
iii). Unions of events must be included

Prop 1.5 . Operating on events

i). $\emptyset \in \mathcal{F}$

ii). $A_1, A_2, \ldots \in \mathcal{F} \Rightarrow \bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$

iii). $A, B \in \mathcal{F} \Rightarrow A \cup B \in \mathcal{F}$

iv). $A, B \in \mathcal{F} \Rightarrow A \cap B \in \mathcal{F}$

Example : Borel σ-Algebra

A set of events $\mathcal{F}$, such that

$$\mathcal{F} := \{ A \subset \Omega \mid \forall x_1, x_2, y_1, y_2 \in [0,1] : A = [x_1, x_2] \times [y_1, y_2] \}$$

$\Rightarrow$ NOTE : $\mathcal{F}$ is the smallest collection of subsets in Ω which satisfies H1 – H3.

Excursion : measure theory
— finding the smallest σ- algebra

Lemma. M - generated σ - algebra

For some $M \subseteq \mathcal{P}(\Omega)$ there exists a smallest σ- algebra which contains M

$\Rightarrow$ where A is σ-algebra

$$\sigma(M) \overset{def}{:=} \bigcap_{M \subseteq A} A$$

Def. Borel σ-algebra

Given a topological space (X, τ) we generate $B(X)$

$\Rightarrow$ Borel σ - algebra

$$B(X) := \sigma(\tau)$$
↳ generated from open sets

EVENT OCCURENCE

Def. Occurrence of an event

Given a possible outcome ω, an event A.

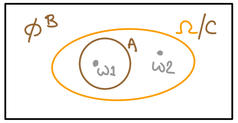
☐ $\omega \in A$

$\Rightarrow$ "The event A occurs for ω"

Occurrence of an event

i). Given an event in an experiment, the probability of the occurrence of an event

POE $\in [0, 1]$



• Event A occurs for $\omega 1$

• Event A does not occur for $\omega 2$

• Event B ($B = \emptyset$) never occurs

• Event C ($C = \Omega$) always occurs

Def. Almost surely (a.s.)

☐ An event $A \in \mathcal{F}$ occurs a.s. $\Leftrightarrow \mathbb{P}[A] = 1$

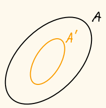
NOTE : Superset of A' (a.s.)

☐ An event $A' \in \mathcal{F}$

☐ A set ~~(event)~~ A

✓ $A' \subset A$

✓ $\mathbb{P}[A'] = 1$



$\Leftarrow$ A occurs almost surely

SYNTAX :

• $X \leq Y$ a.s. $\leftrightarrow \mathbb{P}[X \leq Y] = 1$

• $X \leq a$ a.s. $\leftrightarrow \mathbb{P}[X \leq a] = 1$

Def. Probability measure $\mathbb{P}$

Given a tuple $(\Omega, \mathcal{F})$, a probability measure on it is a map

$$\mathbb{P} : \mathcal{F} \to [0, 1]$$
$$A \mapsto \mathbb{P}[A]$$

associates for each event a number in $[0, 1]$

PROPERTIES (prob. measure)

P1. $\mathbb{P}[\Omega] = 1$

P2. Countable additivity — A as disjoint union

If $A = \bigcup_{i=1}^{\infty} A_i \Rightarrow \mathbb{P}[A] = \sum_{i=1}^{\infty} \mathbb{P}[A_i]$

P3. $\mathbb{P}[A] \geq 0 \quad \forall A \in \mathcal{F}$ for all events A

Prop 1.8 Arithmetics of $\mathbb{P}$

Given a probability measure on $(\Omega, \mathcal{F})$

$\Rightarrow$ i). $\mathbb{P}[\emptyset] = 0$

ii). Additivity [Disjoint]

Given k ($k \geq 1$) pairwise disjoint events $A_1, \ldots, A_k$

$\Rightarrow \mathbb{P}[A_1 \cup \ldots \cup A_k] = \mathbb{P}[A_1] + \cdots + \mathbb{P}[A_k]$

iii). Probability of the complement event

$$\mathbb{P}[A^c] = 1 - \mathbb{P}[A]$$

iv). Pairwise addition ~~[Disjoint]~~ not required

$$\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \cap B]$$

TRICK : Defining the prob. measure

Given Ω finite or countable

STEP 0 : Associate for each outcome ω a probability p_ω

STEP 1 : Adding the probabilities up for an event $A \subset \Omega$

$$\mathbb{P}[A] = \sum_{\omega \in A} p_\omega$$

USEFUL INEQUALITIES

- monotonicity
- union bound

Prop. 1.9. Monotonicity

Let events $A, B \in \mathcal{F}$

☐ $A \subset B \Rightarrow \mathbb{P}[A] \leq \mathbb{P}[B]$

Prop. 1.10 Union bound

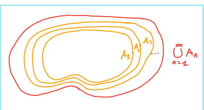
$$\mathbb{P}\left[\bigcup_{i=1}^{\infty} A_i\right] \leq \sum_{i=1}^{\infty} \mathbb{P}[A_i]$$

IDEA: Finding upp. bound of $\mathbb{P}$ using easier sets

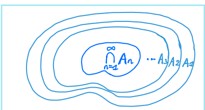
Def. In/decreasing sequence of events

☐ A sequence $(A_n)_{n\geq 1}$ of events is:

- increasing $\longleftrightarrow$ $A_n \subset A_{n+1}$ $\forall n \geq 1$
- decreasing $\longleftrightarrow$ $A_n \supset A_{n+1}$ $\forall n \geq 1$



• increasing



• decreasing

MY NOTES: use equalities

☑ (B_i) decreasing

$$\bigcap_{i=1}^{\infty} B_i = \left(\bigcup_{i=1}^{\infty} B_i^c\right)^c$$

CONTINUITY OF $\mathbb{P}$

Prop. 1.12. Limits

- >> Increasing limit <<

☐ $\forall n$ $A_n \subset A_{n+1}$

$\Rightarrow \lim_{n\to\infty} \mathbb{P}[A_n] = \mathbb{P}\left[\bigcup_{n=1}^{\infty} A_n\right]$

- >> Decreasing limit <<

☐ $\forall n$ $B_n \supset B_{n+1}$

$\Rightarrow \lim_{n\to\infty} \mathbb{P}[B_n] = \mathbb{P}\left[\bigcap_{n=1}^{\infty} B_n\right]$

Laplace models & counting

Def. Laplace model

A tuple $(\Omega, \mathcal{F}, \mathbb{P})$ on a sample space Ω such that:

i). Ω is a finite sample space

ii). $\mathcal{F} = \mathcal{P}(\Omega)$

iii). $\mathbb{P}: \mathcal{F} \to [0,1]$

$$\forall A \in \mathcal{F}: \mathbb{P}[A] = \frac{|A|}{|\Omega|}$$

Estimating the probability for Laplace model $\longleftrightarrow$ counting the number of elements in A and in Ω

CONDITIONAL PROBABILITIES

possesses incomplete info about outcomes of the experiment

Def. Cond. prob. of A given B

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

☐ Events $A, B \in \mathcal{F}$

✓ $\mathbb{P}[B] > 0$

$\Rightarrow$ the cond. prob of A given B:

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

Remark: "conditional on B, B always occurs"

$$\mathbb{P}[B|B] = \frac{\mathbb{P}[B\cap B]}{\mathbb{P}[B]} = 1$$

Multiplication rule

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

Prop. 1.25. induced $\mathbb{P}$ by conditional event

☐ Prob. space $(\Omega, \mathcal{F}, \mathbb{P})$

☐ Event $B \in \mathcal{F}$

✓ $\mathbb{P}[B] > 0$

☐ A map* $\mathbb{P}[\cdot|B]: \mathcal{F} \to [0,1]$

$\Rightarrow$ The map* is a prob. measure on Ω

Prop. 1.26. Total probability

☐ A sample space Ω

✓ with partition

i). B_i pairwise disjoint

ii). $\Omega = \bigcup_{i=1}^{n} B_i$

✓ $\mathbb{P}[B_i] > 0$ $\forall i \in [n]$

} COND1

$\Rightarrow$ Calculating $\mathbb{P}[A]$:

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

Prop. 1.27 Bayes formula

☐ COND1

☐ Event $A \in \mathcal{F}$ with $\mathbb{P}[A] > 0$

$$\Rightarrow \boxed{\mathbb{P}[B_i|A] = \frac{\mathbb{P}[A|B_i]\,\mathbb{P}[B_i]}{\sum_{j\in[n]} \mathbb{P}[A|B_j]\,\mathbb{P}[B_j]}}$$

Def. Indicator function $\mathbb{1}_A$

☐ An event $A \in \mathcal{F}$

$\Rightarrow$ we define the indicator function $\mathbb{1}_A$ of A as

$$\mathbb{1}_A: \Omega \to \mathbb{R}$$

$$\mathbb{1}_A(\omega) = \begin{cases} 0, & \text{if } \omega \notin A \\ 1, & \text{if } \omega \in A \end{cases}$$

NOTE: $\mathbb{1}_A$ is a valid r.v.

ASYMPTOTIC RESULTS

Given an infinite sequence of i.i.d. random variables $X_1, X_2, \ldots$

$$X_i: \Omega \to \mathbb{R}$$

CONSTRAINT

$\forall i_1 < \cdots < i_k$ $\forall x_1, \ldots, x_k \in \mathbb{R}$

$$\mathbb{P}[X_{i_1} \leq x_1, \ldots, X_{i_k} \leq x_k] = F(x_1)\cdots F(x_k)$$

common distri. function

Def. Empirical average

☐ i.i.d. r.v. $X_1, \ldots, X_n: \Omega \to \mathbb{R}$

☐ A r.v. defined by

$$M_n := \frac{\sum_{i=1}^{n} X_i(\omega)}{n} = \frac{X_1(\omega) + \cdots + X_n(\omega)}{n}$$

$\Leftrightarrow$ M_n is the empirical average

LAW OF LARGE NUMBERS

☐ $\mathbb{E}[|X_1|]$ is well-defined and finite

✓ X discrete OR

X contin. integrable

☐ $m = \mathbb{E}[X_1]$

$$\Rightarrow \boxed{\lim_{n\to\infty} \frac{\sum_{i=1}^{n} X_i}{n} = m \quad a.s.} \quad (**)$$

NOTE: view of the event

☐ $E := \left\{\omega \in \Omega \,\middle|\, \lim_{n\to\infty} \frac{\sum_{i=1}^{n} X_i(\omega)}{n} = m\right\}$

$\overset{(**)}{\Rightarrow}$ $\mathbb{P}[E] = 1$

CENTRAL LIMIT THEOREM

☐ Expectation $\mathbb{E}[X_1^2]$

✓ well-defined

✓ finite

☐ Define i). $m = \mathbb{E}[X_1]$

ii). $\sigma^2 = Var(X_1)$

iii). $S_n = X_1 + \cdots + X_n$

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

"How far is $\frac{\sum X_i}{n}$ from $m = \mathbb{E}[X_1]$?"

INDEPENDENCE

- dependency of
  - Events
  - r.v.
    - discrete
    - continuous

Def. Independence of events

☐ Events A, B are independent

$\overset{\text{def}}{\Longleftrightarrow}$ $\boxed{\mathbb{P}[A \cap B] = \mathbb{P}[A]\,\mathbb{P}[B]}$

Prop. 1.30 Equiv. statements

☐ Events $A, B \in \mathcal{F}$

✓ $\mathbb{P}[A], \mathbb{P}[B] > 0$

$\Longrightarrow$ Equiv. statements :

i). $\boxed{\mathbb{P}[A \cap B] = \mathbb{P}[A]\,\mathbb{P}[B]}$

ii). $\mathbb{P}[A \mid B] = \mathbb{P}[A]$ occurence of B has no influence on A

iii). $\mathbb{P}[B \mid A] = \mathbb{P}[B]$ occurence of A has no influence on B

Remark. 1.29

① If $\mathbb{P}[A] \in \{0, 1\}$

$\Longrightarrow \forall B \in \mathcal{F}$:

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

② If $\underbrace{\mathbb{P}[A \cap A] = \mathbb{P}[A]^2}_{\text{Event } A \text{ is independent with itself}}$

$\Longrightarrow \mathbb{P}[A] \in \{0, 1\}$

③ $\boxed{\mathbb{P}[A \cap B] = \mathbb{P}[A]\,\mathbb{P}[B]}$

$\Updownarrow$

$\boxed{\mathbb{P}[A \cap B^c] = \mathbb{P}[A]\,\mathbb{P}[B^c]}$

Def. List of independent events

☐ Events A_i $\forall i \in [n]$

☐ $\forall J \subset \{1, \dots, n\}$:

$$\mathbb{P}\Big[\bigcap_{j \in J} A_j\Big] = \prod_{j \in J} \mathbb{P}[A_j]$$

$\Longleftrightarrow$ The events A_i are independent

Independent test (3 events)

☐ Events A, B, C are independent

$\Longleftrightarrow$ 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). $\mathbb{P}[A \cap B \cap C] = \mathbb{P}[A] \cdot \mathbb{P}[B] \cdot \mathbb{P}[C]$

Def. Independence r.v. [Finite list]

☐ n random variables X_i $\forall i \in [n]$

☐ $\forall a_1, \dots, a_n \in \mathbb{R}$ can be understood as "$\cap$"

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

$= \mathbb{P}[X_1 \le a_1] \cdots \mathbb{P}[X_n \le a_n]$

$\overset{\text{def}}{\Longleftrightarrow}$ $X_1, \dots, X_n$ are independent

$\Longleftrightarrow$ $\boxed{X_i \perp X_j \quad \forall (i,j) \in [n]^2}$

Def. Independence r.v. [∞ - list]

☐ An $\boxed{\infty\text{ - sequence}}$ of r.v.

$X_1, X_2, \dots$

☐ $X_1, \dots, X_n$ are independent $\forall n \in \mathbb{N}$

$\Longleftrightarrow$ $\overbrace{X_1, X_2, \dots}$ are independent.

> RECALL : Given a distribution function F, the existence of a defined prob. space and the r.v. is guaranteed.

Theorem 1.34. Distri. func induced r.v. [List]

☐ n distri. functions $F_1, \dots, F_n$

$\Longrightarrow$ ☐ Existence of prob. space$^{(*)}$ $(\Omega, \mathcal{F}, \mathbb{P})$

☐ n random variables $X_1, \dots, X_n$ on$^{(*)}$

✓ Correspondence $F = F_X$

$\forall a$ $\mathbb{P}[X_i \le a] = F_i(a)$

✓ $X_1, \dots, X_n$ are independent

Def. Indep. & identically distributed r.v

[iid]

☐ Random variables (optional : ∞ / finite)

$X_1, X_2, \dots$

☐ $X_1, X_2, \dots$ are independent

☐ $X_1, X_2, \dots$ have the same distri. func. $\forall i, j$ $F_{X_i} = F_{X_j}$

RANDOM VARIABLES

Def. Random variable (r.v)

A map $X : \Omega \to \mathbb{R}$ s.t.

☐ well-defineness [for $\mathbb{P}[(*)]$]

$\forall a \in \mathbb{R}$:

$\underbrace{\{\omega \in \Omega : X(\omega) \le a\}}_{(*)} \in \mathcal{F}$

> - "X" is measurable.
> - r.v. as a func. mapping Ω to $\mathbb{R}$
> - A map that tells which "info" (outcomes) are of some type.

Remark : Powerset as set of events

i). Random variable easy check ☐ $\mathcal{F} = \mathcal{P}(\Omega)$

$\Longrightarrow$ Every function $X : \Omega \to \mathbb{R}$ is a r.v.

$\{X \le a\} = \{\omega \in \Omega : X(\omega) \le a\}$

$\subseteq \Omega \in \mathcal{P}(\Omega)$

Remark : Valid events (given a r.v.)

☐ $\forall$ r.v. on $(\Omega, \mathcal{F}, \mathbb{P})$

$\Longrightarrow$ Following sets S_i are ensured to be events (i.e. $S_i \in \mathcal{F}$)

$S_1 = \{X > a\}$, $\forall a \in \mathbb{R}$

$S_2 = \{a < X \le b\}$, $\forall a < b$ in $\mathbb{R}$

$S_3 = \{X < a\}$, $\underbrace{\{X \ge a\}}_{\{X < a\}^c}$ $\forall a \in \mathbb{R}$

> **Trick : checking if X is r.v. for $(\Omega, \mathcal{F})$**
>
> STEP 1 :
> Using definition, find out the set
>
> $$\{X \le a\} = \begin{cases} \{\text{CASE } 1\}, & a \in S_1 \\ \vdots & \\ \{\text{CASE } N\}, & a \in S_N \end{cases}$$
>
> (*) case distinction
>
> STEP 2 :
> Check the mapping of X, divide the values into "parts"
>
> STEP 3 :
> Check for each "part" of value a, which outcomes ω should belong to that "part".
>
> STEP 4 : For each case, check if the set is indeed an event
>
> $\rightarrow$ i.e. $S_i \overset{?}{\in} \mathcal{F}$ $\forall i$

CONTINUITY OF R.V.

Def 1. Discrete r.v.

☐ A r.v. $X : \Omega \to \mathbb{R}$

☐ image is at most countable

$X(\Omega) = \{x \in \mathbb{R} : \exists \omega \in \Omega \;\; X(\omega) = x\}$

$\overset{\text{def}}{\Longleftrightarrow}$ The r.v. is discrete

Def 2. Discrete r.v.

☐ r.v. $X : \Omega \to \mathbb{R}$

☐ $\exists E \subset \mathbb{R}$ finite/countable

$\forall \omega \in \Omega$ $X(\omega) \in E$

> **Trick : prove r.v. to be discrete**
>
> ☐ Show Z takes values in the discrete set
>
> ☐ Show $\forall z \in \mathcal{F}$: $\{Z = z\} \in \mathcal{F}$

Def. continuous r.v.

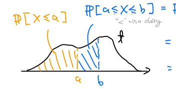
☐ A r.v. $X : \Omega \to \mathbb{R}$

☐ Distri. func of X can be written as

$F_X(a) = \int_{-\infty}^{a} f(x)\,dx \quad \forall a \in \mathbb{R}$

#Density function

f : some non-neg function $f : \mathbb{R} \to \mathbb{R}_+$



$= F_X(b) - F_X(a)$

$= \int_a^b f(x)\,dx$

$\Longleftrightarrow$ r.v. X is continuous

> RECALL : i). $\mathbb{P}[X = a] = F(a) - F(a^-) = 0$
>
> ii). $\forall a \le b$:
>
> $\mathbb{P}[a < X < b] = \mathbb{P}[a \le X \le b] = \int_a^b f(x)\,dx$ (OPTIONAL)

Lemma. Continuity of dis. func

☐ A continuous random var.

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

$\Longrightarrow$ F_X is a continuous function

$\mathbb{P}[X = x] = 0$ $\forall x$ fixed

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

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

$= 0$

> Value evaluation — of a continuous func
>
> Note for X continuous
>
> ①. $F_X(a) = \int_{-\infty}^{a} \boxed{f(x)\,dx}$ probability of X taking a value in $[x, x+dx]$
>
> ②. $\mathbb{P}[X = a] = 0$ $\forall a \in \mathbb{R}$ fixed
>
> $\Longrightarrow$ prob. at one point equals 0 but in an infinitisimal interval calculatable.

RECOGNIZING continuous r.v.

Theorem 3.2.

☐ Distri. func F_X of some r.v. X

✓ F_X is continuous

✓ F_X is p.w. C^1

$\Longleftrightarrow$ $\exists\, x_0 = -\infty < x_1 < \dots < x_{n-1} < x_n = +\infty$

s.t. F_X is C^1 on all $I := (x_i, x_{i+1})$

$\Longrightarrow$ i). r.v. X is continuous

ii). Density func f constructed by

- $\forall x \in (x_i, x_{i+1})$ $f(x) = F_X'(x)$
- setting arbitrary values at $x_1, \dots, x_{n-1}$

DISCRETE DISTRIBUTION

Def. Distribution of X [Discrete]

- ☐ Discrete r.v. $X : E \to \mathbb{R}$
  - ✓ Set E is finite/countable
- ☐ A sequence of numbers $(p_x)_{x\in E}$
  - $\forall x \in E \quad p_x = \mathbb{P}[X = x]$

$\Leftrightarrow$ $(p_x)_{x\in E}$ is the distribution of X

Remark. Calculating $\mathbb{P}$ of a subset

- ☐ A sequence $(p_x)_{x\in E}$ as distribution of X [Discrete]
- ☐ subset $S \subset \mathbb{R}$

$\Rightarrow$ $\mathbb{P}[x \in S] = \sum_{x\in S} p_x$

Prop 2.9. Sum of the distribution

- ☐ Distribution of X (discrete) $(p_x)_{x\in E}$

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

RECALL: Prob. space and r.v. can be induced by a distri. func F that satisfies Properties i)–iii) of dis. fun.

"Let X be a r.v. with distri. $(p_x)_{x\in E}$"

- ☐ A sequence (p_x)
  - ✓ $p_x \in [0,1]$
  - ✓ $\sum_{x\in E} p_x = 1$

$\Rightarrow$ Existence of
- i). $(\Omega, \mathcal{F}, \mathbb{P})$
- ii). r.v. X with distribution (p_x)

TOOLKITS: Approximation

Trick: approx. of ∞-countable set

- F_n approximates E
  - $F_n \uparrow E$

$\Leftrightarrow$ ☐ $\forall n \quad F_n \subset E$ and $F_n \subset F_{n+1}$
☐ $E = \bigcup_{n\in\mathbb{N}} F_n$

Lemma. Countable set is approx-able

- ☐ set E is countable

$\Rightarrow \exists (F_n)$ s.t. $F_n \uparrow E$

Def. Sum of nonneg. numbers

- ☐ Sequence of nonneg. numbers
  - $(a_x)_{x\in E}$ $\quad \forall x \quad a_x \geq 0$ (↳ index set)

$\Leftrightarrow$ Define the sum of the a_x as

$$\sum_{x\in E} a_x := \sup_{\substack{F\subset E \\ F \text{ finite}}} \sum a_x = \lim_{n\to\infty} \sum_{x\in F_n} a_x$$

$\forall$ seq. $F_n \uparrow E$ finite

NOTATION

- CASE 1: Index set $E = \mathbb{N}$

$$\sum_{x\in\mathbb{N}} a_x = \sum_{x=0}^{\infty} a_x = \lim_{n\to\infty}\left(\sum_{x=0}^{n} a_x\right)$$

Def. Sum of an integrable sequence

- ☐ A real sequence $(a_x)_{x\in E}$
- ☐ $\sum_{x\in E} |a_x| < \infty$

$\Leftrightarrow$ sequence is integrable

Lemma. induced sequences

- ☐ Integrable real sequence $(a_x)_{x\in E}$
- ☐ Subsequences
  - $a_x^+ := \max(0, a_x)$ pos. part
  - $a_x^- := \max(0, -a_x)$ neg. part

$\Rightarrow$ sum of the sequence representable as:

$$\sum_{x\in E} a_x = \sum_{x\in E} a_x^+ - \sum_{x\in E} a_x^-$$

NOTE: $a_x^+, a_x^- \geq 0 \Rightarrow$ both sums make sense

Remark. 2.4. Integrability $\Rightarrow$ finite sum

- ☐ A sequence (a_x) is integrable

$\Rightarrow$ $\sum_{x\in E} a_x$ is always finite

Example: Divergent sequence

- set-up:
  - i). $E = \mathbb{Z}$ approx. by $F_n = \{-n, \dots, n\} \uparrow E$
  - ii). Sequence (a_x) with $a_x = (-1)^{|x|}$

Goal: obtain sum of values of $(a_x)_{x\in E}$

$$\sum_{x\in E} a_x = ?$$

R: $\sum_{x\in E} a_x = \sum_{x\in F_n} (-1)^{|x|} = (-1)^n$

Taking the limit

$\nexists \lim_{n\to\infty} \sum_{x\in F_n} (-1)^{|x|} \rightsquigarrow \{-1, 1\}$

FUBINI THEOREMS

set-up:
- ☐ Sets E, F
  - ✓ finite or countable
- ☐ A family of numbers $(u_{xy})_{(x,y)\in E\times F}$

Theorem. Fubini (for nonneg. sequences)

☑ set-up
- ☐ u_{xy} are nonneg. numbers

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

Theorem. Fubini (for integrable seq.)

☑ set-up
- ☐ u_{xy} are real numbers
- ☐ $\sum_{x\in E}\left(\sum_{y\in F} |u_{xy}|\right) < \infty$

$$\Rightarrow \sum_{x\in E}\left(\sum_{y\in F} u_{xy}\right) = \sum_{y\in F}\left(\sum_{x\in E} u_{xy}\right)$$

DISTRIBUTION FUNCTION

Def. Distri. function of X

- ☐ Probability space $(\Omega, \mathcal{F}, \mathbb{P})$
- ☐ random variable X
- ☐ A function $F_X : \mathbb{R} \to [0,1]$
  - $\forall a \in \mathbb{R} : \quad F_X(a) = \mathbb{P}[X \leq a]$

$\overset{\text{def}}{\Leftrightarrow}$ F_X is the distribution function of X.

Prop. 1.17 Basic identity

- ☐ Distr. function F of X
- ☐ $a, b \in \mathbb{R}$ $(a < b)$

$\Rightarrow$ $\mathbb{P}[a < X \leq b] = F(b) - F(a)$

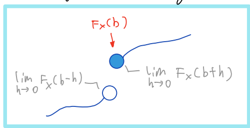
Theorem. Properties of F_X (⚡)

- ☐ Distr. function $F_X : \mathbb{R} \to [0,1]$

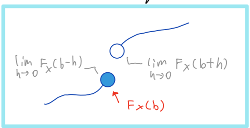
$\Rightarrow$ Properties
- i). F is nondecreasing
- ii). F is right continuous
  - i.e. $\boxed{F(a) = \lim_{h\downarrow 0} F(a+h) \quad \forall a \in \mathbb{R}}$
- iii). In-range $[0,1]$
  - $\lim_{a\to-\infty} F(a) = 0$
  - $\lim_{a\to\infty} F(a) = 1$

CONCEPT: Left/Right continuity

L1. Right continuity



L2. Left continuity



Theorem. Distri. function induced r.v.

- ☐ A function $F : \mathbb{R} \to [0,1]$ satisfies the properties i)–iii). (⚡)

$\Rightarrow \exists$ i). a probability space $(\Omega, \mathcal{F}, \mathbb{P})$
ii). a random variable $X : \Omega \to \mathbb{R}$
s.t. $F = F_X$

IDEA: If F is given
↝ "let X be a r.v. with distribution function F"
↝ enables statements as above

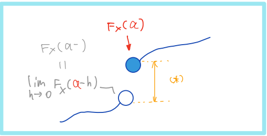
Prop. 1.20. Evaluating $\mathbb{P}[X = a]$

- ☐ A r.v. $X : \Omega \to \mathbb{R}$
- ☐ Distr. func F_X

$\Rightarrow \forall a \in \mathbb{R}$:

(*) $\mathbb{P}[X = a]$

$$= \begin{cases} F(a) - F(a^-), & \text{if } F \text{ dis.con at } a \\ 0, & \text{else} \end{cases}$$



SPECIAL R.V (Discrete)

Def. Bernoulli r.v. $Ber(p)$

☐ A r.v. $X: E \to \mathbb{R}$, $E = \{0, 1\}$

☐ Probability measure

$$\mathbb{P}[X=1] = p \qquad \forall p \in [0,1]$$
$$\mathbb{P}[X=0] = 1-p$$

$\iff$ r.v. is a Bernoulli r.v.

$\iff$ $X \sim Ber(p)$

Expectation

$\mathbb{E}[X] = 1 \cdot \mathbb{P}[X=1] + 0 \cdot \mathbb{P}[X=p] = p$

Variance

$Var[X] = \mathbb{E}[X^2] - \mathbb{E}[X^2] = p \cdot (1-p)$

Def. Binomial r.v. $Bin(n,p)$

☐ A r.v. $X: E \to \mathbb{R}$

✓ $E = \{0, \dots, n\}$

☐ Probability measure

$\forall p \in [0,1]$:

$$\mathbb{P}[X=k] = \underbrace{\binom{n}{k}}_{\text{in how many ways}} \underbrace{p^k (1-p)^{n-k}}_{\text{consecutive event for } k \text{ times}}$$

$\iff$ r.v. is a binomial r.v.

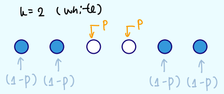
$\iff$ $X \sim Bin(n,p)$

ILLU: BINOMIAL r.v.

"Repeating experiment n times with prob. of success p"

$n = 6$

$k = 2$ (white)



$$(p_k)_{k \in E} = \binom{n}{k} (1-p)^{n-k} \cdot p^k$$

Tipp: Münzwurf $\Leftrightarrow$ Bernoulli

Sum of Bernoullis $\to$ Binomial

$$\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[Y_i] = n \cdot p$$

$\hookrightarrow X = \sum_{i=1}^{n} I_{A_i} = \sum_{i=1}^{n} Y_i$ with weights

$$Var[X] = \sum_{i=1}^{n} Var[Y_i] = n \cdot p \cdot (1-p)$$

NOTE: Guarantees of existence ($Bin(n,p)$)

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

$$\Rightarrow \sum_{k=0}^{n} p_k = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = (p + 1 - p)^n = 1$$

Prop 2.12. Sum of independent $Ber(p)$ & $Bin(n,p)$

☐ $X_i \sim Ber(p)$ $\forall i \in [n]$, $n \in \mathbb{N}$

✓ X_i are independent (iid) r.v.

✓ Parameter $p \in [0,1]$

☐ $S_n := X_1 + \dots + X_n$

$\Rightarrow$ $S_n \sim Bin(n,p)$

NOTE: Given $S \sim Bin(n,p)$ we can find $S_n = \sum_{i \in [n]} X_i$ s.t. S_n and S have the same distribution

Def. Geometric r.v. $Geom(p)$

☐ r.v. $X: E \to \mathbb{R}$

✓ $E = \mathbb{N} \setminus \{0\}$

"1st success in an infinite sequence of Bernoulli experiments with param. p."

☐ Prob. measure

$$\forall k \in E: \mathbb{P}[X=k] = (1-p)^{k-1} \cdot p$$

$\iff$ $X \sim Geom(p)$

NOTE: $X \sim Geom(1)$ with $k=1$

$$\Rightarrow \mathbb{P}[X=k] = (1-1)^{k-1} \cdot 1 = 0^{k-1} \cdot 1 = 0^0 = 1$$

By convention

TOOL: Geometric series

$$\sum_{k=0}^{\infty} a r^k = \frac{a}{1-r} \quad \forall |r| < 1$$

NOTE: Guarantees of existence ($Geom(p)$)

Given $p_k = (1-p)^{k-1} p$

$$\Rightarrow \sum_{k=0}^{\infty} p_k = p \cdot \sum_{k=0}^{\infty} (1-p)^{k-1} = p \cdot \frac{1}{p} = 1$$

$$\mathbb{E}[X] = \sum_{i=0}^{\infty} (1-p)^l = \frac{1}{1-(1-p)} = \frac{1}{p}$$

$$\mathbb{E}[X \cdot (X-1)] = \frac{2(1-p)}{p^2}$$

$$Var[X] = \frac{1-p}{p^2}$$

Prop 2.18. Absence of memory [$Geom(p)$]

☐ $T \sim Geom(p)$ for some $p \in (0,1)$

$\Rightarrow$ $\forall n \geq 0$ $\forall k \geq 1$:

$$\mathbb{P}[T \geq n+k \mid T > n] = \mathbb{P}[T \geq k]$$

Def. Poisson r.v. $Poisson(\lambda)$

☐ r.v. $X: E \to \mathbb{R}$

✓ $E = \mathbb{N}$

☐ Prob. measure $\lambda > 0$

$$\forall k \in \mathbb{N}: \mathbb{P}[X=k] = \frac{\lambda^k}{k!} e^{-\lambda}$$

$\iff$ $X \sim Poisson(\lambda)$

Note: Guarantees of existence (Poisson)

$$\sum_{x \in E} p_x = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1$$

$$\mathbb{E}[X] = \sum_{i=1}^{\infty} i \cdot \frac{\lambda^i}{i!} e^{-\lambda} = \lambda e^{-\lambda} e^{\lambda} = \lambda$$

$$\mathbb{E}[X^2] = \lambda^2 + \lambda$$

$$Var[X] = \lambda$$

Def. Negative binomial

☐ r.v. $X: E \to \mathbb{R}$

✓ $E = \mathbb{N} \setminus \{0\}$

☐ Prob. measure — until "r-th success"

$$\forall k \in E: \mathbb{P}[X=k] = \binom{k-1}{r-1} \cdot p^r \cdot (1-p)^{k-r}$$

$\iff$ $X \sim NB(r,p)$

$$\mathbb{E}[X] = \sum_{i=1}^{r} \mathbb{E}[X_i] = \frac{r}{p}$$

$$Var[X] = \sum_{i=1}^{r} Var[X_i] = \frac{r \cdot (1-p)}{p^2}$$

Def. Hypergeometric

In einer Urne seien n Gegenstände, davon r vom Typ 1 und $n-r$ vom Typ 2. Man zieht ohne Zurücklegen m der Gegenstände. Per Zufallsvariable X beschreibt die Anzahl der Gegenstände vom Typ 1 in der Stichprobe. Der Wertebereich von X ist $\mathcal{W}(X) = \{0, 1, \dots, \min(m,r)\}$ und:

$$\mathbb{P}[X=k] = \frac{\binom{r}{k} \cdot \binom{n-r}{m-k}}{\binom{n}{m}} \quad \text{für } k \in \mathcal{W}(X)$$

$$\mathbb{E}[X] = \sum_{i=1}^{n} i \cdot \mathbb{P}[X=i] = m \cdot \frac{r}{n}$$

$$Var[X] = m \cdot \frac{r}{n} \left(1 - \frac{r}{n}\right) \cdot \frac{n-m}{n-1}$$

RELATION BETWEEN r.v.

①. Binomial & Bernoulli

- Discrete r.v.

i). $Bin(1,p)$ ii). $Ber(p)$

$\Rightarrow$ Distri. sequences $(p_x)_{x \in E}$ of i), ii) are the same.

②. Independent binomial r.v.

- Discrete r.v.

$X \sim Bin(m,p)$

$Y \sim Bin(n,p)$

✓ X, Y are independent ($X \perp Y$)

$\Rightarrow$ $X + Y \sim Bin(m+n, p)$

Prop 2.16. First success

☐ ∞ - sequence of r.v.

$X_i \sim Ber(p)$ $\forall i \in \mathbb{N}$

✓ Independence of X_i

$(X_j)_{j \in J}$ are independent r.v.

$\hookrightarrow \forall J \subset \mathbb{N}$ finite

"$X_1, X_2, \dots$ iid $Ber(p)$"

☐ A map $T := \min\{\overbrace{n}^{n \in \mathbb{N}} \geq 1 : X_n = 1\}$

$\Rightarrow$ $T \sim Geom(p)$

"success at k-th time"

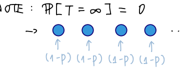
$$\{T = k\} = \{X_1 = 0, \dots, X_{k-1} = 0, X_k = 1\}$$

"first n Bernoulli experiments failed"

$$\{T > n\} = \{w \in \Omega \mid T(w) > n\} = \{w \in \Omega \mid \min\{k \geq 1 : X_k = 1\} > n\}$$

$$\mathbb{P}[T > n] = (\underbrace{1-p}_{\text{failing}})^{\overbrace{n}^{n\text{-times}}}$$

NOTE: $\mathbb{P}[T = \infty] = 0$



"never get a success"

$T = \infty \iff \{T = \infty\}$

POISSON & BINOMIAL

TOOL: Convergence in distribution

☐ A sequence of r.v. $X_1, X_2, \dots$

☐ sequence converges in distribution to a r.v. X

$$X_n \xrightarrow{d} X$$

$\iff$ $\forall x$ at which F_X is continuous

$$\lim_{n \to \infty} F_{X_n}(x) = F_X(x)$$

TOOL: Convergence to $e^{\lambda} / e^{-\lambda}$

i) $\exp(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!}$

ii) $\exp(x) = \lim_{n \to \infty} \left(1 + \frac{x}{n}\right)^n$

prop. Poisson approx. by binomial

☐ $\forall n \geq 1$: $X_n \sim Bin(n, \frac{\lambda}{n})$, $\lambda > 0$

$\Rightarrow$ i). $X_n \xrightarrow{d} N$

i.e. $\forall k \in \mathbb{N}: \lim_{n \to \infty} \mathbb{P}[X_n = k] = \mathbb{P}[N = k]$

ii). $N \sim Poisson(\lambda)$

Example: poisson approx. for large $X \sim Bin(n,p)$

geg. $M \sim Bin(n,p)$

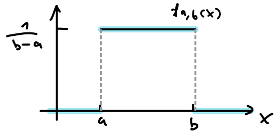
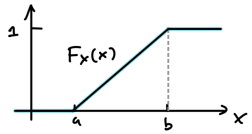
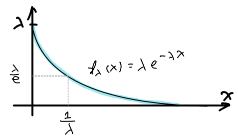
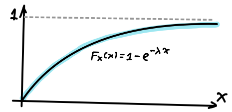
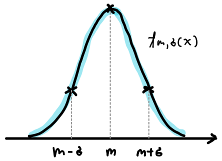
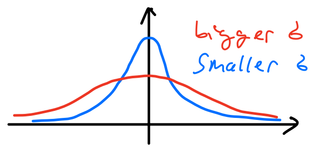
- $n = 10^4$
- $p = \frac{10}{n}$ $\Rightarrow \lambda = 10$

ges. $\mathbb{P}[M=5] \,\hat{=}$ "5 mistakes/page"

$$\mathbb{P}[M=5] \underbrace{\simeq}_{\text{Poisson approx.}} \left(\frac{\lambda}{k!}\right)^k e^{-\lambda} = \left(\frac{10}{5!}\right)^5 e^{-10} \simeq 0{,}0378$$

SPECIAL R.V (Continuous)

Given a continuous r.v. X

TYPES CONT. r.v.		Density func	Ill.		
uniform in $[a,b]$	$X \sim \mathcal{U}([a,b])$	$f_{a,b}(x) = \begin{cases} \frac{1}{b-a} & , x \in [a,b] \\ 0 & , x \notin [a,b] \end{cases}$			$\mathbb{E}[X] = \frac{b-a}{2}$
exponential with $\lambda > 0$	$T \sim Exp(\lambda)$	$f_\lambda(x) = \begin{cases} \lambda e^{-\lambda x} & , x \geq 0 \\ 0 & , x < 0 \end{cases}$			$\mathbb{E}[T] = \frac{1}{\lambda}$ $Var[T] = \frac{1}{\lambda^2}$
normal with $m, \delta^2 > 0$	$X \sim \mathcal{N}(m, \delta^2)$ (m: $\mathbb{E}[X]$, δ^2: $Var(X)$)	$f_{m,\delta}(x) = \frac{1}{\sqrt{2\pi\delta^2}} e^{-\frac{(x-m)^2}{2\delta^2}}$ EVEN FUNC!			$\mathbb{E}[X] = m$

Def. Uniform

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

① "falling" into an interval $[c, c+l] \subset [a,b]$

$$\mathbb{P}[X \in [c, c+l]] = \frac{l}{b-a}$$

② Distri. function

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

③. Standard uniform r.v.

$$X = a + (b-a) Y$$

$\rightsquigarrow Y \sim \mathcal{U}([0,1])$

Def. Exponential

- $T \sim Exp(\lambda)$

① exponentially small waiting prob.

$$\forall t \geq 0 \quad \mathbb{P}[T > t] = e^{-\lambda t}$$

② Absence of memory

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

Def. Normal

- $X \sim \mathcal{N}(m, \delta^2)$

①. ☐ List of independent r.v.

$X_i \sim \mathcal{N}(m_i, \delta_i^2)$

☐ $Z := m_0 + \sum_{i=1}^{n} \lambda_i X_i$

$= m_0 + \lambda_1 X_1 + \cdots + \lambda_n X_n$

$\Rightarrow Z \sim \mathcal{N}(\underbrace{m_0 + \sum_{i=1}^{n} \lambda_i m_i}_{m}, \underbrace{\sum_{i=1}^{n} \lambda_i^2 \delta_i^2}_{\delta^2})$

② Standard normal r.v.

☐ $X \sim \mathcal{N}(0,1) \Rightarrow Z \sim \mathcal{N}(m, \delta^2)$

$Z := m + \delta \cdot X$

$\mathbb{1}_A \sim Ber(p)$

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

EXPECTATION of r.v.

Given discrete r.v. $X : \Omega \to E$

CONDITION	EXPECTATION
i). $X \geq 0$ a.s. (constant sign!)	$\mathbb{E}[X] := \sum_{x \in E} x \cdot \mathbb{P}[X = x]$ → $\mathbb{E}[X] \in \mathbb{R}_+ \cup \{+\infty\}$
ii). X is integrable i.e. $\mathbb{E}[\lvert X \rvert] < \infty$	$\mathbb{E}[X] := \sum_{x \in E} x \cdot \mathbb{P}[X = x]$ → $\mathbb{E}[X] \in \mathbb{R}$

PROPERTIES

- $\mathbb{E}[g(x) + h(x)] = \mathbb{E}[g(x)] + \mathbb{E}[h(x)]$
- $\mathbb{E}[aX] = a \cdot \mathbb{E}[X]$
- $\mathbb{E}[X + c] = \mathbb{E}[X] + c$
- $\mathbb{E}[c] = c$

EXPECTATION LINEARITY

☐ r.v. $X, Y : \Omega \to \mathbb{R}$

✓ integrable

$\Rightarrow$ $\forall \lambda \in \mathbb{R}$: $\lambda \cdot X$, $X + Y$ are integrable discrete r.v

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

REMARK : i) $\forall n \geq 1$,

ii) dis. r.v. $X_1, \dots, X_n : \Omega \to E$ integrable

iii) $\lambda_i \in \mathbb{R}$

$\Rightarrow$ $\mathbb{E}[\lambda_1 X_1 + \dots + \lambda_n X_n] = \lambda_1 \mathbb{E}[X_1] + \dots + \lambda_n \mathbb{E}[X_n]$

Example : Expectation of $S \sim Bin(n, p)$

gey. $S \sim Bin(n, p)$ ($n \geq 1$, $p \in [0,1]$)

ges. $\mathbb{E}[S]$

P ① using $S_n = \sum_{i=1}^{n} X_i$ with same distri.

② make use of linearity

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

Prop 2.30 Tailsum formula

☐ $X : \Omega \to E$ dis. r.v.

✓ $\Omega = \mathbb{N} = \{0, 1, 2, \dots\}$

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

Theorem. Jensen's inequality

☐ Discrete r.v. $X : \Omega \to E$

☐ A convex function $\phi : \mathbb{R} \to \mathbb{R}$

☐ Expectations: well-defineness of — & —

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

NOTE : ①. $\phi(x) = |x|$ (convex) $\Rightarrow$ $|\mathbb{E}[X]| \leq \mathbb{E}[|X|]$

② $\phi(x) = x^2$ (convex) $\Rightarrow$ $\mathbb{E}[|X|] \leq \sqrt{\mathbb{E}[X^2]}$

Theorem. Image random variable $n \in \mathbb{N}$

i) ☐ List of n r.v. $X_1, \dots, X_n : \Omega \to E$

☐ $\phi : \mathbb{R}^n \to \mathbb{R}$

ii) iii) ☐ $\sum_{x_1, \dots, x_n \in E} |\phi(x_1, \dots, x_n)| \cdot \mathbb{P}[X_1 = x_1, \dots, X_n = x_n] < \infty$

$\Rightarrow$ i). Induced discrete r.v.

$Z = \phi(X_1, \dots, X_n)$

ii). Z is integrable i.e. $\mathbb{E}[|Z|] < \infty$

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

Theorem. Image random variable $n = 1$

i) ☐ Discrete r.v. $X : \Omega \to E$

☐ $\phi : \mathbb{R} \to \mathbb{R}$

ii). iii). ☐ $\sum_{x \in E} |\phi(x)| \cdot \mathbb{P}[X = x] < \infty$

$\Rightarrow$ i). Induced discre. r.v. $Z := \phi(X)$

ii). Z is integrable

iii). $\mathbb{E}[Z] = \sum_{x \in E} \phi(x) \cdot \mathbb{P}[X = x]$

NOTE : for $Z = \phi(X)$

☐ $\phi : \mathbb{R}^n \to F \subset [0, +\infty)$

$\Rightarrow$ $Z \geq 0$ a.s.

$\Rightarrow$ (☆) always holds

INDEPENDENCE OF r.v.

Given discrete r.v. X, Y

Equiv. statements

i). X, Y are independent

ii). $\forall a, b \in \mathbb{R}$

$\mathbb{P}[X = a, Y = b] = \mathbb{P}[X = a] \, \mathbb{P}[Y = b]$

iii). $\forall f, g : \mathbb{R} \to \mathbb{R}$

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

with expectations well-defined

Theorem 2.39. Independence [list]

☐ Discre. r.v. $X_1, \dots, X_n$

$\Rightarrow$ Equiv. statements

i). $X_1, \dots, X_n$ are independent

ii). $\forall x_1, \dots, x_n \in \mathbb{R}$:

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

iii). $\forall f_1, \dots, f_n : \mathbb{R} \to \mathbb{R}$

✓ $f_i(X_i)$ integrable $\forall i$

✓ $\mathbb{E}[f_1(X_1) \cdots f_n(X_n)] = \mathbb{E}[f_1(X_1)] \cdots \mathbb{E}[f_n(X_n)]$

COUNTEREXAMPLE : $\mathbb{E}[X]$ not well-defined

Special case : Cauchy distri.

- "X has Cauchy distri."

$\Updownarrow$

- X is continuous with f

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

NOTE : ①. X is not integrable

$\mathbb{E}[|X|] = \frac{1}{\pi} \int_{-\infty}^{\infty} |x| \cdot \frac{1}{1+x^2} \, dx = +\infty$!

②. For bounded intervals

i). $\lim_{N \to \infty} \int_{-N}^{2N} \frac{1}{\pi} \frac{x}{1+x^2} \, dx = \frac{1}{\pi} \log 2$

ii). $\lim_{N \to \infty} \int_{-3N}^{N} \frac{1}{\pi} \frac{x}{1+x^2} \, dx = -\frac{1}{\pi} \log 3$

EXPECTATION [contin.]

Given r.v. $X : \Omega \to \mathbb{R}$

- ✓ X is continuous
- ✓ density func is f

CONDITION	EXPECTATION
i). $X \geq 0$ a.s. (constant sign!)	$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \cdot f(x)\,dx = \int_{0}^{\infty} x \cdot f(x)\,dx$ ($f(x)=0$ $\forall x<0$)
ii). X is integrable i.e. $\mathbb{E}[\lvert X\rvert] < \infty$	$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \cdot f(x)\,dx$
iii). $\phi : \mathbb{R} \to \mathbb{R}$, $\int_{-\infty}^{\infty} \lvert\phi(x)\rvert\, f(x)\,dx < \infty$	$\mathbb{E}[\phi(X)] = \int_{-\infty}^{\infty} \phi(x) f(x)\,dx$

NOTE: Given i). X with f

ii). Y with g, $Y := \phi(X)$

CONSTRAINT:

$$\mathbb{E}[Y] \overset{!}{=} \mathbb{E}[\phi(X)]$$

$Y = \phi(X)$ contin. $\Rightarrow \int_{-\infty}^{\infty} y \cdot g(y)\,dy \overset{!}{=} \int_{-\infty}^{\infty} \phi(x) \cdot f(x)\,dx$

$Y = \phi(X)$ discrete $\Rightarrow \sum_{y \in E} y\, \mathbb{P}[Y = y] \overset{!}{=} \int_{-\infty}^{\infty} \phi(x) \cdot f(x)\,dx$

Theorem 3.14. independent r.v. [cont.]

☐ r.v. X, Y continuous

✓ with densities f_X, f_Y

$\Rightarrow$ Equiva. statements:

i). X, Y are independent

ii). X, Y are jointly cont. with

✓ $f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y)$

iii). $\forall\, \phi, \psi : \mathbb{R} \to \mathbb{R}$ well-defined expec. needed!

$$\boxed{\mathbb{E}[\phi(X)\psi(Y)] = \mathbb{E}[\phi(X)]\, \mathbb{E}[\psi(Y)]}$$

Theorem 3.8. Jensen's inequality [contin.]

☐ r.v. X continuous

☐ $\phi : \mathbb{R} \to \mathbb{R}$ convex

well-definess needed

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

VARIANCE of r.v.

Var(X) large
→ large fluctuations
→ inaccurate measurement

Def. Variance of X

☐ Discre. r.v. X

☐ Expectation bounded $\mathbb{E}[X^2] < \infty$ (△)

$\Leftrightarrow$ i). variance of X $\quad m = \mathbb{E}[X]$

$$\sigma_X^2 := \mathbb{E}[(X - m)^2]$$

ii). "standard deviation of X"

$\longleftrightarrow \sqrt{\sigma_X^2}$ how large the deviation around m are

$$\boxed{\sigma(X) = \sqrt{Var[X]}}$$

NOTE: Well-defineness of m

(△) $\mathbb{E}[X^2] < \infty$

⇓ Jensen's ineq.

$$\mathbb{E}[\lvert X\rvert] \leq \sqrt{\mathbb{E}[X^2]} < \mathbb{E}[X^2] < \infty$$

⇓ well-defineness: integrable X

$\mathbb{E}[X]$ is well-defined

BASIC PROPERTIES σ_X

i). ☐ Discre. r.v. X ✓ $\mathbb{E}[X^2] < \infty$

$$\Rightarrow \boxed{\sigma_X^2 = \mathbb{E}[X^2] - \mathbb{E}[X]^2}$$

ii). ☐ List of r.v. $X_1, \dots, X_n$ ✓ pairwise independent ✓ $S := X_1 + \dots + X_n$

$$\Rightarrow \boxed{\sigma_S^2 = \sum_{i=1}^{n} \sigma_{X_i} = \sigma_{X_1}^2 + \dots + \sigma_{X_n}^2}$$

VARIANCE of X continu.

Given $X : \Omega \to \mathbb{R}$ contin. with f

Def. Variance of X

☐ $\mathbb{E}[X^2] < \infty$

☐ $m = \mathbb{E}[X]$

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

PROPERTIES

Given r.v. X contin.

✓ $\mathbb{E}[X^2] < \infty$

$\Rightarrow$ i). $\sigma_X^2 = \mathbb{E}[X^2] - \mathbb{E}[X]^2$

ii). $\sigma_{\lambda X + \mu}^2 = \lambda^2 \cdot \sigma_X^2$ $(\lambda, \mu \in \mathbb{R})$

Given r.v. $X_1, \dots, X_n$ [List]

✓ X_i pairwise independent

✓ $\mathbb{E}[X_i^2] < \infty$ $\forall i$

Let $S := \sum_{i=1}^{n} \lambda_i X_i = \lambda_1 X_1 + \dots + \lambda_n X_n$

$\Rightarrow Var(S) = \sigma_S^2 = \sum_{i=1}^{n} \lambda_i^2 \sigma_{X_i}^2$

JOINT DISTRIBUTION

Def. Joint distribution

☐ List of n discrete r.v. ⚠ on the same $(\Omega, \mathcal{F}, \mathbb{P})$

$X_i : \Omega \to E_i \quad \forall i \in [n]$

✓ $E_i \subset \mathbb{R}$ finite/countable

☐ An (indexed) family

$$\boxed{(p_{x_1,\dots,x_n})_{x_1 \in E_1, \dots, x_n \in E_n} : \quad p_{x_1,\dots,x_n} \overset{def}{:=} \mathbb{P}[X_1 = x_1, \dots, X_n = x_n]}$$

$\Leftrightarrow (p_{x_1,\dots,x_n})_{x_1 \in E_1, \dots, x_n \in E_n}$ is the joint distribution of the r.v.'s X_i $\forall i \in [n]$

Lemma. joint distri. for independent r.v.

☐ X_i independent $\forall i \in [n]$

$$\Rightarrow p_{x_1,\dots,x_n} = \mathbb{P}[X_1 = x_1, \dots, X_n = x_n] = \mathbb{P}[X_1 = x_1] \cdots \mathbb{P}[X_n = x_n] = \prod_{i=1}^{n} \mathbb{P}[X_i = x_i] \quad := \mathbb{P}^i_{x_i}$$

prop 2.23. Consider r.v. as images of discrete r.v.

☐ A function $\phi : \mathbb{R}^n \to \mathbb{R}$ $n \geq 1$!

☐ List of n discrete r.v.

$X_i : \Omega \to E_i \quad \forall i \in [n]$

✓ E_i finite/countable

✓ $X_i(\Omega) = E_i$

☐ r.v. $Z : \Omega \to \underbrace{\phi(E_1 \times \dots \times E_n)}_{:= F \text{ discrete set}}$

$$Z \overset{def}{:=} \phi(X_1, \dots, X_n)$$

$\Rightarrow$ i). Z is discrete r.v.

ii). Distribution of Z

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

Def. Cont. joint distribution [Contin.]

☐ r.v. $X, Y : \Omega \to \mathbb{R}$ continuous

☐ X, Y have a contin. joint distribution

$\overset{def}{\Leftrightarrow}$ i). Joint density $\exists f : \mathbb{R}^2 \to \mathbb{R}_+$ s.t.

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

For every i). $-\infty \leq a \leq a' \leq \infty$

ii). $-\infty \leq b \leq b' \leq \infty$

RECALL: Distri. of discre. r.v. $\sum_{x \in E} p_x = 1$

⤳ SOUGHT: for f with $\underline{x} \in \Omega \subset \mathbb{R}^n$ $\quad \int_{\Omega} f(\underline{x})\,d\underline{x} = 1$

Lemma 3.13. Correct sum of joint dis.

☐ Joint density $f : \mathbb{R}^2 \to \mathbb{R}^+$ of X, Y

Prob. of $\underline{x} := (X(\omega), Y(\omega))$ lies in $[x, x+dx] \times [y, y+dy]$

$$\Rightarrow \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} f(x, y)\,dy \right) dx = 1 \quad (\star)$$

NOTE: Given $f : \mathbb{R} \to \mathbb{R}^+$ fulfilling $(\star)$ we can construct $(\Omega, \mathcal{F}, \mathbb{P})$ and $X, Y : \Omega \to \mathbb{R}$ with joint distri. f correspondingly

Def. Expectation of $\phi(X, Y)$

☐ Joint density $f_{X,Y}$ of r.v. X, Y

☐ $\phi : \mathbb{R}^2 \to \mathbb{R}$

✓ $Z \overset{def}{:=} \phi(X, Y)$

$\Rightarrow$ Expectation of Z (well-defined integral needed)

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

LINEARITY - Joint distri.

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

Lemma. independent ⇒ jointly con. (r.v.)

☐ r.v. X, Y are independent

$\Rightarrow X, Y$ are jointly continuous

STATISTICS

$\forall i$, X_i i.i.d.

• Discrete models	• continuous models
$X_i \sim Ber(p)$ $p \in [0,1]$	$X_i \sim \mathcal{U}([0,\theta])$, $\theta \geq 0$
$X_i \sim Geom(p)$ $p \in [0,1]$	$X_i \sim Exp(\lambda)$, $\lambda > 0$
$X_i \sim Poisson(\lambda)$ $\lambda > 0$	$X_i \sim \mathcal{N}(m, \sigma^2)$, $m \in \mathbb{R}$, $\sigma^2 > 0$

Terminology : Realization of the model

• Realization $\Leftrightarrow$ a vector $\underline{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$

with possible values for $(X_1, \dots, X_n)$

MAXIMUM LIKELIHOOD ESTIMATOR

finding opt. value to the mean or standard deviation for a distri.

Set-up : Distribution $(p_\theta(x))$ of i.i.d. r.v. $X_1, \dots, X_n$ depends on $\theta \in \mathbb{R}$

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

Def. Likelihood function of $\underline{x}$ [Discr.]

☐ A realization $\underline{x} = (x_1, \dots, x_n) \in E^n$ for $(X_1, \dots, X_n)$

$\Leftrightarrow$ Likelihood function of $\underline{x}$

$$L(\theta) \overset{def}{=} L_{\underline{x}}(\theta) = \underbrace{\mathbb{P}[X_1 = x_1, \dots, X_n = x_n]}_{\text{joint distri.}}$$

$$\overset{\text{i.i.d.}}{=} \mathbb{P}[X_1 = x_1] \cdots \mathbb{P}[X_n = x_n]$$

$$= p_\theta(x_1) \cdots p_\theta(x_n)$$

Def. Likelihood function of $\underline{x}$ [Cont.]

☐ A realization $\underline{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$ for $(X_1, \dots, X_n)$

$\Leftrightarrow$ Likelihood function of $\underline{x}$

$$\boxed{L(\theta) \overset{def}{=} L_{\underline{x}}(\theta) = f_\theta(x_1) \cdots f_\theta(x_n)}$$

Def. Maximum likelihood estimator

☐ A realization $\underline{x} = (x_1, \dots, x_n)$

☐ maximum of the likelihood func

$$L_x(\hat{\theta}) \overset{def}{:=} \max_\theta L_x(\theta)$$

$\Leftrightarrow$ Parameter $\hat{\theta} := \hat{\theta}(x_1, \dots, x_n)$ is the maximum likelihood estimator

NOTE : multivariate likelihood func.

• CASE 1 : $(p_{\theta_1, \theta_2, \dots}(x))$

• CASE 2 : $f_{\theta_1, \theta_2, \dots}(x)$

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

CONFIDENCE INTERVALS

"criterium for measuring how good an estimator is (depends on n)"

Def. z%-confidence interval

☐ A prob. model with param. θ

☐ A realization $\underline{x} = (x_1, \dots, x_n)$

☐ An interval $I = [a(\underline{x}), b(\underline{x})] \subset \mathbb{R}$

✓ I is a z%-confidence interval for θ

$$\Leftrightarrow \forall \theta \ \mathbb{P}[a(X_1, \dots, X_n) \leq \theta \leq b(X_1, \dots, X_n)] \geq \frac{z}{100}$$

STATISTICAL TESTS

null hypo. $H_0 : \theta = \theta_0$ satisfied $\rightsquigarrow \mathbb{P}_{H_0}[\cdot]$

alter. hypo. $H_1 : \theta = \theta_1$ satisfied $\rightsquigarrow \mathbb{P}_{H_1}[\cdot]$

Def. Test function

☐ A realization $\underline{x} = (x_1, \dots, x_n) \in E^n / \mathbb{R}^n$ (Discrete / cont.)

☐ A function $d : E^n / \mathbb{R}^n \longrightarrow \{0, 1\}$

$$\boxed{d(\underline{x}) = \begin{cases} 0, & H_0 \text{ is accepted} \\ 1, & H_0 \text{ is rejected} \end{cases}}$$

$\Leftrightarrow$ d is the test function

TERMINOLOGY

• simple hypo. $\Leftrightarrow$ distri. of $X_1, \dots, X_n$ comple. determined under the hypothesis H_0 or H_1

ERROR TYPES

• Type I error : reject H_0 when H_0 is true

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

relevance level of the test

• Type II error : accept H_0 when H_1 occurs

power of the test

$$1 - \beta = 1 - \mathbb{P}_{H_1}[d(X_1, \dots, X_n) = 0]$$

$$= \mathbb{P}_{H_1}[d(X_1, \dots, X_n) = 1] \ ?$$

$$= \mathbb{E}_{H_1}[d(X_1, \dots, X_n)] \ ?$$

PRIORITIES

Type I error $\geq$ Type II error

Theorem 4.9. Neyman-Pearson's

☐ A stat. framewor

✓ simple hypo. H_0, H_1

☐ Tests with the same relevance level α

✓ Likelihood ratio test $d : E^n / \mathbb{R}^n \longrightarrow \{0, 1\}$

with test func & α, β defined as above

$\Rightarrow$ Likelihood ratio test is the most powerful

$$\boxed{\mathbb{E}_{H_1}[d(X_1, \dots, X_n)] \geq \mathbb{E}_{H_1}[d^*(X_1, \dots, X_n)]}$$

for another test d^*

PROCEDURE: likelihood ratio test

Given $\theta_0, \theta_1, \dots$ (eg. $\mathbb{P}[\{\text{coin on head}\}] = \theta_0 = 0{,}7$

STEP 0: specify the model

- ✓ X_i: ($\#X_i = ?$, $X_i \sim Ber(p) / \mathcal{U}([a,b])$?)

STEP 1: Choose hypotheses for the test

- ✓ Assign θ_i to the hypotheses

$H_0: p = 0{,}7 \overset{\text{correc.}}{\longleftrightarrow}$ Bad coin param $\theta = 0{,}7$

$H_1: p = 0{,}5 \longleftrightarrow$ Good coin param $\theta = 0{,}5$

STEP 2: Understanding error types in the context

Type I: i) $\mathbb{P}_{H_0}[\cdot]$ used $\Rightarrow$ coin actually bad

ii) H_0 rejec. $\Rightarrow$ coin not bad $\Rightarrow$ keep

Type II: i). $\mathbb{P}_{H_1}[\cdot]$ used $\Rightarrow$ coin actually good

ii). H_0 accept. $\Rightarrow$ coin is bad $\Rightarrow$ throw away

STEP 3: Calculating the $\mathbb{P}_{H_0}, \mathbb{P}_{H_1} \rightsquigarrow r(x)$

$$\left.\begin{aligned} \mathbb{P}_{H_0}[X = x] &= L_x(\theta_0) = p_{\theta_0}(x_1)\cdots p_{\theta_0}(x_n) \\ \mathbb{P}_{H_1}[X = x] &= L_x(\theta_1) = p_{\theta_1}(x_1)\cdots p_{\theta_1}(x_n) \end{aligned}\right\} \text{using } |x| \text{ have to count!}$$

$$\Rightarrow r(x) = \frac{L_x(\theta_0)}{L_x(\theta_1)}$$

This can be used to output a table with varying $|x|$

$\|x\|$	0	...	10
θ_0 Bad coin			
θ_1 good coin			
$r(x)$			

TRIGONOMETRY

$$\cosh^2(x) - \sinh^2(x) = 1$$

$$\sin(z) = \frac{1}{2i}\left[e^{iz} - e^{-iz}\right] \qquad \cos(z) = \frac{1}{2}\left[e^{iz} + e^{-iz}\right]$$

$$\sinh(x) = \frac{1}{2}\left[e^{x} - e^{-x}\right] \qquad \cosh(x) = \frac{1}{2}\left[e^{x} + e^{-x}\right]$$

$$\sin^2(x) = \frac{1}{2}\left[1 - \cos(2x)\right] \qquad \sin^3(x) = \frac{1}{4}\left[3\sin(x) - \sin(3x)\right]$$

$$\cos^2(x) = \frac{1}{2}\left[1 + \cos(2x)\right] \qquad \cos^3(x) = \frac{1}{4}\left[3\cos(x) + \cos(3x)\right]$$

$$\sin^4(x) = \frac{1}{8}\left[\cos(4x) - 4\cos(2x) + 3\right]$$

$$\cos^4(x) = \frac{1}{8}\left[\cos(4x) + 4\cos(2x) + 3\right]$$

$$\sin(x)\sin(y) = \frac{1}{2}\left[\cos(x-y) - \cos(x+y)\right]$$

$$\cos(x)\cos(y) = \frac{1}{2}\left[\cos(x-y) + \cos(x+y)\right]$$

$$\sin(x)\cos(y) = \frac{1}{2}\left[\sin(x-y) + \sin(x+y)\right]$$

$$\sin(x)\cos(x) = \frac{1}{2}\left[\sin(2x)\right]$$

$$\sin(x \pm y) = \sin(x)\cos(y) \pm \cos(x)\sin(y)$$

$$\cos(x \pm y) = \cos(x)\cos(y) \mp \sin(x)\sin(y)$$

$$\sin(x+y)\sin(x-y) = \cos^2(y) - \cos^2(x) = \sin^2(x) - \sin^2(y)$$

$$\cos(x+y)\cos(x-y) = \cos^2(y) - \sin^2(x) = \cos^2(x) - \sin^2(y)$$

INTEGRAL

$f(x)$	$F(x)+C$
$\frac{1}{x}$	$\ln\|x\|$
α^x	$\frac{\alpha^x}{\ln\|\alpha\|}$
$\frac{1}{\cos^2 x}$	$\tan x$
$\frac{1}{\sin^2 x}$	$-\cot x = -\frac{1}{\tan x}$
$\frac{1}{\sqrt{1-x^2}}$	$\arcsin x$
$\frac{1}{1+x^2}$	$\arctan x$
$\frac{1}{\cosh^2 x}$	$\tanh x$
$\frac{1}{\sinh^2 x}$	$-\coth x$

$$\int_a^b f(x)G(x)\,dx = \left[F(x)G(x)\right]_a^b - \int_a^b F(x)g(x)\,dx$$

$\cos^2 x$	$\frac{\cos x \sin x + x}{2}$
$\sin^2 x$	$\frac{x - \cos x \sin x}{2}$
$\cos x \sin x$	$\frac{\sin^2 x}{2} = -\frac{1}{2}\cos^2 x$

$$\int_0^{2\pi/\omega} \cos(k\omega x)\sin(l\omega x)\,dx = 0$$

$$\int_0^{2\pi} \cos(x)\,dx = \int_0^{2\pi} \sin(x)\,dx = 0$$

$$\int_0^{2\pi} \cos^3(x)\,dx = \int_0^{2\pi} \sin^3(x)\,dx = 0$$

$$\int_0^{2\pi} \cos^4(x)\,dx = \int_0^{2\pi} \sin^4(x)\,dx = \frac{3\pi}{4}$$

Def. Likelihood ratio $r(x)$

- ☐ realization $\underline{x} = (x_1, \dots, x_n)$
- ☐ Likelihood functions of θ_0, θ_1
  - ✓ Hypo. are satisfied

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

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

$\Longleftrightarrow$ Define the likelihood ratio

$$\boxed{r(x) := \frac{L_x(\theta_0)}{L_x(\theta_1)}} \overset{\text{indu.}}{\Longrightarrow} \boxed{d(\underline{x}) = \begin{cases} 0, & r(x) > c \ (H_0 \text{ acc.}) \\ 1, & r(x) \leq c \ (H_0 \text{ rej.}) \end{cases}} \Longrightarrow$$

ERROR TYPES

Relevance level

$$\alpha = \mathbb{P}_{H_0}[r(X_1, \dots, X_n) \leq 1]$$

Power of the test

$$1 - \beta = 1 - \mathbb{P}_{H_0}[r(X_1, \dots, X_n) > c]$$

$$= \mathbb{P}_{H_1}[r(X_1, \dots, X_n) \leq c]$$